Int. J. Production Economics 129 (2011) 225–231

Contents lists available at ScienceDirect

Int. J. Production Economics

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

A queueing-inventory system with two classes of customers

Ning Zhao [a,b], Zhaotong Lian [b,*]

[a] Faculty of Science, Kunming University of Science and Technology, China
[b] Faculty of Business Administration, University of Macau, Macau SAR, China

ARTICLE INFO

Article history:
Received 10 August 2009
Accepted 8 October 2010

Keywords:
Queueing systems
Inventory control
Priority
Service rule
Performance measures
Cost analysis

ABSTRACT

We consider a queueing-inventory system with two classes of customers. Customers arrive at a service facility according to Poisson processes. Service times follow exponential distributions. Each service uses one item in the attached inventory supplied by an outside supplier with exponentially distributed lead time. We find a priority service rule to minimize the long-run expected waiting cost by dynamic programming method and obtain the necessary and sufficient condition for the priority queueing-inventory system being stable. Formulating the model as a level-dependent quasi-birth-and-death (QBD) process, we can compute the steady state probability distribution by Bright–Taylor algorithm. Useful analytical properties for the cost function are identified and extensive computations are conducted to examine the impact of different parameters to the system performance measures.

1. Introduction

Research on queueing systems with inventory control has captured much attention of researchers over the last decades. In this system, customers arrive at the service facility one by one and require service. In order to complete the customer service, an item from the inventory is needed. A served customer departs immediately from the system and the on-hand inventory decreases by one at the moment of service completion. The inventory is supplied by an outside supplier. This system is called a queueing-inventory system (Schwarz et al., 2006). The queueing-inventory system is different from the traditional queueing system because the attached inventory influences the service. If there is no inventory on hand, the service will be interrupted. Also, it is different from the traditional inventory management because the inventory is consumed at the serving rate rather than the customers' arrival rate when there are customers queued up for service.

Berman and Kim (1999) analyzed a queueing-inventory system with Poisson arrivals, exponential service times and zero lead times. The authors proved that the optimal policy is "never to order when the system is empty". Berman and Sapna (2000) studied queueing-inventory systems with Poisson arrivals, arbitrary distribution service times and zero lead times. The optimal value of the maximum allowable inventory which minimizes the long-run expected cost rate has been obtained. Berman and Sapna (2001) discussed a finite capacity system with Poisson arrivals, exponential distributed lead times and service times. The existence of a stationary optimal service policy has been proved. Berman and Kim (2004) addressed an infinite capacity queuing-inventory system with Poisson arrivals, exponential service times and exponential lead times. The authors identified a replenishment policy which maximized the system profit. Berman and Kim (2001) studied internet-based supply chains with Poisson arrivals, exponential service times and the Erlang lead times and found that the optimal ordering policy has a monotonic threshold structure. Schwarz et al. (2006) derived stationary distributions of joint queue length and inventory processes in explicit product form for $M/M/1$ queuing-inventory system with lost sales under various inventory management policies such as (r,Q) policy and (r,S) policy. The $M/M/1$ queueing-inventory system with backordering was investigated by Schwarz and Daduna (2006). The authors derived the system steady state behavior under $\Pi(1)$ reorder policy which is $(0, Q)$ policy with an additional threshold 1 for the queue length as a decision variable. Krishnamoorthy et al. (2006a) discussed an (s,S) inventory system with service time where the server keeps processing the items even in the absence of customers. Krishnamoorthy et al. (2006b) introduced an additional control policy (N-policy) into (s,S) inventory system with positive service time. In Manuel et al. (2007, 2008) the perishable queueing-inventory systems with Markovian arrival process (MAP) were studied. The joint probability distributions of the number of customers in the system and the inventory level were obtained for the steady state case. The stationary system performance measures and the total expected cost rate were both calculated.

Some related works in the production industry are He and Jewkes (2000) and He et al. (2002a, 2002b). He and Jewkes (2000) developed two algorithms for computing the average total cost per product and other performance measures for a make-to-order inventory-production system with Poisson arrivals, exponential

* Corresponding author. Tel.: +853 83974163; fax: +853 28838320.
E-mail address: lianzt@umac.mo (Z. Lian).

0925-5273/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpe.2010.10.011

production times and zero lead times. He et al. (2002b) studied the inventory replenishment policy of an $M/M/1$ make-to-order inventory-production system with zero lead times. They explored the structure of the optimal replenishment policy which minimizes the average total cost per product. For the $M/PH/1$ make-to-order inventory-production system with Erlang distributed lead times, He et al. (2002a) quantified the value of information used in inventory control.

All the above studies about queueing-inventory systems are limited to one class of customers. To our knowledge, we have not found any literature studying a queueing-inventory system with two or more classes customers with different priority. In fact such system is popular in reality. For example, in an assembly manufactory buyers with long-term supply contracts have higher priority than buyers who do not. In a hospital, accident victims who are seriously injured will get treatment with priority. The real-life problems stimulate us to study the queueing-inventory system with two classes of customers.

An important issue in the queueing-inventory system with two classes of customers is the priority assignment problem. If two classes of customers are both on the queue, the server needs to make a choice between the two classes of customers whenever it would begin a service. In this paper we figure out the optimal service rule to minimize the long-run expected waiting cost. This is different with the previous study about traditional inventory systems with multiple demand classes in which the optimization problem is based on the inventory costs (see Melchiors, 2003; Nahmias and Demmy, 1981; Teng, 2009; Teunter and Klein Haneveld, 2008).

In this paper we consider a queueing system with inventory management in which two classes of customers arrive at a service facility according to Poisson processes, service times follow exponential distributions, each service uses one item in the attached inventory supplied by an outside supplier with exponentially distributed lead times. The rest of the paper is organized as follows. We define the basic queueing-inventory system with two classes of customers in Section 2 and prove a priority service rule to minimize the long-run expected waiting cost in Section 3. In Section 4 we show the stability condition for the priority queueing-inventory system. In Section 5 we compute the joint steady state distributions. In Section 6, we provide some numerical examples. The paper is concluded in Section 7.

2. The model definition

We consider a queueing system with an attached inventory (see Fig. 1). There are two types of customers—class 1 and class 2. Class-i customers arrive at the system independently according to a Poisson process with rate λ_i ($i=1,2$). There is a single server with unlimited waiting room and customers are served one by one. Each customer requires exactly one item in the inventory for service and the service time follows exponential distribution with parameter μ_i ($i=1,2$). A served customer departs immediately from the system and the on-hand inventory decreases by one at the moment of service completion. If there is another customer in the queue and at least one further item in the inventory, the next service starts immediately. If the server is ready to serve a customer and there is no item of inventory, this service starts only at the time instant (and then immediately) when the next replenishment arrives at the inventory. The demands which arrive during the time the inventory is depleted are backordered.

Because (r,Q) policy is the most common ordering policy in connection with inventory control and it was well investigated in queueing-inventory models in Schwarz and Daduna (2006), Schwarz et al. (2006), Sigman and Simchi-Levi (1993), etc., in this paper we also concentrate on the continuous review (r, Q) policy. When a customer finishes getting served and the on-hand inventory drops to a prefixed level r, an order for Q units is placed. In this paper, we assume $r \geq 0, Q > r$. The condition $Q > r$ ensures that there are no perpetual shortages. Hence the maximum on-hand inventory is $r+Q$. The lead time for the order is exponentially distributed with parameter ν. We assume that a replenishment order in process is never interrupted until it is completed and there is at most one outstanding order at any time.

We call this system $M^1,M^2/M_1,M_2/1$ *queueing-inventory system under* (r,Q) *policy* briefly. It can be described as a four-dimensional Markov process

$$Z(t)=\{N_1(t),N_2(t),\theta(t),I(t),t\geq 0\},$$

where $N_i(t)$ is the number of class-i customers in the system (either waiting or in service), $\theta(t)$ is the class of customer being present at the server, and $I(t)$ is the on-hand inventory at time t. The state space of $Z(t)$ is $\mathbb{S}=\{(n_1,n_2,0,0):(n_1,n_2)\in\mathbf{E}^2\}\bigcup\{(0,0,0,i):i\in\mathbf{F}\backslash\{0\}\}\bigcup\{(n_1,0,1,i):n_1\in\mathbf{E}\backslash\{0\},\quad i\in\mathbf{F}\backslash\{0\}\}\bigcup\{(0,n_2,2,i):n_2\in\mathbf{E}\backslash\{0\},i\in\mathbf{F}\backslash\{0\}\}\bigcup\{(n_1,\ n_2,\theta,i):n_1,n_2\in\mathbf{E}\backslash\{0\},\theta\in\{1,2\},i\in\mathbf{F}\backslash\{0\}\}$, where $\mathbf{E}=\{0,1,2,\cdots\}$, $\mathbf{F}=\{0,1,\ldots,r+Q\}$ and

$$\theta=\begin{cases}0 & \text{there is no customer being served,}\\ 1 & \text{class}-1\text{customer being served,}\\ 2 & \text{class}-2\text{customer being served.}\end{cases} \tag{1}$$

3. Priority service rule

We note that there are two classes of customers in the queueing-inventory system. If two classes of customers are both on the queue, the server needs to make a choice between the two classes of customers whenever it would begin a service. In this section we consider the optimal service rule to minimize the long-run expected waiting cost on the condition that there is inventory on hand.

Let $X(t)=(N_1(t),N_2(t))$ be the queue lengths at time t, b_i be the unit waiting cost rate of class-i customer. The waiting cost rate function should be $w(X(t))=b_1N_1(t)+b_2N_2(t)$, which is charged continuously over time. Whenever a server is idle, the service manager at the service facility can take one of the following three actions: not serve, serve a class-1 customer, or serve a class-2 customer. Denote these actions by 0, 1 and 2 and the action taken at time t by $\phi(t)$. Formally, we have $\phi(t)=\phi(N_1(t),N_2(t),I(t))$. For the queueing-inventory system, if there is no inventory, there is no service. Thus we only discuss the service rule when $I(t)>0$.

Let α be the (continuous) interest rate. We want to find the service policy $\phi=\{\phi(t):t\geq 0\}$ that minimizes the following expected discounted cost over an infinite horizon:

$$E_x^{\phi}\int_0^{\infty}e^{-\alpha t}w(X(t))\,dt. \tag{2}$$

Here we use $x=(n_1,n_2)$ to denote the initial state $(N_1(0),N_2(0))$ and E_x^{ϕ} to denote expectation over demand, given policy ϕ and initial

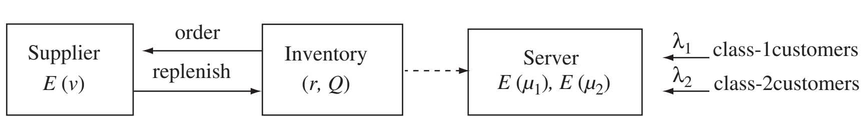

Fig. 1. A queueing-inventory system with two classes of customers.