Journal of Manufacturing Systems 32 (2013) 174–179

Contents lists available at SciVerse ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical paper

Cost analysis of a finite capacity queue with server breakdowns and threshold-based recovery policy

Dong-Yuh Yang*, Ya-Chu Chiang, Ching-Shih Tsou

Institute of Information and Decision Sciences, National Taipei College of Business, Taipei 100, Taiwan

ARTICLE INFO

Article history:
Received 30 December 2011
Received in revised form 10 April 2012
Accepted 3 June 2012
Available online 31 July 2012

Keywords:
Breakdown
Cost
Optimal
Queue
Threshold-based recovery policy

ABSTRACT

This paper analyzes a repairable $M/M/1/N$ queueing system under a threshold-based recovery policy. The threshold-based recovery policy means that the server breaks down only if there is at least one customer in the system, and the recovery can be performed when q $(1 \leq q \leq N)$ or more customers are present. For this queueing system, a recursive method is applied to obtain steady-state solutions in neat closed-form expressions. We then develop some important system characteristics, such as the number of customers in the system, the probability that the server is busy, the effective arrival rate and the expected waiting time in the system, etc. A cost model is constructed to determine the optimal threshold value, the optimal system capacity and the optimal service rate at a minimum cost. In order to solve this optimization problem, the direct search method and Newton's method are employed. Sensitivity analysis is also conducted with various system parameters. Finally, we present some managerial insights through an application example.

1. Introduction

Queueing systems with server breakdowns are very common in stochastic systems, such as computer systems, communication systems, manufacturing systems, and so on. In this paper, we attempt to investigate a finite capacity queue with an unreliable server and threshold-based recovery policy. The threshold-based recovery policy was introduced by Efrosinin and Semenova [4], in which the server may break down unpredictably while providing service for a customer, and the repair can only be performed when q $(1 \leq q \leq N)$ or more customers are present in the system. Furthermore, Efrosinin and Winkler [6] examined an $M/M/1$ retrial queue with constant retrial rate, unreliable server and threshold-based recovery policy. Optimal threshold to minimize the long-run average losses for the given cost structure was also discussed in Efrosinin and Winkler [6]. The recent work of Purohit et al. [15] that focused on an $M/M/1$ retrial queue with constant retrial policy, unreliable server, threshold based recovery and state dependent arrival rates. They utilized the Runge–Kutta method to compute the steady state probabilities as well as various system performance measures.

Threshold policies have been widely investigated by many researchers include Yadin and Naor [24], Heyman [8], Balachandran [1], Gupta [7], etc. The reader is referred to Crabill et al. [2] for an excellent survey on the control of queues. One of the most popular threshold policies is the N-policy studied by Yadin and Naor [24]. The N-policy is to turn the server on whenever N $(N \geq 1)$ or more customers are present in the system, and then turn the server off only when the system becomes empty. Indeed, extensive research works on the N-policy queue with a reliable server has been done in Yadin and Naor [24], Kimura [13], Teghem [17], Wang and Huang [18,19], Wang and Ke [21], and so on. For cases with server breakdowns, exact steady-state solutions of the N-policy $M/E_k/1$, $M/H_2/1$ and $M/H_k/1$ queueing systems were developed by Wang [22], Wang et al. [20] and Wang et al. [23], respectively. Ke and Wang [11] examined the N-policy $M^{[x]}/M/1$ queue with server breakdowns and startup time, in which the arrival rate varies according to the status of the server. In the result of Ke [9], it showed that the stochastic decomposition property holds for the steady-state probabilities and departure point queue size distribution in the N-policy $M^{[x]}/M/1$ vacation queue with an un-reliable server. The study of an N-policy $M/M/1$ queueing system with heterogeneous arrivals, server breakdowns and multiple vacations has been discussed by Ke and Pearn [12]. They not only constructed a cost model to determine the optimal management policy, but also performed a sensitivity analysis on the optimal value of N. Furthermore, Pearn et al. [14] studied the N-policy $M/G/1$ queueing system with server breakdowns, and derived analytical results of the sensitivity analysis. Using the stochastic decomposition approach, Ke

* Corresponding author at: Institute of Information and Decision Sciences, National Taipei College of Business, No. 321, Sec. 1, Jinan Rd., Zhongzheng District, Taipei 100, Taiwan.
E-mail address: yangdy@webmail.ntcb.edu.tw (D.-Y. Yang).

0278-6125/$ – see front matter © 2012 The Society of Manufacturing Engineers. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jmsy.2012.06.002

[10] analyzed the bi-level control policy $M^{[x]}/G/1$ queueing system with server breakdowns. He derived the probability generating function of the number of customers in the queueing system for two different models. Choudhury et al. [3] applied the supplementary variables technique to investigate an $M^{[x]}/G/1$ with an additional second phase of optional service and unreliable server under N-policy. Recently, Efrosinin and Semenova [5] considered the N-policy $M/M/1$ queue with constant retrial rate and non-reliable removable server. In Efrosinin and Semenova [5], they developed analytical steady-state results and derived the explicit formula for calculating the optimal N-policy.

There are many literatures on the N-policy queue in different frameworks. However, to the best of our knowledge, there are few works on the queue with a threshold-based recovery policy. This motivates us to investigate a finite capacity queue with server breakdowns and threshold-based recovery policy. The remainder of this paper is organized as follows. Section 2 describes the basic assumptions of the queueing model. In Section 3, we derive the analytical solutions for the queueing model in neat closed-form expressions. Some important system characteristics are developed in Section 4. In Section 5, an expected cost function per unit time is established to determine the optimal system capacity, the optimal threshold value and the optimal service rate. Moreover, this optimization work is carried out by means of the direct search method and Newton's method. Section 6 provides an application example to demonstrate how the model could be applied in practical situations. Finally, we draw conclusions in Section 7.

2. Basic assumptions

We consider the $M/M/1/N$ queue with server breakdowns and threshold-based recovery policy. The basic assumptions are described as follows:

(1) Customers arrive at the system according to a Poisson process with parameter λ.
(2) Arriving customers form a single waiting line based on the order of their arrivals, that is, the first-come first-served (FCFS) discipline.
(3) Service times during a busy period follow an exponential distribution with mean $1/\mu$.
(4) The server can serve only one customer at a time. If the server is busy, arriving customers must wait until the server is available.
(5) The server can break down only if there is at least one customer in the system. The breakdown times are assumed to be exponential with breakdown rate α.
(6) When the server breaks down, the server cannot be repaired until the number of customers in the system reaches a specified threshold value $q\,(1 \le q \le N)$. Repair times are exponentially distributed with mean $1/\beta$.
(7) After the server is repaired, he returns to the system and provides service until the system becomes empty.
(8) $N\,(N < \infty)$ denotes the system capacity.
(9) Various stochastic processes involved in the system are assumed to be independent of each other.

3. Steady-state results

In this section, we first apply the Markov process theory to obtain the steady-state equations. Next, a recursive method is employed to develop the analytical solutions in a neat closed-form. Let us define some notations in the following:

$N(t) \equiv$ the number of customers in the system at time t,
$Y(t) \equiv$ the server state at time t,

where

$$Y(t) = \begin{cases} 0, & \text{if the server is in busy period,} \\ 1, & \text{if the server is in brokendown period.} \end{cases}$$

Then $\{Y(t), N(t); t \ge 0\}$ is a continuous time Markov process with state space

$$S = \{(0, n) | n = 0, 1, \ldots, N\} \cup \{(1, n) | n = 1, 2, \ldots, N\}.$$

Furthermore, the steady-state probabilities of the system are defined as follows:

$$P_0(n) = \lim_{t\to\infty} \{Y(t) = 0,\ N(t) = n\}, \quad n = 0, 1, \ldots, N.$$

$$P_1(n) = \lim_{t\to\infty} \{Y(t) = 1,\ N(t) = n\}, \quad n = 1, 2, \ldots, N.$$

3.1. Steady-state equations

Referring to the state-transition-rate diagram for the $M/M/1/N$ queue with server breakdowns and threshold-based recovery policy as shown in Fig. 1, we obtain the steady-state equations for $P_0(n)$ and $P_1(n)$ in the following:

$$\lambda P_0(0) = \mu P_0(1) \tag{1}$$

$$(\lambda + \mu + \alpha) P_0(n) = \lambda P_0(n-1) + \mu P_0(n+1), \quad n = 1, 2, \ldots, q-1, \tag{2}$$

$$(\lambda + \mu + \alpha) P_0(n) = \lambda P_0(n-1) + \mu P_0(n+1) + \beta P_1(n), \quad n = q, q+1, \ldots, N-1, \tag{3}$$

$$(\mu + \alpha) P_0(N) = \lambda P_0(N-1) + \beta P_1(N), \tag{4}$$

$$\lambda P_1(1) = \alpha P_0(1), \tag{5}$$

$$\lambda P_1(n) = \lambda P_1(n-1) + \alpha P_0(n), \quad n = 2, 3, \ldots, q-1, \tag{6}$$

$$(\lambda + \beta) P_1(n) = \lambda P_1(n-1) + \alpha P_0(n), \quad n = q, q+1, \ldots, N-1, \tag{7}$$

$$\beta P_1(N) = \lambda P_1(N-1) + \alpha P_0(N). \tag{8}$$

3.2. Recursions for the $P_i(n)$

Solving Eqs. (1)–(8), recursively, we obtain analytical solutions $P_i(n)\,(i = 0, 1)$ in neat closed-form as follows:

$$P_0(1) = \frac{\lambda}{\mu} P_0(0), \tag{9}$$

$$P_0(n) = \frac{\lambda}{\mu} P_0(n-1) + \frac{\alpha}{\mu} \sum_{k=1}^{n-1} P_0(k), \quad n = 2, \ldots, q, \tag{10}$$

$$P_0(n) = \frac{\lambda}{\mu} P_0(n-1) - \frac{\beta}{\mu} \sum_{k=q}^{n-1} P_1(k) + \frac{\alpha}{\mu} \sum_{k=1}^{n-1} P_0(k), \quad n = q+1, q+2, \ldots, N-1, \tag{11}$$

$$P_0(N) = \frac{\lambda}{\mu} [P_0(N-1) + P_1(N-1)], \tag{12}$$

$$P_1(n) = \frac{\alpha}{\lambda} \sum_{k=1}^{n} P_0(k), \quad n = 1, 2, \ldots, q-1, \tag{13}$$