Research Article

Binomial Schedule for an M/G/1 Type Queueing System with an Unreliable Server under N-Policy

Lotfi Tadj[1] and K. Paul Yoon[2]

[1] *Department of Information Systems and Decision Sciences, Silberman College of Business, Fairleigh Dickinson University, Vancouver, BC, Canada V6B 2P6*

[2] *Department of Information Systems and Decision Sciences, Silberman College of Business, Fairleigh Dickinson University, Teaneck, NJ 07666, USA*

Correspondence should be addressed to Lotfi Tadj; lotftadj@yahoo.com

We consider in this paper an M/G/1 type queueing system with the following extensions. First, the server is unreliable and is subject to random breakdowns. Second, the server also implements the well-known N-policy. Third, instead of a Bernoulli vacation schedule, the more general notion of binomial schedule with K vacations is applied. A cost function with two decision variables is developed. A numerical example shows the effect of the system parameters on the optimal management policy.

1. Introduction

Queueing systems where the server uses her/his idle time to perform some secondary job such as maintenance are called systems with server vacations. These systems have received a lot of attention due to their wide applications in different domains such as telecommunications, computer systems, service systems, and production and quality control problems. Survey papers have been written on this subject; the most recent one being that of Ke et al. [1].

Keilson and Servi [2] introduced a class of vacation models called the Bernoulli vacation schedule. When a customer has just been served and other customers are present, the server serves the next customer in line with probability p or takes a vacation of random duration with probability $(1 - p)$. The Bernoulli vacation schedule has been extensively considered. Among the most recent references we cite Kumar et al. [3], Choudhury and Ke [4], Gao and Liu [5], Tao et al. [6, 7], and Wu and Lian [8].

Kella [9] generalized the Bernoulli vacation schedule to a more general scheme according to which the server goes on

i consecutive vacations with probability p_i if the queue upon her/his return is empty. Ba-Rukab et al. [10] propose another generalization. They argued that since the server may attend different activities while idle, a binomial vacation schedule may be more appropriate than a Bernoulli vacation schedule. In that case, instead of taking just one vacation, the server may take many vacations, for a maximum number of, say, K vacations.

Yadin and Naor introduced the N-policy in which, following an idle period, the server resumes his service only when the number of waiting customers reaches the level N. This policy is efficient in that it reduces setup costs. The N-policy too has been extensively studied by researches. We refer the reader to the following recent references: Kumar and Jain [11], Lee and Yang [12], Lim et al. [13], and Wei et al. [14].

Another characteristic of servers in a queueing system is that they may break down while providing service. White and Christie [15] were the first to study a queueing system with an unreliable server. Since then, many authors have incorporated this feature in their studies. We cite the recent papers of Dimitriou [16], Wu and Lian [8], Choudhury and Ke [4],

Ke et al. [17], Yang et al. [18], Yarmand and Down [19], Kumar et al. [3], and Zhang and Wang [20].

The goal of this paper is to study an M/G/1 type queueing system where the server implements the notion of a binomial vacation schedule. Under this policy, at a service completion and before serving the next customer, the server takes a series of vacations. The number of vacations follows a binomial distribution with parameters p and K. Each vacation has a random duration and corresponds to some auxiliary activity. Also, we assume the N-policy discipline, so that, following a busy period, the server remains idle and does not resume work until the number of units in the queue reaches the threshold level N. Furthermore, we assume that the server is unreliable and may breakdown while providing service to a customer. In that case, the server is repaired and the repair period has a random duration. Following repair, service is resumed.

There are various reasons for studying this queueing system. First, from a theoretical point of view, although some aspects of this system have been discussed separately, no work has been found that combines N-policy, binomial vacation schedule, and all service interruptions. Hence this paper is an attempt to fill up the gap. Second, from a practical point of view, combining these features gives the decision maker a better control over the system. As pointed out earlier, the N-policy reduces the number of setups and thus reduces setup costs. The binomial schedule allows the server to accomplish more secondary tasks than a regular Bernoulli schedule. Secondary tasks could include actions such as virus scans, maintenance operation, quality control tests, and attending other queues.

The notation used to describe this system is given in the next section. In Section 3, we obtain the probability generating function of the system size in the steady-state. We also derive the main performance measures. These measures are used to prescribe an optimal management policy for the system by finding the optimal values of the thresholds K and N in Section 4. Section 5 summarizes the paper and provides some future research directions.

2. Model Description

We consider an M/G/1 type queueing system with positive arrival rate λ. The service time random variable B has cumulative distribution function (CDF) $B(t)$, Laplace-Stieltjes transform (LST) $B^*(\theta) = E[e^{-\theta B}]$, and finite first and second moments $b_j, j = 1, 2$.

The server is unreliable and the time to failure has an exponential distribution with positive rate α. Following a breakdown, a repair of the server takes place. The repair time random variable R has CDF $R(t)$, LST $R^*(\theta) = E[e^{-\theta R}]$, and finite first and second moments $r_j, j = 1, 2$.

We introduce the *modified service time H* which includes the actual service time and possible repairs. The random variable H has CDF $H(t)$, LST $H^*(\theta) = E[e^{-\theta H}]$, and finite and second moments $h_j, j = 1, 2$. It is easy to see that the modified service times, actual service times, and repair times are related through the following expression:

$$H^*(\theta) = \sum_{n=0}^{\infty} \int_0^{\infty} e^{-\theta x} e^{-\alpha x} \frac{(\alpha x)^n}{n!} [R^*(\theta)]^n dB(x)$$
$$= B^*(\theta + \alpha(1 - R^*(\theta))). \tag{1}$$

From this expression, the first two moments of the modified service time are found as

$$h_1 = b_1(1 + \alpha r_1),$$
$$h_2 = b_2(1 + \alpha r_1)^2 + \alpha b_1 r_2. \tag{2}$$

The server implements the binomial vacation schedule. At the end of a service, if no customer is present in the system, the server takes k $(k = 0, \ldots, K)$ vacations of length V_k with probability

$$y_k = \binom{K}{k} p^k (1 - p)^{K-k}, \quad k = 0, \ldots, K. \tag{3}$$

The vacation time random variable V has CDF $V(t)$, LST $V^*(\theta) = E[e^{-\theta V}]$, and finite first and second moments $v_j, j = 1, 2$.

We now introduce the *generalized service time G*. The random variable G has CDF $G(t)$, LST $G^*(\theta) = E[e^{-\theta G}]$, and finite first and second moments $g_j, j = 1, 2$. It is easy to see that the generalized service times, modified service times, and vacation times are related through the following expression:

$$G = H + V_0 + \cdots + V_k \quad \text{with probability } y_k. \tag{4}$$

Therefore,

$$G^*(\theta) = H^*(\theta) \sum_{k=0}^{K} y_k [V^*(\theta)]^k. \tag{5}$$

From this expression, the first two moments of the generalized service time are found as

$$g_1 = h_1 + Kpv_1,$$
$$g_2 = h_2 + 2Kph_1 + K(K-1)p^2. \tag{6}$$

We are assuming that the server implements also the N-policy discipline. Thus, if there are N or more customers in the queue at a service completion epoch, then the server serves the next customer in line. However, if there are less than N customers in the queue at a service completion epoch, then the server remains idle and waits for the queue to reach the level N.

Finally, we denote by $Q(t)$ the number of customers in the system at any instant of time $t \geq 0$. We are first interested in the distribution of this process in the steady-state.

3. Model Analysis

In this section, we derive the PGF of the queueing process $Q(t)$ in the steady-state along with the main performance measures required to develop an optimal management policy of the system.

3.1. Probability Generating Function. Let $p_i = \lim_{t \to \infty} P\{Q(t) = i\}$, $i = 0, 1, \ldots$, denote the steady-state probability of state i. Since our model is of M/G/1 type with a modified service time, see, for example, Çinlar [21], we can readily generalize Pollaczek-Khinchine formula to obtain the probability generating function $P(z) = \sum_{i=0}^{\infty} p_i z^i$ as

$$
P(z) = \frac{1 - \rho}{N}
$$

$$
\times \left(\left(z^N - 1 \right) B^* \left(\lambda - \lambda z + \alpha \left(1 - R^* \left(\lambda - \lambda z \right) \right) \right) \right.
$$

$$
\times \sum_{k=0}^{K} y_k [V^* \left(\lambda - \lambda z \right)]^k \right)
$$

$$
\times \left(z - B^* \left(\lambda - \lambda z + \alpha \left(1 - R^* \left(\lambda - \lambda z \right) \right) \right) \right.
$$

$$
\left. \times \sum_{k=0}^{K} y_k [V^* \left(\lambda - \lambda z \right)]^k \right)^{-1},
$$

$$
(7)
$$

where $\rho = \lambda [b_1 (1 + \alpha r_1) + K p v_1]$ must satisfy the ergodicity condition

$$
\rho < 1. \tag{8}
$$

3.2. System Performance Measures

(a) The expected number of customers in the system at an arbitrary instant of time in the steady-state is given by

$$
L = \frac{\lambda [b_1 (1 + \alpha r_1) + K p v_1]}{1 - \lambda [b_1 (1 + \alpha r_1) + K p v_1]}
$$

$$
+ \left(\lambda^2 \left\{ b_2 (1 + \alpha r_1)^2 + \alpha b_1 r_2 \right. \right.
$$

$$
\left. \left. + 2 K p b_1 (1 + \alpha r_1) + K (K - 1) p^2 \right\} \right) \tag{9}
$$

$$
\times \left(2 \left\{ 1 - \lambda [b_1 (1 + \alpha r_1) + K p v_1] \right\} \right)^{-1}
$$

$$
+ \frac{N - 1}{2}.
$$

(b) The expected length of an idle period of the server in the equilibrium is given by

$$
I = \frac{N}{\lambda}. \tag{10}
$$

(c) The expected length of the busy period is

$$
B = \frac{N [b_1 (1 + \alpha r_1) + K p v_1]}{1 - \rho}. \tag{11}
$$

(d) The expected length of the busy cycle $C = I + B$ is given by

$$
C = \frac{N}{\lambda (1 - \rho)}. \tag{12}
$$

Using (10)–(12), one can derive the following probabilities.

(e) The probability that the server is idle is as follows:

$$
P_I = \frac{I}{C} = 1 - \rho. \tag{13}
$$

(f) The probability that the server is busy is as follows:

$$
P_B = \frac{B}{C} = \rho. \tag{14}
$$

4. Optimal Management Policy

We now are ready to develop an expression for the total expected cost per unit of time. The decision variables in this expression are K and N. The goal is to find the optimal values of these two parameters. This would optimize the performance of the system, as it would allow the server to know exactly when to end an idle period and how many auxiliary jobs should be performed before moving to the next customer in line.

4.1. Total Expected Cost per Unit of Time. Using a linear cost structure, the total expected cost function per unit time is given by

$$
TC(K, N) = c_h L + c_o \frac{B}{C} + c_a \frac{I}{C} + c_s \frac{1}{C}
$$

$$
= c_h \left[\rho + \frac{\lambda^2 g_2}{2(1 - \rho)} + \frac{N - 1}{2} \right] + c_o \rho + c_a (1 - \rho)
$$

$$
+ \frac{c_s \lambda (1 - \rho)}{N},
$$

$$
(15)
$$

where g_2 is given by (6), ρ is given by (8), c_h is the holding cost per unit for each customer present in the system, c_o is the cost per unit time for keeping the server on and in operation, c_a is the startup cost per unit time for the preparatory work of the server before starting, and c_s is the setup cost per busy cycle.

Treating the decision variables as continuous, the optimal values of K and N are found by solving the system of two equations

$$
\frac{\partial TC(K, N)}{\partial K} = 0,
$$

$$
\frac{\partial TC(K, N)}{\partial N} = 0 \tag{16}
$$

and checking that the Hessian matrix is positive definite. These calculations are done numerically since expression (15) is nonlinear and closed form expressions for the optimal values K^* and N^* are difficult to obtain. To determine these values, we present the following procedure. For a given K, the optimal value $N^*(K)$ of N is given by the first ℓ such that

$$
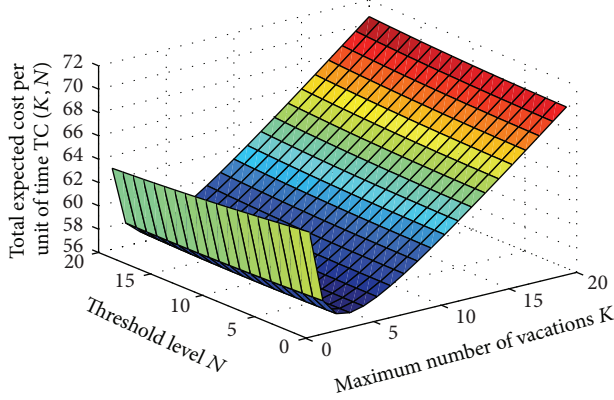TC(K, \ell + 1) > TC(K, \ell); \tag{17}
$$

FIGURE 1: Variations of the total expected cost per unit of time.

that is,

$$N^*(K) = \min \{N \geq 1 \mid I(K, N) > 0\}, \qquad (18)$$

where

$$I(K, N) = TC(K, N + 1) - TC(K, N). \qquad (19)$$

We summarize the procedure to find the optimal values (K^*, N^*) as follows.

(1) Set $K = 1$. Determine $N^*(K)$ using (18) and compute $TC(K, N^*(K))$ using (15).

(2) Compute $N^*(K + 1)$ using (18) and $TC(K + 1, N^*(K + 1))$ using (15).

(3) If $TC(K + 1, N^*(K + 1)) > TC(K, N^*(K))$, stop. The optimal values are $(K^*, N^*) = (K, N^*(K))$. Otherwise, go to step 2.

4.2. Numerical Illustration. We now present some numerical computations to show the practicality of the results obtained. For illustration, we will assume that the service, repair, and vacation times are exponentially distributed. We recall that if X follows the exponential distribution with mean $E[X]$, then

$$P(X < x) = 1 - e^{(x/E[X])}, \qquad B^*(\theta) = \frac{1}{1 + \theta E[X]}, \\ E[X^2] = 2E[X]^2. \qquad (20)$$

As an illustration, we assume the system parameters listed in Table 1.

Figure 1 shows the variations of the total expected cost per unit of time as K and N vary. The curve is convex and the optimal thresholds are found to be $(K^*, N^*) = (5, 4)$. The optimal cost is $TC(K^*, N^*) = 56.67$. Therefore, to minimize the total expected cost per unit of time, following an idle period, the server of this system should not be activated until the number of customers waiting in the queue reaches 4 customers. Also, at a service completion, when the queue is not empty, the server may take up to 5 vacations or perform 5 auxiliary tasks, before serving the next customer in line. It

TABLE 1: Data.

Parameters	Values
Unit costs	$c_h = 2, c_o = 50, c_a = 50, c_s = 300$
Mean arrival rate	$\lambda = 0.05$
Mean breakdown rate	$\alpha = 0.05$
Probability of a vacation	$p = 0.60$
Mean service time	$b_1 = 0.7$
Mean repair time	$r_1 = 0.5$
Mean vacation time	$v_1 = 0.25$

is worth noting that the optimal value of K is not $K^* = 2$, which would correspond to the well-known and widely used Bernoulli vacation schedule. This shows the effectiveness of the binomial over the Bernoulli schedule.

4.3. Sensitivity Analysis. We also performed a sensitivity analysis by looking at the effect of the system parameters on the optimal solution. The sensitivity analyses with the monetary parameters are given in Tables 2–5. We note that as customer holding cost c_h increases, both K^* and N^* decrease while they both increase as operation setup cost c_s increases (Tables 2 and 5). However, as server keeping cost c_o increases, K^* decreases, while server idle cost c_a has the opposite effect on K^* (Tables 3 and 4). It is interesting to see that N^* remains unchanged; that is, the amount of server costs does not influence the threshold level of waiting customers.

Concerning the nonmonetary parameters, we note from Tables 6, 7, 8, 9, 10, and 11 that their effect is almost negligible on N^*. Also, α has very slight effect on K^*. However, K^* increases as v_1 increases and decreases as either of the other parameters (p, b_1, or r_1) increases. Finally, the nonmonetary parameters affect very little the optimal cost.

It is interesting to find that K^* is much more sensitive than N^* to the parameter changes. Actually, N^* remains around 4 customers under most parameter values.

5. Conclusion

We have considered in this paper an unreliable queueing system where the server implements a binomial up to K-vacation schedule and an N-customer waiting policy. The system characteristics are obtained and an optimal management policy is described. The effect of the system parameters on the optimal threshold level N^* and the optimal maximum number of auxiliary jobs K^* that the server should perform is shown in a numerical example.

It is interesting to find that K^* is much more sensitive than N^* to the parameter changes. Actually, N^* remains around 4 customers under most parameter values. The manager of this queueing system thus should be more concerned with limiting vacations than with increasing the maximum number of customers allowed to wait.

This work can be generalized by assuming, for example, a batch arrival Poisson process. Binomial schedule may also be worth investigating in retrial queueing systems.

TABLE 2: Effect of the unit cost c_h on the optimal solution.

c_h	1	2	3	4	5	6	7	8	9	10
K^*	11	5	3	1	2	1	1	1	1	1
N^*	5	4	3	3	2	2	2	2	2	2
$TC(K^*, N^*)$	54.85	56.67	57.92	58.97	59.91	60.46	61.01	61.56	62.11	62.65

TABLE 3: Effect of the unit cost c_o on the optimal solution.

c_o	50	60	70	80	90	100	110	120	130	140
K^*	5	1	1	1	1	1	1	1	1	1
N^*	4	4	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	56.67	57.12	57.55	57.98	58.42	58.85	59.29	59.72	60.15	60.59

TABLE 4: Effect of the unit cost c_a on the optimal solution.

c_a	5	6	7	8	9	10	20	30	40	50
K^*	1	1	1	1	1	1	1	1	1	5
N^*	4	4	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	13.63	14.59	15.55	16.50	17.46	18.42	27.98	37.55	47.12	56.67

TABLE 5: Effect of the unit cost c_s on the optimal solution.

c_s	100	200	300	400	500	600	700	800	900	1000
K^*	1	3	5	8	8	11	11	13	15	15
N^*	2	3	4	4	5	5	6	6	6	7
$TC(K^*, N^*)$	53.49	55.28	56.67	57.81	58.81	59.70	60.55	61.28	62.00	62.69

TABLE 6: Effect of the arrival rate λ on the optimal solution.

λ	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045	0.050	0.055
K^*	1	2	8	2	5	6	7	4	5	5
N^*	2	2	2	3	3	3	4	4	4	4
$TC(K^*, N^*)$	52.50	53.25	53.98	54.49	54.97	55.44	55.91	56.32	56.67	57.01

TABLE 7: Effect of the breakdown rate α on the optimal solution.

α	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
K^*	5	5	5	5	5	4	4	4	4	4
N^*	5	4	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	56.67	56.67	56.67	56.67	56.67	56.67	56.67	56.67	56.67	56.67

TABLE 8: Effect of the probability of a vacation p on the optimal solution.

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
K^*	26	13	9	7	6	5	4	4
N^*	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	56.67	56.67	56.67	56.67	56.67	56.67	56.67	56.67

TABLE 9: Effect of the mean service time b_1 on the optimal solution.

b_1	0.35	0.70	1.05	1.40	1.75	2.10	2.45	2.80	3.15	3.50
K^*	6	5	4	3	2	1	1	1	1	1
N^*	5	4	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	56.68	56.67	56.65	56.64	56.62	56.61	56.60	56.60	56.60	56.60

TABLE 10: Effect of the mean repair time r_1 on the optimal solution.

r_1	0.125	0.250	0.500	1.000	2.000	4.000	8.000	16.000	32.000	64.000
K^*	5	5	5	5	5	4	4	3	1	1
N^*	4	4	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	56.67	56.67	56.67	56.67	56.62	56.68	56.67	56.69	56.82	57.45

TABLE 11: Effect of the mean vacation time v_1 on the optimal solution.

v_1	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28
K^*	2	2	2	2	4	4	4	5	5	5
N^*	4	4	4	4	4	4	4	4	4	4
$TC(K^*, N^*)$	56.69	56.69	56.69	56.68	56.68	56.68	56.67	56.67	56.67	56.66

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] J.-C. Ke, C.-H. Wu, and Z. G. Zhang, "Recent developments in vacation queueing models: a short survey," *International Journal of Operations Research*, vol. 7, no. 4, pp. 3–8, 2010.

[2] J. Keilson and L. D. Servi, "Oscillating random walk models for GI/G/1 vacation system with Bernoulli schedules," *Journal of Applied Probability*, vol. 23, no. 3, pp. 790–802, 1986.

[3] B. K. Kumar, R. Rukmani, and S. R. A. Lakshmi, "Performance analysis of an M/G/1 queueing system under Bernoulli vacation schedules with server setup and close down periods," *Computers & Industrial Engineering*, vol. 66, no. 1, pp. 1–9, 2013.

[4] G. Choudhury and J.-C. Ke, "A batch arrival retrial queue with general retrial times under Bernoulli vacation schedule for unreliable server and delaying repair," *Applied Mathematical Modelling*, vol. 36, no. 1, pp. 255–269, 2012.

[5] S. Gao and Z. Liu, "An M/G/1 queue with single working vacation and vacation interruption under Bernoulli schedule," *Applied Mathematical Modelling*, vol. 37, no. 3, pp. 1564–1579, 2013.

[6] L. Tao, Z. Wang, and Z. Liu, "The GI/M/1 queue with Bernoulli-schedule-controlled vacation and vacation interruption," *Applied Mathematical Modelling*, vol. 37, no. 6, pp. 3724–3735, 2013.

[7] L. Tao, L. Zhang, X. Xu, and S. Gao, "The GI/Geo/1 queue with Bernoulli-schedule-controlled vacation and vacation interruption," *Computers & Operations Research*, vol. 40, no. 7, pp. 1680–1692, 2013.

[8] J. Wu and Z. Lian, "A single-server retrial G-queue with priority and unreliable server under Bernoulli vacation schedule," *Computers & Industrial Engineering*, vol. 64, no. 1, pp. 84–93, 2013.

[9] O. Kella, "The threshold policy in the M/G/1 queue with server vacations," *Naval Research Logistics*, vol. 36, no. 1, pp. 111–123, 1989.

[10] O. Ba-Rukab, L. Tadj, and J.-C. Ke, "Binomial schedule for an M/G/1 queueing system with an unreliable server," *International Journal of Modelling in Operations Management*, vol. 3, no. 3-4, pp. 206–218, 2013.

[11] K. Kumar and M. Jain, "Threshold F-policy and N-policy for multi-component machining system with warm standbys," *Journal of Industrial Engineering International*, vol. 9, no. 28, 9 pages, 2013.

[12] D. H. Lee and W. S. Yang, "The N-policy of a discrete time Geo/G/1 queue with disasters and its application to wireless sensor networks," *Applied Mathematical Modelling*, vol. 37, no. 23, pp. 9722–9731, 2013.

[13] D.-E. Lim, D. H. Lee, W. S. Yang, and K. Chae, "Analysis of the GI/Geo/1 queue with N-policy," *Applied Mathematical Modelling*, vol. 37, no. 7, pp. 4643–4652, 2013.

[14] Y. Wei, M. Yu, Y. Tang, and J. Gu, "Queue size distribution and capacity optimum design for image-policy image queue with setup time and variable input rate," *Mathematical and Computer Modelling*, vol. 57, no. 5-6, pp. 1559–1571, 2013.

[15] H. White and L. G. Christie, "Queueing with preemptive priority or breakdown," *Operations Research*, vol. 6, no. 1, pp. 79–96, 1958.

[16] I. Dimitriou, "A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations," *Applied Mathematical Modelling*, vol. 37, no. 3, pp. 1295–1309, 2013.

[17] J.-C. Ke, Y.-L. Hsu, T.-H. Liu, and Z. G. Zhang, "Computational analysis of machine repair problem with unreliable multi-repairmen," *Computers & Operations Research*, vol. 40, no. 3, pp. 848–855, 2013.

[18] D.-Y. Yang, Y.-C. Chiang, and C.-S. Tsou, "Cost analysis of a finite capacity queue with server breakdowns and threshold-based recovery policy," *Journal of Manufacturing Systems*, vol. 32, no. 1, pp. 174–179, 2013.

[19] M. H. Yarmand and D. G. Down, "Server allocation for zero buffer tandem queues," *European Journal of Operational Research*, vol. 230, no. 3, pp. 596–603, 2013.

[20] F. Zhang and J. Wang, "Performance analysis of the retrial queues with finite number of sources and service interruptions," *Journal of the Korean Statistical Society*, vol. 42, no. 1, pp. 117–131, 2013.

[21] E. Çinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1975.