Available online at www.sciencedirect.com
SCIENCE DIRECT®
Int. J. Production Economics 90 (2004) 169–186
international journal of production economics
www.elsevier.com/locate/dsw

Performance analysis of make-to-order manufacturing systems under different workload control regimes

A. Haskose*, B.G. Kingsman, D Worthington

Department of Management Science, Lancaster University, Management School, Bailrigg, Lancaster LA1 4YX, UK

Received 13 May 2002; accepted 28 February 2003

Abstract

This paper firstly discusses the make-to-order (MTO) manufacturing sector to show the different types of queuing network that may exist and need to be covered. The workload control (WLC) production planning method is modelled as a queuing network with limited buffer capacities in front of each workstation. Exact solutions for a general network with more than 3 or 4 workstations are not possible. An approximation algorithm, as an extension of earlier work on simple tandem queuing networks, has been developed to cope with any number of workstations and to allow flows forwards and backwards between the workstations. The essentials of the model and solution algorithm are briefly described. The second half of the paper presents the results of using the model and algorithm to analyse four issues in WLC in MTO. The first set of experiments examines the relative value of two WLC mechanisms for controlling manufacturing lead times, job release and order acceptance. The second set of experiments is to gain insight into how increased complexity in production layouts and the product variety impact on manufacturing performance measures. The third set of experiments examines the differential effects of extra buffer capacity at earlier or later workstations in the main path flow; whilst the final set of experiments examines the impact of having groups of high- and low-priority jobs.

Keywords: Manufacturing; Job shops; Queuing theory; Markov processes; Workload control; Make-to-order production

1. Introduction–the produce-to-order manufacturing environment

Manufacturing companies differ in the way they meet their demand. Some deliver products to their clients from finished goods inventories as their production anticipates customers' orders; others, however, manufacture only in response to customers' orders. Orders for the products tend to be on a *make-to-order, make-to-print* or *engineer-to-order* basis (MTO), often being specific to a particular customer with intermittent or no repetition of demands for the same product.

The typical company in the produce-to-order manufacturing sector has to supply a wide variety of products, usually in small quantities, ranging from a range of standard products to all orders requiring a customised product. The arrival of customer enquiries is a stochastic process over time. Each potential order from the enquiry tends

*Corresponding author. Tel.: +44-1524-594471.
E-mail address: a.haskose@lancaster.ac.uk
(A. Haskose).

0925-5273/03/$ - see front matter
doi:10.1016/S0925-5273(03)00052-5

to be for a differing number of units and requires varying routings and processing times through the production facilities.

Each order requires processing (transformation work) on a series of workstations. Jobs enter the production system and go to the first workstation in their routing sequence. They typically join a queue of other jobs waiting their turn for their processing work to be carried out. Once the work on a job at a workstation is completed, the job is transported to the next workstation in its routing sequence, where it again joins a queue of jobs awaiting processing. The manufacturing lead time is thus the sum of the set-up and processing times at each of the workstations in the job's routing sequence plus all of the time spent waiting in queues in front of the workstations. It is well known that in the produce-to-order sector an order can spend up to 90% of the total time in production waiting in front of or between workstations. It is reported that manufacturing lead times are often long and unreliable almost entirely due to the large proportion of time spent in the queues, see Stommel (1976) and Stalk and Hout (1992).

A general model of the shop floor is a network of workstations each with a set of orders (jobs) queuing waiting their turn to be processed. Clearly, the production process can be viewed as a network of queues, of different types as illustrated in Fig. 1.

MTO companies differ in the degree of product customisation, covering pure customisation, tailored customisation, standard customisation and non-customisation, and the amount of product variety. Some companies offer standard products that are expensive to produce to meet intermittent low demands. So in such cases they would not be kept in stock but only produced to order. As there is repetitive production, albeit on an intermittent basis, the products are likely to be manufactured on a single production line (Fig. 1a) or on assembly lines. The assembly line is better known as a flow shop. In a flow shop, the products move in one direction only with only one operation at each workstation, as illustrated in Fig. 1b. There may be multiple entry and leaving points in the system. A purely customised product is one that is developed from scratch for each customer. Since almost every product is unique and produced in a different way, production tends to be via a job shop layout, see Fig. 1c. Most of the literature on the production planning problems of produce-to-order companies has focused on the individual workstation level. The bulk of this has concentrated on priority dispatching; the order in which jobs should be scheduled through each workstation. Surveys show that hundreds of such priority rules have been devised for application on the shop floor. However, experience has demonstrated that priority dispatching is a relatively weak mechanism for the control of queues. If used alone, it has little effect in reducing the long lengths and high variability of the queues of jobs in front of machines. A stronger instrument, controlled job release, entails maintaining a 'pool' of unreleased jobs in the production planner's office, which are only released onto the shop floor if doing so would not cause the planned queues to exceed some pre-determined norms. This in turn reduces the work-in-process and the task of

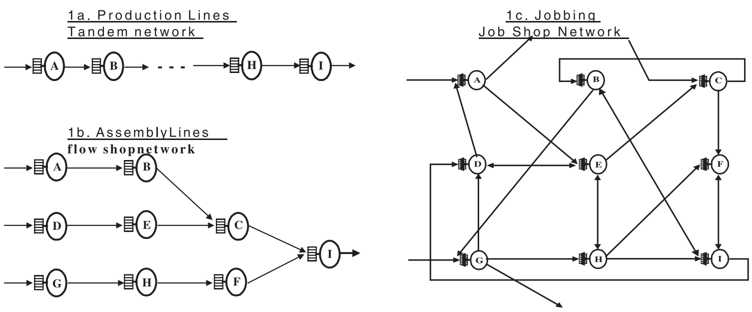

Fig. 1. In MTO manufacturing the production system will be an arbitrary queuing network, which can take a variety of forms, where A, B, C,…,I represent the different workstations and the arrows denote job flows.