Int. J. Production Economics 144 (2013) 44–56

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

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

Dynamic pooling of make-to-stock and make-to-order operations

Zhe George Zhang [a], Ilhyung Kim [a,b,*], Mark Springer [a], Gangshu (George) Cai [c], Yugang Yu [d]

[a] College of Business and Economics, Western Washington University, 516 High Street, Bellingham, WA 98225-9077, USA
[b] Ewha School of Business, Ewha Womans University, 11-1 Daehyun-dong, Seoul 120-750, South Korea
[c] Department of Operations and Management Information Systems, Leavey School of Business, Santa Clara University, Santa Clara, CA 95053, USA
[d] School of Management, University of Science and Technology of China, Hefei 230026, Anhui, PR China

ARTICLE INFO

Article history:
Received 24 July 2012
Accepted 6 January 2013
Available online 23 January 2013

Keywords:
Make-to-stock
Make-to-order
Queueing
Capacity sharing
Service interruption

ABSTRACT

It has become increasingly common for companies to offer make-to-stock (MTS) and make-to-order (MTO) versions of the same product through different sales channels. To satisfy these two distinct demand streams, one strategy is to produce both the MTS and MTO items in a single "hybrid" facility partially comprised of flexible servers or machines. We develop a multi-server queueing model of this system, where a subset of the servers or machines is *dynamically* switched between MTS and MTO production via a congestion-based switching policy. We develop analytical formulae for quantifying all major performance measures of the system. We also present a search procedure to find the capacity and inventory control parameters which minimize the total costs of the system while satisfying the customer service constraints. Numerical results are used to illustrate the general behavior of the dynamic hybrid system and to compare its performance to that of a more conventional *static* hybrid facility with *dedicated* MTS and MTO servers. For high levels of traffic intensity, the dynamic system is shown to provide superior customer service for both sales channels with lower finished goods inventory levels.

1. Introduction

As noted by Soman et al. (2004), researchers on production/inventory control systems have traditionally characterized manufacturing operations as either make-to-stock (MTS) or make-to-order (MTO). In an MTS system, production is initiated before demand occurs and items enter finished goods inventory before they are sold to customers. The MTS system is therefore appropriate for mass production where products can be easily standardized; the benefits of the system usually include greater economies of scale and shorter lead times. In an MTO system, on the other hand, production is initiated only after demand is known and each item is delivered directly to the customer after production is completed. Thus, the MTO system, though it could be used for any type of demand, is especially appropriate for mass customization, where consumers place orders catering to their specific needs. The company Shoes of Prey, for example, provides a Web-based design engine allowing potential customers to design their own women's shoes (Shoes of Prey, 2011). After submitting their finalized design, consumers can expect to receive their custom-built shoes in approximately 5 weeks. In general, lead times are longer in an MTO system than in an MTS system and production efficiencies may not be as great, but the appeal of customization to the consumer yields increased demand.

Between these two extremes there can be an intermediate option or a *hybrid* system which combines both MTS and MTO operations. Such hybrid systems have become increasingly common for firms facing both standardized and customized product demand (Williams, 1984; Adan and Van der Wal, 1998; Chang et al., 2003; Soman et al., 2004; Wu et al., 2008; Cattani et al., 2010). Soman et al. (2007) note that the food processing industry provides a good example of coping with both types of demand in a hybrid system; as the demand for customer-specific products has increased relative to the sales of standardized products, companies have been forced to add MTO production to facilities that once were exclusively MTS. Timbuk2, a manufacturer of bicycle messenger bags, produces both standard bags destined for retail outlets and customer-designed bags sold over the internet at the same San Francisco facility (Cattani et al., 2010; Timbuk2, 2011).

Producing MTS and MTO items in the same facility is a challenging task. Firms must avoid stocking out of MTS items; they must satisfy MTO demand within a competitive time window; and they must maintain overall production efficiency. One approach to achieving these goals in a hybrid facility is to

* Corresponding author at: College of Business and Economics, Western Washington University, 516 High Street, Bellingham, WA 98225-9077, USA. Tel.: +1 360 650 2428; fax: +1 360 650 4844.
E-mail addresses: george.zhang@wwu.edu (Z.G. Zhang), ilhyung.kim@wwu.edu, ilhyunk@wwu.edu, i.kim@ewha.ac.kr (I. Kim), mark.springer@wwu.edu (M. Springer), gcai@scu.edu (G. Cai), ygyums@gmail.com (Y. Yu).

0925-5273/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2013.01.012

divide the facility's machines into two groups, one dedicated to MTS production and the other committed to satisfying MTO demand. This *static* approach effectively separates the facility into two independent production units, with each unit having its own distinct demand and the responsibility for meeting that demand. While this approach has the advantage of relative organizational simplicity, a changing mix of MTS and MTO demand can result in the simultaneous underutilization of one unit and below-average customer service for the other unit. By dedicating capacity to either the MTS or MTO part of the facility, a level of inflexibility is built into the static hybrid facility.

An alternative approach to organizing a hybrid MTS–MTO facility is to include, along with machines dedicated to either MTS or MTO production, a group of flexible machines which can switch between production of MTS and MTO products. In the face of a changing product mix, this *dynamic* approach offers the potential for superior customer service *and* lower inventory levels than a corresponding static hybrid facility possessing only dedicated MTS and MTO capacity. In a dynamic hybrid system, the set of flexible machines would switch between MTS and MTO production depending upon demand and inventory status. For example, if MTS inventory dropped perilously low, the flexible capacity could be assigned to help build MTS inventory, while reaching sufficient levels of MTS inventory could switch the flexible machines to assist in reducing the MTO backlog. Notwithstanding its intuitive appeal, however, the superiority of the dynamic hybrid system across all performance measures is not guaranteed. Although moving from a static to a dynamic hybrid system is likely to reduce the slack capacity of machines, the increased variability in processing rates generated by the dynamic system may lead to *decreased* performance of the MTS or MTO shop depending on the values of certain system parameters and decision variables. Of critical importance in setting up such a dynamic hybrid facility, therefore, is determining the desired number of dedicated and flexible machines, as well as the appropriate inventory control parameters, given the arrival and service rates for the MTO and MTS shops.

We develop an analytical model of a multiple-machine dynamic hybrid MTS–MTO facility which is capable of efficient standardized production and mass customization. To the best of our knowledge, this is the first attempt to use the multi-server model to analyze the operations of a hybrid MTS–MTO system. We propose a simple policy for switching a select group of flexible machines between MTS and MTO production, and show that functions defining key system performance measures are unimodal functions of the same system parameter. This enables us to develop an algorithm which finds the minimum number of overall machines necessary to satisfy customer service requirements, and then determines the machine allocation and inventory policy parameters needed to minimize the production switching and inventory holding costs per unit of time.

For situations where reducing the number of machines required is not the most important objective, our model can be used to evaluate the costs and benefits of switching from a static to a dynamic system. We consequently compare the performance of the two systems and find that the dynamic system's relative performance is best when MTO and MTS demand is high in relation to overall system capacity: in such circumstances, moving to a dynamic system could result in a decrease in finished goods inventory and an improvement in customer service for both MTO and MTS customers.

The reminder of the paper is organized as follows. In Section 2, we provide a brief overview of the existing research in hybrid MTS–MTO systems. In Section 3, we propose a dynamic hybrid model which combines make-to-stock and make-to-order operations. Major performance measures, including service level requirements, are introduced. This section also outlines an algorithm to determine the optimal capacity configuration and inventory control decisions for the dynamic hybrid system. In Section 4, we compare the relative performance of a dynamic hybrid MTS–MTO system vis-à-vis its static equivalent and discuss the managerial implications of these results. Finally, conclusions and directions for future research are presented in Section 5.

2. Literature review

Comparatively little research has been conducted on multiple-machine hybrid MTS–MTO systems. Several authors have investigated a related but somewhat different problem: given a single machine capable of producing a finite set of products possessing different demand rates, which products should be produced to stock and which should only be produced upon receiving a customer order (Williams, 1984; Carr et al., 1993; Sox et al., 1997; Arreola-Risa and DeCroix, 1998; Rajagopalan, 2002; Soman et al., 2007). These single-machine hybrid systems may be properly considered dynamic since they enable switching between MTS and MTO production on a single machine. However, the MTO demand considered by these authors is not customized demand, but rather represents low-volume standardized products that a producer may choose, for cost and efficiency reasons, not to stock in inventory. Work in this area has focused on partitioning the set of products into those which are most effectively produced to stock and those which are not.

In addition, a handful of authors have examined hybrid single-machine MTS–MTO systems where the identification of a product as MTS or MTO is stated as given and not a decision variable for management. Consequently, researchers taking this approach have focused on building tractable approximations of single-machine systems for performance analysis and developing insights into system policies and management. Nguyen (1998) develops performance measure approximations for a single-machine system which produces a fixed number of MTS and MTO products. Federgruen and Katalan (1999) examine a system where a single MTO product has been introduced onto a single machine previously dedicated to MTS production.

Some authors, however, have considered a hybrid multi-machine MTS–MTO system for a given set of MTS and MTO products. Chang et al. (2003) and Wu et al. (2008) investigate, using simulation, various types of job release and dispatching rules for a hybrid wafer construction facility. Cattani et al. (2010) compare three different hybrid strategies: focused, spackling, and layered spackling strategies. The focused strategy utilizes two dedicated plants, one for MTS and the other for MTO products; the spackling strategy employs a single flexible plant which can produce both MTS and MTO products; and the layered spackling strategy utilizes a flexible plant and a dedicated plant for MTS products only. They develop an analytical model to determine the optimal capacity decision using a cost-minimization criterion, and show which strategy is optimal under which conditions. Because of the nature of the spackling or layered spackling strategies, which utilizes the slack capacity of the flexible plants as much as possible, the utilization of the plants tends to reach near 100%. However, the queueing effect, which increases substantially as the traffic intensity approaches one, is not considered in their analysis.

Our work is similar to these earlier efforts in that we investigate a hybrid MTS–MTO system that contains machines capable of dynamically switching between MTS and MTO production. Similar to previous studies (Nguyen, 1998; Federgruen and Katalan, 1999; Chang et al., 2003; Wu et al., 2008; Cattani et al., 2010), we assume that the distinction between MTS and MTO products is determined by the customer's desired level of