Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 63 (2017) 639 – 644

www.elsevier.com/locate/procedia

The 50th CIRP Conference on Manufacturing Systems

Delivery lot splitting as an enabler for cross-docking and fast delivery

Georg Bienert[a,*], Bernard Kornfeld[a], Sami Kara[a]

[a] *Sustainable Manufacturing & Life Cycle Engineering Research Group, School of Mechanical & Manufacturing Engineering, The University of New South Wales, Sydney, NSW 2052, Australia*

* Corresponding author. Tel.: +61 (2) 9385-6851. *E-mail address:* g.bienert@unsw.edu.au

Abstract

In business sectors like the pharmaceutical industry fast customer order deliveries can be crucial both for the businesses' and the customers' viability. To accomplish this, finished products need to be stocked close to the customers at the expense of higher inventory holding and obsolescence costs in the distribution centres. Cross-docking of full orders as an approach for increasing the inventory efficiency is not feasible where delivery distances for replenishment of the distribution centres are long.

An approach that could enable cross-docking despite rigid delivery lead time constraints is the "partial shipment strategy". This delivery strategy splits large order volumes into smaller delivery lot sizes that are shipped to the customers at different points in time. This is based on the knowledge that many customers do not consume the entire ordered quantity at once. Several numerical studies have shown that the optimal delivery lot size can reduce inventory levels and increase the general supply chain efficiency. However, they fail to capture the details associated within complex distribution systems. Therefore, in this paper an agent-based distribution system model is developed in order to simulate the partial shipment strategy. A real industry case is used to investigate implications of implementing such a strategy in a warehouse environment. The developed model allows users to analyse the impact of split deliveries on the operations and inventory levels within distribution centres as well as the viability of cross-docking in a warehouse environment without compromising customer lead time.

Keywords: Cross-docking; Partial Shipment; Agent-based Simulation Modelling

1. Introduction

Many production companies face the same simple problem: if they want to guarantee a high fill rate and quick delivery lead times, they need to stock a huge amount of finished products in their customer facing warehouses. Especially industries with slow supply chains and volatile demands like the pharmaceutical industry suffer from this problem [1].

However, storing finished goods bonds working capital, causes storage costs, consumes resources and can lead to operational disruption if storing space is limited. Instead of adding value to the product, storing products over time may even do the opposite and deteriorate perishable products, which is the case for many pharmaceuticals [2].

A strategy that can decrease inventory storage is cross-docking. With cross-docking, replenishment shipments are not put onto shelf racks in storages [5]. Instead, inbound shipments from upstream warehouses arrive at the distribution centres just-in-time to be sorted, consolidated, reloaded and sent out to customers [3].

Cross-docking is enabled by logistics postponement. The inbound shipment to distribution centres is postponed until actual customer demand is issued [4]. As a result, the logistics activities upstream the distribution centres' are initiated by an actual demand signal. They do not follow a risky forecasting based plan [5].

However, due to the additional time for inbound shipment, logistics postponement can extend the customer order lead time [6]. This is not viable where customers demand quick delivery.

doi:10.1016/j.procir.2017.03.151

Fortunately, there are countermeasures that prevent longer order lead times of logistics postponement. One such method is cooperation between warehouses [2]. Another measure is increase in flexibility [3], i.e. the flexibility to use faster means of shipment for reducing transportation lead time.

A cooperative supply chain strategy that uses flexibility in the order fulfilment is the partial shipment strategy. The partial shipment strategy permits the splitting of orders in smaller batches that are delivered at different points in time. This strategy was developed by Banerjee et al. and Dawande et al. to fulfil orders at least partially in case of insufficient on-hand inventory [7, 8]. It can prevent lost sales [8, 9] or avoid stock out penalties [7]. The authors pre-suppose that some customers order larger quantities than they will consume at once. Instead they will need it over (a known) time [7, 8]. If an order is split, the first batch could be sent out from stock with short delivery lead times. While it is consumed by the customer, the distribution center could source the remaining quantity from the upstream supplier and use cross-docking for shipping it to the customer just-in-time. Where this strategy is feasible, a distribution centre will need to carry sufficient inventory only to cover the first partial deliveries.

Unfortunately, express delivery increases the shipping cost per delivery and partial shipment requires multiple deliveries for each order. Hence, both measures are likely to increase transportation expenses and a trade-off between additional transportation cost and possibly lower warehousing cost is evident. Supply chain managers need to balance this trade-off carefully. Therefore, they need decision support tools.

The aim of this study is to develop such a tool. It shall be able to predict for any given distribution system, if cross-docking of customer orders will be feasible without delaying customer order lead time if strategies for increasing flexibility or cooperation are applied. It shall also compare these strategies with the conventional distribution strategy regarding their order fulfilment cost.

Research has provided many decision support approaches for distribution planning applications. Some of them are reviewed in the following section.

2. Literature review

Generally, logistics postponement affects the entire production-distribution system which comprises multiple echelons. Multi-echelon distribution chains are complex systems due to the high number of independent actors, their interdependencies and the vast numbers of degrees of freedom. Thus, determining a feasible batch size is a complex task. Therefore, most numerical solution approaches for the partial shipment problem have to use complex algorithms often combined with advanced mathematical solution methods such as heuristics [5]. Yet, the underlying models of many of these approaches have limited scope, use simplistic assumptions and minimal details. These simplifications may be necessary to make the complex systems mathematically tractable. Nonetheless, these models do not represent a real distribution system very well. They still may provide insights into supply chain performance and dependencies but they might not be suitable as a decision support tool [8].

A more promising approach to cope with the mathematical complexity of multi-echelon supply chain systems is multi agent simulation (MAS). A multi-agent system is a network of autonomously operating entities (agents) that can communicate with each other. Agents can be structured in hierarchies and act towards local and global objectives [9]. Thus, they can represent autonomous echelons in a hierarchical supply chain. Research has provided a number of MAS frameworks for the coordinated supply chain context (i.e. [10][11]). These frameworks have in common that supply chain entities and echelons are created as generic agents in an object-oriented manner. As a consequence, agents can be replicated at any given number. All replicates share the same categories of attributes, feature characteristics and communication interfaces. Thus, MAS models can represent any supply chain network with little modelling effort.

MAS frameworks can be combined with discrete event simulation techniques. The applicability of these approaches in the coordinated SC context was successfully proven by several studies. Lee and Kumara used this technique for dynamic determination of feasible and cost-effective shipment batch sizes [12]. Umeda and Zhang compared centralized push- and decentralized pull-models in a complex supply chain model regarding their impact on order lead time, inventory levels and shortage [13]. Essoussi compared centralized and decentralized replenishment decisions in a hospital environment. In this study replenishments are cross-docked in a central pharmacy and delivered to the individual departments. However, transportation lead times within the hospital are not regarded [14]. Mishra and Wadhwa simulate different postponement decisions in a complex supply chain model. They show that transshipment as a supply chain coordination initiative counters the negative lead time performance of postponement. Moreover, they conclude that MAS is a superior methodology for modelling complex, stochastic supply chain systems [2].

Concluding the reviewed literature, there have been several numerical studies that calculate optimal shipment quantities in a partial shipment policy. Yet, there is no work that analyses the use of partial shipments to make cross-docking or logistics postponement viable. Furthermore, it can be concluded that MAS is a promising method for this analysis. Therefore, it is the proposed methodology of this study.

In the following section, the methodology to assess the feasibility of the partial-shipment enabled cross-docking strategy is explained. Based upon a theoretical framework of the distribution chain, the characteristics of the generic and hierarchical MAS model are shown. This method is applied to a simulation study of a pharmaceutical company in section 4.

3. Methodology

3.1. The theoretical distribution chain framework

In this paper a hierarchical multi-echelon distribution chain network is considered. The number of levels is variable as well as the number of echelons per level. Consumers order variable quantities of multiple products at random points in time. Their consumption frequency is known from previous