Two-Echelon Spare Parts Inventory System Subject to a Service Constraint*

Deniz Caglar†

Chung-Lun Li‡

David Simchi-Levi§

February 2002

Revised December 2002

Abstract

In this paper, we consider a spare parts inventory problem faced by a manufacturer of electronic machines with expensive parts that are located at various customer locations. The parts fail infrequently according to a Poisson process. To serve customers when a failure occurs, the manufacturer operates a central warehouse and many field depots that stock spare parts. The central warehouse acts as a repair facility and replenishes stock at the field depots. There is a centralized decision maker who manages the inventory in both the central warehouse and the field depots.

We develop a continuous review, base stock policy for this two-echelon, multi-item spare parts inventory system. We formulate a model to minimize the system-wide inventory cost subject to a response time constraint at each field depot. We present an efficient heuristic algorithm and study its computational effectiveness.

*Research supported in part by ONR Contracts N00014-95-1-0232 and N00014-01-1-0146, and by NSF Contracts DMI-9732795 and DMI-0085683.

†Booz-Allen & Hamilton, Chicago, IL

‡Department of Management, The Hong Kong Polytechnic University, Kowloon, Hong Kong

§Corresponding author; Engineering Systems Division and Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139-4307

1 Introduction

We consider a spare parts inventory problem faced by a manufacturer of an expensive electronic machine such as mainframe computer. The manufacturer produces and sells the machine and provides service contracts to geographically dispersed customers. To support the service process, the manufacturer operates a central warehouse and many field depots that are in close proximity to the customers. Both types of facilities stock spare parts, which are highly reliable and very expensive. Furthermore, the warehouse acts as a repair facility and replenishes stock at the field depots. When a machine fails, the customer reports the failure and the part that failed to the field depot serving the customer. If the field depot has the spare part on hand and a technician is available, the technician travels to the customer site to fix the machine. Otherwise, the repair is delayed until either a technician is available to fix the machine or the spare part becomes available at the field depot. In either case, the delay is very costly to the customer.

A measure of the service quality used by the manufacturer is the *response time*, defined as the time it takes for a technician to arrive at the customer site with a spare part to fix the machine after the customer reports a failure. To provide high-quality service, the manufacturer prefers to keep the response time to each customer short. However, the spare parts are expensive and electronic components have high depreciation and obsolescence costs. Therefore, it is imperative that the manufacturer maintains the inventory level as low as possible at the central warehouse and the field depots.

Since the cost of providing a technician to each field depot is dominated by the inventory cost of holding the expensive parts and the failure rates of the parts are quite low, a technician is usually available for service whenever a repair is required. Also, since the field depots are typically close to the customer sites, the technician's travel time is negligible. Thus, the ability to meet the response time constraint depends mainly on the inventory policy at the field depots and the central warehouse. If the required spare part is in stock at the field depot, the customer is served immediately and the response time is negligible. On the other hand, if the depot is out of stock of the specific part requested, the response time includes the time until the field depot receives the part from the central warehouse and the time it takes for the technician to bring it to the customer, which includes the repair time as well as the travel time to and from the central warehouse.

The problem described and analyzed in this paper was motivated by a project conducted by the authors for a large electronics manufacturing company. The model constructed and analyzed for our client was characterized by hundreds of parts and customers, very low part failure rates, tight response times, many field depots around the U.S., and a base stock policy for each part at each field depot. To control the quality of service, the company prefers to keep the average response time to each customer below a threshold level, say 4 hours. Our model, therefore, has an objective of determining an inventory policy to minimize system-wide inventory holding costs such that the *average response time* is no greater than the threshold.

Multi-echelon spare parts inventory systems have been analyzed quite extensively in the literature. One of the earliest works in this area was Sherbrooke's METRIC model. In his classical paper, Sherbrooke [18] considered a two-echelon spare parts inventory system for repairable items. All the facilities in the system had ample repair capacity and operated according to a continuous review $(S-1, S)$ policy. Unlike in our model, Sherbrooke considered the minimization of the total expected backorders at the depots subject to a budget constraint. Muckstadt [13] presented the MOD-METRIC, which generalized Sherbrooke's METRIC model to include multi-indentures, i.e., hierarchical parts structures. Other related works that study multi-echelon resupply systems with budget considerations include Hausman and Erkip [7] and Muckstadt and Thomas [14].

Some studies have focused on analyzing multi-echelon inventory systems with a single product. This includes the work of Moinzadeh and Lee [12], who considered a single-product model and developed a search routine for the stocking levels. They also derived a decision rule to select an $(S-1, S)$ versus an (r, Q) policy. Axsäter [1] considered a single-product, two-echelon, one-for-one replenishment model. He developed a recursive procedure to determine average holding and shortage costs and discussed the determination of optimal inventory base stock levels.

Another line of research has focused on characterizing the service performance of a multi-echelon system for given stocking values, rather than developing solution methods. Graves [6] analyzed a two-echelon spare parts inventory system similar to ours. In his model, the warehouse acted as a centralized repair facility, while the depots faced Poisson demand and followed a continuous review $(S-1, S)$ policy. He presented exact and approximation methods to determine steady-state inventory level distributions at the depots and warehouse. Sherbrooke [19] used Graves' approximation to improve Muckstadt's MOD-METRIC model. Simon [20] derived exact expressions for the stationary distributions of backorders at each facility for a system very similar to that of Sherbrooke's. Simon's model was more general than Sherbrooke's as Simon considered items that are either completely recoverable, completely consumable, or recoverable with some rate of condemnation. Shanker [17] extended Simon's analysis to allow compound Poisson demand at the depots. Lee and Moinzadeh [11] developed a two-parameter approximation to the distribution of backorders when the depots follow an (r, Q) ordering policy. Svoronos and Zipkin [21] considered a multi-echelon system with exogenously generated stochastic transportation times. They approximated the steady-state behavior of the system and showed that transit-time variances significantly affect the system performance. Wang *et al.* [22] considered a two-echelon, multi-item, stochastic demand spare parts system with stocking-center-dependent replenishment lead-times, and they characterized the system performance of the stocking policies.

There are a few researchers who have considered multi-echelon spare parts inventory systems with service constraints. Cohen *et al.* [4] developed and implemented Optimizer to determine inventory policies for IBM's periodic review, multi-item, multi-echelon spare parts inventory system. They solved the problem by using level-by-level decomposition of facilities and by assuming infinite supply at the replenishment sources. Cohen *et al.* [5] reported a successful implementation of two basic inventory models to improve a complex spare parts system.

In a more recent paper, Hopp *et al.* [9] considered a system very similar to the one that we are analyzing. In their case, field depots followed a continuous review $(S-1, S)$ policy and faced Poisson customer demand for consumable parts. The depots were replenished by a central warehouse that followed an (r, Q) inventory policy. Their focus was on devising effective and easily implementable heuristics to minimize system-wide inventory holding costs while keeping the average total delay at each facility below a threshold level. They approximated the lead-time demand distribution with a negative binomial distribution. Their heuristic decomposed the problem hierarchically. First, they solved the warehouse level problem to minimize inventory holding costs at the warehouse subject to a service level constraint using the heuristics they had previously developed in [8]. Once the values of the parameters r and Q were set for each part at the warehouse, the problem was decomposed for each depot. The depot-level problem was to minimize the inventory holding costs at the depot subject to the average total delay constraint. They used some simplifying approximations for steady state inventory and backorder expressions. They solved each depot problem by using these approximations and a Lagrangian relaxation of the depot problem. They also performed computational analysis to study the performance of their heuristic on small-size problems, and their results showed that their heuristic was very effective.

In this paper, our focus is on developing a near-optimal heuristic to minimize system-wide inventory holding costs subject to a response time constraint. Our approach is different from most previous researchers' due to the explicit consideration of a response time constraint and our focus on the development of heuristics. Also, our system considers repairable items, and consequently, we use a base stock level at every facility. However, as mentioned by Richards [16], if we set the order quantity Q to one, set the reorder point r equal to the base stock level less one,

and interpret recoverable failures as demands with repair times corresponding to resupply times, then the model with consumable parts is equivalent to the model with repairable parts. Thus, besides some of the minor differences in the modeling assumptions, our model is indeed a special case of Hopp *et al.*'s. However, our work focuses on the development of an efficient and effective solution method customized for the case with repairable parts. We first derive the expressions for the expected inventory levels and backorder levels and then develop a heuristic algorithm for the problem. We show via computational experiments that our solution approach is significantly more effective than that of Hopp *et al.*'s when it is applied to large-size repairable parts systems.

The rest of our paper is organized as follows. In the next section, we describe the system that we are analyzing, state our assumptions, and set up a model representing the system. In Section 3, we discuss the special case with a single depot. Section 4 presents a heuristic to solve the general problem and develops a lower bound on the optimal value of the inventory holding cost. The performance of the heuristic is tested against the lower bound. Finally, in Section 5, we state our conclusions and discuss possible extensions of the model.

2 The Model

2.1 Model Description

Consider a two-echelon spare parts inventory system consisting of a central warehouse and M field depots as illustrated in Figure 1. The field depots serve customers, each of whom owns exactly one machine, e.g., a mainframe computer, which is very expensive and highly reliable. The machine consists of a set I of n parts that fail infrequently and independently. When a part at a customer site fails, it is replaced by a spare part from the depot serving the customer, if the depot has the part on hand. Otherwise, the part is backordered and the customer has to wait until a part becomes available at the depot. The failed part is shipped to the central warehouse, where all the failed parts are repaired. At the same time, the warehouse ships a spare part to the field depot from its inventory, if it has an available part. Otherwise, the replacement order is backordered at the warehouse until a part is repaired and becomes available.

Our goal is to determine inventory policies at the warehouse and the field depots to minimize system-wide inventory holding costs while maintaining an average response time below a given threshold. For this reason, we develop a model based on the following assumptions:

- The spare parts supply network consists of the warehouse, M field depots, and the customers.
- The shipment time between the warehouse and a field depot j is stochastic with mean T_j. The travel time from a field depot to a customer served by that depot is negligible, and we assume that it is zero.
- Since the parts are very expensive with low failure rates and the ordering cost is negligible, the field depots employ a continuous review, base stock policy with the base stock level for part i at depot j set at S_{ij}, which cannot exceed a limit $\hat{S}_{ij}$ specified by management.
- All inventory is kept at the warehouse and the field depots; customers keep no inventory.
- Every part is crucial for operating the machine at the customer site, that is, the mainframe computer is *down* if one of its parts fails.
- The "time to failure" of a part i is exponentially distributed with mean $1/l_i$, independent of the

machine it is in. This assumption is justified for electronic components of electronic machines (see, for example, [10]).

- When a machine fails, the customer knows which part failed and places an order for that part to the field depot from which it is served. Furthermore, there is always an available technician at the field depot. As mentioned later in Section 5, an interesting future extension of this work is to incorporate the capacity decision, i.e., the availability of the technician, into the model.

- The technician's travel time from the field depot to the customer site is negligible and is assumed to be zero.

- We assume that K_j, the number of customers served by depot j, is sufficiently large and we model the demand for part i at depot j as a Poisson arrival process with rate $\lambda_{ij} = K_j l_i$. Although this assumption is typically violated whenever there are failed machines in the field, it is common in the literature (see, for example, Sherbrooke [18] and Graves [6]). As pointed out by Graves [6], this assumption is reasonable when the expected number of failed machines is small relative to the total number of machines.

- We assume ample repair capacity at the warehouse, that is, no queuing occurs and successive repair times for part i are i.i.d. random variables with mean R_i.

- There are no emergency lateral shipments among the field depots; the depots are resupplied only from the warehouse. As mentioned later in Section 5, relaxing this assumption is a future direction of this work.

We use the following notation in our model:

$I = \{1, 2, \ldots, n\}$ = set of spare parts;

$J = \{1, 2, \ldots, M\}$ = set of field depots; the index 0 is reserved for the central warehouse;

l_i = failure rate of part i;

$\lambda_{ij} = K_j l_i$ = demand rate faced by part i at depot j;

S_{ij} = base stock level for part i at facility j (decision variable);

$\hat{S}_{ij}$ = upper limit on S_{ij};

R_i = mean repair time of part i at the warehouse;

T_j = mean transportation time between the warehouse and depot j;

L_{ij} = lead-time to replenish part i at facility j (random variable); the lead-time at a depot is the transportation time from the warehouse plus any delay due to stockouts; the lead-time at the warehouse is the transportation time plus the repair time;

θ_{ij} = demand rate of part i at facility j during lead-time (note: θ_{i0} is a constant, while θ_{ij} $(j \neq 0)$ depends on S_{i0});

$\bar{B}_{ij}(S_{ij}, S_{i0})$ = expected backorder level at depot j when the base stock level of part i is set at S_{ij} at depot j and at S_{i0} at the warehouse;

$\bar{I}_{ij}(S_{ij}, S_{i0})$ = expected inventory on hand at depot j when the base stock level of part i is set at S_{ij} at depot j and at S_{i0} at the warehouse;

$\bar{B}_i(S_{i0})$ = expected backorder level at the warehouse when the base stock level of part i is set at S_{i0} at the warehouse;

$\bar{I}_i(S_{i0})$ = expected inventory on hand at the warehouse when the base stock level of part i is set at S_{i0} at the warehouse;

h_i = inventory holding cost rate for part i;

τ_j = response time threshold specified by the manufacturer for depot j;

W_{ij} = average time a customer waits to receive an order of part i at depot j;

W_j = average response time to a customer at depot j.

Using the notation described above, we can derive a formulation for the problem of minimizing total inventory investment subject to a response time constraint at each depot:

$$\begin{aligned}
\text{Minimize} \quad & \sum_{i \in I} h_i \bar{I}_i(S_{i0}) + \sum_{i \in I} \sum_{j \in J} h_i \bar{I}_{ij}(S_{ij}, S_{i0}) \\
\text{subject to} \quad & W_j \leq \tau_j \quad (j \in J) \\
& 0 \leq S_{ij} \leq \hat{S}_{ij}, \quad S_{ij} \text{ integer} \quad (i \in I;\ j \in J) \\
& 0 \leq S_{i0} \leq \hat{S}_{i0}, \quad S_{i0} \text{ integer} \quad (i \in I).
\end{aligned}$$

First, we will focus on the constraint that the average response time to a customer must be within the specified response time threshold. The response time to a customer will be a function of the part that fails and the stocking policies of that part at the field depot serving the customer and at the central warehouse. Recall that we assume there is always a technician available and the lead-time from a field depot to a customer is zero. Together, these imply that the response time to a customer is exactly the time that the customer waits to receive an order. The expected waiting time W_{ij} for part i at depot j can easily be found by an application of Little's law:

$$W_{ij} = \frac{\bar{B}_{ij}(S_{ij}, S_{i0})}{\lambda_{ij}}.$$

Thus, the expected waiting time for a customer is

$$W_j = \sum_{i \in I} W_{ij} \cdot Prob(\text{Part } i \text{ fails}) = \sum_{i \in I} \frac{\bar{B}_{ij}(S_{ij}, S_{i0})}{\lambda_{ij}} \cdot \frac{\lambda_{ij}}{\sum_{\ell \in I} \lambda_{\ell j}} = \sum_{i \in I} \frac{\bar{B}_{ij}(S_{ij}, S_{i0})}{\sum_{\ell \in I} \lambda_{\ell j}}.$$

Hence, the response time constraint becomes

$$\sum_{i \in I} \frac{\bar{B}_{ij}(S_{ij}, S_{i0})}{\sum_{\ell \in I} \lambda_{\ell j}} \leq \tau_j,$$

or equivalently,

$$\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \leq \tau_j \sum_{i \in I} \lambda_{ij}.$$

For simplicity, we will refer to $\tau_j \sum_{i \in I} \lambda_{ij}$ as α_j. Hence, the response time constraint becomes:

$$\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \leq \alpha_j \quad (j \in J).$$

Our next important task is to identify the expected backorder quantity $\bar{B}_{ij}(S_{ij}, S_{i0})$ for a part i at a depot j given the stocking quantities of S_{ij} and S_{i0}.

2.2 Finding Backorder Levels at the Depots

Graves [6] described an exact method to determine backorder quantities at depots. However, he pointed out that the computational requirements of this exact method are not trivial and this methodology would be impractical for optimizing an inventory system that consists of many parts. To ease the computational burden, researchers have suggested approximations to backorder quantities, such as the METRIC approximation by Sherbrooke [18] and the negative binomial approximation by Graves [6].

The METRIC model provides an approximate distribution for inventory on hand and backorders at each depot for a two-echelon system with compound Poisson failure processes and ample repair capacity at the warehouse. METRIC allows general replenishment times for depots that are stochastic random variables, although this model ends up replacing the stochastic lead-time by its mean. The approximation uses Palm's theorem [15] to find the outstanding orders at each depot; outstanding orders at a depot can be interpreted as the occupancy level in an $M/G/\infty$ queuing system. According to Palm's theorem, the occupancy level in an $M/G/\infty$ queuing system is Poisson with mean λ/μ, where λ is the arrival rate and $1/\mu$ is the average service time. In our system, the arrival rate is simply λ_{ij}, which is the demand rate for part i at depot j. The service time is the lead-time L_{ij}, which is the sum of the transportation time from the warehouse to the depot and the delay at the warehouse due to stockouts. Let Q_{ij} be the outstanding orders for a part at depot j. Once the distribution for Q_{ij} is obtained, it is straightforward to determine backorders at the depot: If S_{ij} is the base stock level, then the backorder level in the steady state is

$$E[(Q_{ij} - S_{ij})^+] = \sum_{k=S_{ij}+1}^{\infty} (k - S_{ij}) Prob(Q_{ij} = k),$$

where Q_{ij} is Poisson with mean $\lambda_{ij} E[L_{ij}]$.

METRIC, however, is only an approximation because it ignores the dependence between successive lead-times from the warehouse to a depot. These lead-times are not independent since they depend on the inventory situation at the warehouse. Axsäter [2] pointed out that "the METRIC approximation will, in general, work well as long as the demand at each [depot] is low relative to the total demand, for example in a case with many small [depots]." The approximation works well in such a case essentially because the dependence between successive lead-times to a depot is reduced due to many

orders being placed at the warehouse by other depots in the system. This observation suggests that the METRIC approximation may actually work well for the system we are considering, since none of the depots in our system face more than 5% of the total demand, which implies that there are more than 20 orders, on average, between two successive orders from a depot. Thus, the dependence between successive lead-times would be small and METRIC should be a very good approximation. In fact, numerical experiments have been conducted to show that the METRIC approximation works very well in our model (see Caglar [3]).

2.3 METRIC-Like Model

To use the METRIC approximation, we first identify steady state expressions for the backorder and inventory on hand levels for the warehouse using Palm's theorem. This requires that we identify the mean demand rate and the mean lead-time to the warehouse. The demand faced by the warehouse for part i is the superposition of the Poisson demands faced by the depots, with rate $\lambda_{i0} \equiv \sum_{j \in J} \lambda_{ij}$. The lead-time for part i to the warehouse, L_{i0}, is the sum of the transportation time from a depot to the warehouse and the repair time at the

warehouse. The repair time is independent of which depot that part i is originating from and has a mean R_i. However, the transportation time, which has a mean T_j, does depend on the particular depot. Then

$$E[L_{i0}] = \sum_{j \in J} T_j \cdot Prob(\text{demand originates from depot } j) + R_i = \sum_{j \in J} T_j \cdot \frac{\lambda_{ij}}{\sum_{\ell \in J} \lambda_{i\ell}} + R_i.$$

By Palm's theorem, the outstanding orders at the warehouse for part i, Q_{i0}, is Poisson with mean $\theta_{i0} = \lambda_{i0} E[L_{i0}]$. Now, we can obtain the expected backorder level at the warehouse as follows:

$$\begin{aligned}
\bar{B}_i(S_{i0}) &= \sum_{k=S_{i0}+1}^{\infty} (k - S_{i0}) Prob(Q_{i0} = k) \\
&= \sum_{k=0}^{\infty} k\, Prob(Q_{i0} = k) - \sum_{k=0}^{S_{i0}} k\, Prob(Q_{i0} = k) - S_{i0}\Big[1 - \sum_{k=0}^{S_{i0}} Prob(Q_{i0} = k)\Big] \\
&= E[Q_{i0}] - S_{i0} + \sum_{k=0}^{S_{i0}} (S_{i0} - k) Prob(Q_{i0} = k) \\
&= E[Q_{i0}] - S_{i0} + \sum_{k=0}^{S_{i0}-1} F_{i0}(k) \\
&= \theta_{i0} - \sum_{k=0}^{S_{i0}-1} \Big[1 - F_{i0}(k)\Big], \qquad (1)
\end{aligned}$$

where

$$F_{i0}(k) = \sum_{\ell=0}^{k} Prob(Q_{i0} = \ell) = \sum_{\ell=0}^{k} \frac{e^{-\theta_{i0}} \theta_{i0}^{\ell}}{\ell!}.$$

The expected inventory on hand at the warehouse can also be obtained:

$$\begin{aligned}
\bar{I}_i(S_{i0}) &= \sum_{k=0}^{S_{i0}} (S_{i0} - k) Prob(Q_{i0} = k) \\
&= \sum_{k=0}^{S_{i0}-1} F_{i0}(k) \\
&= S_{i0} - \theta_{i0} + \bar{B}_i(S_{i0}) \qquad \text{(by (1))}. \qquad (2)
\end{aligned}$$

Because the successive lead-times to the warehouse are independent, the above expected backorder and inventory expressions for the warehouse are exact. Note that θ_{i0} is independent of the stocking decisions at the warehouse.

Next, we use the METRIC approximation to compute the backorder quantities for the depots. That is, we ignore the dependence of successive lead-times from the warehouse to a depot to satisfy the demand and use Palm's theorem to find the backorders at a depot. The backorders at a depot depend not only on the demand and the inventory policy at that depot, but also on the inventory policy at the warehouse as the stockouts at the warehouse affect the lead-time to a depot. Specifically, the lead-time from the warehouse to depot j is the sum of the transportation time and the delay at the warehouse due to stockouts. The expected delay at the warehouse, according to Little's law, is the expected backorder level divided by the demand rate at the warehouse. Thus,

$$E[L_{ij}] = T_j + \frac{\bar{B}_i(S_{i0})}{\lambda_{i0}}.$$

Hence, for $j \in J$,

$$\theta_{ij}(S_{i0}) = E[Q_{ij}] = \lambda_{ij}\Big(T_j + \frac{\bar{B}_i(S_{i0})}{\lambda_{i0}}\Big).$$

Using Palm's theorem and following the derivation of the warehouse backorder and inventory levels, we obtain

$$\bar{B}_{ij}(S_{ij}, S_{i0}) = \theta_{ij}(S_{i0}) - \sum_{k=0}^{S_{ij}-1} \Big[1 - F_{ij}(k)\Big] \tag{3}$$

and

$$\bar{I}_{ij}(S_{ij}, S_{i0}) = S_{ij} - \theta_{ij}(S_{i0}) + \bar{B}_{ij}(S_{ij}, S_{i0}), \tag{4}$$

where $F_{ij}(k)$ is the probability that the outstanding orders of part i at depot j is at most k. Clearly,

$$F_{ij}(k) = \sum_{\ell=0}^{k} \frac{e^{-\theta_{ij}(S_{i0})}\theta_{ij}^{\ell}(S_{i0})}{\ell!}.$$

By (4), the objective function of our problem can be rewritten as $\sum_{i\in I} h_i[\bar{I}_i(S_{i0}) - \sum_{j\in J} \theta_{ij}(S_{i0})] + \sum_{i\in I}\sum_{j\in J} h_i[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0})]$. It is easy to check that $\bar{I}_i(S_{i0}) - \sum_{j\in J} \theta_{ij}(S_{i0}) = S_{i0} - \theta_{i0} - \sum_{j\in J} \lambda_{ij} T_j$. Thus, our problem can be rewritten as:

$$\text{Problem } P\text{:} \quad \text{Minimize} \quad \sum_{i \in I} h_i \Big[S_{i0} - \theta_{i0} - \sum_{j \in J} \lambda_{ij} T_j \Big] + \sum_{i \in I} \sum_{j \in J} h_i \Big[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0}) \Big]$$

$$\text{subject to} \quad \sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \le \alpha_j \quad (j \in J)$$

$$0 \le S_{ij} \le \hat{S}_{ij}, \quad S_{ij} \text{ integer} \quad (i \in I;\ j \in J)$$

$$0 \le S_{i0} \le \hat{S}_{i0}, \quad S_{i0} \text{ integer} \quad (i \in I).$$

3 The Single Depot Subproblem

In this section, we consider the special case in which the base stock levels at the warehouse, S_{i0} ($i \in I$), are known. The heuristic developed for this case will be extended to solve problem P later. In this special case, problem P can be decomposed into n single-depot subproblems. The j-th subproblem is:

$$\text{Problem } P_j\text{:} \quad \text{Minimize} \quad \sum_{i \in I} h_i \Big[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0}) \Big]$$

$$\text{subject to} \quad \sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \le \alpha_j$$

$$0 \le S_{ij} \le \hat{S}_{ij}, \quad S_{ij} \text{ integer} \quad (i \in I).$$

From equation (3), it is easy to see that $\bar{B}_{ij}(S_{ij}, S_{i0})$ decreases but $S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0})$ increases as S_{ij} increases. This implies that the objective function value of P_j is increasing in S_{ij}. In the following analysis, we assume that $\sum_{i \in I} \bar{B}_{ij}(\hat{S}_{ij}, S_{i0}) \le \alpha_j$. Otherwise, the problem is infeasible. We also assume that $\sum_{i \in I} \bar{B}_{ij}(0, S_{i0}) > \alpha_j$. Otherwise, $S_{1j} = \cdots = S_{nj} = 0$ is the optimal solution and the problem is trivial.

We now present a heuristic algorithm for solving problem P_j using Lagrangian relaxation. For a given real number $\pi_j > 0$, consider the following nonlinear integer program:

$$\text{Problem } P_j^L(\pi_j)\text{:} \quad \text{Minimize} \quad \sum_{i \in I} h_i \Big[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0}) \Big] + \pi_j \Big[\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) - \alpha_j \Big]$$

$$\text{subject to} \quad 0 \le S_{ij} \le \hat{S}_{ij}, \quad S_{ij} \text{ integer} \quad (i \in I).$$

Clearly, the optimal solution value of problem $P_j^L(\pi_j)$ is a lower bound on that of problem P_j. The objective function of problem $P_j^L(\pi_j)$ can be rewritten as

$$\begin{aligned}
&\sum_{i\in I} h_i\Big[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0})\Big] + \pi_j\Big[\sum_{i\in I}\bar{B}_{ij}(S_{ij}, S_{i0}) - \alpha_j\Big] \\
&= \sum_{i\in I}\Big\{(\pi_j + h_i)\Big[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0})\Big] - \pi_j S_{ij}\Big\} - \pi_j\alpha_j \\
&= \sum_{i\in I}\Big\{(\pi_j + h_i)\Big[\theta_{ij}(S_{i0}) + \sum_{k=0}^{S_{ij}-1} F_{ij}(k)\Big] - \pi_j S_{ij}\Big\} - \pi_j\alpha_j \qquad \text{(by (3))} \\
&= \sum_{i\in I}\Big\{\sum_{k=0}^{S_{ij}-1}\Big[(\pi_j + h_i)F_{ij}(k) - \pi_j\Big] + (\pi_j + h_i)\theta_{ij}(S_{i0})\Big\} - \pi_j\alpha_j.
\end{aligned}$$

Thus, minimizing this objective function is equivalent to minimizing $\sum_{k=0}^{S_{ij}-1}[(\pi_j + h_i)F_{ij}(k) - \pi_j]$ for each $i \in I$. Such a minimum can be attained by the following procedure:

Solution procedure SP for problem $P_j^L(\pi_j)$:

For $i = 1, \ldots, n$,

if $F_{ij}(\hat{S}_{ij} - 1) \le \pi_j/(\pi_j + h_i)$, set S_{ij} to $\hat{S}_{ij}$;

if $F_{ij}(\hat{S}_{ij} - 1) > \pi_j/(\pi_j + h_i)$, select S_{ij} as the smallest integer such that $F_{ij}(S_{ij}) > \pi_j/(\pi_j + h_i)$.

The solution generated by procedure SP is optimal to problem $P_j^L(\pi_j)$. This is because F_{ij} is an increasing function, and therefore, $(\pi_j + h_i)F_{ij}(k) - \pi_j$ becomes positive when $F_{ij}(k) > \pi_j/(\pi_j + h_i)$. In case $F_{ij}(k) = \pi_j/(\pi_j + h_i)$ for some integer $j < \hat{S}_{ij}$, we may select S_{ij} such that $F_{ij}(S_{ij} - 1) = \pi_j/(\pi_j + h_i)$ or select S_{ij} such that $F_{ij}(S_{ij}) = \pi_j/(\pi_j + h_i)$ and the objective function will be minimized either way.

The main idea of our heuristic is to identify a value of π_j as well as an optimal solution to $P_j^L(\pi_j)$. We use this optimal solution to $P_j^L(\pi_j)$ as the heuristic solution to problem P_j, if it is feasible for P_j. Clearly, if π_j is sufficiently large, then $\sum_{i\in I}\bar{B}_{ij}(S_{ij}, S_{i0}) \le \alpha_j$ and the heuristic solution is feasible for problem P. However, the heuristic solution value, $\sum_{i\in I} h_i[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0})]$, is nondecreasing as π_j increases. Therefore, we search for the smallest value of π_j such that $\sum_{i\in I}\bar{B}_{ij}(S_{ij}, S_{i0}) \le \alpha_j$ in the optimal solution of $P_j^L(\pi_j)$. Let π_j^* be such a value of π_j. We have the following lemma.

Lemma 1 *$F_{ij}(k) = \pi_j^*/(\pi_j^* + h_i)$ for some $i \in I$ and some $k \in \{0, 1, \ldots, \hat{S}_{ij}-1\}$.*

The validity of Lemma 1 is obvious. It is because the distributions $F_{ij}(k)$ ($i \in I$ and $k \in$

$\{0, 1, \ldots, \hat{S}_{ij}-1\}$) have discrete values and the "newsboy ratio" $\pi_j^*/(\pi_j^* + h_i)$ is continuous in π_j^*. Hence, any changes in π_j^*, such that the newsboy ratio remains between the same two consecutive points, do not change the solution of problem P_j.

For $i = 1, \ldots, n$ and $k = 0, 1, \ldots, \hat{S}_{ij}-1$, we define

$$\pi_j(i, k) = h_i F_{ij}(k)/[1 - F_{ij}(k)]$$

and

$$\Pi = \{\pi_j(i, k) \mid i = 1, \ldots, n;\ k = 0, 1, \ldots, \hat{S}_{ij}-1\}.$$

Lemma 1 implies that π_j^* should be equal to one of the $\pi_j(i, k) \in \Pi$. In fact, π_j^* should be the smallest $\pi_j(i, k)$ that gives $\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \leq \alpha_j$ when the solution procedure SP is used. Therefore, we propose the following heuristic for solving problem P_j:

Heuristic $H1$:

Step 1: Initialize $\Pi' \leftarrow \Pi$ and $S_{ij} \leftarrow 0$ for $i \in I$.

Step 2: Let $\pi_j(r, s)$ be the smallest element in Π' (in case of a tie, precedence is given to part i with the lower subscript). Set $S_{rj} \leftarrow S_{rj} + 1$.

Step 3: If $\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \leq \alpha_j$, then stop; otherwise, let $\Pi' \leftarrow \Pi' \setminus \{\pi_j(r, s)\}$ and go to Step 2.

Let $\pi_j(r, s)$ be the element of Π selected in the last iteration of heuristic $H1$. It is easy to check that after Step 2 of this heuristic, S_{rj} is the smallest integer such that $F_{rj}(S_{rj}) > \pi_j(r, s)/[\pi_j(r, s) + h_r]$. For other parts $i \in I \setminus \{r\}$, if $S_{ij} < \hat{S}_{ij}$ then S_{ij} is either the smallest integer that satisfies $F_{ij}(S_{ij}) > \pi_j(r, s)/[\pi_j(r, s) + h_i]$, or in case a tie occurs in Step 2, $F_{ij}(S_{ij}) = \pi_j(r, s)/[\pi_j(r, s) + h_i]$. Thus, heuristic $H1$ produces the same solution as the solution procedure SP would produce for $P_j^L(\pi_j)$ with $\pi_j = \pi_j(r, s)$. Hence, $\pi_j(r, s)$ is the smallest element of set Π that produces $\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) \leq \alpha_j$ when we use the solution procedure SP. Therefore, we have the following result:

Theorem 2 *π_j^* is equal to the final value of $\pi_j(r, s)$ obtained in Step 2 of heuristic $H1$.*

Heuristic $H1$ enables us to search efficiently for an approximate solution to problem P_j. It only requires a linear search for the optimal value of π_j, whereas a complete enumeration approach would require a simultaneous search of n different S_{ij} values. In fact, it is easy to see that the optimal value of π_j can also be obtained via a binary search and that the efficiency of this heuristic can be improved.

4 Solution to the General Problem

In this section, we provide a solution method for problem P and study its effectiveness via computational experiments. We also provide a procedure for determining a lower bound on the optimal solution value of problem P.

4.1 A lower bound procedure

For any given vector $\pi = (\pi_1, \ldots, \pi_M)$ such that $\pi_j > 0$ $(j = 1, \ldots, M)$, consider the following Lagrangian relaxation of problem P:

$$\begin{aligned}
\text{Problem } P^L(\pi): \quad \text{Minimize} \quad & \sum_{i \in I} h_i \Big[S_{i0} - \theta_{i0} - \sum_{j \in J} \lambda_{ij} T_j\Big] + \sum_{i \in I} \sum_{j \in J} h_i \Big[S_{ij} + \bar{B}_{ij}(S_{ij}, S_{i0})\Big] \\
& + \sum_{j \in J} \pi_j \Big[\sum_{i \in I} \bar{B}_{ij}(S_{ij}, S_{i0}) - \alpha_j\Big] \\
\text{subject to} \quad & 0 \le S_{ij} \le \hat{S}_{ij}, \quad S_{ij} \text{ integer} \quad (i \in I;\ j \in J) \\
& 0 \le S_{i0} \le \hat{S}_{i0}, \quad S_{i0} \text{ integer} \quad (i \in I).
\end{aligned}$$

Note that the constant term $-\sum_{i \in I} h_i[\theta_{i0} + \sum_{j \in J} \lambda_{ij} T_j] - \sum_{j \in J} \pi_j \alpha_j$ can be removed from the objective function, and the resulting problem can be decomposed into n independent subproblems, one for each part i. The i-th subproblem is:

$$\begin{aligned}
\text{Problem } P_i^L(\pi): \quad \text{Minimize} \quad & h_i S_{i0} + \sum_{j \in J} \Big[h_i S_{ij} + (h_i + \pi_j) \bar{B}_{ij}(S_{ij}, S_{i0})\Big] \\
\text{subject to} \quad & 0 \le S_{ij} \le \hat{S}_{ij}, \quad S_{ij} \text{ integer} \quad (j \in J) \\
& 0 \le S_{i0} \le \hat{S}_{i0}, \quad S_{i0} \text{ integer}.
\end{aligned}$$

A lower bound on the optimal objective function value of problem P can be obtained by solving each of these n subproblems. To solve subproblem $P_i^L(\pi)$, we search over all possible values of S_{i0}, where $0 \leq S_{i0} \leq \hat{S}_{i0}$. For each value of S_{i0}, the optimal values of $S_{i1}, \ldots, S_{in}$ can be obtained by using procedure SP. Caglar [3] has proven that, in the i-th subproblem, the optimal value of S_{i0} must be no greater than

$$\tilde{S}_{i0} = \min\left\{ S_{i0} \;\middle|\; C_i^*(S_{i0}) < C_i^*(\hat{S}_{i0}) + h_i \right\},$$

where

$$C_i^*(S_{i0}) = \min\left\{ \sum_{j \in J} \left[h_i S_{ij} + (h_i + \pi_j)\bar{B}_{ij}(S_{ij}, S_{i0}) \right] \;\middle|\; 0 \leq S_{ij} \leq \hat{S}_{ij} \text{ for } j = 1, \ldots, M \right\}.$$

Hence, the computational time of the lower bound procedure can be reduced through limiting the search of S_{i0} over the interval $[0, \tilde{S}_{i0}]$.

4.2 The heuristic

We now propose a heuristic algorithm for solving problem P. This heuristic has two main steps. We first fix the warehouse base stock levels S_{i0} ($i \in I$) and solve for each depot individually to obtain S_{ij} ($i \in I$; $j \in J$) and π_j ($j \in J$) using heuristic $H1$. Next, using the π_j values obtained, we apply the lower bound procedure to determine an updated set of warehouse base stock levels S_{i0}. We repeat these two steps until the values of π_j ($j \in J$) are identical to those in the previous iteration (that is, the algorithm converges) or until a certain number of iterations have been carried out. Throughout the computational process, we keep track of the best solution and the best lower bound that we have obtained. Here is a formal description of the heuristic:

Heuristic $H2$:

Step 1: Set $S_{i0}^* \leftarrow \hat{S}_{i0}$ for $i \in I$. Set $Z^{LB} \leftarrow 0$ and $Z^{H2} \leftarrow +\infty$. Set $\pi_j' \leftarrow 0$ for $j \in J$.

Step 2: Using S_{i0}^* ($i \in I$), apply heuristic $H1$ to determine S_{ij}^* ($i \in I$) and π_j^*, for every $j = 1, \ldots, M$. If this new solution has an objective value less than Z^{H2}, then replace

Z^{H2} by the objective value of this solution and record this solution as the incumbent solution. If $\sum_{j \in J} \left| \pi_j^* - \pi_j' \right| = 0$ or the maximum iteration limit is reached, then stop. Otherwise, set $\pi_j' \leftarrow \pi_j^*$ ($j \in J$) and go to Step 3.

Step 3: Using π_j^* ($j \in J$), apply the lower bound procedure developed in subsection 4.1 to determine S_{i0}^* ($i \in I$) and a lower bound value. If this new lower bound is larger than Z^{LB}, then replace Z^{LB} by the new lower bound. Go to Step 2.

4.3 Computational Experiments

In this subsection, we test the effectiveness of heuristic $H2$ via various sets of test data. We also compare the performance of heuristic $H2$ with that of Hopp *et al.*'s [9] heuristic when applied to our model.

For each set of test data, we obtain the heuristic solution value Z^{H2} and the lower bound Z^{LB} from heuristic $H2$. Meanwhile, we compute the solution value of Hopp *et al.*'s heuristic and denote it as Z^{HZS}. For small-size problems, we also determine the optimal solution value Z^*. The relative errors of the two above-mentioned heuristics are given by $\frac{Z^{H2}-Z^*}{Z^*} \times 100\%$ and $\frac{Z^{HZS}-Z^*}{Z^*} \times 100\%$. Because the optimal solution values of large-size problems are unavailable, lower bounds are used to estimate the performance of the heuristics, and the estimates of the relative errors of the two heuristics are denoted as $e^{H2} = \frac{Z^{H2}-Z^{LB}}{Z^{LB}} \times 100\%$ and $e^{HZS} = \frac{Z^{HZS}-Z^{LB}}{Z^{LB}} \times 100\%$. To simplify the analysis, for each set of test data, we set $\hat{S}_{ij}$ and $\hat{S}_{i0}$ to be sufficiently large such that the constraints "$S_{ij} \leq \hat{S}_{ij}$" and "$S_{i0} \leq \hat{S}_{i0}$" are not binding in the heuristic solutions. Some initial runs showed that when heuristic H2 was applied, the solution converged within three iterations in most cases. Therefore, throughout our computational study, we set the maximum number of iterations of heuristic H2 to three.

We first consider the four small examples with two parts and two depots given by Hopp *et al.* (i.e., Case 8 – Case 11 in [9]). The data for these four examples are shown in Table 1. These examples are small enough that the optimal solutions could be obtained via complete enumeration. The computational results are reported in Table 2. We observe that, for these four sets of data, the performance of heuristic H2 is very close to that of Hopp *et al.*'s heuristic. We also observe that the lower bound Z^{LB} is not tight, and therefore, e^{H2} and e^{HZS} are only conservative estimates of the relative errors. Note that the performance of Hopp *et al.*'s heuristic reported in Table 2 is different from that in their paper. This is because in our model, we use a base stock policy throughout the system, while in their paper, a (Q, r) policy is used and an additional constraint is imposed on the average order frequency at the central warehouse.

Next, we consider the situation that our client faced. There were about 40 depots, each serving approximately 40 customers. There were around 200 parts and a 4-hour average response time was considered as desirable. The average transportation lead-time was about a week, and the average repair time was within a day. The mean time to failure (MTTF) of each part is estimated to be about 500 weeks on the average. Hence, on the average, a depot has a demand rate of around $\frac{40}{500\times7\times24} = 0.00048$ units/hour and a one-way travel time to the warehouse of $7 \times 24 = 168$ hours. On average, a part has an expected lead-time of about $(7+1) \times 24 = 196$ hours.

To evaluate the effectiveness of the heuristic under such an environment, we created 24 test cases as follows. We set the response time threshold τ_j to 4 for all $j \in J$. We set the average demand rate $\bar{\lambda}$ to 0.0005, the average lead-time at the central warehouse $\bar{L}$ to 200, the average transportation time between the warehouse and the depot $\bar{T}$ to 160, and the average holding cost $\bar{h}$ to 500. Note that the relative error of a heuristic is independent of the value of $\bar{h}$.

For one-third of the test cases, we set all λ_{ij} equal to $\bar{\lambda}$. These cases cover the scenario with homogeneous demand rates across all depots and all parts. For another one-third of the cases, we set $\lambda_{ij} = \frac{2i-1}{n} \cdot \bar{\lambda}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, M$. These cases cover the scenario with balanced demand rates among depots but unbalanced demand among different parts. For the remaining one-third of the cases, we set $\lambda_{ij} = \frac{2j-1}{M} \cdot \bar{\lambda}$ for $i = 1, \ldots, n$ and

$j = 1, \ldots, M$. These cases cover the scenario with balanced demand among different parts but unbalanced demand among depots. We set $E[L_{i0}] = \bar{L}$ $(i = 1, \ldots, n)$ for half of the test cases. They represent the scenarios of having homogeneous lead-times at the warehouse across all parts. We set $E[L_{i0}] = \frac{2i-1}{n} \cdot \bar{L}$ $(i = 1, \ldots, n)$ for the other half of the cases. They represent the scenarios of having unbalanced lead-times at the warehouse. Similarly, we set $h_i = \bar{h}$ $(i = 1, \ldots, n)$ for half of the test cases and set $h_i = \frac{2i-1}{n} \cdot \bar{h}$ $(i = 1, \ldots, n)$ for the other cases. We set $T_j = \bar{T}$ $(j = 1, \ldots, M)$ for half of the test cases and set $T_j = \frac{2j-1}{M} \cdot \bar{T}$ $(j = 1, \ldots, M)$ for the other cases. The parameter settings of these 24 cases are shown in Table 3.

We tested the performance of the heuristics for these 24 cases and for three different sets of values of n and M, with the results reported in the table. We have several observations. First, the performance of heuristic $H2$ (with respect to the lower bound that it generates) grows worse as the problem size gets smaller. Second, it performs worse in Cases 1, 2, 5, 6, 17, 18, 21, and 22 than in the other 16 cases. These 8 cases have one characteristic in common: all parts have identical parameters. This suggests that the performance of heuristic $H2$ improves as the variety of customers increases. Next, we observe that the average value of e^{H2} is 3.2% while the average value of e^{HZS} is 7.5%. Since e^{H2} and e^{HZS} are estimates of relative errors obtained from a conservative lower bound, the above percentages suggest that heuristic H2 outperforms Hopp *et al.*'s heuristic substantially when applied to our setting, i.e., when a base stock policy is used at the warehouse and when the problem size is large. This is because heuristic H2 searches for the optimal values of the set of Lagrangian multipliers $(\pi_1, \ldots, \pi_M)$ in order to determine the warehouse base stock levels $(S_{10}, \ldots, S_{n0})$. However, Hopp *et al.*'s heuristic uses a single Lagrangian multiplier to determine the warehouse reorder points. Thus, our heuristic allows us to have a larger degree of freedom to select the base stock levels at the central warehouse. Note that in some of the test instances, Hopp *et al.*'s heuristic performed better than heuristic H2, but in those instances the difference in performance between the two heuristics is relatively small.

Furthermore, heuristic H2 is highly efficient. It takes less than 2 minutes to solve each of problem instances with a 2GHz processor.

To investigate further the performance of our heuristic on different problem sizes, we repeated the 24 test cases for various values of n and M. For each combination of n and M, we report the average value of e^{H2} in Table 4. The results indicate that the performance of heuristic H2 improves monotonically as the number of parts increases. Also, the performance of H2 tends to improve as the number of depots increases. This further confirms the effectiveness of our heuristics for large-size problems.

5 Conclusions

In this paper, we analyzed a continuous review spare parts inventory system in which the parts are highly reliable and very expensive. We considered a multi-echelon system that includes a warehouse, many field depots, and many customers supported by the field depots. We developed a model to minimize the system-wide inventory holding costs while meeting a service constraint at each of the field depots. The service constraint we considered was based on the average response time, defined as the average time it takes a customer to receive a spare part after a failure is reported.

We first developed a heuristic to solve the single-depot problem and then extended it to solve the general problem. Extensive computational experiments were performed to evaluate the performance of the heuristic. The analysis demonstrated that the heuristic works very well on large-size problems. In the future, we would like to extend our analysis here to a model that allows emergency lateral shipments among the field depots in case of stockouts.

This paper was motivated by the research the authors have conducted for an electronic machine manufacturer. The problem the manufacturer faced was further complicated by the ability to transfer inventory between field depots depending on inventory levels and response time requirements. Thus, the results in this paper are important building blocks and a benchmark that will be used when we analyze the problem with lateral shipments.

Note that, in our model, the service constraint provides a threshold on the average response time of a depot. An alternative formulation is to provide an upper limit on the percentage of demand that violates a given due date requirement. Such a formulation is applicable to the case when a due date requirement is specified in the customer service contract. Developing efficient and effective solution methods for such a model is another interesting future research direction.

In the current model, we assumed that the technician is always available. However, in some applications, the cost of maintaining excess repair capacity is significant. Therefore, another possible future research direction is to incorporate the technician availability into the analysis.

Acknowledgments

The authors would like to thank Professor Izak Duenyas and two anonymous referees for their helpful comments and suggestions.

References

[1] Axsäter, S. (1990). Simple Solution Procedures for a Class of Two-Echelon Inventory Problems. *Operations Research* **38**, pp. 64–69.

[2] Axsäter, S. (1993). Continuous Review Policies for Multi-Level Inventory Systems with Stochastic Demand, in *Logistics of Production and Inventory,* S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin (eds.), North-Holland, New York, pp. 175–197.

[3] Caglar, D. (2001). A Multi-Echelon Spare Parts Inventory System with Emergency Lateral Shipments Subject to a Response Time Constraint. PhD Dissertation, Northwestern University, Evanston, IL.

[4] Cohen, M., P.V. Kamesam, P. Kleindorfer, H. Lee and A. Tekerian (1990). Optimizer: IBM's Multi-Echelon Inventory System for Managing Service Logistics. *Interfaces* **20**, *1*, pp. 65–82.

[5] Cohen, M.A., Y.-S. Zheng and Y. Wang (1999). Identifying Opportunities for Improving Teradyne's Service-Parts Logistics Systems. *Interfaces* **29**, *4*, pp. 1–18.

[6] Graves, S.C. (1985). A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment. *Management Science* **31**, pp. 1247–1256.

[7] Hausman, W.H. and N.K. Erkip (1994). Multi-echelon vs. Single-echelon Inventory Control Policies for Low-demand Items. *Management Science* **40**, pp. 597–602.

[8] Hopp, W.J., M.L. Spearman and R.Q. Zhang (1997). Easily Implementable Inventory Control Policies. *Operations Research* **45**, pp. 327–340.

[9] Hopp, W.J., R.Q. Zhang and M.L. Spearman (1999). An Easily Implementable Hierarchical Heuristic for a Two-Echelon Spare Parts Distribution System. *IIE Transactions* **31**, pp. 977–988.

[10] Kececioglu, D. (1991). *Reliability Engineering Handbook*, Prentice-Hall, Englewood Cliffs, NJ.

[11] Lee, H.L. and K. Moinzadeh (1987). Two-Parameter Approximations for Multi-Echelon Repairable Inventory Models with Batch Ordering Policy. *IIE Transactions* **19**, pp. 140–149.

[12] Moinzadeh, K. and H.L. Lee (1986). Batch Size and Stocking Levels in Multi-Echelon Repairable Systems. *Management Science* **32**, pp. 1567–1581.

[13] Muckstadt, J.A. (1973). A Model for a Multi-Item, Multi-Echelon, Multi-Indenture Inventory System. *Management Science* **20**, pp. 472–481.

[14] Muckstadt, J.A. and L.J. Thomas (1980). Are Multi-Echelon Inventory Methods Worth Implementing in Systems with Low-Demand-Rate Items? *Management Science* **26**, pp. 483–494.

[15] Palm, C. (1938). Analysis of the Erlang Traffic Formula for Busy Signal Arrangements. *Ericsson Technics* **5**, pp. 39–58.

[16] Richards, F.R. (1976). A Stochastic Model of a Repairable-Item Inventory System with Attrition and Random Lead Times. *Operations Research* **24**, pp. 118–130.

[17] Shanker, K. (1981). Exact Analysis of a Two-Echelon Inventory System for Recoverable Items Under Batch Inspection Policy. *Naval Research Logistics Quarterly* **28**, pp. 579–601.

[18] Sherbrooke, C.C. (1968). METRIC: A Multi-Echelon Technique for Recoverable Item Control. *Operations Research* **16**, pp. 122–141.

[19] Sherbrooke, C.C. (1986). VARI-METRIC: Improved Approximations for Multi-Indenture, Multi-Echelon Availability Models. *Operations Research* **34**, pp. 311–319.

[20] Simon, R.M. (1971). Stationary Properties of a Two-Echelon Inventory Model for Low Demand Items. *Operations Research* **19**, pp. 761–777.

[21] Svoronos, A. and P. Zipkin (1991). Evaluation of One-for-One Replenishment Policies for Multiechelon Inventory Systems. *Management Science* **37**, pp. 68–83.

[22] Wang, Y., M.A. Cohen and Y.-S. Zheng (2000). A Two-Echelon Repairable Inventory System with Stocking-Center-Dependent Depot Replenishment Lead Times. *Management Science* **46**, pp. 1441–1453.

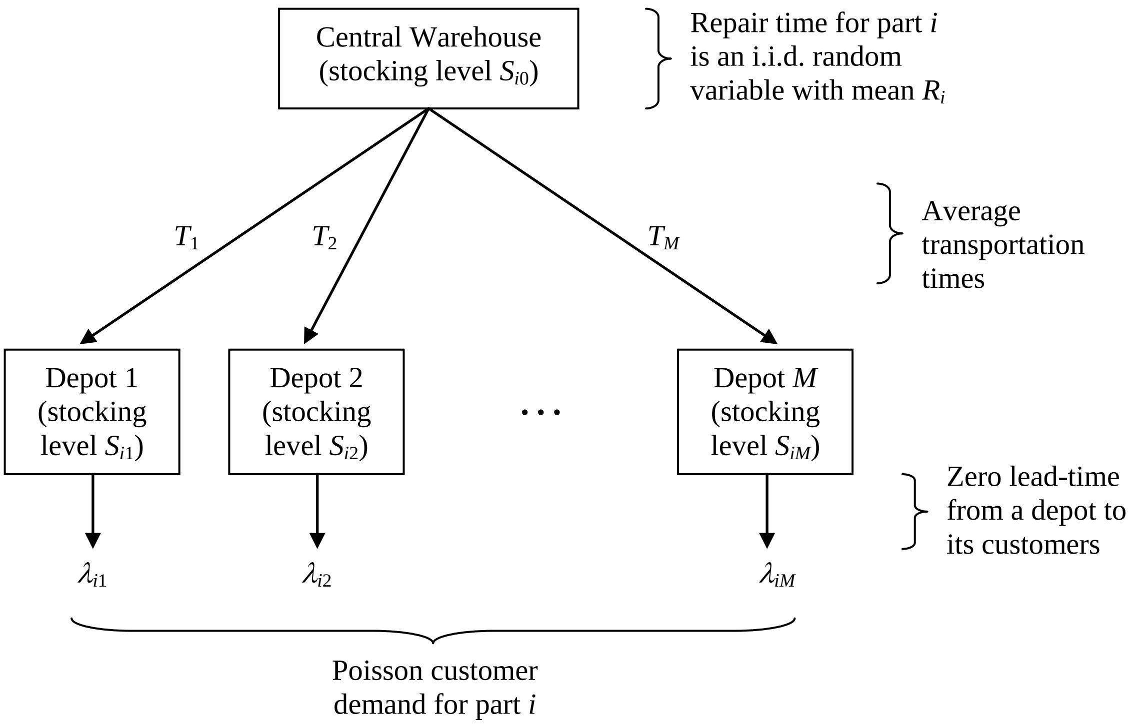

Figure 1. The two-echelon spare parts system

Table 1. Data of cases 8 – 11 in Hopp *et al.* (1999)

	HZS case 8		HZS case 9		HZS case 10		HZS case 11	
n	2		2		2		2	
M	2		2		2		2	
T_1 (hours)	10		5		10		24	
T_2 (hours)	10		24		10		48	
	i = 1	i = 2	i = 1	i = 2	i = 1	i = 2	i = 1	i = 2
h_i ($)	10	20	10	20	10	20	10	20
$E[L_{i0}]$ (days)	50	100	50	100	50	100	50	100
λ_{i1} (units/year)	10	5	10	5	15	2	15	2
λ_{i2} (units/year)	10	5	10	5	5	8	5	8

Note: 1 day = 24 hours; 1 year = 365 days

Note: $\tau_1 = \tau_2$ = 1 hour in all four cases

Table 2. Heuristic and optimal solutions to cases 8 – 11 in Hopp *et al.* (1999)

	HZS case 8	HZS case 9	HZS case 10	HZS case 11
Z^{H2}	137.411	157.166	157.369	166.150
Z^{HZS}	137.412	157.170	157.367	166.154
Z^*	137.411	157.166	147.400	156.164
Z^{LB}	136.638	137.995	131.135	142.441
$(Z^{H2} - Z^*)/Z^*$	0.0%	0.0%	6.8%	6.4%
$(Z^{HZS} - Z^*)/Z^*$	0.0%	0.0%	6.8%	6.4%
$e^{H2} = (Z^{H2} - Z^{LB})/Z^{LB}$	0.6%	13.9%	20.0%	16.6%
$e^{HZS} = (Z^{HZS} - Z^{LB})/Z^{LB}$	0.6%	13.9%	20.0%	16.6%

Table 3. Computational results

	λ_{ij}	$E[L_{i0}]$	h_i	T_j	n = 50; M = 10		n = 100; M = 20		n = 200; M = 40	
					e^{H2}	e^{HZS}	e^{H2}	e^{HZS}	e^{H2}	e^{HZS}
Case 1	$\bar{\lambda}$	$\bar{L}$	$\bar{h}$	$\bar{T}$	12.0%	32.3%	5.4%	32.2%	4.4%	35.6%
Case 2	$\bar{\lambda}$	$\bar{L}$	$\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	11.7%	29.2%	5.5%	23.0%	4.3%	23.4%
Case 3	$\bar{\lambda}$	$\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\bar{T}$	0.2%	2.5%	0.0%	1.7%	0.3%	1.1%
Case 4	$\bar{\lambda}$	$\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	2.4%	4.5%	1.2%	2.2%	0.4%	1.3%
Case 5	$\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\bar{h}$	$\bar{T}$	6.3%	5.0%	5.3%	4.2%	3.7%	3.4%
Case 6	$\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	7.8%	5.0%	6.2%	3.9%	4.6%	3.3%
Case 7	$\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\bar{T}$	0.2%	2.8%	0.5%	1.5%	0.3%	1.0%
Case 8	$\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	3.2%	5.4%	1.4%	3.1%	0.5%	1.7%
Case 9	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\bar{L}$	$\bar{h}$	$\bar{T}$	1.4%	5.0%	1.0%	3.9%	0.3%	3.3%
Case 10	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\bar{L}$	$\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	0.9%	4.7%	0.5%	4.2%	0.2%	3.3%
Case 11	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\bar{T}$	3.3%	3.2%	1.5%	2.7%	1.7%	2.2%
Case 12	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	2.9%	2.7%	3.4%	3.1%	2.5%	2.4%
Case 13	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\bar{h}$	$\bar{T}$	1.3%	5.1%	0.9%	4.0%	0.5%	3.5%
Case 14	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	0.8%	2.9%	0.4%	2.7%	0.2%	2.2%
Case 15	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\bar{T}$	1.9%	3.8%	2.8%	3.1%	1.7%	2.3%
Case 16	$\frac{2i-1}{n}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	3.3%	3.1%	2.2%	2.3%	1.6%	2.2%
Case 17	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\bar{L}$	$\bar{h}$	$\bar{T}$	12.2%	38.4%	5.8%	31.5%	4.6%	32.2%
Case 18	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\bar{L}$	$\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	12.3%	26.6%	5.7%	19.8%	4.7%	22.4%
Case 19	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\bar{T}$	1.3%	3.3%	0.5%	1.7%	0.3%	1.1%
Case 20	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	6.0%	5.5%	2.0%	2.9%	0.6%	1.7%
Case 21	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\bar{h}$	$\bar{T}$	6.7%	4.9%	5.2%	4.0%	4.0%	3.2%
Case 22	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	8.6%	5.3%	6.2%	3.8%	4.6%	3.3%
Case 23	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\bar{T}$	1.8%	4.1%	0.6%	2.0%	0.2%	1.1%
Case 24	$\frac{2j-1}{M}\cdot\bar{\lambda}$	$\frac{2i-1}{n}\cdot\bar{L}$	$\frac{2i-1}{n}\cdot\bar{h}$	$\frac{2j-1}{M}\cdot\bar{T}$	5.6%	6.8%	2.6%	3.9%	1.0%	2.1%

Table 4. Average e^{H2} values for various problem sizes

	M = 2	M = 5	M = 10	M = 20	M = 40
n = 10	8.6%	6.7%	7.3%	5.7%	5.8%
n = 25	6.9%	5.0%	5.7%	3.9%	3.6%
n = 50	6.5%	4.3%	4.7%	3.2%	2.6%
n = 100	5.9%	3.9%	4.4%	2.8%	2.2%
n = 200	5.7%	3.8%	4.2%	2.5%	2.0%