Optimal Policy for an Unreliable Service System

Gautam Choudhury
Institute of Advanced Study in Science and Technology, India

Lotfi Tadj
Fairleigh Dickinson University, Canada

INTRODUCTION

Services play a vital role in economies of nations and in world commerce. An economy cannot operate without the services provided by service firms. In its turn, a service firm cannot survive without a sound competitive strategy. Management of service systems is a continuous challenge, due to the dynamic and random nature of the service environment.

The theory of waiting lines provides insight and identifies management options for improving customer service. A wide variety of queueing models have been developed and successfully exploited for very complex service situations. This chapter describes one such queueing model. For a comprehensive classification of various control policies applied in queueing systems, see the survey by Tadj and Choudhury (2005).

An understanding of the queueing phenomenon is necessary before creative approaches to the management of service systems can be considered. Some of the main concern in dealing with waiting line problems are the customer arrival process, the service process, the procedure or priority rule to use in selecting the next product or customer to be served, etc.

The service system we consider in this chapter is characterized by an unreliable server. Random breakdowns, when they occur, prevent the server from providing service to the customers. To model a realistic system, we assume that a repair may not be immediate. We also make assumptions regarding the server discipline in order to model a wide range of real-world systems. The first assumption is called a Bernoulli vacation schedule. It specifies that, at the end of a service, the server may either take a vacation or serve the next customer. The term server vacation is used when the server is away from the work station, attending some secondary job. The second assumption is called T-policy. It specifies that when the queue is empty, the server again takes vacations and scans the queue periodically, every T units of time, to check if some customer has arrived while he was away. For example in an inventory system, a T-policy would correspond to a periodic-review policy.

The primary purpose of waiting lines theory is to obtain the performance measures of the service system. We thus obtain probability generating functions of the system state and the main performance measures such as the mean number of customers in the system, and the mean lengths of the idle period, busy period, and busy cycle. These measures are then used to design an optimal management policy of the system, by balancing various types of costs. We illustrate the results obtained by presenting a numerical example and perform some sensitivity analysis to show the effect of the system parameter on the optimal policy.

BACKGROUND

The first feature of the system studied in this chapter is the unreliability of the server. Queueing systems prone to failure are commonly encountered in the real world. The earliest reference on

DOI: 10.4018/978-1-4666-5202-6.ch153

queueing systems with server breakdown that came to our attention is that of White and Christie (1958). Since then, queueing systems with unreliable servers have been extensively studied by many researchers; see Tadj, Choudhury, and Rekab (2012) for a comprehensive survey on the subject.

The next feature of the system of interest to us is the Bernoulli vacation schedule. The classical vacation scheme with Bernoulli service discipline was introduced and developed by Keilson and Servi (1986). Various aspects of Bernoulli vacation models have been discussed by a number of authors; see the survey of Ke, Wu, and Zhang (2010).

The other important feature considered in our service system is the server T-policy. The M/G/1 queue with a T-policy was first studied by Heyman (1977). Many variants of the T-policy discipline model have been considered in the literature since then. We list here with some detail the latest contributions, since there is no recent survey on the subject. Wang and Ke (2002) consider a single non-reliable server in the ordinary M/G/1 queueing system operating under the N-policy, the T-policy and the Min(N, T)-policy. They show that the optimal N-policy and the optimal Min(N, T)-policy are always superior to the optimal T-policy. Tadj (2003) studies an M/G/1 quorum queueing system under T-policy. The quorum or q-policy means that the server does not start service unless a specified number of customers are in the queue, and service is always rendered to groups of a fixed size. Ke (2005) studies an M/G/1 queueing system with an unreliable server, startup, and the following modified T-policy: After all the customers are served in the queue exhaustively, the server deactivates and takes at most J vacations of constant time length T repeatedly until at least one customer is found waiting in the queue upon returning from a vacation. If no customers arrive by the end of the Jth vacation, the server remains dormant in the system until at least one customer arrives. This model is generalized by Ke (2008) to the case of compound Poisson arrival process. More complex scenarios for the server are considered by Ke (2006). Kim and Moon (2006) study an M/G/1 queueing system where the server can take a vacation time T after the system becomes empty. At that time, the server is switched off with probability p and takes a vacation or remains on serving the arriving customers with probability (1-p). Wang, Wang, and Pearn (2009a) investigate the T-policy M/G/1 queue with server breakdowns, and startup times. The server is turned on after a fixed length of time T repeatedly until at least one customer is present in the waiting line. The same model is studied by Wang, Wand, and Pearn (2009b) who use the maximum entropy approach to solve for the steady-state probabilities. Zhang, Tadj, and Bounkhel (2011) clarify the concept of regeneration cycle used in evaluating the average operating cost of the M/G/1 queue with T-policy. Two ways of defining the regeneration cycle are compared and advantages and disadvantages of each way are pointed out.

Randomized control policies, with random control of the server at the beginning of the service when at least one customer appears, have also been combined recently with the T-policy. Yang, Wang, Ke, and Pearn (2008) study the randomized T-policy in an unreliable M/G/1 queueing system with second optional service to the customers and server startup. Ke and Chu (2008) compare the randomized T-policy and the randomized N-policy for an M/G/1 queueing system with second optional service. Ke and Chu (2009) deal with a variation of the randomized T-policy of Ke and Chu (2008). Wang, Yang, and Pearn (2010) compare the randomized T-policy and the randomized N-policy for an M/G/1 queueing system with second optional service and server startup. They show that the optimal randomized N-policy outperforms the optimal randomized T-policy.

In this paper, we follow the method of Zhang et al. (2011) to obtain the system performance measures. These will be used to find the vacation length that will optimize the system performance.

Related Content