AN APPLICATION OF AN HIERARCHICAL MODELING FRAMEWORK TO THE COMPUTER MAINTENANCE INDUSTRY¹

by

Laura DeNardis, Patricia McCrink, and John Muckstadt

¹This research was supported in part by NSF Grant DDM-8819542.
ABSTRACT

The purpose of this paper is to show how the modeling framework presented by Maxwell et al. [15] can be applied to address various issues facing the management of a computer servicing company. The hierarchical system we describe consists of five models. The first two models are planning models, which address issues of replenishing parts inventory over a relatively long time horizon. Deciding on this long term inventory sourcing strategy for the computer servicing company corresponds to developing a production plan in the Maxwell et al. [15] framework. The third model is a spare parts model, which addresses the issue of establishing stocking levels at each location in the system. By holding spare parts stock at various locations, the company buffers itself against different types of uncertainty. The final two models address issues that require real-time decision making. These are resource allocation models which are needed due to the dynamic nature of the business.

The modeling hierarchy provides a comprehensive decision-making system that takes a complex and interrelated series of questions and organizes them into a manageable framework. The system breaks down the unwieldy sets of interrelated problems into an easily understood sequence of decisions. This approach has been implemented successfully by a major international company.
I. INTRODUCTION

The purpose of this paper is to describe how the hierarchical modeling framework developed by Maxwell et al. [15] has been applied to address a complex set of issues faced by the management of a computer maintenance company. The hierarchical system we discuss consists of five interrelated models: two are similar to production planning models, one is used to address the effects of uncertainty, and the final two deal with real-time decision making. These models have been designed to take a complicated series of issues and to organize them into a workable management system.

The remainder of the paper is organized as follows. In the following section we review the key elements of the Maxwell et al. modeling framework. We then describe the environment that stimulated the development of the models we subsequently describe, and state the specific issues that must be examined for this environment within each phase of the modeling hierarchy. The next several sections provide descriptions of the models used to address the issues that were posed. Finally, we summarize the contents of the paper.

II. THE MODELING HIERARCHY

The hierarchical modeling framework presented in reference [15] was developed for planning and controlling the production and distribution of discrete parts. The authors of that paper proposed an approach, consisting of a set of interacting models, that is divided into three phases: the creation of a production plan, the specification of plan for buffering against uncertainty, and the development of methods for real-time resource allocation and decision making.

Phase one, the production planning phase, addresses the problem of determining what product mix should be produced and distributed over a relatively long period of time. The models used in this phase help establish, for example, the desired capacity over time at each location, the structure of the distribution system, the preferred alternatives for sourcing materials and finished products, and the production lead times and lot sizes for
each product or product family. These planning models are usually deterministic in nature and view much of the planning process in an aggregate manner. For example, capacity is often thought of in terms of machine hours per quarter or per year. Exact sequencing of products through a manufacturing plant may not be of concern at this phase. One instance of such a model is given in reference [13]. However, a model used at this stage must produce a plan that is implementable in the short run. Otherwise, it is of no real value. The integration of decisions produced by these models with real-time decisions is the key to an effective management support system.

The second phase deals with planning for uncertainty. The production planning models are all based on the assumption that input data are known with certainty. Since this is rarely, if ever, true, there is a need to provide buffer inventories or safety stocks and safety time to insure that a system will operate smoothly. The amount of safety stock or time needed for each item or component in a system depends on the physical nature of the production process, lead times, value added characteristics, the sources of uncertainty (yield losses, fluctuating demand, equipment failures, etc.), and the structure of the distribution system. Unlike deterministic planning models, the models developed to address the variety of issues that arise due to the presence of uncertainty can be of several types. Simulation models may be important when studying how much machine capacity and work-in-process inventory are needed throughout a manufacturing system to provide certain flow or cycle times. (See [4] for a discussion of such models.) In other cases, analytic models may be appropriate, such as when determining inventory levels in distribution systems. All these models are used to study the behavior of integrated systems so that the plans put forth in phase one are either appropriately modified or are verified to be adequate.

The third phase of the modeling hierarchy is concerned with real-time resource allocation and decision making. The models used in the first two phases are used to create and allocate capacity allowing a company to achieve its strategic goals in a cost effective
manner. However, these models do not consider the day-to-day dynamic behavior of the systems. Real-time decisions must establish what products to make on what machines and in what quantities considering current inventory levels, available machines and labor, financial projections and demand forecasts. The third phase model must consider such factors. The importance and form of these real-time models must depend on the nature of the operating environments and the level of service that will be provided to the customer.

As we mentioned earlier, our goal here is to illustrate how this three-phase modeling hierarchy has been extended and applied in a specific instance. In particular, we will show how it has been effectively employed by a company that services computers and related equipment found in customer locations.

III. PROBLEM ENVIRONMENT

The company we studied provides repair services for computers and related computing equipment. The company's objective is to provide timely repair at a customer site when equipment fails. To do this, a field engineer is dispatched to diagnose the problem and, if required, to remove and replace failed parts. When a part is required, the field engineer must be provided with the appropriate parts. The hierarchy of models described in the following section addresses the key issues relating to the management of the inventories needed to complete these repairs: procurement of inventory, repair of failed parts, and allocation of inventory investment among parts and stocking locations.

The system we examined to resupply the field engineer consists of a number of parts stocking locations and a central parts repair facility. There are three types of stocking locations. First, the field engineer has a small inventory of commonly failing items. When a field engineer requires a part, he or she obtains the part from a supporting branch. This is normally done on a one-for-one replacement basis, that is, the engineers follow an (s−1,s) policy, except for very low cost and high demand rate items for which the reorder quantity is a fixed batch size. Thus, the branch is the second echelon of the resupply chain.
At the top of this chain is a central warehouse, which normally resupplies the branches. Branch replenishment is also normally done on a one-for-one basis.

If a field engineer requires a part and he or she does not have one, and the supporting branch also does not have one, an emergency order is placed to the central warehouse. If the central warehouse has a unit on hand, it sends it immediately to the field engineer. On the other hand, if the central warehouse does not have one in stock, then each branch is queried in a prescribed order to determine whether or not a unit is available somewhere in the system. If none exists, an emergency order is placed by the central warehouse to the original equipment manufacturer, who normally has the part in stock, but at a very high price. The part is then sent directly to the field engineer.

Part repair and all external procurement is managed through the central warehouse. The central warehouse, as described, is the ultimate source for all parts acquisitions. Resupply of the central warehouse occurs from one of three sources: part repair, machine teardown, or individual part procurement. Part repair is accomplished either internally or externally by a vendor. Machine teardown is the primary means for acquiring the parts that enter the system. Machines (computers, primarily) are purchased on the open market for used computers as a parts source. Parts are removed from the machines, tested and, if appropriate, placed into inventory. The final way to acquire parts is to purchase them singly either from a so-called fourth party source or from the original parts manufacturer. Fourth party vendors normally charge far less than does the original equipment manufacturer. Hence, the original equipment manufacturer is normally a source of last resort.

Figure 1 contains a summary of the above ideas. It shows how the material flows throughout the system.

The description of the system's operation suggests a number of strategic and tactical operational issues that must be considered if customer service is to be achieved in a cost-effective manner.
Figure 1. Material Flow in the Multi-Location Parts Resupply System.
A major strategic question to the company concerns what machine types it will contract to support. This is a non-trivial problem which can only be answered by measuring the logistics consequences of the decision. However, once the decision is made, the company must acquire spare parts to service the machines. Since repair of many types of parts is very cost-effective, an important decision is to determine how much repair capacity is appropriate for different points in time and what parts should be repaired within the facility.

Since some parts cannot be repaired and some safety stock is needed to provide protection against the demand variation for parts over the repair, procurement and transportation lead times, it is necessary to procure parts on a continuing basis. As mentioned, a primary source for these parts comes through machine teardown. Since some parts can be found on many models of computers, the question arises as to which machines should be purchased for teardown. This decision affects unit costs, holding costs and teardown costs, and consequently it greatly determines the cost of repairing a customer's machine. There is also a question of determining when parts should be procured. The decisions as to when to buy a machine and what machine to buy depend on the situation at a particular moment in time. As the situation changes, the decisions may also change. Hence, the modeling framework must address this part sourcing issue from both the longer planning viewpoint, where budgeting and strategic planning are important, and the short-term decision viewpoint, where real-time decisions must be made to establish which machines should be purchased and torn down to obtain which parts.

The final issue relates to the stock levels that must be determined for each spare part type at each location in the system (central warehouse, branches, and field engineers). The more inventory, the better the service, but the higher the initial and subsequent investment. Determining these stock level values is a difficult task. Many papers have been written on this complex subject. See, for example, [1,2,3,5,6,7,8,9,11,12,14,16,17]. We will not concentrate on this issue in this paper. Although the modeling approach we have taken to
solve the problem is new, we will not describe it here, since the discussion would be lengthy. It is the subject of another paper. Nonetheless, establishing stock levels at each location to provide the desired levels of service at minimum cost requires careful modeling. This model must take the repair and procurement decisions and their consequences into account to provide stock that yields the desired level of customer service.

IV. HIERARCHICAL MODELING FRAMEWORK

The hierarchical modeling framework we have developed to address the range of issues associated with the repair and procurement of spare parts is divided into three categories: planning models, spare parts stocking models, and real-time decision models. As shown in Figure 2, there are five models in total that were constructed. The purpose of

<table>
<thead>
<tr>
<th>PLANNING</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL 1. DETERMINE REPAIR CAPACITY</td>
</tr>
<tr>
<td>MODEL 2. DETERMINE AGGREGATE PARTS PROCUREMENT STRATEGY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPARE PARTS DETERMINATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL 3. ESTABLISH STOCK LEVELS FOR EACH PART TYPE AT EACH STOCKING LOCATION</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REAL TIME DECISION MODELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL 4. ESTABLISH A PRIORITY RANKING FOR THE REPAIR OF PARTS CONSIDERING CURRENT CONDITIONS</td>
</tr>
<tr>
<td>MODEL 5. DETERMINE THE LEAST-COST WAY TO MAINTAIN THE OPTIMAL INVENTORY LEVELS CONSIDERING CURRENT CONDITIONS</td>
</tr>
</tbody>
</table>

Figure 2. Modeling Hierarchy.
each of the models is also given in the figure. The next five sections describe each of these
models in turn.

V. REPAIR PLANNING MODEL

Repairing parts that have failed in a customer's machine can provide a low cost
source of replacement parts. This recycling activity has the potential for greatly reducing
the number of parts that must be procured externally. However, to carry out these repairs
requires both equipment and human resources. At a planning level, a decision must be
made as to what this repair capacity should be within the manufacturing facility and what
classes of parts should be repaired within the facility. The repair planning model addresses
both of these questions.

The parts that are candidates for repair are divided into classes. The parts in a given
class require the same types of equipment and labor resources. Different classes require
different types of resources. Thus, for example, there are CRTs, keyboards, memory
boards, hard drives, etc. Each of these types of parts is in a different class, and the
equipment in limited supply needed to repair a hard drive is different from that needed to
repair a CRT. Labor categories also exist. We assume there is one labor category
corresponding to each class of parts that restricts the quantity of parts that can be repaired
within a class. We also assume there is only one type of equipment in each class that
restricts the number of parts that could be repaired.

Those parts that are not repaired internally must be either repaired externally or must
be replaced through procurement.

The goal of the repair planning model is to select which set of parts to repair
internally, purchase externally or repair externally over a planning horizon of six months to
a year, given a constraint on the total amount of available labor and equipment over this
horizon. The model considers the costs associated with internal repair, external repair, or
procurement. Furthermore, it considers the effect on safety stock and lead time stock
associated with each possible source of supply. Since lead times can differ substantially depending on the source of supply, it is essential that this incremental inventory cost be considered in the model. Although some part type may be selected for internal or external repair as the primary source of supply, it is not always possible to repair a failed part. Hence, it may be necessary to procure some parts of a given type even though the primary source of supply is through the repair process. This situation is represented in the model, too.

The repair planning model can be stated as the following integer linear program.

\[
\begin{align*}
\text{min} & \quad \sum_j c_j x_j + \sum_j d_j (1-x_j) \\
& \quad \sum_{j \in S_k} a_j x_j \leq l_k , \quad k = 1,\ldots,K \\
& \quad \sum_{j \in S_k} b_j x_j \leq e_k , \quad k = 1,\ldots,K \\
& \quad x_j = 0,1 ,
\end{align*}
\]

(1)

where

\[
x_j = \begin{cases} 
1, & \text{if part type } j \text{ is repaired internally} \\
0, & \text{otherwise}
\end{cases}
\]

\(K\) = the number of classes of parts,

\(S_k\) = the set of parts in class \(k\), \(k = 1,\ldots,K\),

\(c_j\) = the expected cost to repair part type \(j\) internally

\[= \text{demand rate for a defined period of time for part type } j \times \left( \text{parts cost per unit to repair a type } j \text{ part} \times \text{fraction repairable of part type } j \text{ removed from customer machines} + \text{fraction of part type } j \text{ that are not repairable} \times \text{average unit procurement cost for part type } j \right) + \text{incremental inventory costs} \]

\(d_j\) = minimum of the cost to repair part type \(j\) externally or to procure it
\[ = \min \{ \text{expected procurement cost to satisfy the period's demand for} \]
\[ \text{part type } j \]
\[ + \text{the incremental inventory cost, external repair cost per unit} \]
\[ \times \text{demand rate for part type } j \text{ for a period} \]
\[ \times \text{fraction repairable of part type } j \text{ demand} \]
\[ + \text{the fraction not repairable} \times \text{demand rate for part type } j \text{ demand} \]
\[ \times \text{average procurement cost per unit of type } j \text{ part} \]
\[ + \text{the incremental inventory cost} \} \]

\[ a_j = \text{total expected hours in a period of the appropriate labor type} \]
\[ \text{needed to repair part type } j \]
\[ = \text{period demand rate for part type } j \]
\[ \times \text{average labor hours per unit repaired of type } j \]
\[ \times \text{expected fraction of type } j \text{ parts that are repairable} \]

\[ b_j = \text{total expected hours in a period of the appropriate limited} \]
\[ \text{equipment type needed to repair part type } j \]
\[ = \text{period demand rate for part type } j \]
\[ \times \text{average hours of the equipment type required} \]
\[ \text{to repair a unit of part type } j \]
\[ \times \text{expected fraction of type } j \text{ parts that are repairable}. \]

\[ l_k = \text{available hours of class } k \text{ labor in the period} \]
\[ e_k = \text{available hours of class } k \text{ equipment in the period}. \]

Note that the model does not consider the costs of labor and equipment directly. For a given value of \( l_k \) and \( e_k \), these costs are fixed. In the model's implementation we considered various levels of \( l_k \) and \( e_k \) along with their associated fixed costs. The software we developed generates appropriate combinations of the \( l_k \) and \( e_k \) values. The total cost of a solution, then, includes both these fixed costs and the expected costs associated with the solution to Problem (1).
The incremental inventory costs mentioned earlier are found based on the need to have two types of inventory. First, there is inventory needed to cover the expected demand over the administrative lead time plus the expected procurement lead time plus the transportation lead time plus the expected repair lead time. The average lead time is calculated considering the fraction of units that are repairable for the particular part type. Second, the safety stock needed to provide a specified service level (either fill rate or probability of not running out of stock over a lead time) is established. The total inventory requirement is the sum of the lead time requirement and the safety stock requirement.

The model presented in Problem (1) has a very special structure which makes this problem relatively easy to solve. Observe that Problem (1) is decomposable by part class. That is, it can be written as

\[
\sum_{k=1}^{K} \min_{x_j=0,1} \left\{ \sum_{j \in S_k} c_j x_j + \sum_{j \in S_k} d_j (1-x_j) : \sum_{j \in S_k} a_j x_j \leq l_k ; \sum_{j \in S_k} b_j x_j \leq e_k \right\}
\]  

(2)

The inner problem, which is associated with class k only, can be written as

\[
\min_{j \in S_k} \sum_{j \in S_k} (c_j - d_j) x_j + \sum_{j \in S_k} d_j ,
\]

\[
\sum_{j \in S_k} a_j x_j \leq l_k ,
\]

\[
\sum_{j \in S_k} b_j x_j \leq e_k ,
\]

\[
x_j = 0,1 .
\]

(3)

We developed a specialized algorithm for solving a relaxation to the problem. The purpose of this procedure was to generate a range of combinations of resource levels \( l_k \) and \( e_k \) in addition to obtaining the values for the \( x_j \) variables corresponding to various amounts of available resources. We also note that in the application we considered, no single part type in a class consumes more than a small amount of the available capacity for either labor
or equipment. Hence, we could relax the integrality constraint and solve Problem (3) as if it were a linear program. Since the right-hand side parameter values were adjusted, and costed separately as fixed costs, the linear relaxation would be an appropriate problem to solve. We chose to use the specialized approximation procedure rather than some standard linear programming algorithm to solve this relaxation. Since the software is intended for interactive use, we chose to implement a specialized Lagrangian relaxation algorithm.

The resource levels were not included as decision variables in this linear model. If resource costs were proportioned to their levels, then it would have been appropriate to do so. Since we did not want to restrict ourselves in this manner, we chose the approach we have described instead.

VI. PROCUREMENT PLANNING MODEL

One goal of the repair planning model was to establish which parts should be repaired primarily internally or externally. Given that decision, a number of parts will still have to be procured. This will occur because there is insufficient repair capacity or there are parts of a given type that cannot be repaired. In any case, once the repair capacity decision has been made and the repair decision is made for each part type, there remains a net requirement for each part type that must be satisfied through external procurement. The main goal of the procurement planning model is to determine the most cost-effective strategy for fulfilling these net requirements for each part type. A second goal is to provide an expected procurement budget for management to use in planning business costs and in projecting cash flow throughout the planning horizon. A third goal is to determine the expected average unit cost for each part type. These costs can be used in the spare parts model, which is described subsequently. These estimates can be compared with historical averages, and, if they differ significantly from these averages, it may be necessary to re-evaluate the repair planning decisions.
As in the repair planning model, we consider a relatively long planning horizon, say 6 months to one year. This planning horizon is subdivided into periods, perhaps three months in duration. This allows us to evaluate the impact of current decisions on future expenditures and to estimate cash flows and inventory balances over the planning horizon. Since inventory can be purchased at one time and held for use in a subsequent period, it is important to understand the long term consequences and opportunities of purchasing parts early in the planning horizon.

The objective is to select a procurement strategy over the planning horizon that minimizes the sum of machine purchase costs, machine teardown costs, individual parts procurement costs, inventory holding costs, and inventory pipeline costs due to the procurement source. The model indicates

1. The expected number of machines of each type to purchase in each time period,
2. The expected number of parts of each type to purchase individually in each period, and
3. The expected number of parts of each type to carry over from one period to the next.

There are two types of constraints imposed. First, the number of units of a part type carried from one period into the next is limited. This type of constraint limits the exposure to obsolescence and excess inventories that often allows inventories to grow and write-offs to occur in practice. Any inventory in excess of the maximum at the end of a period is assumed to be scrapped.

The second type of constraint pertains to each part type. It states that the number of parts of that type taken from all types of machines that are torn down in that period plus the number of units of that part type carried into this period from the previous one minus the number of parts of that type carried over to the following period minus the number of units of that part type that are scrapped must equal the net requirement for that part type in that period.
Let

\[ g_{it} = \text{the teardown and procurement cost plus the inventory costs implied for buying a machine of type } i \text{ in period } t, \]
\[ c_{jt} = \text{procurement and inventory costs for a part of type } j \text{ in period } t, \]
\[ b_{jt} = \text{net demand for part type } j \text{ in period } t \text{ (after considering on-hand stock at the beginning of the horizon plus repair center production during the period),} \]
\[ a_{ij} = \text{expected number of usable parts of type } j \text{ per machine of type } i \text{ (parts could be missing or broken beyond their ability to be repaired),} \]
\[ f_{jt} = \text{the cost to scrap a unit of a type } j \text{ part in period } t, \]
\[ h_{jt} = \text{the holding costs of carrying one unit of part type } j \text{ from period } t \text{ into the following period,} \]
\[ z_{it} = \text{number of machines of type } i \text{ purchased in period } t, \]
\[ x_{jt} = \text{number of parts of type } j \text{ purchased as single parts in period } t, \]
\[ y_{jt} = \text{number of parts of type } j \text{ carried over from period } t \text{ to the following period, } (y_{j0} = 0), \]
\[ w_{jt} = \text{number of parts of type } j \text{ that are scrapped in period } t, \]
\[ u_j = \text{maximum number of units of part type } j \text{ that can be carried from one period to the next.} \]

The procurement planning model is the following linear program:

\[
\min \sum_{t} \left\{ \sum_{i} g_{it} z_{it} + \sum_{j} (c_{jt} x_{jt} + f_{jt} w_{jt} + h_{jt} y_{jt}) \right\}
\]

subject to

\[ y_{jt} \leq u_j, \]
\[ y_{jt-1} + \sum_{i} a_{ij} z_{it} + x_{jt} - y_{jt} - w_{jt} = b_{jt}, \]
\[ z_{it}, x_{jt}, y_{jt}, w_{jt} \geq 0. \]
VII. SPARE PARTS MODELS

The investment in spare parts made by computer service companies is, in our experience, very substantial. While this investment is high, the level of service provided to customers is often only marginally satisfactory. Inventory turns of one or less and fill rates of 70% or less are not uncommon. Thus, one of the most effective ways to reduce cost and to improve customer satisfaction is to manage spare parts efficiently.

To address this problem, we developed a model and a heuristic that calculates the approximate optimal inventory levels for each part at each location in the system subject to a constraint on the investment budget. While the model we developed and implemented considers the multi-echelon nature of the distribution system, we will present a model as if there is only a single location in the system. The complete modeling approach is the topic of another paper. The model we will describe is similar to others found in the literature mentioned earlier. Our purpose here is not to provide a detailed description of a spare parts stocking model, but rather to show how the modeling framework was applied in a particular environment.

The model we describe is similar to the one we have implemented; however, this model is simplified to finding the optimal stock levels for all parts stocked at a single location given a budget constraint. As in the model we implemented, these stock levels, once determined, are targets to be maintained at all times. The real-time models we subsequently describe are designed to ensure that the target values are maintained in a cost-effective and timely manner.

As in the model we implemented, we assume an (s-1,s) continuous review policy is followed. As we described previously, if a demand for a part occurs when no stock is on hand, an emergency order is placed with the original equipment manufacturer to provide the part. However, there is often a substantial premium over the normal cost that must be paid to obtain the parts through emergency resupply requisitions.
The model's objective is to minimize the expected annual incremental procurement cost (emergency procurement cost minus normal replenishment cost) for all stocked items. When computing the optimal stock levels, the following factors were considered:

1. Procurement and repair lead times.
2. Percentage of units repaired or procured.
3. Emergency and normal procurement costs.
4. Demand rates (we assume a Poisson demand process).

Furthermore, the following constraints were considered. First, the cost of stocking all parts to their optimal level should not exceed the given budget. Second, no additional parts should be added once the fill rate of each part (i.e. the percentage of time that a request for that part will be filled with no delay) has attained or exceeded a management-specified fill rate. Finally, the optimal inventory level for each part should be at least as large as the current starting inventory level, and it should also be no less than the floor for that item, that is, the integer part of the expected lead time demand.

The mathematical model, based on these assumptions and constraints is

$$\min \sum_{i=1}^{n} \pi_i \mu_i \sum_{x \geq s_i} p(x \mid \mu_i T_i)$$

subject to

$$\sum_{i} c_i (s_i - l_i) \leq b,$$

$$F(s_i) \leq f_i,$$

$$s_i \geq l_i \geq 0,$$

$$s_i \geq \lfloor \mu_i T_i \rfloor, \quad i = 1, \ldots, n$$

where

- $\mu_i$ is the demand rate for part $i$,
- $c_i$ is the normal procurement cost for part $i$,
- $R_i$ is the repair lead time for part $i$, 

---

16
\( \alpha_i \) is the probability that a failed part \( i \) is repaired,
\( P_i \) is the normal procurement lead time for part \( i \),
\( T_i = \alpha_i R_i + (1 - \alpha_i)P_i \), the average resupply lead time for part \( i \),
\( l_i \) is the current inventory level for part \( i \),
b is the maximum incremental investment in inventory,
f_1 is the maximum desired fill rate for part \( i \),
s_i is the stock level for part \( i \) (the decision variable),
\( F(s_i) \) is the fill rate for part \( i \) given \( s_i \),
\( \pi_i \) is the incremental emergency procurement cost for part \( i \), and
\( n \) is the number of part types in the system.

The function \( p(x \mid \mu_i T_i) \) measures the probability that \( x \) units are in resupply—on-order or in repair. This probability can be calculated as a truncated Poisson distribution as follows:
\[
p(x \mid \mu_i T_i) = \frac{e^{\mu_i T_i} (\mu_i T_i)^x / x!}{\sum_{k=0}^{s_i} e^{-\mu_i T_i} (\mu_i T_i)^k / k!}.
\]

This expression is developed in reference 9. The truncation occurs because the above model is a lost sales type model. That is, if a demand occurs for an item when there is no stock on-hand, the demand is met from outside the normal resupply system. Consequently, there will never be more than \( s_i \) units in normal resupply.

Note that initial investment in safety and "pipeline" stock is \( c_i (s_i - l_i) \) for part \( i \). The annual expected normal procurement cost will be \((1 - \alpha_i) \mu_i c_i \cdot F_i(s_i)\) and the annual expected repair cost will be \( \alpha_i \mu_i r_i F_i(s_i) \) for part \( i \).

The spare parts problem is a knapsack type problem and could be solved using a dynamic programming approach. In practice, however, the number of items is large and the cost of one unit of any part type is very small compared with the total budget \( b \). Hence,
we have used a simple greedy marginal analysis method to find the $s_i$ values. A Lagrangian relaxation approach, such as described by Fox and Landi [12], could also be employed.

VIII. REAL-TIME REPAIR SCHEDULING

The repair planning model provides management with a guide as to which parts to repair and what capacity to provide in the repair facility. As a planning model, it does not address the daily dynamics of repair scheduling. Consequently, a weekly or daily dispatching model is needed to aid in detailed decisions. There are many situations that make such a model necessary. For example, suppose a part is normally repaired but at a given moment there is an excess supply of this part in inventory such that the probability of stocking out is almost zero. In this case, we may wish to scrap or hold the part for repair at a later time and use the repair capacity at the present time for other parts. In another case, suppose the same part is in great need (i.e. the probability of stocking out is high) but the necessary components to fix the part are not available. Thus, we may choose to acquire a part of that type from an outside vendor, the original supplier, or from a machine teardown. Finally, the unit may be needed to raise the part type's inventory position to a value at or below the target level. However, the queue of waiting repair jobs at that time may contain many parts whose probability of running out of stock may be much higher or whose expected cost of running out of stock is significantly higher. Consequently, it may be economically desirable to obtain this part from one of the alternative sources of supply, due to the capacity limitation.

The real-time repair scheduling model provides a mechanism for determining which part types should be repaired over a short planning horizon (e.g. a few days or a week) so as to minimize the expected penalty costs over the horizon subject to a constraint on repair capacity. The cost of alternative supply in the case a shortage occurs (i.e. the penalty cost per unit), the current on-hand inventory level plus due in from other sources over the repair
planning horizon for each part in the repair queue, the repair time needed to repair each candidate part type, the number of failed parts of each type in the repair queue, the part demand rate, and the repair capacity are all assumed known. For ease of presentation, assume the demand process for each part type over the horizon is a Poisson process.

The following nomenclature is used in the statement of the model. Let

\[ y_i = \text{the amount on-hand plus due-in from other sources over the repair scheduling horizon plus the amount repaired in the repair scheduling horizon for part } i, \]

\[ \pi_i = \text{penalty cost for part } i, \text{ the difference between the emergency procurement and repair costs for part } i, \]

\[ \lambda_i = \text{demand rate for part } i, \]

\[ T = \text{length of the repair scheduling horizon,} \]

\[ g_i = \text{on-hand plus due-in plus failed parts presently in the queue for part } i \text{ awaiting repair,} \]

\[ f_i = \text{on-hand plus due-in during the scheduling horizon for part } i, \]

\[ a_i = \text{hours required to repair one unit of part } i, \]

\[ A = \text{the available hours of repair capacity over the planning horizon measured in hours, and} \]

\[ p(x|\lambda_i T) = e^{-\lambda_i T}(\lambda_i T)^x / x!. \]

The model is formulated as follows:

\[
\min \sum_i \pi_i \sum_{x>y_i} (x - y_i) p(x|\lambda_i T) 
\]

subject to

\[
\sum_i a_i(y_i - f_i) \leq A, \\
\]

\[
f_i \leq y_i \leq g_i, \\
y_i \text{ integer.}
\]
Note the similarity of this mathematical model with the spare parts model discussed earlier. We also propose solving this non-linear integer program using a simple greedy heuristic since the objective function is discretely convex. The greedy heuristic works well in practice since the proportion of the total repair capacity required to repair any single unit is very small. Consequently the heuristic should produce a solution close to optimal. Furthermore, the additional expected benefit provided by an optimal solution is of little practical value since there are modeling approximations and averages used in the data. We have found that the solution obtained using the simple heuristic provides a useful and effective guide to repair scheduling.

Once the set of parts is identified for repair, the sequence in which to repair the units must be determined. The solution to the above problem provides this sequence. The greedy heuristic used to solve that problem generates a list of parts to repair based on the following ratio: the expected per unit reduction in penalty cost divided by the time needed to accomplish the repair of one unit. The units that have the highest ratios are selected by the greedy heuristic as the ones to repair. By sorting the units that will be repaired according to these ratios from the largest value to the smallest value we will produce the desired repair sequence. Thus the unit that reduces the expected penalty costs the most per unit of repair time should be repaired first.

**IX. REAL TIME PROCUREMENT MODEL**

Recall that the procurement planning model provides an expected number of each machine type and individual part type to procure over the planning horizon. That model ignores the daily requirements for inventory replenishment; rather, it provides useful budget planning information and identifies the types of machines that will be cost-effective to procure throughout the planning horizon. The number of each part type that should be procured to raise each part's inventory position to the desired level varies considerably over time. Furthermore, machines on the secondary market vary in price over time. Thus it is
an extremely difficult task to determine the best course of action to take at any given moment since the combination of parts that need to be purchased, machine availability, and price changes from day to day. The procurement planning model does not address this type of question, but the real-time procurement model does.

As mentioned previously, to satisfy immediate needs, parts may be procured individually or may be obtained from a machine purchased for teardown. Any particular part type may be found on a variety of machines. Furthermore, each type of machine typically contains many parts that could be of use at some time. If a machine is purchased for teardown, some parts will be needed to satisfy current needs, that is, to raise the inventory position for these parts towards their desired levels. Other parts may be stripped off of the machine and placed into inventory even though the resulting inventory position will be above the target level. Some parts that could be of future use may not be stripped from the machine because the cost of doing so exceeds the value gained. Since the parts contained in each machine type differ, and the need for and value of parts changes over time, it is necessary to have a model that properly addresses the real-time parts procurement problem.

The real-time procurement model considers several critical factors that aid in determining which machines should be bought at a point in time and which parts should be stripped off of these machines. As a side benefit, the software implementation of the model provides real-time part costing information that properly allocates machine purchase costs to stripped parts. This feature turns out to be of significant value when determining the true cost of satisfying past and future customer demand. By providing an adequate method for determining the value and cost of a part, it helps eliminate the desire to strip off all parts that have only a negligible chance of being used at some future time. This was previously a common practice in the company we examined. It resulted in the company's having unnecessarily large inventory write-offs and incorrect costing of service calls.
Given the state of the inventory at a point in time, the real-time procurement model suggests which machines to purchase and which parts to buy on an individual basis by examining the economic consequences of purchasing decisions. If a machine is considered for purchase, it is necessary to determine its expected value compared against the cost of buying and tearing down the machine. Once this calculation has been made, the resulting value is compared to the costs of procuring individually those parts on the machine that are required at that time.

The expected value of the machine is comprised of several items. The most obvious benefit arises because parts that are torn down and used immediately do not need to be procured individually. Also, there are possible future savings since any parts from teardown will not need to be procured in the future. In addition, the company may avoid an emergency stockout condition by stocking excess parts. However, there are also costs associated with the machine. Any parts that are torn down and are found to be defective will need to be repaired. Thus, the cost of this operation needs to be taken into account. Also, any parts that are torn down but are considered surplus will have holding costs until they are removed from inventory. Of course, the timing all of these future costs and benefits must be determined and need to be discounted to present worth.

The mathematical model is based on the assumption that the demand process for procured parts for each part type is a Poisson process. A brief summary of some of the calculations is given in the Appendix.

This model was implemented by considering the possible purchase of all available and appropriate machine types on a particular day. The calculations outlined in the Appendix were made for each machine type. The machine type providing the highest net value (expected value minus machine procurement and teardown costs versus individual parts procurement costs) was suggested to be purchased. After adjusting the inventory levels to reflect the addition of this machine type, the evaluation process is repeated to establish the best machine type to purchase next. This greedy type heuristic is
repeated until all immediate parts requirements have been satisfied or until it is not economically justifiable to procure any additional machines of any type. This procedure can clearly find a less than optimal solution. However, the solutions generated in practice using this procedure have appeared to be quite reasonable. By using this model, the company has altered its procurement decisions while reducing its inventory requirements and improving its service level. The software also allows the user to try any sequence of suggested machine purchases and evaluates the economic consequences so that if a solution produced by the greedy marginal analysis approach appears to be questionable, an alternative solution can easily be evaluated.

X. SUMMARY

We have developed and described an hierarchical approach that addresses various issues facing the management of a company that maintains computers and related equipment. This approach was based on the modeling framework described by Maxwell et al. [15]. Our system consists of five interconnected models. The first two are planning models, which address issues associated with parts replenishment over a relatively long time horizon. The sourcing strategy provided through these models corresponds to developing a production plan in the Maxwell et al. modeling framework. The third model is a spare parts stocking model, which addresses the issue of establishing inventory levels for each part at each location in the system. By maintaining spare parts inventories at various locations, the company buffers itself against different types of uncertainty. Finally, we discussed two models for real-time decision making. They are real-time resource allocation models which are needed to address the daily dynamics of the business environment.

The modeling hierarchy provides a comprehensive system that takes a series of complex and interrelated questions and organizes them into a manageable decision system. The system reduces these complicated problems into an easily understood sequence of
decisions. The hierarchical modeling approach we have described has been successfully implemented by a major international company.

ACKNOWLEDGEMENTS

We are pleased to acknowledge the support provided to us by Mr. David Briskman, Professor Peter Jackson, and Mr. Ronald Nawrocki. This research was supported in part by the National Science Foundation under Grant DDM 8819542.

REFERENCES


APPENDIX

The purpose of this appendix is to develop elements of the real-time procurement model. We begin by stating some nomenclature.

Let

\[ \lambda_i = \text{the demand rate for some part } i, \]
\[ h_i = \text{the holding cost rate for part } i \text{ exclusive of the interest rate,} \]
\[ \pi_i = \text{the penalty cost per shortage incident for part } i, \]
\[ \text{the original equipment manufacturers procurement cost} \]
\[ \text{plus emergency administrative costs}, \]
\[ \tau_i = \text{the nominal average lead time for part } i, \]
\[ s_i = \text{the target stock level for part } i \text{ determined by the spare parts model,} \]
\[ g_j = \text{the teardown and procurement cost for a machine of type } j, \]
\[ I_j = \text{the set of parts in machine } j \text{ that require procurement to raise} \]
\[ \text{their stock levels towards their desired target values,} \]
\[ c_i = \text{average procurement cost for part } i, \]
\[ b_i = \text{vendor procurement cost for part } i, \]
\[ \alpha = \text{interest rate.} \]

The model is also based on the assumption that the demand process for procured parts for type i is a Poisson process.

The decision to be made is as follows: Should we procure part i now as an individual part or should we purchase a machine, tear it down, and obtain part i and other parts through this process? Since there may, in fact, be several machine types that contain the desired part, the following types of calculations would be made for each candidate machine type.

Suppose we are considering whether or not to purchase a machine of type j. If the machine is not purchased and the needed parts are procured individually, the cost would be
\[
\sum_{i \in I_j} b_i ,
\]
assuming there is only one of each part type on a machine. On the other hand, if a machine of type j is purchased, then we must calculate the expected cost and value of doing so.

This expected value is

\[ g_j \text{ (the machine procurement and teardown cost)} \]

\[ + \text{the expected discounted holding costs for all parts whose resultant stock levels would exceed their target values} \]

\[ - \text{expected discounted reduction in parts shortage costs due to the machine increasing stock levels beyond the target values} \]

\[ - \text{expected discounted value associated with the avoidance of future procurement costs for parts stocked above their target values due to the acquisition of the machine.} \]

We will show how each of the above terms can be calculated.

Suppose the acquisition of the machine results in an excess of stock for part type i. Further, assume the desired stock level for part type i will be exceeded by several units if the machine of type j is procured. Then the expected discounted holding cost for the j-th unit of part type i in excess of s_i is

\[
\int_0^\infty (\text{holding cost rate}) \cdot (\text{discounting effect given this j-th unit is required at time t in the future}) \cdot (\text{probability density that the j-th unit must be purchased t time units in the future}) \, dt
\]

\[ = \int_0^\infty h_i \int_0^t e^{-\alpha u} \lambda_i e^{-\lambda_i t} \frac{(\lambda_i t)^{j-1}}{(j-1)!} \, du \, dt \quad \text{(after some algebra)}
\]

\[ = \frac{h_i}{\alpha} \left( 1 - \left[ \frac{\lambda_i}{\alpha + \lambda_i} \right]^j \right).
\]

The total expected discounted cost is found by summing over the excess stock amount for each item and across items found in the machine. If the part stock level is not affected by
the calculation or as a result of the acquisition is not in excess of the desired stock level, then this incremental expected holding cost is zero.

Next we give an approximate expression for the expected discounted value of the reduction in the penalty cost due to adding the \( l \)-th unit of part \( i \) in excess of \( s_i \). This penalty cost avoidance occurs if the demand for the \((s_i + l)\)-th unit occurs following the lead time for acquiring the machine, which we denote by \( L \), and prior to the time at which the \((l-1)\)-th unit is demanded plus the nominal lead time \( \tau \). That is, if \( T_j \) is the time of the \( j \)-th subsequent demand, then the events \( T_{l-1} + T_{s_i+1} \geq L \) and \( T_{l-1} + \tau_i \geq T_{l-1} + T_{s_i+1} \) must occur if the penalty is to be avoided, recalling that the first demand occurs at the present time, time 0. The exact expression for this expectation requires evaluating

\[
E \left[ \pi_i e^{-\alpha(T_{l-1} + T_{s_i+1})} \right] \frac{1}{\{T_{l-1} + T_{s_i+1} \geq L\}} \frac{1}{\{T_{l-1} + \tau_i \geq T_{l-1} + T_{s_i+1}\}},
\]

where \( 1(\{\}) \) is the indicator function. This expectation is very tedious to compute. After some experimentation we found that the following expression adequately approximates this expectation:

\[
\pi_i \left( \frac{\lambda_i}{\lambda_i + \alpha} \right)^{s_i+1} \left\{ 1 - \left( 1 - \sum_{w=0}^{l-2} e^{-\lambda_i L \frac{\left(\lambda_i L\right)^w}{w!}} \right) \cdot \left[ 1 - \sum_{w=0}^{s_i} e^{-\beta(L - \tilde{T}_{l-1}(L))} \frac{(L - \tilde{T}_{l-1}(L))^w}{w!} \right] \right. \\
\left. - \sum_{w=0}^{s_i} \frac{(\lambda_i + \alpha)^w \tau_i}{w!} \right\},
\]

where \( \tilde{T}_{l-1} = (l-1)/\lambda_i \)

and

\[
\tilde{T}_{l-1}(L) = \left( \frac{l-1}{\lambda_i} \right) \left( 1 - \sum_{w=0}^{l-2} e^{-\lambda_i L \frac{\left(\lambda_i L\right)^w}{w!}} \right).
\]
The final term measures the expected discounted future procurement cost that could be avoided if the machine is purchased now. For the \( l \)-th unit in excess of \( s_i \), this expected discounted cost is

\[
\int_0^\infty c_i \left( e^{-\alpha t} \frac{(\lambda_i t)^{l-1} \lambda_i e^{-\lambda_i t}}{(l-1)!} \right) dt
= c_i \left[ \frac{\lambda_i}{\lambda_i + \alpha} \right]^l.
\]

The total expected reduction in procurement cost is found by making these calculations for all appropriate values of \( l \) and for all appropriate part types and summing the resultant values.
Figure 1. Material Flow in the Multi-Location Parts Resupply System.