Int. J. Production Economics 143 (2013) 546–552

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

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

Near-optimal heuristics to set base stock levels in a two-echelon distribution network

R.J.I. Basten [a,*], G.J. van Houtum [b]

[a] University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[b] Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Available online 28 October 2011

Keywords:
Service logistics
Spare parts inventories
Heuristic
Two-echelon
Distribution network

ABSTRACT

We consider a continuous review two-echelon distribution network with one central warehouse and multiple local stock points, each facing independent Poisson demand for one item. Demands are fulfilled from stock if possible and backordered otherwise. We assume base stock control with one-for-one replenishments and the goal is to minimize the inventory holding and backordering costs. Although this problem is widely studied, only enumerative procedures are known for the exact optimization. A number of heuristics exist, but they find solutions that are far from optimal in some cases (over 20% error on realistic problem instances). We propose a heuristic that is computationally efficient and finds solutions that are close to optimal: 0.2% error on average and less than 5.0% error at maximum on realistic problem instances in our computational experiment.

1. Introduction

Capital goods are expensive, technologically advanced products or systems that are critical for the companies that are using them. Examples are MRI-scanners in hospitals, baggage handling systems at airports, or radar systems on board naval vessels. Since the availability of these systems is so important, spare parts are generally located close to the installed base. However, since spare parts are often expensive, also more central stocking locations at higher echelon levels are used to obtain risk pooling effects. Although in principle any number of echelon levels is possible, spare parts networks for capital goods often have a two-echelon structure (Cohen et al., 1997). In such networks, many different items are stocked: Gallego et al. (2007) give the example of General Motors' spare parts organization, which manages over four million stock-keeping units.

We consider a single-item, two-echelon distribution system with continuous review base stock control, one-for-one replenishments, and first-come first-served (FCFS) allocation at the central warehouse. We assume Poisson demand processes at the local stock points. Demands are backordered if they cannot be fulfilled immediately, and a penalty is paid per unit time for items that are on backorder. This model is widely studied and can be evaluated exactly. However, only enumerative procedures are known for the exact optimization of the base stock levels. Because in practice this optimization has to be executed for thousands of components, efficient heuristics are required.

Our main contribution is that we develop two such heuristics. One always finds the optimal solution in our computational experiment. The other finds base stock levels that lead to costs that are much closer to optimal than the costs found by existing heuristics, while it is still faster than most existing heuristics. A second contribution is that we derive new properties of the total cost function.

The remainder of this paper is structured as follows. We give a literature review in Section 2. In Section 3, we give the mathematical problem formulation, and in Section 4, we show some analytical properties of the cost function. In Section 5, we describe the heuristics that we propose. We show the results of the computational experiment in Section 6. In Section 7, we give conclusions and recommendations for further research.

2. Literature review

Because of the practical relevance, there has been a lot of interest in distribution networks, starting with the seminal papers by Clark and Scarf (1960) for *serial* (*N*-echelon) *networks*, which are networks in which each stock point has at most one successor and at most one predecessor, and Sherbrooke (1968) for general *distribution networks*, which are networks in which each stock point has (at most) one predecessor.

The paper by Clark and Scarf (1960) started the research on *periodic review* models. For distribution networks, Eppen and Schrage (1981) introduce the balance assumption, which

* Corresponding author. Tel.: +31 53 489 4007.
E-mail addresses: r.j.i.basten@utwente.nl (R.J.I. Basten), g.j.v.houtum@tue.nl (G.J. van Houtum).

0925-5273/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpe.2011.10.018

effectively means that the inventory position of a local stock point immediately after ordering is allowed to be lower than the inventory position just before ordering. In other words, after ordering, the inventory is balanced over the various stock points even if it was unbalanced before ordering. Using this assumption, the optimal policy can be characterized under various circumstances, see, e.g., Diks and de Kok (1998, 1999) or Van der Heijden et al. (1997).

Instead of periodic review, we consider *continuous review*. We further assume base stock control and first-come first-served (FCFS) allocation at the central warehouse. This stream of research started with the paper by Sherbrooke (1968) and overviews are given in Sherbrooke (2004) and Muckstadt (2005). Van Houtum (2006) gives a more general overview of literature on multi-echelon inventory systems.

We will discuss the most relevant literature in more detail: First, we discuss two papers in which the model that we use is evaluated exactly and second, we discuss a number of papers in which heuristics are developed to optimize the base stock levels.

For the model that we consider, Axsäter (1990) gives recursive formulas to evaluate the inventory holding and penalty costs. Furthermore, he finds that the costs are convex in each of the local base stock levels, but not in the central base stock level. Axsäter (1990) gives a lower and upper bound on the optimal central base stock level, between which he enumerates all possible values to set optimal base stock levels in the network. Because of his recursive formulas, this enumeration is still relatively fast, e.g., in comparison with the exact evaluation that Graves (1985) proposes.

Graves (1985) develops both an exact and approximate procedure to determine the distribution of the items in inventory and on backorder at the local stock points. If items are on backorder at the central warehouse, they delay the replenishments at the local stock points. In the exact evaluation, Graves (1985) calculates a convolution, which is computationally intensive. In the approximate procedure, he fits a negative binomial distribution on the first two moments of the distribution of the number of items on backorder at the central warehouse. Optimization of the base stock levels is not considered.

Heuristics for this problem exist as well. Gallego et al. (2007) compare FCFS allocation with other forms of allocation in which central information is required. Although this is very relevant in certain business situations, we consider FCFS allocation only here. Gallego et al. (2007) develop the *restriction-decomposition* heuristic (*RD*), in which they restrict the central base stock level to have one of three possible values: zero ('cross-docking'), the expected lead time demand at the central warehouse ('zero safety stock'), or the maximum value at the central warehouse if each local stock point would hold no inventory ('stock-pooling'). Using each of these three values, the base stock levels at each of the local stock points can be optimized independently using simple newsvendor equations.

Rong et al. (2010) propose two different heuristics for distribution systems with a *general* number of echelon levels. In the *recursive optimization* (*RO*) heuristic, first the base stock levels of all local stock points (lowest echelon level) are optimized assuming that there are no delays at higher echelon levels. These problems reduce to simple newsvendor equations. Then the next higher echelon's base stock levels are optimized using the base stock level at the lowest echelon level as an input. This cost function is convex and can be solved easily. Then the next higher echelon level can be solved (if existent), etc.

In the *decomposition-aggregation* (*DA*) heuristic of Rong et al. (2010), the problem is decomposed into one serial system per local stock point. Each of these serial systems consists of the path from the most central location to (and including) the local stock point. Each non-local stock point is part of a number of serial systems. The backorders at such a stock point that result from the optimization of each of the serial systems are summed (aggregated) and it is determined which base stock level is required at that stock point to achieve the summation of the backorders. This is the final base stock level.

3. Problem description and notation

We consider a two-echelon distribution network consisting of one central warehouse and N local stock points in steady state. The central warehouse has index 0 and the local stock points are indexed $1,\dots,N$. The demand process at local stock point i is a Poisson process with rate $\lambda_i(>0)$. A demand is backordered if it cannot be fulfilled immediately. Local stock point i uses a base stock policy with base stock level $S_i(\geq 0)$. Hence, each demand at a local stock point is immediately followed by a replenishment order at the central warehouse. As a result, the demand process at the central warehouse is also a Poisson process, with rate $\lambda_0=\sum_{i=1}^{N}\lambda_i$. Demands that cannot be fulfilled immediately at the central warehouse are backordered and fulfilled first come first served (FCFS). The central warehouse uses a base stock policy with base stock level $S_0(\geq 0)$.

The central warehouse orders components at an external supplier with infinite supply or, equivalently, repairs components using an uncapacitated repair facility. The total lead time from the occurrence of a demand until the central warehouse receives a new or repaired component is a random variable with mean $L_0(>0)$. Successive lead times are i.i.d. Each shipment from the warehouse to local stock point i takes $L_i(\geq 0)$ time units (deterministic).

We consider holding costs $h_0(>0)$ at the central warehouse and $h_i(>0)$ at each of the local stock points i for each unit time that a spare part is in stock (considering holding costs for spare parts in transit is straightforward since the average number of spare parts in transit is a constant given the fact that we use backordering). Furthermore, we consider penalty costs $\beta_i(>0)$ per unit per unit time for backordered demand at local stock point i.

We need the following additional notation for the cost function that we want to minimize and the analysis in Section 4. The number of items on order at the central warehouse is X_0. This number is Poisson-distributed with parameter $\lambda_0 L_0$. The corresponding number of backorders is $B_0(S_0)=(X_0-S_0)^+$ and the number of items on hand is $I_0(S_0)=(S_0-X_0)^+$. $B_0^{(i)}(S_0)$ is the number of backorders at the central warehouse that is due to local stock point i ($\sum_{i=1}^{N}B_0^{(i)}(S_0)=B_0(S_0)$). The number of items on order at local stock point i, $X_i(S_0)$, is equal to the number of demands over the lead time, Y_i, which is Poisson-distributed with parameter $\lambda_i L_i$, plus the number of backorders at the central warehouse that are due to stock point i, $B_0^{(i)}(S_0)$. (Y_i and $B_0^{(i)}(S_0)$ are mutually independent.) The corresponding number of backorders is $B_i(S_0,S_i)=(X_i(S_0)-S_i)^+$ and the number of items on hand is $I_i(S_0,S_i)=(S_i-X_i(S_0))^+$. Finally, Z_i is a Bernoulli-distributed random variable for all $i\in\{1,\dots,N\}$: $Z_i=1$ with probability λ_i/λ_0 and 0 otherwise. If there is a queue at the central warehouse, then $Z_i=1$ can be interpreted as indicating that the first backorder in that queue is due to local stock point i.

The goal of the optimization is to set the base stock levels S_i for $i\in\{0,\dots,N\}$ such that total average costs per unit time in steady state are minimized:

$$C(S_0,S_1,\dots,S_N)=h_0\mathbb{E}I_0(S_0)+\sum_{i=1}^{N}[h_i\mathbb{E}I_i(S_0,S_i)+\beta_i\mathbb{E}B_i(S_0,S_i)] \quad (1)$$