Simulation Modelling Practice and Theory 43 (2014) 54–66

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

The effect of backlog queue and load-building processing in a multi-echelon inventory network

Yisha Xiang [a,*], Manuel D. Rossetti [b]

[a] *Sun Yat-sen Business School, Sun Yat-sen University, 135 W. Xingang Road, Guangzhou, Guangdong 510275, China*
[b] *Department of Industrial Engineering, University of Arkansas, 4207 Bell Engineering Center, Fayetteville, AR 72701, USA*

ARTICLE INFO

Article history:
Received 9 June 2013
Received in revised form 2 January 2014
Accepted 27 January 2014
Available online 21 February 2014

Keywords:
Multi-echelon
Supply chain management
Simulation
Inventory control
Supply chain performance

ABSTRACT

A continuous-review, two-echelon inventory system with one central warehouse and an arbitrary number of non-identical retailers is considered in this study. Retailers face independent Poisson demands and apply standard (r, Nq) policies. Filled orders at the central warehouse must be consolidated into loads before shipping to the retailer level. New modeling options for the backorder processing and load-building processes are considered. Employing simulation, a set of experiments is performed to illustrate how different processing rules for the backlogging and load-building queues affect the lead-time experienced at the retailer level. Simulation results indicate that there are cases where considerable improvements can be gained from using different processing rules in backlogging and load-building queues.

1. Introduction

The focus of this paper concerns the processing of backorders and the building of replenishment loads within a multi-echelon inventory system. Multi-echelon inventory systems consist of two or more echelons where locations at upper echelons act as suppliers to locations at lower echelons. Fig. 1 illustrates a three-echelon system, in which a set of warehouses at the 2nd level are supplied by a single external supplier. The warehouses, in turn, re-supply the replenishment requests from their assigned retailers at the lowest level. Inventory for multiple types of items may be stocked at each location.

Customer demand occurs at the lowest level of the network and is propagated upwards through the network as replenishment requests. This tree-like structure is widely seen in distribution inventory systems which are divergent in general, and particularly in a pure distribution system, or arborescent system, each stock has at most a single immediate predecessor [9]. The arborescent system is considered in many multi-echelon supply chain networks (see examples in Caggiano et al. [10] and Caggiano et al. [11]). The tree-like structure allows for the modeling of complex supply chains of practical importance to industry. In general, other support patterns (e.g. multiple suppliers, lateral shipments, etc.) are possible.

Farasyn et al. [20] provide an illustration of multi-echelon inventory networks within Proctor & Gamble. They describe tools that were used to save over \$1.5 billion in costs. Thus, inventory networks of this form are of important practical interest. The support structure can be specified for each item. For example, Farasyn et al. [20] describe networks as large as 4 to 5 thousand stages (locations) with 6 to 10 thousand arcs, which has motivated the development of specialized algorithms to handle such complexity (see, for example, Al-Rifai and Rossetti [1]).

---

[*] Corresponding author. Tel.: +86 20 84112601.
E-mail addresses: xiangysh@mail.sysu.edu.cn (Y. Xiang), rossetti@uark.edu (M.D. Rossetti).

http://dx.doi.org/10.1016/j.simpat.2014.01.006
1569-190X/

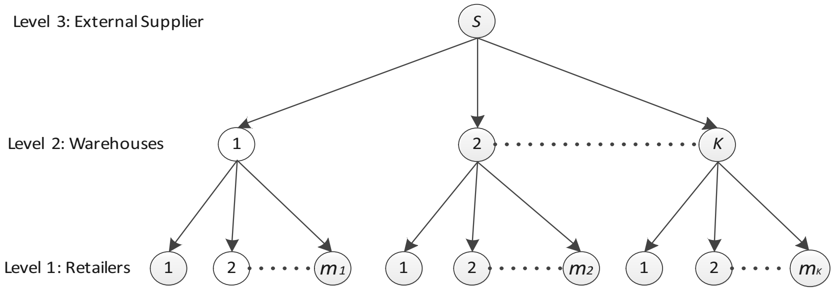

Fig. 1. Multi-echelon inventory system.

The interactions between the echelons make it difficult to adequately model the performance characteristics of the locations, because the item's lead time at any location is a random variable that is a function of the inventory control policies at the supplier location, which in turn depends on its supplier, and so on. At any location the expected waiting time for a replenishment to be filled depends upon any delay due to a stockout. This implies that the replenishment lead-time is a function of the waiting time experienced at the next higher echelon. Thus, the locations are tied together through the transport time between the locations and the stochastic waiting that may occur for replenishment requests. In addition to the waiting time associated with a backorder, a replenishment request may experience additional delays as a location builds an economical shipment size. The modeling and effects of these two delays (backorder delays and load-building delays) are the focus of this paper. The objective is to understand how the processing of these queues interacts with the inventory policies specified for each item.

In practice, most of the backlogging and load building queues are processed using a first-in-first-out (FIFO) rule. Following the FIFO rule, a backlogged order requiring only a small amount of a certain product might have to wait for an earlier order that requires large amount of the same product, which results in long average waiting time and poor customer response times. The same problem also exists in the load building process using the FIFO rule. Motivated by the smallest processing time rule used in scheduling [33], we are interested in exploring the potential benefits from implementing similar priority rules in both backlogging and load building queues. Our goal of implementing these priority rules is to make the multi-echelon inventory networks more effective in terms of lead time and service levels by giving priorities to orders with small quantities in the backlogging queues and to orders with larger shipping weight in the load building queues. Multi-echelon inventory network is often mathematically complex. Gümüs and Güneri [21] indicate that simulation is the most used research technique in their review of over 92 references in this area. The incorporation of priorities rules in backlogging and load building queues further complicates the problem. Therefore, a simulation approach is more desirable.

This paper is organized as follows. Section 2 provides a brief discussion of the relevant literature. Section 3 describes details of the simulation model used in this paper. The emphasis will be to ensure that the credibility of the model is accepted, so that the results provided in Section 4 are understood. The results in Section 4 illustrate the significant findings of the series of experiments. Section 5 provides managerial insights gained from the simulation results. Finally, the last section describes some areas for future research.

2. Background and literature

Overviews of inventory modeling can be found in Axsäter [9], Zipkin [47], Muckstadt [32], Tempelmeier [43], and Silver et al. [41]. Research on multi-echelon inventory systems has been an active area in the literature for decades (see for example Clark and Scarf [18], Sherbrooke [39], and Sherbrooke [40]). Gümüs and Güneri [21] provide a survey of the area with over 92 references. The model described in Clark and Scarf [18] sets the stage for determining the safety stock within the multi-echelon inventory system based on decomposition. First, the most downstream echelons meet customer demand. Shortage at the next echelon leads to stochastic delay which creates an additional cost. This additional cost affects the process of determining the optimal policy for the next upstream installation. For stochastic multi-echelon inventory systems, the seminal paper by Sherbrooke [39] presents a model for determining the optimal stock levels at the bases (warehouse) and the depots (retailers) in order to minimize the total number of outstanding backorders at the depot level for a given amount of investment. Additionally, Deuermeyer and Schwarz [19] develop analytical models to approximate the performance of a multi-echelon supply chain network by assuming (r, Q) policies with stationary Poisson demand. The model was applied to a system consisting of one warehouse that supplies N retailers to obtain the expected service level, including fill rate and backorders.

Building on the Deuermeyer and Schwarz model, Svoronos and Zipkin [42] refine the model in order to derive a different approximation for the mean and variance of the warehouse lead-time demand. They approximate the warehouse lead-time demand using a mixture of two translated Poisson distributions (MTP). Using the MTP, they approximate the performance measures at the warehouse, such as the expected number of backorders, which they then used to calculate the delay at the warehouse due to stockout. This approximation works well for the identical retailer case, especially for Poisson customer