Computers & Operations Research 53 (2015) 118–127

Contents lists available at ScienceDirect

Computers & Operations Research

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

A simulation-based decision support system for a multi-echelon inventory problem with service level constraints

Shing Chih Tsai*, Chung Hung Liu

National Cheng Kung University No. 1, University Road, Tainan City, Taiwan

ARTICLE INFO

Available online 13 August 2014

Keywords:
Decision support system
Multi-echelon inventory problem
Service level
Computer simulation

ABSTRACT

In this paper, we present a simulation-based decision support system for solving the multi-echelon constrained inventory problem. The goal is to determine the optimal setting of stocking levels to minimize the total inventory investment costs while satisfying the expected response time targets for each field depot. We derive new decision support algorithms to be applied in different scenarios, including small-sample and large-sample cases. The first case requires that the set of alternative solutions is known at the beginning of the experiment, and the number of evaluated solutions may depend on the simulation budget (i.e., the time available to solve the problem). In the second case, the alternative solutions are generated sequentially during the searching process, and we may terminate the algorithm when the specified sampling budget is exhausted. Empirical studies are conducted to compare the performance of the proposed algorithms with other conventional optimization approaches.

1. Introduction

The majority of real-life processes are difficult to study via analytical methods due to their complicated features, and thus there is a lack of practically tractable solutions. In contrast, a simulation model can almost always be constructed and implemented to obtain useful statistical estimators on system performance measures. Therefore, simulation modeling is used extensively in industry as a decision-support tool to solve numerous practical problems, including estimation of system capacities, obtaining insights about the interactions among variables, and predicting the impact of alternative system designs (e.g., [12,13,17]). Equipment-intensive industries such as airlines, nuclear power plants, and manufacturers of expensive electronic machines often require large quantities of spare parts to guarantee high system availability, which in turn results in excessive holding costs. On the other hand, an insufficient stock of spare parts when demand occurs can also lead to excessive downtime costs. An implicit assumption in the existing literature is that inventory holding costs dominate (e.g., [6–8,31]). The focus is often on the optimization of stocking levels of spare parts rather than the use of lateral transshipments and emergency supplies. In this paper, we consider a multi-item, multi-echelon spare parts inventory system, which consists of a warehouse repair center and a warehouse inventory center in the higher echelon and multiple field depots (i.e., inventory stocking centers) in the lower echelon. The specific service measure that we use in this paper is response time, which is defined as the time it takes to obtain service parts after the customer reports a failure. To maintain high-quality service, the analyst prefers to keep the expected response time for each depot below a given target level. In the capital goods industry common practice is to follow a base stock $(S-1, S)$ policy. Further, the average inventory investment cost is high and dominates the total cost because of the high-reliability and low-demand nature of spare parts. Therefore, our goal is to determine base-stock levels for all items at all locations so that the service-level requirements are met at the minimum inventory investment cost. There are alternative supply chain designs that have been studied in the literature on this topic. For instance, remanufacturing can be considered at the local depots [20]. Cattani et al. [10] studied a dual-role central warehouse structure, which is common in practice. In this design, the central warehouse not only replenishes other field depots but also meets demand from customers in the region near the central warehouse. The other possible supply chain mechanism is to employ a decentralized control policy that allows individual divisions or organizational entities to make their own inventory decisions [9].

Finding the optimal design parameters for a multi-item, multi-echelon service parts constrained inventory problem is generally difficult. Most of the existing works employ approximate evaluation methods, such as METRIC (Multi-Echelon Technique for Recoverable Item Control; see Muckstadt [23]) or negative binomial approximation, in which the distributions of the number of outstanding orders at lower echelon facilities are approximated (see [15]). METRIC approximation applies the results derived from queueing theory (i.e., Palm [24]'s theorem and Little [21]'s law) to characterize the inventory process at each retailer (i.e., depots in

* Corresponding author.

http://dx.doi.org/10.1016/j.cor.2014.07.018
0305-0548/

our problem setting), in which the assumption of independent replenishment lead times for retailers is required. However, these lead times are obviously not independent, since they depend on the same inventory situation at the warehouse inventory center. The METRIC approximation is made primarily for analytical convenience, and is reasonable only when the correlation between successive lead times is negligible. If the correlation is high, the errors in the METRIC approximation may result in an incorrect stocking level decision in the optimization procedure [15]. Wong et al. [32] provided numerical results to demonstrate that the solutions returned by either METRIC or Grave's negative binomial approximation procedures might be highly suboptimal or even infeasible. Recently, more research effort has been expended to develop efficient solution approaches for multi-echelon service parts systems. For instance, Van Ommeren et al. [30] proposed a local search heuristic to determine appropriate locations for repair shops and calculated their capacity in order to insure a minimal expected cost given by a small probability that a customer has to wait. Rappold and Van Roo [25] presented an approach to solve the joint problem of facility location, inventory allocation, and capacity investment when demand is stochastic. Their approach extends the results of Graves [15] to incorporate finite repair capacity and is validated by a simulation study. Lieckens et al. [20] developed a profit maximizing model to simultaneously decide the optimal network design and the optimal service delivery strategy for a multi-echelon service parts network. They formulated a mixed integer non-linear model that integrates queueing relationships and is solved by a differential evolution search procedure.

In this paper we develop simulation-based decision support algorithms to solve the multi-echelon constrained inventory problem. For simulation-based algorithms, common and critical questions are related to how much sampling cost should be allocated or how many candidate solutions should be simulated. Therefore we propose two algorithms that can be applied to different scenarios, including small and large-sample cases. The first algorithm is based on ranking and selection (R&S) procedures, in which we adopt the assumption of ordinary R&S for which the set of candidate solutions to be evaluated is fixed and known before the simulation experiment. In this case, the number of solutions to be compared or the required sample size is usually small. The second algorithm is based on a stochastic genetic algorithm (SGA), which is more appropriate when the candidate solutions are generated sequentially during the simulation experiment. The GA-type procedure provides a searching mechanism to explore the entire solution space. In this case we are allowed to evaluate more candidate solutions, and thus more simulation replications are required as compared to the first algorithm. Further, these two simulation algorithms attempt to solve the inventory problem from different perspectives. Since the first algorithm focuses on a small set of candidate solutions inside a large solution space, it is very likely that no feasible solutions can be found. Therefore we hope to select a compromise solution that minimizes the amount by which each service level constraint is violated. On the other hand, the second algorithm searches the solution space extensively for promising solutions. It allows a heuristic search procedure (specifically a genetic algorithm), which is an originally deterministic optimization procedure that does not rely heavily on problem structure, to function effectively in a stochastic environment.

Recently, a number of researchers in the field of simulation have focused on solving expected value constrained stochastic problems via the technique of sample average approximation. The expected value function is not analytically tractable but can be estimated by a sample average approximation of some random observations. The sample average approximation problem is then solved by standard deterministic optimization approaches. The process is repeated several times to obtain a final solution. In each iteration the sample average is also used to determine the feasibility of any candidate solution. Therefore it is very likely to obtain infeasible solutions if the estimator is very noisy and the sample size is not large enough. For instance, Cezik and L'Ecuyer [11] formulated a call center staffing problem as an optimization problem with expected value constraints, and used sample average approximation together with a cutting plane method to solve it. Subsequently, Tsai and Zheng [29] used a similar approach to solve the two-echelon constrained inventory problem, in which they needed to assume that the expected response time at each field depot is nonincreasing and jointly convex componentwise with regard to the stocking level vector. However, experience with time-based service level constraints demonstrates that this assumption is in general not true [7], and may lead to infeasible solutions because of invalid cutting planes. By contrast, the decision support algorithms proposed in this paper are effective under very mild conditions, and we compare our algorithms with the sample-average-approximation method in a numerical study.

The remainder of this paper is organized as follows. In Section 2 we briefly review the literature related to the two major algorithms that constitute our decision support system to solve the inventory problem. Section 3 presents our modeling framework in detail, and formulates the inventory problem in the presence of service level constraints. In Section 3 we also describe the details of the simulation algorithms to be applied for different scenarios, including small and large-sample cases. In Section 4 we evaluate the performance of these algorithms in comparison with other existing ones on example problems of various configurations. The paper then ends with some concluding remarks in Section 5.

2. Related work

The following sections introduce the two component algorithms that constitute our decision support system to solve the inventory problem for different scenarios. In Section 2.1, we provide a brief review of R&S procedure considering multiple stochastic constraints, which is extended later to solve the inventory problem when the set of alternative solutions is known before the experiment. In Section 2.2, we review the SGA specifically designed for the optimization problem with a single stochastic objective, which is extended later to handle multiple stochastic constraints. This extended SGA plays a key role in decision support when we are allowed to evaluate more candidate solutions (i.e., given a greater sampling budget).

2.1. Ranking and selection considering stochastic constraints

Ranking and selection (R&S) procedures have proven to be quite useful for choosing a solution with the best (or near the best) expected performance among a finite number of simulated alternatives (see Kim and Nelson [19] for a survey of R&S). The number of candidate solutions is often so small that we may need to simulate all of them to attain a pre-specified confidence level of correct selection. We are particularly interested in fully sequential selection procedures, where decision makers obtain a single observation at a time from each solution still in contention, and then eliminate solutions from continued sampling when they are statistically inferior. When we encounter a single stochastic objective function, the fully sequential selection procedure (denoted as FSP) presented in Kim and Nelson [18] has been shown to be more efficient compared to other R&S procedures in terms of the total number of observations required to reach a decision. In practice, however, decision makers usually need to