INVITED PAPER

THE M/G/1 RETRIAL QUEUE WITH NONPERSISTENT CUSTOMERS

T. YANG, M.J.M. POSNER and J.G.C. TEMPLETON

Department of Industrial Engineering, University of Toronto, Toronto, Canada M5S 1A4

Received 18 September 1989; revised 17 October 1989

Abstract

We consider an M/G/1 retrial queue in which blocked customers may leave the system forever without service. Basic equations concerning the system in steady state are established in terms of generating functions. An indirect method (the method of moments) is applied to solve the basic equations and expressions for related factorial moments, steady-state probabilities and other system performance measures are derived in terms of server utilization. A numerical algorithm is then developed for the calculation of the server utilization and some numerical results are presented.

Keywords: Factorial moments, nonpersistent customers, retrial queues.

1. Introduction

Queueing systems in which arriving customers who find all servers and waiting positions occupied may retry for service after a period of time are called retrial queues or queues with repeated calls. Retrial queues are useful in modelling many problems in telephone switching systems, telecommunication networks, computer networks and computer systems.

One of the most extensively studied single-server retrial queues is the M/G/1 retrial queue without customer loss. Analytic results such as the distribution of the number of customers in the system, the number of retrials per customer, and the waiting time, etc. can be found in Keilson et al. [8], Falin [3,4], Choo and Conolly [1], and Yang and Templeton [10]. Lubacz and Roberts [9] and Falin [4] have studied a modified M/G/1 retrial queue in which customers, blocked on their first attempts, are allowed to balk (or leave the system without service). Lubacz and Roberts presented a direct derivation of the mean number of customers in orbit, the mean waiting time, and the mean number of retrials made by orbiting customers. Falin obtained the probability generating function of the number of customers in the system, and other major system performance measures.
Recently, attention has been drawn to a more general, but much more complicated model (see, for example, Falin [4] and Greenberg [5]) in which each time that a customer is blocked on his attempt, he will either leave the system without service with probability $1 - \alpha$ or stay in orbit for later retrial with probability $\alpha$. However, exact results for the steady-state distribution of the number of customers in the system have not been obtained except for the special case of exponential service time distribution (Jonin and Sedol [7] and Falin [4]) and performance measures are available only in the form of upper bounds (Greenberg [5]).

In this paper we present a semi-analytic treatment of this model. Specifically, we consider a single-server queueing system with no waiting space, to which customers arrive according to a Poisson process with rate $\lambda$. An arriving (or primary) customer obtains service immediately if he finds the server idle, and otherwise he will decide either to leave the system without service with probability $1 - \alpha$ or stay in orbit for later retrial with probability $\alpha$, where $0 < \alpha < 1$. Customers in the process of making retrials are said to be in orbit and are called orbiting customers. On retrial, an orbiting customer obtains service immediately if the server is idle, otherwise he will decide either to leave without service with probability $1 - \alpha$ or to return to the orbit with probability $\alpha$ for later retrial. The retrial time (the time interval between two consecutive attempts made by a customer) is exponentially-distributed with mean $1/\theta$ and is independent of all previous retrial times and all other stochastic processes in the system. The service times of customers are identically, independently distributed with CDF $B(\cdot)$, pdf $b(\cdot)$ and mean $\bar{x}$. We derive explicit formulae for steady-state probabilities, their generating functions, related factorial moments, and other performance measures in terms of the server utilization. We also develop a practically useful algorithm to calculate the exact numerical value of the server utilization.

2. The steady-state equations

As in ordinary $M/G/1$ queues, the stochastic process $\{N(t); t \geq 0\}$, where $N(t)$ is the number of customers in the system (both in orbit and in service) at time $t$, is not Markovian. Therefore, we introduce random variables $N_0(t)$, $I(t)$ and $X(t)$, where $N_0(t)$ is the number of customers in the orbit at time $t$, $I(t) = 0$ if the server is idle at time $t$ and $I(t) = 1$ otherwise, and $X(t)$ is the elapsed service time of the customer in service at time $t$ if $I(t) = 1$ and $X(t) = 0$ if $I(t) = 0$. Thus the stochastic process $\{(I(t), N_0(t), X(t)); t \geq 0\}$ is Markovian with state space $\{(n, k, x); n = 0, 1, k = 0, 1, \ldots, 0 \leq x < \infty\}$. Now we define for $k = 0, 1, \ldots$ and $0 \leq x < \infty$:

$p_{0,k}(t)$ the probability that the process is in state $(0, k, 0)$ at time $t$;

$p_{1,k}(t, x)$ the probability density that the process is in state $(1, k, x)$ at time $t$. 