ASSEMBLY-LIKE QUEUES WITH FINITE CAPACITY: BOUNDS, ASYMPTOTICS AND APPROXIMATIONS

E.H. LIPPER and B. SENGUPTA

AT&T Bell Laboratories, Crawfords Corner Road, Holmdel, NJ 07733, USA

Received 13 August 1985
(Revised 13 January 1986)

Abstract

In this paper we study a queueing model of assembly-like manufacturing operations. This study was motivated by an examination of a circuit pack testing procedure in an AT&T factory. However, the model may be representative of many manufacturing assembly operations. We assume that customers from $n$ classes arrive according to independent Poisson processes with the same arrival rate into a single-server queueing station where the service times are exponentially distributed. The service discipline requires that service be rendered simultaneously to a group of customers consisting of exactly one member from each class. The server is idle if there are not enough customers to form a group. There is a separate waiting area for customers belonging to the same class and the size of the waiting area is the same for all classes. Customers who arrive to find the waiting area for their class full, are lost. Performance measures of interest include blocking probability, throughput, mean queue length and mean sojourn time. Since the state space for this queueing system could be large, exact answers for even reasonable values of the parameters may not be easy to obtain. We have therefore focused on two approaches. First, we find upper and lower bounds for the mean sojourn time. From these bounds we obtain the asymptotic solutions as the arrival rate (waiting room, service rate) approaches zero (infinity). Second, for moderate values of these parameters we suggest an approximate solution method. We compare the results of our approximation against simulation results and report good correspondence.

Keywords

Assembly queue, manufacturing, approximations.
1. Introduction

In this paper we study a queueing model of an assembly-like work center in which customers from several classes must be processed as 'complete units', i.e. a customer of each type must be present before service can begin. Our examination of this problem was motivated by a study of a circuit-pack testing procedure in an AT&T factory. In this test, several types of packs are inserted into frames and then subjected to a range of temperatures and voltages. The test can only be performed with the right 'mix' of circuit packs present in the frames. The model may be representative, however, of many assembly-type manufacturing operations in which a number of components must be combined into a single assembly. In such operations, the 'service' may be the actual combining of the components into an assembly or it may be that additional processing is needed after the assembly operation (as in the circuit-pack test mentioned above). We assume that customers (corresponding to the different components of an assembly operation) from \( n \) classes arrive into a queueing station, and their arrival streams correspond to independent Poisson processes each of rate \( \lambda \). The variability in interarrival times may be due to uncertain lead times from vendors, uncertainty in production of the components at upstream operations in an assembly line, etc. Each customer class has its own finite waiting room and there can be at most \( B \) customers of a given class present in the system, where \( B \geq 1 \). There is a single server whose service rate is \( \mu \) and successive service times are independent, identically distributed exponential random variables. The variability in the service time may correspond to some real variability in the processing time, or it may take into account the fact that the assembly process may be subject to breakdown. Only one group at a time can be served, where a group consists of exactly one member from each customer class. The server is idle whenever the customers present do not comprise a group. Arriving customers who find their waiting room full are lost to the queueing system. If the assembly-like operation represents a factory receiving components from vendors, this would approximately correspond to controlling the arrival of components from vendors so that orders for a component are canceled when the inventory for the component reaches some prespecified threshold. Such a policy may stem from a desire not to allow the inventory for any component to become excessive. If the model represents an assembly line within a factory, where \( n \) stations manufacture at a rate \( \lambda \) and feed components to a station that assembles at a rate of \( \mu \), this could represent shutting-off a station when it has accumulated \( B \) components downstream.

A variation of this problem was first studied by Harrison [5], who showed that under fairly general arrival and service mechanisms, the queueing station is unstable when there is no bound on the number of customers in the queue. Latouche [7] studied the problem for two queues and characterized the distribution of the number in the system under a variety of arrival mechanisms in which the number of customers