

This article was downloaded by: [National Chiao Tung University 國立交通大學]

On: 28 April 2014, At: 15:36

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## International Journal of Systems Science

Publication details, including instructions for authors and subscription information:  
<http://www.tandfonline.com/loi/tsys20>

### A batch arrival queue under randomised multi-vacation policy with unreliable server and repair

Jau-Chuan Ke<sup>a</sup>, Kai-Bin Huang<sup>b</sup> & Wen Lea Pearn<sup>b</sup>

<sup>a</sup> Department of Applied Statistics, National Taichung Institute of Technology, No. 129, Sec. 3 Sanmin Rd, Taichung 404, Taiwan, Republic of China

<sup>b</sup> Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan, Republic of China

Published online: 14 Nov 2010.

To cite this article: Jau-Chuan Ke, Kai-Bin Huang & Wen Lea Pearn (2012) A batch arrival queue under randomised multi-vacation policy with unreliable server and repair, International Journal of Systems Science, 43:3, 552-565, DOI: [10.1080/00207721.2010.517863](https://doi.org/10.1080/00207721.2010.517863)

To link to this article: <http://dx.doi.org/10.1080/00207721.2010.517863>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## A batch arrival queue under randomised multi-vacation policy with unreliable server and repair

Jau-Chuan Ke<sup>a\*</sup>, Kai-Bin Huang<sup>b</sup> and Wen Lea Pearn<sup>b</sup>

<sup>a</sup>Department of Applied Statistics, National Taichung Institute of Technology, No. 129, Sec. 3 Sanmin Rd, Taichung 404, Taiwan, Republic of China; <sup>b</sup>Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan, Republic of China

(Received 27 May 2009; final version received 27 June 2010)

This article examines an  $M^{[x]}/G/1$  queueing system with an unreliable server and a repair, in which the server operates a randomised vacation policy with multiple available vacations. Upon the system being found to be empty, the server immediately takes a vacation. If there is at least one customer found waiting in the queue upon returning from a vacation, the server will be activated for service. Otherwise, if no customers are waiting for service at the end of a vacation, the server either remains idle with probability  $p$  or leaves for another vacation with probability  $1 - p$ . When one or more customers arrive when the server is idle, the server immediately starts providing service for the arrivals. It is possible that an unpredictable breakdown may occur in the server, in which case a repair time is requested. For such a system, we derive the distributions of several important system characteristics, such as the system size distribution at a random epoch and at a departure epoch, the system size distribution at the busy period initiation epoch, and the distribution of the idle and busy periods. We perform a numerical analysis for changes in the system characteristics, along with changes in specific values of the system parameters. A cost effectiveness maximisation model is constructed to show the benefits of such a queueing system.

**Keywords:** batch arrival queue; cost effectiveness; randomised vacation; supplementary variable technique; reliability

### 1. Introduction

We consider an  $M^{[x]}/G/1$  queueing system in which an unreliable server operates a randomised vacation policy with multiple vacations; the term *unreliable server* refers to a server which is typically subject to unpredictable breakdowns. The randomised vacation policy presented in this article is as follows: when the system is empty, the server immediately takes a vacation. If there is at least one customer found waiting in the queue upon returning from a vacation, the server will be activated for service. Otherwise, if no customers are waiting for service at the end of a vacation, the server remains idle in the system with probability  $p$  and leaves for another vacation with probability  $1 - p$ . In practical applications, there are instances where the server may suffer from unpredictable breakdowns which interrupt service (such as power outages or when the server actually breaks down) during busy periods or where it must perform secondary (vacations) tasks during idle periods.

The modelling analysis for vacation queueing models has been undertaken previously by a number of researchers and has successfully been used in various applied problems, such as production/inventory

systems, communication systems and computer networks (Doshi 1986). A comprehensive study on vacation models can be found in Levy and Yechiali (1975) and Takagi (1991). Past work regarding the vacation queueing models may be divided into two parts, depending on whether the server is reliable or unreliable. We will first look at the work which deals with a reliable server. Baba (1986) studied the  $M^{[x]}/G/1$  queueing system with multiple vacations. The first study on vacation models with a control policy was done by Kella (1989). Madan (2001) dealt with a single-server queue with two-stage heterogeneous service and deterministic server vacations. Variations and extensions of these vacation models with control policies have been studied by numerous authors (Lee, Lee, Park, and Chae 1994; Lee, Lee, and Chae 1995; Ke 2001; Hur, Kim, and Kang 2003; Tadj 2003; and Ke 2008). Lee et al. (1994, 1995) analysed batch arrival queues with the N-policy under single and multiple vacations. Ke (2001) examined the control policy of an  $M/G/1$  queueing system with server startup and two vacation types. Hur et al. (2003) optimised the operating cost of an  $M/G/1$  queueing system using the  $\text{Min}(N, T)$  policy. They derived the

\*Corresponding author. Email: jauchuan@ntit.edu.tw

steady-state system size distribution, established a cost function to reveal the characteristics of the cost function and determined the optimal operating policy. The T policy for the M/G/1 quorum queueing system was first proposed by Tadj (2003), who obtained the probability-generating function (PGF) of the number of customers in the system, the expected length of the idle period, busy period and busy cycle, and the determination of the optimum value  $T$ . Later, Ke (2008) examined the two thresholds of a batch arrival  $M^{[x]}/G/1$  queueing system under a modified T vacation policy with startup and closedown. The developments and applications on the optimal control of vacation queueing models with different considerations are rich and varied (Tadj and Choudhury 2005). Choudhury and Madan (2005) and Choudhury and Paul (2006) investigated the behaviour of batch arrival queues with an additional second optional service under the N-policy where all of the arrivals demand the first essential service, whereas only some of them demand the second 'optional' service. In addition, Takagi (1991) first proposed the concept of a variant vacation (a generalisation of single and multiple vacations) for the single arrival M/G/1 regular system. Zhang and Tian (2001) treated the discrete time Geo/G/1 system with variant vacations, wherein the server takes a random maximum number of vacations after serving all customers in the system. Ke and Chu (2006) examined the variant policy for an  $M^{[x]}/G/1$  queueing system using a stochastic decomposition property. Ke (2007) recently used the supplementary variable technique to study an  $M^{[x]}/G/1$  queueing system with balking under a variant vacation. As for the second category, when the server is unreliable, Wang and Ke (2002) dealt with three parametric control policies (N, T and Min(N,T)) for a single removable and unreliable server M/G/1 queueing system. Ke (2003a) investigated Kella's (1989) system by considering vacations and startup of an unreliable server. Ke (2003b) examined the optimal strategy policy for an unreliable server  $M^{[x]}/G/1$  queueing system with multiple vacations. Ke (2005) studied a variant T policy for an unreliable server M/G/1 queueing system. The NT policy for an unreliable server M/G/1 queueing system with server startup and closedown was investigated by Ke (2006a), who derived the explicit formulae for various system performance measures such as the expected number of customers in the system, the expected waiting time in the queue, the expected lengths of the idle, busy and breakdown periods, as well as the expected length of the busy cycle. Ke (2006b) proposed a hierarchical vacation policy for an unreliable server M/G/1 queueing system with an early startup. Ke (2007) studied two vacation policies for an unreliable server  $M^{[x]}/G/1$

queueing system with startup and closedown times. Yang, Wang, Ke, and Pearn (2008) analysed the optimal control of a  $(T,p)$  policy of an unreliable server M/G/1 system with second optional service and startup. We should note that, in the above-listed works, there are immediately available repair services when a server breaks down. However, in many real-life situations it may not be feasible to start the repairs immediately due to non-availability of the server or if the system is turned off. Choudhury and Tadj (2009) recently investigated the steady-state behaviour of an unreliable M/G/1 queue with an additional second phase of optional service and delayed repair. Choudhury, Ke, and Tadj (2009) proposed the N-policy for an unreliable server  $M^{[x]}/G/1$  queueing system with delayed repair and two phases of service.

The existing literature on queueing problems has focused on vacation policies that depend on queue size or time. To date, very few authors have studied vacation queueing systems where an unreliable server takes a sequence of randomised vacations in the idle time and the repair is requested. This motivates us to develop a randomised vacation policy for an  $M^{[x]}/G/1$  queueing system where the unreliable server operates a randomised vacation policy when the system is empty. Conveniently, we represent this randomised vacation system as an  $M^{[x]}/(G_1, G_2)/1/VAC(\infty)$  queueing system, where  $G_1$  and  $G_2$  represent the service time and repair time, respectively, and  $VAC(\infty)$  implies that there is no limit to the number of vacations. Our study is also motivated by some real-world problems. For example, the proposed model can be applied to wafer fabrication problems. Specifically, the wafer fabrication process consists of a series of operations such as epitaxy, oxidation, diffusion, ion implantation, etching and photolithography that build layers of circuitry on a wafer of silicon or gallium arsenide. Depending on the function of the product, the manufacturing process is composed of 200–500 steps. The wafer fabrication problem has received a lot of research attention due to the diverse characteristics of the process, especially in the photolithography process (Uzsoy, Lee, and Martin-Vega 1992, 1994). The photolithography process uses masks/reticles to transfer circuit patterns onto a wafer, and the etching process forms tangible circuit patterns onto the wafer chip. With the required number of processes in the photolithography, integrated circuitry products with preset functions are developed on the wafer (Toktay and Uzsoy 1998). From the concept of the theory of constraints (TOC) proposed by Goldratt and Cox (1992), the performance of a system is determined by the bottleneck resource in that system. In the wafer fabrication photolithography area, the stepper machine (which can be referred to as the server) used to process a wafer

lot is the most expensive machine in a wafer fabrication factory and usually has the highest utilisation rate; hence, the photolithography workstation is a critical resource in the wafer fabrication process. Consequently, optimised photolithography capacity allocation is very important in the process. There are three main methods used for investigating capacity allocation problems: simulation-based, mathematical and hybrid. Under the mathematical method, a mathematical model, such as linear programming, or queueing theory is utilised. Notably, several important system characteristics are derived in this article, such as the system size distribution or busy cycle, which can be helpful for better understanding the impact of the machine loading balance problem.

In the photolithography process, the arrival of jobs at the workstation can be modelled as a compound Poisson process. The service/process time of each job (provided by a stepper machine) is a random variable with a general distribution. Each job enters the queue of a prescribed machine within the workstation of the photolithography area based on the load allocation (Toktay and Uzsoy 1998). Whenever all jobs are completed and no new jobs arrive, the service is stopped and the stepper machine takes a vacation immediately (to execute preventive maintenance). The primary goal of preventive maintenance is to prevent the breakdown and failure of equipment before it actually occurs. As mentioned above, the photolithography workstation is a critical resource in the wafer fabrication process, so in order to ensure the performance of the system, the machine in the photolithography workstation should utilise idle periods to undertake a sequence of preventive maintenance (such as minimal repairs, overhaul, etc.) to prevent failures. So, when the machine finishes the primary preventive maintenance and returns to the workstation, finding that there are no jobs needed to be processed (perhaps the jobs have been processed by the previous workstation), the machine will either remain idle in the workstation or leave for another vacation (to execute another type of preventive maintenance). Moreover, the machine may be interrupted due to some unpredictable events. When the unpredictable events occur, it is repaired immediately. The service of jobs will start again when the interruption is resolved.

The objectives of this article are as follows: first, we develop the PGF of the number of customers present in the system at a random epoch and at a departure epoch; second, we derive other system characteristics such as the system size distribution at a busy period initiation epoch, and the busy and idle period distributions; third, we deduce reliability indices such as availability and failure frequency of the server; fourth,

the effect of parameters on the system characteristics is studied numerically and finally, a cost effectiveness maximisation model is developed to illustrate the benefits from the investment.

## 2. The system

We consider an  $M^{[x]}/G/1$  system with an unreliable server and a repair, in which the server operates a randomised vacation policy when all customers have been served. The detailed description of the model is given as follows: customers arrive in batches according to a compound Poisson process with an arrival rate  $\lambda$ . Let  $X_k$  denote the number of customers belonging to the  $k$ th arrival batch, where  $X_k, k = 1, 2, 3, \dots$ , are with a common distribution.

$$\Pr(X_k = n) = \chi_n, \quad n = 1, 2, 3, \dots$$

The service time provided by a single server is an independent and identically distributed random variable  $S$  with distribution function  $S(t)$  and Laplace–Stieltjes transform (LST)  $S^*(\theta)$ . Arriving customers who join the system form a single waiting line based on the order of their arrival; that is, they are queued according to the first-come, first-served (FCFS) discipline. The server can serve only one customer at a time and the service is independent of the arrival of the customers. If the server is busy or on vacation, arrivals in the queue must wait until the server is available. Whenever the system becomes empty, the server deactivates and leaves for a vacation of random length  $V$ . If at least one customer is found waiting in the queue upon returning from the vacation, the server is immediately activated for service. Alternatively, if no customers are found in the queue at this time, the server either remains idle in the system with probability  $p$  or leaves for another vacation with probability  $\bar{p}$  ( $= 1 - p$ ). The vacation time  $V$  has distribution function  $V(t)$  and LST  $V^*(\theta)$ . The server is subject to breakdowns at any time with a Poisson breakdown rate  $\alpha$  when it is working. As soon as the server fails, it is sent for repair during which time the server stops serving the arriving customers and waits for the repair to begin. The repair time of the malfunctioning server is an independent and identically distributed random variable  $R$  with a general distribution function  $R(t)$  and LST  $R^*(\theta)$ . A customer who arrives and finds the server busy or broken down must wait in the queue until the server is available. Although no service occurs during the repair period of a broken server, customers continue to arrive according to a compound Poisson process. In the case where the server breaks down while serving a customer, the server is sent for repair and the customer who is currently being served waits for

the server to return to complete the service. Immediately after the server is fixed, it starts to serve customers until the system is once again empty. Notably, the various stochastic processes involved in the system are independent of each other.

Define  $G$  as the generalised service time random variable representing the completion of a customer service, which consists of both the service time of a customer and the repair time of a server. The LST of  $G$  can be expressed as follows:

$$G^*(\theta) = \int_0^\infty \sum_{n=0}^\infty \frac{e^{-\alpha t} (\alpha t)^n}{n!} e^{-\theta t} [R^*(\theta)]^n dS(t) = S^*(\theta + \alpha(1 - R^*(\theta))). \tag{1}$$

From (1), we obtain the first moment of  $G$ ,

$$E[G] = -\frac{d}{d\theta} [G^*(\theta)]|_{\theta=0} = E[S](1 + \alpha E[R]), \tag{2}$$

where  $E[S] = -\frac{d}{d\theta} [G^*(\theta)]|_{\theta=0}$  is the mean service time and  $E[R] = -\frac{d}{d\theta} [R^*(\theta)]|_{\theta=0}$  is the mean repair time.

### 3. The analysis

We first develop the steady-state differential-difference equations for the  $M^{[X]}(G_1, G_2)/1/VAC(\infty)$  queueing system by treating the elapsed service time, the elapsed repair time and the elapsed vacation time as supplementary variables. Then we solve the system equations and derive the PGFs of various server states at a random epoch.

#### 3.1. System size distribution at a random epoch

In the steady state, let us assume that  $S(t) = 0$ , for  $t \leq 0$ ,  $S(\infty) = 1$ ;  $R(t) = 0$ , for  $t \leq 0$ ,  $R(\infty) = 1$  and  $V(t) = 0$ , for  $t \leq 0$ ,  $V(\infty) = 1$ , and these distribution functions are continuous at  $t = 0$ , so that  $\mu(t)dt = \frac{dS(t)}{1-S(t)}$  can be interpreted as the probability density function of the remaining service time, given that the elapsed time is  $x$ .  $\eta(t)dt = \frac{dR(t)}{1-R(t)}$  and  $\omega(t)dt = \frac{dV(t)}{1-V(t)}$  can be referred to the corresponding repair and vacation probabilities, respectively.

The following random variables we define are used for the development of  $M^{[X]}(G_1, G_2)/1/VAC(\infty)$  queueing system:

- $N(t)$  the number of customers in the system,
- $S^-(t)$  the elapsed service time,
- $R^-(t)$  the elapsed repair time,
- $V_j^-(t)$  the elapsed time of the  $j$ th vacation.

We introduce the following random variable for further development of the randomised vacation

queueing model:

$$\Delta(t) = \begin{cases} 0, & \text{if the server is idle in the system at time } t \\ 1, & \text{if the server is busy at time } t \\ 2, & \text{if the server is under repair at time } t \\ 3, & \text{if the server is on the 1st vacation at} \\ & \text{time } t \\ \vdots & \\ j+2, & \text{if the server is on the } j\text{th vacation} \\ & \text{at time } t \\ \vdots & \end{cases}$$

Thus, the supplementary variables  $S^-(t)$ ,  $R^-(t)$  and  $V_j^-(t)$  are introduced in order to obtain a trivariate Markov process  $\{N(t), \Delta(t), \delta(t)\}$ , where  $\delta(t) = 0$  if  $\Delta(t) = 0$ ,  $\delta(t) = S^-(t)$  if  $\Delta(t) = 1$ ,  $\delta(t) = R^-(t)$  if  $\Delta(t) = 2$ , and  $\delta(t) = V_j^-(t)$  if  $\Delta(t) = j + 2$  ( $j = 1, 2, \dots$ ).

Furthermore, let us define the following probabilities:

$$\begin{aligned} P_0(t) &= P_r\{N(t) = 0, \delta(t) = 0\}, \\ P_n(x, t)dx &= P_r\{N(t) = n, \delta(t) = S^-(t); \\ & \quad x < S^-(t) \leq x + dx\}, \quad x > 0, n \geq 1, \\ Q_n(x, y, t)dx &= P_r\{N(t) = n, \delta(t) = R^-(t); \\ & \quad y < R^-(t) \leq y + dy | S^-(t) = x\}, \\ & \quad (x, y) > 0, n \geq 1, \\ \Omega_{j,n}(x, t)dx &= P_r\{N(t) = n, \delta(t) = V^-(t); \\ & \quad x < V_j^-(t) \leq x + dx\}, \\ & \quad x > 0, n \geq 0, j = 1, 2, \dots \end{aligned}$$

In the steady state, we can set the following limit probabilities:

$$P_0 = \lim_{t \rightarrow \infty} P_0(t); \quad P_n(x) = \lim_{t \rightarrow \infty} P_n(x, t); \\ Q_n(x, y) = \lim_{t \rightarrow \infty} Q_n(x, y, t) \quad \text{and} \quad \Omega_{j,n}(x) = \lim_{t \rightarrow \infty} \Omega_{j,n}(x, t).$$

According to Cox (1955), the Kolmogorov forward equations that govern the system under steady-state conditions can be written as follows:

$$\lambda P_0 = p \sum_{j=1}^\infty \int_0^\infty \Omega_{j,0}(x) \omega(x) dx, \tag{3}$$

$$\begin{aligned} \frac{d}{dx} P_n(x) + [\lambda + \alpha + \mu(x)] P_n(x) \\ = \lambda \sum_{k=1}^{n-1} \chi_k P_{n-k}(x) + \int_0^\infty Q_n(x, y) \eta(y) dy, \\ x > 0, y > 0, n \geq 1, \end{aligned} \tag{4}$$

$$\begin{aligned} & \frac{d}{dy} Q_n(x, y) + [\lambda + \eta(y)]Q_n(x, y) \\ &= \lambda \sum_{k=1}^{n-1} \chi_k Q_{n-k}(x, y), \quad x > 0, y > 0, n \geq 1, \end{aligned} \quad (5)$$

$$\frac{d}{dx} \Omega_{j,0}(x) + [\lambda + \omega(x)]\Omega_{j,0}(x) = 0, \quad x > 0, j = 1, 2, \dots, \quad (6)$$

$$\begin{aligned} & \frac{d}{dx} \Omega_{j,n}(x) + [\lambda + \omega(x)]\Omega_{j,n}(x) \\ &= \lambda \sum_{k=1}^{n-1} \chi_k \Omega_{j,n-k}(x), \quad x > 0, n \geq 1, j = 1, 2, \dots \end{aligned} \quad (7)$$

We solve the above equations by means of the following boundary conditions at  $x = 0$ .

$$\begin{aligned} P_n(0) &= \sum_{j=1}^{\infty} \int_0^{\infty} \Omega_{j,n}(x)\omega(x)dx + \int_0^{\infty} P_{n+1}(x)\mu(x)dx \\ &+ \lambda \chi_n P_0, \quad n \geq 1, \end{aligned} \quad (8)$$

$$\Omega_{1,n}(0) = \begin{cases} \int_0^{\infty} P_1(x)\mu(x)dx, & n = 0, \\ 0, & n \geq 1. \end{cases} \quad (9)$$

$$\Omega_{j,n}(0) = \begin{cases} \bar{p} \int_0^{\infty} \Omega_{j-1,n}(x)\omega(x)dx, & n = 0, j = 2, 3, \dots \\ 0, & n \geq 1, j = 2, 3, \dots \end{cases} \quad (10)$$

and at  $y = 0$  and fixed values of  $x$

$$Q_n(x, 0) = \alpha P_n(x), \quad x > 0, n \geq 1, \quad (11)$$

and the normalisation condition

$$\begin{aligned} P_0 + \left\{ \sum_{n=1}^{\infty} \int_0^{\infty} P_n(x)dx + \int_0^{\infty} \int_0^{\infty} Q_n(x, y)dx dy \right\} \\ + \sum_{j=1}^{\infty} \left[ \sum_{n=0}^{\infty} \int_0^{\infty} \Omega_{j,n}(x)dx \right] = 1. \end{aligned} \quad (12)$$

We define the PGFs for  $\{\chi_n\}$ ,  $\{P_n(\cdot)\}$ ,  $\{Q_n(\cdot)\}$  and  $\{\Omega_{j,n}(\cdot)\}$ , as follows:

$$X(z) = \sum_{n=1}^{\infty} z^n \chi_n, \quad |z| \leq 1, \quad P(x; z) = \sum_{n=1}^{\infty} z^n P_n(x), \quad |z| \leq 1,$$

$$Q(x, y; z) = \sum_{n=1}^{\infty} z^n Q_n(x, y), \quad |z| \leq 1,$$

$$\Omega_j(x; z) = \sum_{n=0}^{\infty} z^n \Omega_{j,n}(x), \quad |z| \leq 1, j = 1, 2, \dots$$

Now multiplying (4) by  $z^n$  ( $n = 1, 2, 3, \dots$ ) and then adding the equations up term by term, we obtain

$$\begin{aligned} & \frac{\partial P(x; z)}{\partial x} + [a(z) + \mu(x) + \alpha]P(x; z) \\ &= \int_0^{\infty} \eta(y)Q(x, y; z)dy, \end{aligned} \quad (13)$$

where  $a(z) = \lambda(1 - X(z))$ .

Similarly, proceeding in the usual manner with (5)–(8), we have

$$\frac{\partial Q(x, y; z)}{\partial y} + [a(z) + \eta(y)]Q(x, y; z) = 0, \quad (14)$$

$$\frac{\partial \Omega_j(x; z)}{\partial x} + [a(z) + \omega(x)]\Omega_j(x; z) = 0, \quad (15)$$

and

$$\begin{aligned} P(0; z) &= \sum_{j=1}^{\infty} \int_0^{\infty} \Omega_j(x; z)\omega(x)dx \\ &+ \frac{1}{z} \int_0^{\infty} P(x; z)\mu(x)dx + \lambda X(z)P_0 \\ &- \sum_{j=1}^{\infty} \Omega_j(0; z) - \lambda P_0, \end{aligned} \quad (16)$$

where  $x > 0$ .

Solving the partial differential Equations (13)–(15), we obtain

$$P(x; z) = P(0; z)[1 - S(x)]e^{-A(z)x}, \quad (17)$$

$$Q(x, y; z) = Q(x, 0; z)[1 - R(y)]e^{-a(z)y} \quad (18)$$

and

$$\Omega_j(x; z) = \Omega_j(0; z)[1 - V(x)]e^{-a(z)x}, \quad j = 1, 2, \dots, \quad (19)$$

where  $A(z) = a(z) + \alpha(1 - R^*(a(z)))$ .

Solving the differential Equation (6) yields

$$\Omega_{j,0}(x) = \Omega_{j,0}(0)[1 - V(x)]e^{-\lambda x}, \quad j = 1, 2, \dots \quad (20)$$

Now, (20) is multiplied by  $\omega(x)$  on both sides and integrating with over  $x$  from 0 to  $\infty$ ; we then have

$$\int_0^{\infty} \Omega_{j,0}(x)\omega(x)dx = \Omega_{j,0}(0)\gamma_0, \quad (21)$$

where  $\gamma_0 = V^*(\lambda)$ . Using (21) and (10), we can recursively obtain

$$\Omega_{j,0}(0) = (\bar{p}\gamma_0)^{j-1} \Omega_{1,0}(0), \quad j = 2, 3, \dots \quad (22)$$

Using (22) in (3), and after doing some algebraic manipulation, we have

$$\Omega_{1,0}(0) = \frac{\lambda P_0(1 - \bar{p}\gamma_0)}{p\gamma_0}. \tag{23}$$

From (22) and (23), we finally obtain

$$\Omega_j(0; z) = \Omega_{j,0}(0) = (\bar{p}\gamma_0)^{j-1} \frac{\lambda P_0(1 - \bar{p}\gamma_0)}{p\gamma_0}, \quad j = 1, 2, \dots \tag{24}$$

Integrating (20) with respect to  $x$  from 0 to  $\infty$ , we obtain

$$\Omega_{j,0} = \Omega_{j,0}(0) \int_0^\infty [1 - V(x)]e^{-\lambda x} dx = \frac{1}{\lambda} \Omega_{j,0}(0)(1 - \gamma_0). \tag{25}$$

Equations (24) and (25) yield

$$\Omega_{j,0} = (\bar{p}\gamma_0)^{j-1} \frac{P_0(1 - \bar{p}\gamma_0)(1 - \gamma_0)}{p\gamma_0}, \quad j = 1, 2, \dots \tag{26}$$

Note that  $\Omega_{j,0}$  represents the steady-state probability that there are no customers in the system when the server is on the  $j$ th vacation. Let  $\Omega_0$  be the probability that no customers appear in the system when the server is on vacation. We then have

$$\Omega_0 = \sum_{j=1}^\infty \Omega_{j,0} = \frac{P_0(1 - \gamma_0)}{p\gamma_0}. \tag{27}$$

Substituting (17), (19) and (24) into (16), we get

$$P(0; z) = \frac{\lambda P_0 V^*(a(z))}{p\gamma_0} + \frac{P(0; z)S^*(A(z))}{z} + \lambda X(z)P_0 - \sum_{j=1}^\infty \Omega_{j,0}(0) - \lambda P_0. \tag{28}$$

Solving  $P(0; z)$  from (28) and using (24) yields

$$P(0; z) = \frac{\lambda P_0 z \left( \frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z) \right)}{z - S^*(A(z))}. \tag{29}$$

It follows from (17) and (29) that

$$P(x; z) = \frac{\lambda P_0 z \left( \frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z) \right)}{z - S^*(A(z))} \times [1 - S(x)]e^{-A(z)x}, \tag{30}$$

which leads to

$$P(z) = \int_0^\infty P(x; z) dx = \frac{\lambda P_0 z \left( \frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z) \right)}{z - S^*(A(z))} \times \frac{1 - S^*(A(z))}{A(z)}. \tag{31}$$

Inserting (18) with the boundary conditions (11),  $Q(x, 0; z)$  can be expressed as

$$Q(x, 0; z) = \alpha P(x; z), \tag{32}$$

Inserting (17) and (32) into (18) yields

$$Q(x, y; z) = \alpha P(0; z)[1 - S(x)]e^{-A(z)x}[1 - R(y)]e^{-a(z)y}. \tag{33}$$

Inserting (29) into (33) we obtain

$$Q(x, y; z) = \frac{\alpha \lambda P_0 z \left( \frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z) \right)}{z - S^*(A(z))} \times [1 - S(x)]e^{-A(z)x}[1 - R(y)]e^{-a(z)y}, \tag{34}$$

Calculating the double integral  $\int_0^\infty \int_0^\infty W(x, y; z) dx dy$  and  $\int_0^\infty \int_0^\infty Q(x, y; z) dx dy$  we finally obtain

$$Q(z) = \frac{\lambda P_0 z \left( \frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z) \right)}{z - S^*(A(z))} \times \frac{1 - S^*(A(z))}{A(z)} \times \frac{\alpha [1 - R^*(a(z))]}{a(z)} = P(z) \times \frac{\alpha [1 - R^*(a(z))]}{a(z)}. \tag{35}$$

Using (19) and (24) results in

$$\Omega_j(z) = (\bar{p}\gamma_0)^{j-1} \frac{P_0(1 - \bar{p}\gamma_0)(V^*(a(z)) - 1)}{p\gamma_0(X(z) - 1)}, \quad j = 1, 2, 3, \dots \tag{36}$$

The unknown constant  $P_0$  can be determined by using the normalisation condition (12), which is equivalent to  $P_0 + P(1) + Q(1) + W(1) + \sum_{j=1}^\infty \Omega_j(1) = 1$ . Thus, we obtain:

$$P_0 = \frac{1 - \rho_H}{1 + \frac{\lambda E[V]}{p\gamma_0}}, \tag{37}$$

where  $\rho_H = \rho(1 + \alpha E[R])$  and  $\rho = \lambda E[X]E[S]$ .

Note that Equation (37) represents the steady-state probability that the server is idle but available in the system. Also from Equation (37), we have  $\rho_H < 1$ , which is the necessary and sufficient condition under which steady-state solution exists.

Let  $\Phi(z) = P_0 + P(z) + Q(z) + \sum_{j=1}^\infty \Omega_j(z)$  be the PGF of the system size distribution at a stationary point of time, we then have

$$\Phi(z) = \frac{(1 - \rho_H)S^*(A(z))(z - 1)}{z - S^*(A(z))} \times \frac{(V^*(a(z)) - 1) + (X(z) - 1)p\gamma_0}{(\lambda E[V] + p\gamma_0)(X(z) - 1)}. \tag{38}$$

**Remark 1:** By setting  $p = 1$  and  $\alpha = 0$ , our model can be simplified to the M/G/1 queueing system with single vacation.  $\Phi(z)$  can be rewritten as

$$\left(\frac{(1-\rho)(1-z)S^*(a(z))}{S^*(a(z))-z}\right)\left(\frac{1-V^*(a(z))+\gamma_0(1-X(z))}{(1-X(z))(\lambda E[V]+\gamma_0)}\right),$$

which confirms the results in Section 6 of Choudhury's (2002) system.  $\square$

**3.2. The expected number of customers in the system and the expected waiting time**

In (38), we evaluate  $\frac{d}{dz}\Phi(z)|_{z=1}$  by using L'hopital's rule which leads to the expected number of customers in the system, which  $\bar{L}_s$  is given by

$$\begin{aligned} \bar{L}_s = \rho_H + & \frac{\lambda E[X(X-1)]E[G] + (\lambda E[X](1 + \alpha E[R]))^2 E[S^2]}{2(1-\rho_H)} \\ & + \frac{\alpha(\lambda E[X])^2 E[R^2]E[S]}{2(1-\rho_H)} + \frac{\lambda^2 E[X]E[V^2]}{2(\lambda E[V] + p\gamma_0)}. \end{aligned} \quad (39)$$

By using Little's formula, we obtain the expected waiting time in the queue,  $\bar{W}_q$ , which is given by

$$\begin{aligned} \bar{W}_q = & \frac{\alpha(E[D] + E[R])}{\mu} \\ & + \frac{E[X(X-1)]E[G] + \lambda(E[X](1 + \alpha E[R]))^2 E[S^2]}{2E[X](1-\rho_H)} \\ & + \frac{\alpha\lambda E[X](E[R^2]E[S])}{2(1-\rho_H)} + \frac{\lambda E[V^2]}{2(\lambda E[V] + p\gamma_0)}. \end{aligned} \quad (40)$$

**Remark 2:** Suppose that we have  $p = 1$  and  $\alpha = 0$ ; then if we set  $\Pr(X = 1) = 1$ , our model can be reduced to the ordinary M/G/1 queueing system with a single vacation. It follows from (40) that the expected waiting time in the system is given by

$$\bar{W}_s = \frac{\lambda E[V^2]}{2(\lambda E[V] + \gamma_0)} + \frac{\lambda E[S^2]}{2(1-\rho)},$$

which is in accordance with Takagi's system (1991, sec. 2.2, p. 126).

**Remark 3:** Allowing for  $p = 0$  and  $\alpha = 0$ , our model reduces to the ordinary M<sup>[x]</sup>/G/1 queueing system with multiple vacations. From (40), we have the expected waiting time in the system as

$$\bar{W}_s = \frac{\lambda E[V^2]}{2\lambda E[V]} + \frac{\lambda E[X]E[S^2]}{2(1-\rho)} + \frac{E[X(X-1)]E[S]}{2E[X](1-\rho)},$$

which agrees with those of Takagi's system (1991), or Ke and Chu's system (2006) for multiple vacations (i.e.  $J = \infty$ ).

**3.3. Queue size distribution at a departure epoch**

We derive the PGF of the steady-state distribution of the number of units in the queue at a departure epoch for the M<sup>[x]</sup>/(G<sub>1</sub>,G<sub>2</sub>)/1/VAC( $\infty$ ) queueing system. Following the arguments by Wolff (1982), we state that a departing customer will see  $l$  customers in the queue just after a departure if and only if there were  $(l + 1)$  customers in the queue just before the departure. Thus we may write

$$\Phi_l^+ = K_0 \int_0^\infty \mu(x)P_{l+1}(x)dx, \quad l = 0, 1, \dots, \quad (41)$$

where  $\Phi_l^+ = \Pr$  (A departing customer will see  $l$  customers in the queue), and  $K_0$  is the normalising constant. The PGF of  $\Phi_l^+$  is given by

$$\Phi^+(z) = K_0 \times \frac{\lambda P_0 \left(\frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z)\right) S^*(A(z))}{z - S^*(A(z))}. \quad (42)$$

Using the normalisation condition  $\Phi^+(1) = 1$  results in

$$K_0 = \frac{1 - \rho_H}{\lambda P_0 E[X] \left(1 + \frac{\lambda E[V]}{p\gamma_0}\right)}, \quad (43)$$

which leads to the PGF of the departure point queue size distribution as

$$\Phi^+(z) = \frac{(1 - \rho_H) \left(\frac{V^*(a(z))-1}{p\gamma_0} - 1 + X(z)\right) S^*(A(z))}{E[X] \left(1 + \frac{\lambda E[V]}{p\gamma_0}\right) (z - S^*(A(z)))}. \quad (44)$$

From (44), it is obvious that  $\Phi^+(z)$  can be decomposed into two independent terms:

$$\Phi^+(z) = \frac{1 - X(z)}{E[X](1 - z)} \times \Phi(z). \quad (45)$$

It should be noted that the departure point queue size distribution given by Equation (45) can be decomposed into two independent random variables: one (the first term) is the number of customers placed before a tagged customer in a batch in which the tagged customer arrives and the other (the second term) is the stationary system size of the M<sup>[x]</sup>/(G<sub>1</sub>,G<sub>2</sub>)/1/VAC( $\infty$ ) queueing system.

**Remark 4:** Suppose we let  $p = 1$  and  $\alpha = 0$  in (44). Our system can then be reduced to the M<sup>[x]</sup>/G/1 queue with single vacation. Equation (44) can be rewritten as

$$\begin{aligned} \Phi^+(z) = & \left(\frac{1 - V^*(a(z)) + V^*(\lambda)(1 - X(z))}{E[X](\lambda E[V] + V^*(\lambda))(1 - z)}\right) \\ & \times \left(\frac{(1 - \rho)(1 - z)S^*(a(z))}{S^*(a(z)) - z}\right) \\ = & \beta(z)\Phi^+(z; M^{[x]}/G/1), \end{aligned}$$



where

$$\beta(z) = \frac{1 - V^*(a(z)) + V^*(\lambda)(1 - X(z))}{E[X](\lambda E[V] + \gamma_0)(1 - z)},$$

and

$$\Phi^+(z; M^{[x]}/G/1) = \frac{(1 - \rho)(1 - z)S^*(a(z))}{S^*(a(z)) - z}$$

is the PGF of the stationary queue size distribution of an ordinary  $M^{[x]}/G/1$  queue. This is in accordance with the stochastic decomposition property demonstrated in Choudhury's (2002) system.

**Remark 5:** If we consider the ordinary  $M/G/1$  system with server breakdown without vacations (i.e.  $\Pr(X=1)=1$  and  $\Pr(V=0)=1$ ), Equation (44) can be reduced to

$$\Phi^+(z) = \frac{(1 - \rho_H)(1 - z)S^*(A(z))}{S^*(A(z)) - z},$$

which is consistent with Choudhury and Tadj (2009) for no optional service and no delayed repair.

### 3.4. System size distribution at busy period initiation epoch

First, we define  $\varphi_n$  ( $n = 1, 2, \dots$ ) as the steady-state probability such that an arbitrary (tagged) customer finds  $n$  customers in the system at the busy initiation epoch (or completion epoch of the idle period). This implies that  $t_l$  ( $l = 0, 1, 2, \dots$ ) are the initiation epochs of the busy period and  $N(t_l)$  is the number of customers in the system at the time instant  $t_l$ ; then we have

$$\varphi_n = \lim_{l \rightarrow \infty} \Pr(N(t_l) = n), \quad n = 1, 2, \dots$$

Conditional upon the number of customers who arrive during the first vacation, from the concept of *Poisson Arrivals See Time Average* (Wolff 1982), we have the following steady-state equation:

$$\begin{aligned} \varphi_n &= (1 + \bar{p}\gamma_0 + \bar{p}^2\gamma_0^2 + \dots) \sum_{k=1}^n \gamma_k \chi_n^{(k)*} \\ &\quad + p(\gamma_0 + \bar{p}\gamma_0^2 + \dots) \chi_n, \\ &= \sum_{m=0}^{\infty} (\bar{p}\gamma_0)^m \sum_{k=1}^n \gamma_k \chi_n^{(k)*} + \left( p \sum_{m=0}^{\infty} \bar{p}^m \gamma_0^{m+1} \right) \chi_n, \end{aligned} \quad (46)$$

where  $\chi_n^{(k)*} = \Pr(X_1 + X_2 + \dots + X_k = n)$  is the  $k$ -fold convolution of  $\chi_n$ ,  $\chi_n^{(0)}$  is defined to be 1 and  $\gamma_k = \Pr(k \text{ batches arrive during a vacation time})$ .

Now multiplying (46) by appropriate powers of  $z$  and summing over all possible values of  $n$ , we get the PGF of  $\{\varphi_n\}$ :

$$\varphi(z) = \frac{V^*(a(z)) - \gamma_0}{1 - \bar{p}\gamma_0} + \frac{p\gamma_0}{1 - \bar{p}\gamma_0} X(z), \quad (47)$$

which leads to

$$E[\varphi] = \frac{\lambda E[X]E[V]}{1 - \bar{p}\gamma_0} + \frac{p\gamma_0}{1 - \bar{p}\gamma_0} E[X]. \quad (48)$$

Equation (47) represents the PGF of the number of customers in the system at the completion epoch of the idle period; this is equivalent to the PGF of the system size distribution at the busy period initiation epoch.

**Remark 6:** Substituting  $p = 1$  and  $\alpha = 0$  into (47), our system can be reduced to the ordinary  $M^{[x]}/G/1$  single vacation policy queue. In this case,  $\varphi(z)$  can be rewritten as

$$\varphi(z) = V^*(a(z)) + \gamma_0[X(z) - 1],$$

which is in accordance with Choudhury's (2002) system.

**Remark 7:** As  $p = 0$  and  $\alpha = 0$ , our system can be simplified to the  $M^{[x]}/G/1$  queue and with multiple vacations. In this case,  $\varphi(z)$  can be rewritten as

$$\varphi(z) = \frac{V^*(a(z)) - \gamma_0}{1 - \gamma_0},$$

which is consistent with Ke and Chu (2006) for  $J = \infty$ .

### 3.5. Expected length of the completion period and idle period

Let  $H^*(\theta)$  and  $I^*(\theta)$  represent the LST of the completion period (including busy period and breakdown period) and idle period for the  $M^{[x]}/(G_1, G_2)/1/VAC(\infty)$  queueing system. Utilising the arguments by Takagi (1991, sec. 2.2) and Tang (1997),  $H^*(\theta)$  and  $I^*(\theta)$  can be expressed as

$$\begin{aligned} H^*(\theta) &= \frac{V^*[\lambda(1 - X(H_0^*(\theta)))] - \gamma_0}{1 - \bar{p}\gamma_0} \\ &\quad + \left( \frac{p\gamma_0}{1 - \bar{p}\gamma_0} \right) X(H_0^*(\theta)), \end{aligned} \quad (49)$$

and

$$I^*(\theta) = \frac{V^*(\theta) - V^*(\theta + \lambda)}{1 - \bar{p}V^*(\theta + \lambda)} + \left( \frac{pV^*(\theta + \lambda)}{1 - \bar{p}V^*(\theta + \lambda)} \right) \left( \frac{\lambda}{\lambda + \theta} \right), \quad (50)$$

where  $H_0^*(\theta) = G^*[\theta + \lambda - \lambda X(H_0^*(\theta))]$  is the LST of the completion period in the ordinary  $M^{[x]}/G/1$  queueing model with an unreliable server.

Now, we further define the following:

- $E[H]$  the expected length of completion period,
- $E[I]$  the expected length of idle period,
- $E[C]$  the expected length of busy cycle.

Using (49) and (50), this elicits:

$$E[H] = \left( \frac{\lambda E[V]}{1 - \bar{p}\gamma_0} + \frac{p\gamma_0}{1 - \bar{p}\gamma_0} \right) \left( \frac{E[X]E[G]}{1 - \rho_H} \right), \quad (51)$$

$$E[I] = \frac{E[V]}{(1 - \bar{p}\gamma_0)} + \frac{p\gamma_0}{\lambda(1 - \bar{p}\gamma_0)}, \quad (52)$$

and

$$E[C] = E[H] + E[I] = \left( \frac{\lambda E[V]}{1 - \bar{p}\gamma_0} + \frac{p\gamma_0}{(1 - \bar{p}\gamma_0)} \right) \cdot \frac{1}{\lambda(1 - \rho_H)}. \quad (53)$$

**Remark 8:** Letting  $\Pr(X = 1) = 1$  and  $\Pr(V = 0) = 1$ , Equation (51) can be rewritten as

$$E[H] = \frac{E[S](1 + \alpha E[R])}{1 - \rho_H},$$

which is in accordance with Choudhury and Tadj (2009) for no optional service and no delayed repair.

#### 4. Reliability indices

In this section, we develop two main reliability indices of the presented model, namely, the system availability and failure frequency under the steady-state conditions. Let us define  $Av(t)$  as the system availability at time  $t$ , that is, the probability that the server is working for a customer, being on vacation or remaining idle in the system. The steady-state availability of the server is given by  $Av = \lim_{t \rightarrow \infty} Av(t)$ .

**Theorem 1:** *The steady-state availability of the server is given by*

$$Av = \rho + \frac{1 - \rho_H}{1 + \frac{\lambda E[V]}{p\gamma_0}}. \quad (54)$$

**Proof:** We first consider the following equation:

$$Av = P_0 + \int_0^\infty P(x, 1) dx = P_0 + \lim_{z \rightarrow 1} P(z),$$

then using Equations (31) and (37), we get the result (54).  $\square$

**Theorem 2:** *The steady-state failure frequency of the server is given by*

$$M_f = \alpha \rho. \quad (55)$$

**Proof:** Following the argument by Li, Shi, and Chao (1997), we obtain

$$M_f = \alpha \int_0^\infty P(x, 1) dx,$$

then using (31), we can get the result (55).  $\square$

#### 5. The cost effectiveness maximisation model

In this section, we develop the cost effectiveness, which is defined as (availability)/(the expected out of pocket cost rate) to be an alternative cost criterion which reflects the efficiency per dollar outlay (Park and Park 1986). This criterion is useful for the effective allocation of available funds. That is, the availability is evaluated under a given investment (restricted money). This criterion is helpful for management/practitioner in practical use such that benefits are obtained from the investment.

Let  $Cr$  be the expected cost rate per busy cycle, then

$$Cr = \frac{C_s}{E[C]}, \quad (56)$$

where  $C_s$  is the out of pocket cost per cycle. The cost effectiveness is defined as

$$Ce = \frac{Av}{Cr}, \quad (57)$$

which represents the effective use for a given cost (money). This is why we maximise Equation (57).

#### 6. Numerical illustration

The first purpose of this section is to study the effects of various parameters on the system characteristics such as the expected number of customers/jobs in the system and ( $\bar{L}_s$ ) the expected length of completion period ( $E[H]$ ). An example (as the wafer fabrication scenario mentioned in Section 1) is provided to visualise the numerical investigation:

- The jobs arrive in batches according to a compound Poisson process with an arrival rate of  $\lambda = 0.4$ .
- The stepper machine may be interrupted due to some unpredictable accident with a Poisson breakdown rate of  $\alpha = 0.05$ .
- The number of jobs ( $X$ ) belonging to each arrival obeys the Geometric distribution with the parameter set to 0.5 (denoted by  $Geo(0.5)$ ).
- The probability that the machine will go on preventative maintenance when no jobs are queued in the system at the end of a maintenance routine is  $1 - p = 0.5$ .
- The service/process time of the machine in the workstation of photolithography area follows a 4-stage Erlang distribution (denoted by  $E_4$  with a mean  $E[S] = 0.5$ ).
- The time taken to perform the preventative maintenance for the stepper machine is distributed according to an exponential distribution (denoted by  $M$  with a mean  $E[V] = 1.0$ ).

- The repair time of the broken-down machine in the workstation of the photolithography area has an exponential distribution (denoted by  $M$  with a mean  $E[R] = 0.5$ ).

In our first set of numerical investigations, we use the above parameters, and vary  $\alpha$  from 0.0 to 1.0 and  $p$  from 0.0 to 1.0 to investigate the effects of different values of  $p$  and  $\alpha$  on  $\bar{L}_s$  and  $E[H]$ . From Figures 1 and 2, we observe that (1)  $\bar{L}_s$  and  $E[H]$  slowly decrease as  $p$  increases for fixed  $\alpha$ ; and (2)  $\bar{L}_s$  and  $E[H]$  increase as  $\alpha$  increases for fixed  $p$ . As expected, a larger  $p$  implies that the number of jobs (in the system) and the completion period become smaller, due to ongoing preventative maintenance having a lower probability. In addition, when the breakdown rate increases, the server is often unable to provide service for the jobs, which leads to the expected number of jobs in the

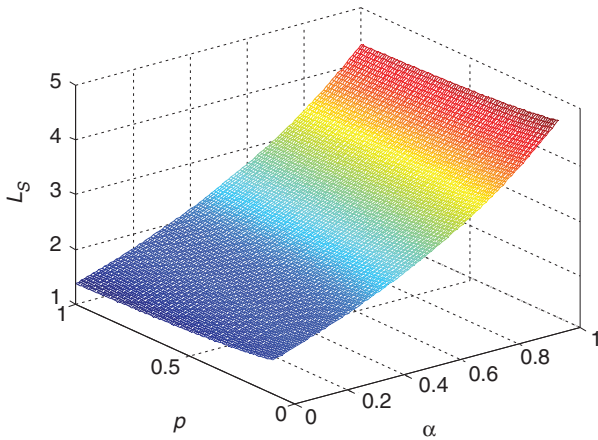


Figure 1. The expected number of customers in the system for different values of  $p$  and  $\alpha$ .

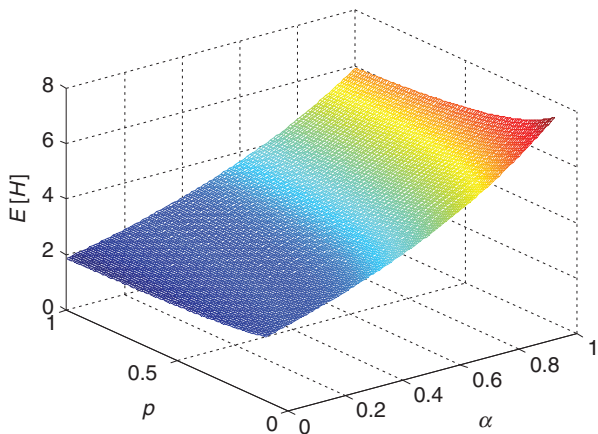


Figure 2. The expected lengths of completion period for different values of  $p$  and  $\alpha$ .

system becoming larger and the completion period longer.

A second numerical investigation deals with the effects of various service time distributions and different service rates on  $\bar{L}_s$ . The service time distributions considered are the exponential ( $M$ ), 2-stage Erlang ( $E_2$ ), and 2-stage hyper-exponential ( $H_2$ ) distributions. The effects of the service time distributions and different service rates on  $\bar{L}_s$  are shown in Table 1. From Table 1, the comparison of  $\bar{L}_s$  for  $M$ ,  $E_2$  and  $H_2$ , shows the results of service rate changes in accordance with the selected values 1.0, 5.0 and 10.0: (i) the three service distributions ordered by their relative magnitudes on  $\bar{L}_s$  produce  $H_2 > M \geq E_2$ ; and (ii) the expected number of jobs increases as  $1/E[S]$  decreases. This suggests that the larger the  $CV_{\text{service}}$  (coefficient of variation of the service time), the larger  $\bar{L}_s$  becomes.

For the third set of numerical investigations, we study the effects of changing the vacation time distribution and the vacation rate on  $\bar{L}_s$ , the results of which are summarised in Table 2. The comparison of  $\bar{L}_s$  for the three vacation time distributions  $M$ ,  $E_2$ , and  $H_2$  shows the results when the vacation rate changes from the selected values 1.0, 5.0 and 10.0: (i) the three vacation distributions by their relative magnitudes on  $\bar{L}_s$  yield  $H_2 > M \geq E_2$ ; and (ii) the expected number of jobs increases as  $1/E[V]$  decreases.

Table 1. System characteristics for different service time distributions and service rates.

$1/E[S]$		$S \equiv M$	$S \equiv E_2$	$S \equiv H_2$
1.0	$\bar{L}_s$	9.83	8.84	26.66
5.0	$\bar{L}_s$	0.53	0.53	0.83
10.0	$\bar{L}_s$	0.32	0.31	0.43

Notes:  $\lambda = 0.4$ ,  $\alpha = 0.05$ ,  $X \equiv Geo(0.5)$ ,  $p = 0.5$ ,  $V \equiv M$  with  $E[V] = 1$ ,  $R \equiv M$  with  $E[R] = 0.5$ .

Table 2. System characteristics for different vacation distributions and vacation rates.

$1/E[V]$		$V \equiv M$	$V \equiv E_2$	$V \equiv H_2$
1.0	$\bar{L}_s$	1.75	1.48	2.57
5.0	$\bar{L}_s$	1.34	1.33	1.40
10.0	$\bar{L}_s$	1.32	1.32	1.34

Notes:  $\lambda = 0.4$ ,  $\alpha = 0.05$ ,  $X \equiv Geo(0.5)$ ,  $p = 0.5$ ,  $S \equiv E_4$  with  $E[S] = 0.5$ ,  $R \equiv M$  with  $E[R] = 0.5$ .

For the fourth set of numerical investigations, we examine the effects of various repair time distributions and different repair rates on  $\bar{L}_s$ . The effects of the different repair time distributions and repair rates on  $\bar{L}_s$  are shown in Table 3. From Table 3, the comparison of  $\bar{L}_s$  for the three repair time distributions  $M$ ,  $E_2$  and  $H_2$ , showed the results when the repair rate changes from the selected values 1.0, 5.0 and 10.0: (i) the three repair distributions ordered by their relative magnitudes on  $\bar{L}_s$  produce  $H_2 > M \geq E_2$  and (ii) the expected number of jobs decreases as  $1/E[R]$  increases. It is interesting that the larger  $CV_{\text{repair}}$  (coefficient of variation of the repair time), the larger  $\bar{L}_s$  becomes.

For the last set of numerical investigations, we assume that  $E[R] = 1$ , and investigate the effect of different values of  $p$  and vacation time distributions on  $\bar{L}_s$  and  $\bar{W}_s$ . Three vacation distributions with  $E[V] = 0.5$  are considered. Table 4 clearly shows that  $\bar{L}_s$  and  $\bar{W}_s$  decrease as  $p$  increases for the vacation time distributions:  $M$ ,  $E_2$  and  $H_2$ . This also reveals that when  $p$  changes from selected values 0.0, 0.5 and 1.0, the three vacation time distributions ordered by their relative magnitudes on  $\bar{L}_s$  and  $\bar{W}_s$  produce  $H_2 > M > E_2$ . This implies that the larger  $CV_{\text{vacation}}$

Table 3. System characteristics for different delay time distributions and repair rates.

$1/E[R]$		$R \equiv M$	$R \equiv E_2$	$R \equiv H_2$
1.0	$\bar{L}_s$	1.83	1.81	2.04
5.0	$\bar{L}_s$	1.71	1.71	1.74
10.0	$\bar{L}_s$	1.70	1.70	1.71

Notes:  $\lambda = 0.4$ ,  $\alpha = 0.05$ ,  $X \equiv \text{Geo}(0.5)$ ,  $p = 0.5$ ,  $S \equiv E_4$  with  $E[S] = 0.5$ ,  $V \equiv M$  with  $E[V] = 1$ .

Table 4. The expected system sizes  $\bar{L}_s$  and expected waiting time  $\bar{W}_s$  for different values of  $p$  and different vacation distributions.

$p$		$V \equiv M$	$V \equiv E_2$	$V \equiv H_2$
0.0	$\bar{L}_s$	2.97	1.97	4.65
	$\bar{W}_s$	3.71	2.46	5.81
0.5	$\bar{L}_s$	2.56	1.82	4.31
	$\bar{W}_s$	3.20	2.28	5.39
1.0	$\bar{L}_s$	2.31	1.74	4.03
	$\bar{W}_s$	2.89	2.18	5.04

Notes:  $\lambda = 0.4$ ,  $\alpha = 0.05$ ,  $X \equiv \text{Geo}(0.5)$ ,  $S \equiv E_4$  with  $E[S] = 0.5$ ,  $R \equiv M$  with  $E[R] = 1$ .

(coefficient of variation of the vacation time), the larger  $\bar{L}_s$  becomes.

Our numerical investigations indicate that (i) when all parameters are given, the impact of the service (or vacation) distribution on the system characteristics is not significant for large service rate (or vacation rate) and (ii) the vacation time distribution of the stepper machine has a much more significant effect on the system characteristics than  $p$  does.

The second purpose of this section is to examine the effects of various parameters on the cost rate and cost effectiveness discussed in the previous section. For convenience, the settings of system's parameters are given as follows:

- $C_s = 100$ .
- Geometric batch size with a mean  $E[X] = 1.0$ .
- Exponential service time with a mean  $E[S] = 0.5$ .
- Exponential repair time with a mean  $E[R] = 1.0$ .

For illustrative purposes, an exponential distribution is considered as the vacation time with a vacation rate of 2. The results of cost rate and effectiveness are

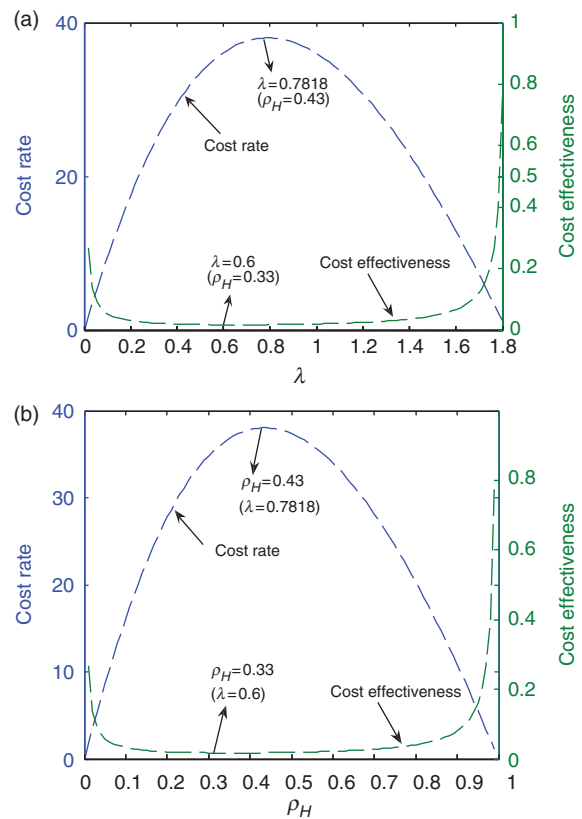


Figure 3. Cost rate and cost effectiveness for various values of (a)  $\lambda$  and (b)  $\rho_H$ .

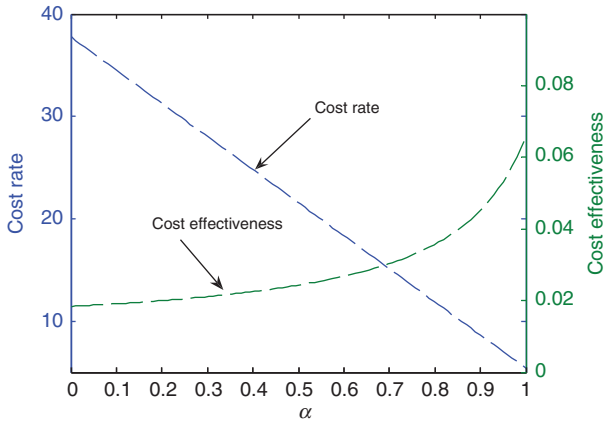


Figure 4. Cost rate and cost effectiveness for various values of  $\alpha$ .

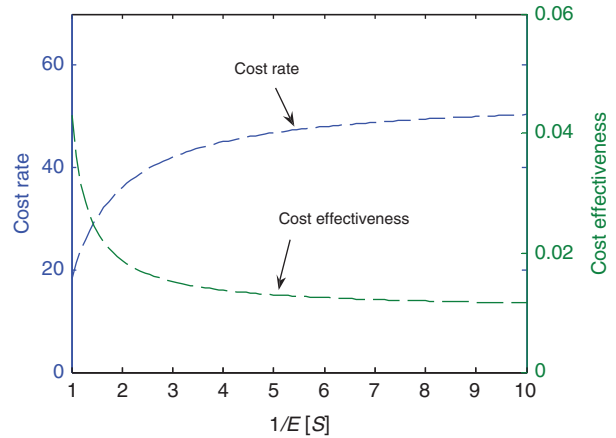


Figure 6. Cost rate and cost effectiveness for various values of service rate  $1/E[S]$ .

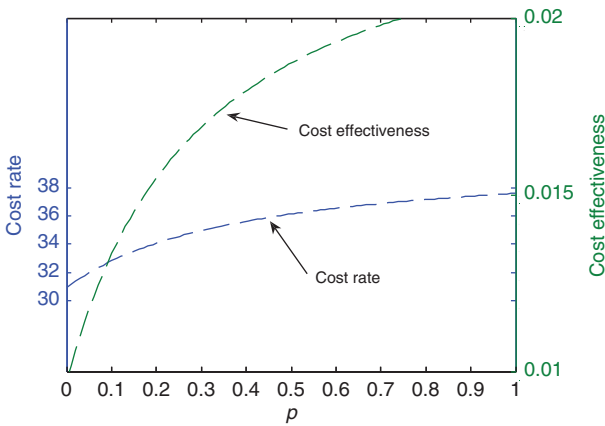


Figure 5. Cost rate and cost effectiveness for various values of  $p$ .

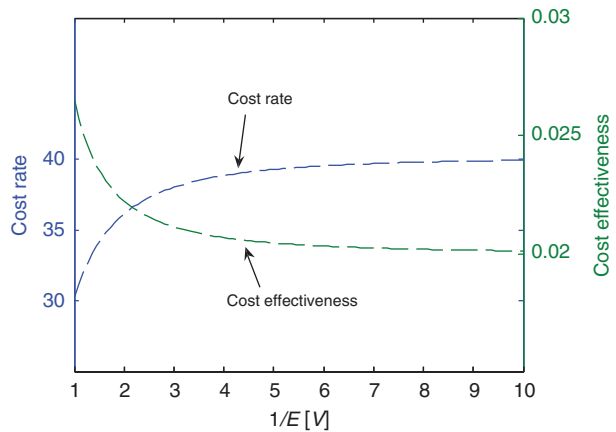


Figure 7. Cost rate and cost effectiveness for various values of vacation rate  $1/E[V]$ .

shown, respectively, in Figures 3–7 for the following five cases.

**Case 1:** We choose  $\alpha = 0.05$ ,  $p = 0.5$  and vary the values of  $\lambda$  from 0.0 to 1.8.

**Case 2:** We choose  $\lambda = 0.6$ ,  $p = 0.5$  and vary the values of  $\alpha$  from 0.0 to 0.05.

**Case 3:** We choose  $\lambda = 0.6$ ,  $\alpha = 0.05$  and vary the values of  $p$  from 0.0 to 1.0.

**Case 4:** We choose  $\lambda = 0.6$ ,  $\alpha = 0.05$  and vary the values of the service rate ( $1/E[S]$ ) from 0.0 to 1.0.

**Case 5:** We choose  $\lambda = 0.6$ ,  $\alpha = 0.05$  and vary the values of the vacation rate ( $1/E[V]$ ) from 0.0 to 1.0.

Figure 3(a) and (b) reveals that (i) the cost rate first increases ( $\rho_H \leq 0.43$ ) and then decreases ( $\rho_H > 0.43$ ) with increasing  $\lambda$ ; and (ii) the cost effectiveness first decreases ( $\rho_H \leq 0.33$ ), and then increases slightly

( $\rho_H > 0.33$ ) with increasing  $\lambda$  from 0.0 to 1.0. One can also observe that the cost rate has a minimum and cost effectiveness has a maximum when  $\rho_H$  closes to 0.0 or 1.0. From practice aspects, the cost rate reaches a minimum when the zero loading ( $\rho_H \rightarrow 0$ ) or overloading ( $\rho_H \geq 1$ ) of the stepper machine occurs; this is due to the busy cycle becoming longer. In contrast, the cost effectiveness in a busy period is large for the zero loading or overloading. One sees from Figures 4 and 5 that the cost rate decreases as  $\alpha$  increases but increases as  $p$  increases from 0.0 to 1.0. Moreover, we also observe that the cost effectiveness increases as  $\alpha$  or  $p$  increases from 0.0 to 1.0. It is interesting that for a given cost, the larger  $\alpha$  (or  $p$ ), the larger  $Ce$ . That is, when the cost per cycle is fixed, the stepper machine has larger cost effectiveness in a busy cycle for a larger breakdown rate (or maintenance probability). From Figures 6 and 7, we observe that the

cost rate increases as  $1/E[S]$ , or  $1/E[V]$  increases; on the other hand, the cost effectiveness decreases as  $1/E[S]$ , or  $1/E[V]$  increases. This implies that the stepper machine has a larger cost effectiveness in a busy cycle for smaller service rates (or vacation rates) when the cost is restricted.

## 7. Conclusions

In this article, we analysed an  $M^{[x]}/(G_1, G_2)/1/VAC(\infty)$  queueing system, in which the unreliable server applies a randomised vacation policy with multiple vacations in the idle period and a repair is requested when the server breaks down. The distributions of some important system characteristics for such a system were derived. A cost effectiveness maximisation model was also developed to demonstrate the efficiency per dollar outlay. Some extensive numerical computations were performed to study the effects of the system parameters on the system characteristics. This research presents an extension of the vacation model theory and the analysis of the model will provide a useful performance evaluation tool for more general situations arising in practical applications, for example, the wafer fabrication problems in manufacturing systems presented in Section 1.

## Acknowledgements

We thank the editors and referees for their thoughtful contributions to preliminary and final versions of the paper and for their constructive suggestions towards improving its presentation.

## Notes on contributors



**Jau-Chuan Ke** is a Professor of Applied Statistics at National Taichung Institute of Technology, Taiwan. He received his MS in Applied Mathematics at National Cheng-Kung University, Taiwan, and his PhD in Applied Mathematics at National Chung-Hsing University, Taiwan. His research interests include

stochastic processes, queueing theory, reliability and applied statistics.



**Kai Bin Huang** received his PhD degree in Industrial Engineering and Management from National Chiao-Tung University, Hsinchu, Taiwan, China. He is currently an IT engineer in the Engineering Information System Development Department, Ardentec Co. Ltd., Hsinchu. His research interests include queueing theory, quality engineering and applied statistics.



**Wen Lea Pearn** is a Professor of Operations Research and Quality Assurance at National Chiao Tung University, Taiwan. He received his PhD degree in Operations Research from the University of Maryland, College Park, Maryland, USA. He worked at AT&T Bell Laboratories as a Quality Research Scientist before joining National Chiao Tung University. His research interests include process capability, network optimisation, production management and queueing theory. His publications have appeared in the *Journal of the Royal Statistical Society*, *Journal of Quality Technology*, *Journal of Applied Statistics*, *Statistics*, *Journal of the Operational Research Society*, *European Journal of Operations Research*, *Operations Research Letters*, *Omega*, *Networks*, *International Journal of Production Research*, *Journal of Computational and Applied Mathematics*, *Quality and Reliability Engineering International*, *IEEE Transactions on Semiconductor Manufacturing* and *IEEE Transactions on Electronics Packaging Manufacturing* and others.

## References

- Baba, Y. (1986), 'On the  $M^{[x]}/G/1$  Queue With Vacation Time', *Operations Research Letters*, 5, 93–98.
- Choudhury, G. (2002), 'A Batch Arrival Queue with a Vacation Time Under Single Vacation Policy', *Computers and Operations Research*, 29, 1941–1955.
- Choudhury, G., Ke, J.-C., and Tadj, L. (2009), 'The  $N$ -Policy for an Unreliable Server With Delaying Repair and Two Phases of Service', *Journal of Computational and Applied Mathematics*, 231, 349–364.
- Choudhury, G., and Madan, K.C. (2005), 'A Two-Stage Batch Arrival Queueing System With a Modified Bernoulli Schedule Vacation Under  $N$ -Policy', *Mathematical and Computer Modelling*, 42, 71–85.
- Choudhury, G., and Paul, M. (2006), 'A Batch Arrival Queue With a Second Optional Service Channel Under  $N$ -Policy', *Stochastic Analysis and Applications*, 24, 1–21.
- Choudhury, G., and Tadj, L. (2009), 'An  $M/G/1$  Queue With Two Phases of Service Subject to the Server Breakdown and Delayed Repair', *Applied Mathematical Modelling*, 33, 2699–2709.
- Cox, D.R. (1955), 'The Analysis of Non-Markovian Stochastic Processes by the Inclusion of Supplementary Variables', *Proceedings of the Cambridge Philosophical Society*, 51, 433–441.
- Doshi, B.T. (1986), 'Queueing Systems With Vacations: A Survey', *Queueing Systems*, 1, 29–66.
- Goldratt, E.M., and Cox, J. (1992), *The Goal: A Process of Ongoing Improvement* (2nd ed.), New York: North River Press.
- Hur, S., Kim, J., and Kang, C. (2003), 'An Analysis of the  $M/G/1$  System With  $N$  and  $T$  Policy', *Applied Mathematical Modelling*, 27, 665–675.

- Ke, J.-C. (2001), 'The Control Policy of an  $M^{[x]}/G/1$  Queueing System with Server Startup and Two Vacation Types', *Mathematical Methods of Operations Research*, 54, 471–490.
- Ke, J.-C. (2003a), 'The Optimal Control of an  $M/G/1$  Queueing System With Server Vacations, Startup and Breakdowns', *Computers and Industrial Engineering*, 44, 567–579.
- Ke, J.-C. (2003b), 'Optimal Strategy Policy in Batch Arrival Queue With Server Breakdowns and Multiple Vacations', *Mathematical Methods of Operations Research*, 58, 41–56.
- Ke, J.-C. (2005), 'Modified T Vacation Policy for an  $M/G/1$  Queueing System With an Un-reliable Server and Startup', *Mathematical and Computer Modelling*, 41, 1267–1277.
- Ke, J.-C. (2006a), 'On  $M/G/1$  System Under NT Policies With Breakdowns, Startup and Closedown', *Applied Mathematical Modelling*, 30, 49–66.
- Ke, J.-C. (2006b), 'An  $M/G/1$  Queue Under Hysteretic Vacation Policy With an Early Startup and Un-Reliable Server', *Mathematical Methods of Operations Research*, 63, 357–369.
- Ke, J.-C. (2007), 'Operating Characteristic Analysis on the  $M^{[x]}/G/1$  System With a Variant Vacation Policy and Balking', *Applied Mathematical Modelling*, 31, 1321–1337.
- Ke, J.-C. (2008), 'Two Thresholds of a Batch Arrival Queueing System Under Modified T Vacation Policy With Startup and Closedown', *Mathematical Methods in the Applied Sciences*, 31, 229–247.
- Ke, J.-C., and Chu, Y.-K. (2006), 'A Modified Vacation Model  $M^{[x]}/G/1$  System', *Applied Stochastic Models in Business and Industry*, 22, 1–16.
- Kella, O. (1989), 'The Threshold Policy in the  $M/G/1$  Queue With Server Vacations', *Naval Research Logistics*, 36, 111–123.
- Lee, S.S., Lee, H.W., and Chae, K.C. (1995), 'Batch Arrival Queue With N Policy and Single Vacation', *Computers and Operations Research*, 22, 173–189.
- Lee, H.W., Lee, S.S., Park, J.O., and Chae, K.C. (1994), 'Analysis of  $M^{[x]}/G/1$  Queue With N Policy and Multiple Vacations', *Journal of Applied Probability*, 31, 467–496.
- Levy, Y., and Yechiali, U. (1975), 'Utilization of Idle Time in an  $M/G/1$  Queueing System', *Management Science*, 22, 202–211.
- Li, W., Shi, D., and Chao, X. (1997), 'Reliability Analysis of  $M/G/1$  Queueing System With Server Breakdowns and Vacations', *Journal of Applied Probability*, 34, 546–555.
- Madan, K.C. (2001), 'On a Single Server Queue With Two-Stage Heterogeneous Service and Deterministic Server Vacations', *International Journal of Systems Science*, 32, 837–844.
- Park, Y.T., and Park, K.S. (1986), 'Generalized Spare Ordering Policies With Random Lead Time', *European Journal of Operational Research*, 23, 320–330.
- Tadj, L. (2003), 'A Quorum Queueing System Under T-Policy', *Journal of the Operational Research Society*, 54, 466–471.
- Tadj, L., and Choudhury, G. (2005), 'Optimal Design and Control of Queues', *TOP*, 13, 359–412.
- Takagi, H. (1991), *Queueing Analysis: A Foundation of Performance Evaluation, Vol. I, Vacation and Priority Systems, Part I*, North-Holland: Amsterdam.
- Tang, Y.H. (1997), 'A Single-Server  $M/G/1$  Queueing System Subject to Breakdowns – Some Reliability and Queueing Problem', *Microelectronics and Reliability*, 37, 315–321.
- Toktay, L.B., and Uzsoy, R. (1998), 'A Capacity Allocation Problem with Integer Side Constraints', *European Journal of Operational Research*, 109, 170–182.
- Uzsoy, R., Lee, C.-Y., and Martin-Vega, L.A. (1992), 'A Review of Production Planning and Scheduling Models in the Semiconductor Industry (I): System Characteristics, Performance Evaluation and Production Planning', *IIE Transactions*, 24, 47–60.
- Uzsoy, R., Lee, C.-Y., and Martin-Vega, L.A. (1994), 'A Review of Production Planning and Scheduling Models in the Semiconductor Industry (II): Shop-Floor Control', *IIE Transactions*, 26, 44–55.
- Wang, K.-H., and Ke, J.-C. (2002), 'Control Policies of an  $M/G/1$  Queueing System with a Removable and Non-reliable Server', *International Transactions in Operations Research*, 9, 195–212.
- Wolff, R.W. (1982), 'Poisson Arrival See Time Average', *Operations Research*, 30, 223–231.
- Yang, D.-Y., Wang, K.-H., Ke, J.-C., and Pearn, W.L. (2008), 'Optimal Randomized Control Policy of an Unreliable Server System With Second Optional Service and Startup', *Engineering Computations: International Journal for Computer-Aided Engineering and Software*, 25, 783–800.
- Zhang, Z.G., and Tian, N. (2001), 'Discrete Time Geo/G/1 Queue With Multiple Adaptive Vacations', *Queueing Systems*, 38, 419–429.