Expert Systems with Applications 37 (2010) 4276–4282

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Genetic algorithm for optimal thresholds of an infinite capacity multi-server system with triadic policy

Chuen-Horng Lin [a], Jau-Chuan Ke [b,*]

[a] Graduate School of Computer Science and Information Technology, National Taichung Institute of Technology, Taichung 404, Taiwan, ROC
[b] Department of Applied Statistics, National Taichung Institute of Technology, Taichung 404, Taiwan, ROC

ARTICLE INFO

Keywords:
Cost
Genetic algorithm
Rouche's theorem
Queue
Triadic policy

ABSTRACT

We consider an M/M/c queue with $c = 2$, in which the number of working servers can be adjusted one at a time at arrival epochs or at service completion epochs depending on the number of customers in the system. Analytic closed-form solutions of the infinite capacity M/M/2 queueing system operating under the triadic $(0,Q,N,M)$ policy are derived. The total expected cost function per unit time is developed, to obtain the optimal operating $(0,Q,N,M)$ policy and the optimal service rate, at minimum cost. Some illustrative examples are provided and the genetic algorithm is employed to search for the optimal management policy of the multi-server queueing system.

1. Introduction

Yadin and Naor (1963) first introduced the concept of a single-threshold control policy (i.e., N policy) which turns the server on whenever N $(N \geqslant 1)$ or more customers are present, the server is turned off when none is present. After the server is turned off, the server may not activate until N customers are present in the system. Rhee and Sivazlian (1990) derived the busy period distribution of the two-server queueing system operating under the three-threshold control policy, where inter-arrival times and service times are all exponentially distributed (i.e., triadic $(0,Q,N,M)$ policy). The definition of the triadic $(0,Q,N,M)$ policy is described as follows: whenever there are no customers in the system, both servers are inactive temporarily until certain specified conditions arise. Initially, we assume that both servers are inactive. When the number of customers waiting for service reaches a specific number, denoted by N, which is a decision variable, one of the two servers will be active instantly. At a later time, when the number of customers waiting for service increases to another specified level, say $M(N < M)$, then the remaining server will also be active instantly. However, if the number of customers in the system decreases to Q where $Q < N$, while both servers are active simultaneously, the server just finishing a service will be removed from the system at that time. In addition, if the number of customers in the system drops down to zero while one server is active, that server deactivates until the above conditions are met.

In this paper, we consider the three-threshold control policy for the infinite capacity M/M/2 queue system. It is assumed that customers arrive according to a Poisson process with parameter λ. Customers arrive at the service faculty form a single waiting line and are served in the order of their arrivals; that is, the first-come, first-served discipline. The service faculty operates the triadic $(0,Q,N,M)$ policy by two servers that provide for the arrivals, in which the service times are assumed to be exponentially distributed with mean $1/\mu$. Each server can serve only one customer at a time, and that the service is independent of the arrival of the customers. If customer is in service, then arriving customers have to waiting in the queue until the working server is available.

During the last one decade, the queueing systems with control (threshold) policy have been investigated by many researchers. Some application examples of such models can be found in Tadj and Choudhury (2005). Past work may be divided into two categories: (i) single-threshold, and (ii) two or more thresholds. In the case of single-threshold, the readers can be referred to an excellent survey by Tadj and Choudhury (2005). Later, Choudhury and Madan (2005) investigated the queue size distribution at a random epoch as well as at a departure epoch for the batch arrival M/G/1 queueing system with Bernoulli vacation schedule and two phases of service under an N policy. They also derived a simple procedure to obtain optimal stationary policy under a suitable linear cost structure. Choudhury and Paul (2006) studied an N policy of model for the batch arrival queue with second optional channel. Yang, Wang, Ke, and Pearn (2008) gave an analytical optimization analysis of a randomized T policy for an unreliable server M/G/1 queueing system with second optional service. Ke and Chu (2008) derived the explicit distributions of important system characteristics of the randomized T and N policies for the M/G/1 queueing system, respectively. They also made some comparisons between both policies and then found the randomized N policy is superior to the

* Corresponding author.
E-mail addresses: linch@ntit.edu.tw (C.-H. Lin), jauchuan@ntit.edu.tw (J.-C. Ke).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.11.074

randomized T policy based on the minimum cost. Recently, Lin and Ke (2009) examined the optimal N policy of a queueing system with fuzzy parameters using Zadeh's extension principle. As for the second category, Rhee and Sivazlian (1990) derived the busy period distribution of an infinite capacity M/M/2 queueing system operating under triadic $(0,Q,N,M)$ policy. Wang and Wang (2002) studied the expected characteristics (such as expected number of customers in the system) and performed optimization of an finite capacity M/M/2 queueing system under triadic $(0,Q,N,M)$ policy. An M/G/1 queueing system with server breakdowns and startup under NT policy was investigated by Ke (2006a), who provided an optimization analysis of the optimal two-threshold NT policy. An M/G/1 queueing system under hysteretic vacation policy (the first threshold) with an early startup (the second threshold) and breakdown server by was analyzed by Ke (2006b), who developed important system characteristics such as, the distribution of queue size, waiting time, busy period, and completion period, etc. Recently, Ke (2008) examined system characteristics and optimal two-threshold policy of a batch arrival M/G/1 queueing system with a modified T policy.

In the study of queueing systems with control policies, existing literature focuses mainly on either single-threshold or multi-threshold with finite capacity. We should note that in multi-threshold (Ke, 2006a, 2006b, 2008; Ke & Chu, 2008; Lin & Ke, 2009; Rhee & Sivazlian, 1990; Wang & Wang, 2002), no *infinite capacity* system for optimization analysis is obtained. This would motivate us to investigate the infinite capacity M/M/2 queueing system with triadic $(0,Q,N,M)$ policy. This paper is organized as follows. In Section 2, we present the model formulation using birth-and death process, and then derive the stationary probabilities using Rouche's theorem. In Section 3, we derive the stationary distributions of the system characteristics, such as the expected number of customers in the system, the expected number of one server being busy, etc. In Section 4, by constructing a cost model, we develop an genetic algorithm to find the joint optimal value of (Q,N,M,μ) for the system of our interest, including some numerical examples. Section 5 concludes.

The primary objectives of this paper are:

(i) to derive the steady-state solutions for the infinite capacity M/M/2 queueing system with triadic $(0,Q,N,M)$ policy;
(ii) to develop a cost model to determine the optimum values of three discrete thresholds (Q,N,M) and one continuous service rate (μ), simultaneously, which is in order to minimize the steady-state expected cost per unit time;
(iii) we develop an efficient approach to find the joint optimal values of three discrete variables (Q,N,M) and one continuous variable (μ), simultaneously, using genetic algorithm;
(iv) to investigate the effect of arrival rate and cost structure on the optimum values of Q,N,M, and μ.

2. Steady-state results

For the infinite capacity $M/M/2/\infty$ queueing system operating under the triadic $(0,Q,N,M)$ policy, we describe the states of the system by the pairs $(i,n), i = 0,1,2,\ldots,Q,\ldots,M,\ldots,\infty$, where $i = 0$ denotes that the server is turned off, $i = 1$ denotes that one of the servers is turned on and active, $i = 2$ denotes that both servers are turned on and active, and n is the number of customers in the system.

In steady-state, the following probabilities are used.

$p_{0,n} \equiv$ probability that the server is turned off and there are n customers in the system, where $n = 0,1,2,\ldots,N-1$;

$p_{1,n} \equiv$ probability that one of the servers is turned on and active and there are n customers in the system, where $n = 1,2,\ldots,Q,\ldots,N,\ldots,M-1$;

$p_{2,n} \equiv$ probability that both servers are turned on and active and there are n customers in the system, where $n = Q+1, Q+2,\ldots,N,\ldots,M,\ldots,\infty$.

Using birth-and-death process, the steady-state equations for the infinite capacity M/M/2 queueing system under the triadic $(0,Q,N,M)$ policy gives as follows:

(i) $n = 0$

$$\lambda p_{0,n} = \mu p_{1,n+1}. \tag{1}$$

(ii) $1 \leqslant n \leqslant N-1$

$$p_{0,n} = p_{0,n-1}. \tag{2}$$

(iii) $n = 1$

$$(\lambda+\mu)p_{1,n} = \mu p_{1,n+1}. \tag{3}$$

(iv) $2 \leqslant n \leqslant Q-1$

$$(\lambda+\mu)p_{1,n} = \lambda p_{1,n-1} + \mu p_{1,n+1}. \tag{4}$$

(v) $n = Q$

$$(\lambda+\mu)p_{1,n} = \lambda p_{1,n-1} + \mu p_{1,n+1} + 2\mu p_{2,n+1}. \tag{5}$$

(vi) $Q+1 \leqslant n \leqslant N-1$

$$(\lambda+\mu)p_{1,n} = \lambda p_{1,n-1} + \mu p_{1,n+1}. \tag{6}$$

(vii) $n = N$

$$(\lambda+\mu)p_{1,n} = \lambda p_{1,n-1} + \mu p_{1,n+1} + \lambda p_{0,n-1}. \tag{7}$$

(viii) $N+1 \leqslant n \leqslant M-2$

$$(\lambda+\mu)p_{1,n} = \lambda p_{1,n-1} + \mu p_{1,n+1}. \tag{8}$$

(ix) $n = M-1$

$$(\lambda+\mu)p_{1,n} = \lambda p_{1,n-1}. \tag{9}$$

(x) $n = Q+1$

$$(\lambda+2\mu)p_{2,n} = 2\mu p_{2,n+1}. \tag{10}$$

(xi) $Q+2 \leqslant n \leqslant M-1$

$$(\lambda+2\mu)p_{2,n} = \lambda p_{2,n-1} + 2\mu p_{2,n+1}. \tag{11}$$

(xii) $n = M$

$$(\lambda+2\mu)p_{2,n} = \lambda p_{2,n-1} + 2\mu p_{2,n+1} + \lambda p_{1,n-1}. \tag{12}$$