European Journal of Operational Research 242 (2015) 828–841

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Production, Manufacturing and Logistics

Managing an assemble-to-order system with after sales market for components

Mohsen Elhafsi*, Essia Hamouda

School of Business Administration, University of California, Riverside, CA 92521, United States

ARTICLE INFO

Article history:
Received 24 May 2014
Accepted 21 October 2014
Available online 3 November 2014

Keywords:
Dynamic programming
Assemble-to-order
Production/inventory management
Markov decision process
Markov chains

ABSTRACT

In this paper, we consider an assemble-to-order manufacturing system producing a single end product, assembled from n components, and serving an after sales market for individual components. Components are produced in a make-to-stock fashion, one unit at a time, on independent production facilities. Production times are exponentially distributed with finite production rates. The components are stocked ahead of demand and therefore incur a holding cost rate per unit. Demand for the end product as well as for the individual components occurs continuously over time according to independent Poisson streams. In order to characterize the optimal production and inventory rationing policies, we formulate such a problem using a Markov decision process framework. In particular, we show that the optimal component production policy is a state-dependent base-stock policy. We also show that the optimal component inventory rationing policy is a rationing policy with state-dependent rationing levels. Recognizing that such a policy is generally not only difficult to obtain numerically but also is difficult to implement in practice, we propose three heuristic policies that are easier to implement in practice. We show that two of these heuristics are highly efficient compared to the optimal policy. In particular, we show that one of the two heuristics strikes a balance between high efficiency and computational effort and thus can be used as an effective substitute of the optimal policy.

1. Introduction

Assemble-to-order (ATO) production is a common manufacturing practice used by many firms. It allows a firm to shorten its response time to its customers by stocking inventory of components ahead of demand while delaying the assembly of the end products until demand materializes. Such a practice is especially common for systems where component production lead-times are much significant compared to the assembly time of the end product. Examples of ATO systems include Dell with its online ordering segment as well as the build-to-order strategies of companies such as Toyota, General Motors and BMW, to cite a few. ATO strategies are also used in systems where demand is correlated across several products. Examples include retailers that sell bundled items, order fulfillment at e-retailers, and mail order catalogs.

In this paper, we consider the optimal production and inventory allocation policy for an ATO system facing demand for its end product as well as its individual components. Several manufacturing systems operate under this setting such as white goods manufacturers who not only sell end products such as dish washers, ranges and refrigerators but also sell individual components as spare or repair parts to support their after sales market. According to Cohen et al. (2006), the after sales market is a significant revenue generator, especially in the auto manufacturing and white goods industries, averaging 45 percent of gross profits. As such, managing a system facing demand for both the end product and its individual components becomes an important task. In particular, a manager of such a system faces two types of decisions: components production decision and components inventory rationing or allocation decision. Specifically, a manager has to decide on the number of components to produce and when; and also has to decide how to allocate components inventory in order to satisfy the demand for both the end product and the individual components.

The literature on ATO systems is quite extensive and can be broadly categorized into two classes: one deals with periodic review models while the other deals with continuous review models. In most of the literature within the periodic review class there was no attempt to characterize the optimal operating policy rather the focus was on the performance evaluation of fixed level base-stock policies; example of research papers within this class include Hausman, Lee, and Zhang (1998), De Kok and Visschers (1999), Cheng, Ettl, Lin, and Yao (2002), Zhang (1997), Agrawal and Cohen (2001), and Frank, Ahn, and Zhang (2004). Research within the continuous review literature class is also split into two classes: one class deals with pure inventory systems in which components lead-times are assumed to be exogenous and where the system is modeled as a set of queues with infinite servers

* Corresponding author. Tel.: +1 9518274557.
E-mail addresses: mohsen.elhafsi@ucr.edu (M. Elhafsi), Essia.hamouda@ucr.edu (E. Hamouda).

http://dx.doi.org/10.1016/j.ejor.2014.10.055
0377-2217/

and correlated arrivals; while the second class deals with systems where components lead-times are assumed to be load-dependent and the system is viewed as a set of finite capacity queues with correlated arrivals. These systems involve making decisions about both production and inventory levels of components. Papers within the continuous review category include the work of Song (1998), Song and Yao (2002), Gallien and Wein (2001), Glasserman and Wang (1998), Song, Xu, and Liu (1999), Dayanik, Song, and Xu (2003), Zhao (2009), Lu, Song, and Yao (2010) and the references therein. An extensive review of the ATO literature is provided by Song and Zipkin (2003, chap. 13).

Our paper fits the literature on ATO systems within the continuous review category. However, in our case, we model our problem using an integrated production and inventory control framework and focus on characterizing the optimal policy. Papers dealing with optimal control of integrated production and inventory systems include Ha (1997a, 1997b) and De Véricourt, Karaesmen, and Dallery (2002). These papers dealt with a single product system with multiple demand classes with either the demand being lost if not fulfilled immediately or backordered. The work in this paper is close to the work of Song et al. (1999) who studied a multi-component multi-product integrated production and inventory system where components' inventory is controlled via independent base stock levels, the work of Benjaafar and Elhafsi (2006) who studied a single-product ATO system with multiple demand classes, the work of Elhafsi (2009) who extended Benjaafar and Elhafsi (2006) to the case of compound demands, the work of Benjaafar, ElHafsi, Lee, and Zhou (2011) who studied a general assembly system, and the work of Ceryan, Duenyas, and Koren (2012) who studied a two-stage assembly system with intermediate components demand.

In particular, we consider an ATO system producing a single end product subject to demand for both the end product and individual components. Any demand, whether for the end product or the individual components, that cannot be fulfilled immediately is backordered and satisfied later when stock is available. This assumption usually reflects systems where there are contractual agreements between the manufacturer and the retailer(s) where the demand has to be satisfied (by the manufacturer) sooner or later. Since the parts are exclusively used (after sale) to repair or replace defective components of the end product, customers have no choice but to obtain the parts from the same manufacturer. Hence, in case the manufacturer cannot satisfy the demand immediately, these customers have to wait until components are available again, leading to backordered demand. As we show in the paper, such ATO system involves a sophisticated and challenging mathematical formulation requiring a much larger number of properties that are needed to characterize the optimal policy compared to the relevant literature. Our main contributions, in this paper, consist of the following: because the optimality equation turns out to be very complex, we first show that the optimal value function satisfies a set of preliminary properties. Such properties allow a great simplification of the optimality equation resulting in a major improvement in the execution of the value iteration algorithms used to numerically obtain the optimal policy. A second set of properties is then used to fully characterize the optimal policy structure. The heuristics we propose in this paper, although in the same spirit as those in Benjaafar and Elhafsi (2006), are more sophisticated and efficient. In particular, we decompose the ATO problem into single-component problems that are easier to solve individually which we then link through a correction term to account for demand correlations. We do this by approximating the distribution of the backorder level of the end product in the original n-component system through the use of the distribution of the components inventory level in the single-component system. Such distributions are obtained using effective matrix geometric techniques. This methodology results in performances within 2 percent of the optimal cost, for two of the proposed heuristics, for a very large number of randomly generated problems.

The rest of the paper is organized as follows. In Section 2, we formulate the problem using a Markov Decision Process (MDP) framework. In Section 3, we characterize the structure of the optimal Policy. In Section 4, we propose three heuristic policies and compare their performance against the optimal policy. We conclude the paper with future research directions in Section 5.

2. The mathematical model

We consider a system with similar characteristics as those in Song et al. (1999), Dayanik et al. (2003), and Benjaafar and Elhafsi (2006). Similar to these papers, we assume that demand forms a Poisson process, assembly is instantaneous, and components are produced on separate single-server production facilities with exponentially distributed production times. The assumption of instantaneous assembly is in part motivated by the fact that production or procurement leadtimes are much longer relative to the assembly time. For example, this is the case for Dell where electronic components take months to procure while order processing and assembly takes only few hours. We consider a system consisting of component production as well as final assembly of a single end product. Components are independently produced, in a make-to-stock fashion in anticipation of future demand, on separate production facilities one unit at a time. Production time of components is assumed exponentially distributed with rate μ_k units per unit of time. Assembly time of the end product is assumed negligible compared to the production times of individual components. Similar to other papers that treat ATO systems, this assumption is a natural one as is typical for these systems to have very long component production or procurement lead times compared to the actual assembly time of the end product. The demand for the end product occurs continuously, over time, according to an independent Poisson stream with rate λ. In addition, the system is subject to demand for individual components which occur continuously, over time, according to independent Poisson streams with rate λ_k for component k ($k = 1, \ldots, n$). If demand cannot be satisfied immediately (whether it is for the end product or it is for a component), it is backordered. Completed units of Component k are either placed in stock or used to reduce backorders of Component k (if any) or assembled with other components (if available) to produce a unit of the end product to satisfy its backorder (if any). In this situation, the system manager is faced with two decisions: a production decision and an inventory allocation decision. The manager must decide when to produce Component k and when to idle its production facility if there is too much stock of the component in order to reduce the inventory carrying charge. On the other hand, the manager should produce Component k when the stock level is low to avoid large penalties due to backorders of either the end product or Component k or both. Also, when demand for Component k occurs, and if stock for Component k is available, the system manager must decide whether to satisfy it or backorder it in order to reserve inventory for the end product and avoid larger future backorder penalties of the latter. Hence, a problem of inventory rationing arises.

Since it is conceivable that the system may not immediately satisfy a demand for a component in order to save stock for future demand of the end product, it is possible to have stock and backorder for a component at the same time. Hence, one needs to keep track of both inventory and backorder for all components. Since the end product is assembled instantaneously, we only need to keep track of its number of units that are backordered. Therefore, we define the state of the system by the $2n + 1$ dimensional vector $(\mathbf{X}(t), \mathbf{Y}(t), Z(t))$, where $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ and $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_n(t))$. Here, $X_k(t)$ and $Y_k(t)$ are non-negative integers denoting the inventory and backorder level, respectively, of Component k at time t. $Z(t)$ denotes the backorder level of the end product. Hence, the state space, S, is the Cartesian product $(Z^+)^{2n+1}$ where Z^+ denotes the set of non-negative integers. Let h_k denote the per unit inventory holding cost rate of Component