Optimal Capacity Scheduling-I
Author(s): Arthur F. Veinott Jr. and Harvey M. Wagner
Source: *Operations Research*, Vol. 10, No. 4 (Jul. - Aug., 1962), pp. 518-532
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/168053

OPTIMAL CAPACITY SCHEDULING—I

Arthur F. Veinott, Jr. and Harvey M. Wagner

Stanford University, California

(Received October 20, 1961)

Efficient algorithms are developed optimizing an important class of capacity scheduling models. The specific problem considered can be simply described in terms of contracting for warehousing capacity. Contracts must be let for warehouse capacity over n time periods, with the minimum capacity to be provided in each time period being specified. Savings may be achieved by long-term leasing arrangements or by contracting at favorable periods of time, even though this creates idle capacity at certain time periods. A minimum cost solution to this problem is sought. The mathematical model also applies to problems of equipment replacement and overhaul; checkout, repair, and replacement of stochastically failing equipment; determination of economic lot size, product assortment, and deterministic batch queuing policies; labor-force planning; and multicommodity warehouse decisions. For some of these problems, such as equipment replacement, the computing algorithms presented are even more efficient than schemes heretofore proposed for simpler versions of the same problem.

THIS PAPER has a twofold purpose: (1) to exhibit simple and efficient algorithms for solving a particular class of optimization problems, and (2) to demonstrate the wide applicability of this class, which includes, as significant models, capacity scheduling; equipment replacement and overhaul; automatic checkout, repair, and replacement of stochastically failing mechanisms; determination of economic lot sizes, product assortments, and batch queuing policies; labor-force planning; and multicommodity warehouse decisions. Of some importance is the fact that not only do the algorithms assist in solving generalized versions of these models, but in many cases, such as equipment replacement, they actually improve on computational schemes heretofore proposed for simplified versions.

Because of the entire paper's length, it is published in two parts, the second part immediately following this one. In containing the number of pages to a reasonable amount, discussions of certain detailed and relatively minor considerations have been omitted here; the reader interested in these aspects will find them treated in full in reference 15, which is available from The Rand Corporation.

In the next two sections confusion that would be engendered by simultaneously referring to several of the models has been avoided by keeping the exposition in terms of one particular problem, capacity scheduling; in the final section, attention is turned to the other interpretations of the model. The specific capacity scheduling problem is described as follows: a decision maker must contract for warehousing capacity over n time periods, the minimal capacity requirement for each period being deterministically specified. His economic problem arises because savings may possibly accrue by his undertaking long-term leasing or contracting at favorable periods of time, even though such commitments may necessitate leaving some of the capacity idle during several periods. Clearly this programming model might also apply to other types of capacity, such as transport facilities, insurance protection, and leased telephone lines.

The next section starts with a precise mathematical statement of the model, shows the equivalence of the model to a special type of transshipment network model, and states the qualitative properties of an optimal program. Following is an examination of how the model may be solved with linear programming techniques, either as an ordinary or as a reduced transshipment problem. For the latter approach, a rule is given for efficient computation of the required cost coefficients; the rule is based on a dynamic programming recursion that fully exploits the particular structure of the model. A comparison is made of the two transshipment formulations and a test is provided for determining which of the two approaches is more promising in any particular application. The third section concentrates on special patterns for the minimal requirements, and in particular, horizon planning procedures are derived for monotonic patterns. In the final section several of the alternative interpretations of the model are discussed.

In Part II special computing algorithms are presented that, when applicable, are significantly more efficient than the transshipment calculations. The results are also extended to situations in which initial capacity exists and to the multicommodity warehousing model.

MATHEMATICAL FORMULATION

Description as a Transshipment Problem

WE NOW formalize the discussion in the preceding section. Let D_k be the minimal capacity requirement during period k, $k=1, 2, \cdots, n$. Capacity acquired at the beginning of period k, available for possible use during periods $k, k+1, \cdots, t-1$, and relinquished at the beginning of period t is said to be *available* for the interval $[k, t]$. Let X_{kt} denote the amount of capacity available during the interval $[k, t]$ and let c_{kt} be the associated unit cost. Then the problem is to find X_{kt} that minimize

$$\sum_k \sum_{t>k} c_{kt} X_{kt}, \tag{1}$$

subject to the capacity constraints

$$\sum_{k\leqq j} \sum_{t>j} X_{kt}-S_j=D_j, \qquad (j=1, 2, \cdots, n; X_{kt}\geqq 0; S_j\geqq 0) \tag{2}$$

where the S_j may be interpreted as slack variables representing unused capacity in period j.† We denote optimal program values by asterisks, e.g., X_{kt}^*.

It is plausible to expect in most real problems that

$$c_{kt} \leqq c_{rj} \quad \text{for every} \quad r \leqq k < t \leqq j, \ddagger \tag{3}$$

which states that if $[r, j]$ encompasses $[k, t]$, then the unit cost of the activity for $[r, j]$ is at least as great as the unit cost of the activity for $[k, t]$. Unless otherwise stated we assume in the remainder of this paper that the model is so formulated that (3) is satisfied.§

By a simple manipulation of the restrictions (2), we show that the problem described above is a transshipment linear programming model.[12] We give the details for the case $n=4$, but the reader will observe the proof is general. Writing (2) in full we have

$$X_{12}+X_{13}+X_{14}+X_{15} \qquad -S_1 = D_1, \tag{4}$$

$$X_{13}+X_{14}+X_{15}+X_{23}+X_{24}+X_{25} \qquad -S_2 = D_2, \tag{5}$$

$$X_{14}+X_{15} \quad +X_{24}+X_{25}+X_{34}+X_{35} \qquad -S_3 = D_3, \tag{6}$$

$$X_{15} \quad +X_{25} \quad +X_{35}+X_{45} \qquad -S_4 = D_4. \tag{7}$$

Subtracting (6) from (7), (5) from (6), (4) from (5), and multiplying (7) by (-1) yields the equivalent system written in matrix form (with one redundant equation):

$$\begin{Vmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 1 \end{Vmatrix} \begin{Vmatrix} X_{12} \\ X_{13} \\ X_{14} \\ \cdots \\ S_3 \\ S_4 \end{Vmatrix} = \begin{Vmatrix} D_1 \\ D_2-D_1 \\ D_3-D_2 \\ D_4-D_3 \\ -D_4 \end{Vmatrix}. \tag{8}$$

The structure (8) is that of a transshipment model, and it follows as one consequence, that if the D_k are integer valued, there exists an integer valued optimal program.

In reference 12 two alternative procedures are proposed for solving a transshipment problem with p restrictions. The first procedure requires that a related transportation problem (TP) with $2p$ restrictions be solved. The second method requires solving up to $p/2$ 'shortest-route problems'

† If there are several ways of providing capacity for a given interval, we choose one with the lowest unit cost.

Clearly if any $c_{kt}<0$, then an unbounded solution exists; throughout we assume that $c_{kt} \geqq 0$.

‡ For purposes of testing whether (3) holds, we adopt the convention that $c_{ij}=+\infty$ for each interval $[i, j]$ for which no activity is defined.

§ It is shown in reference 15 that if initially (3) does not hold, it is possible to redefine the c_{ij} so that (3) holds and so that any optimal program with the new unit costs $\hat{c}_{ij}=\min_{k \leqq i, j \leqq t} c_{kt}$ is after a simple transformation optimal for the original problem.

and then solving a transportation problem with at most p restrictions; this latter method we call the reduced transportation method (RTP). Within the context of our capacity problems, we explore each of these methods with a view toward determining the more efficient approach for any specific application.

The TP Method

If we use the standard approach to convert the transshipment model to a transportation problem, we have, for our example with $n=4$, Table I as representing the possible activities and Table II as the associated costs.

TABLE I
ACTIVITIES AND RESTRICTIONS

						Row sum
	Y_1	X_{12}	X_{13}	X_{14}	X_{15}	D_1+G
	S_1	Y_2	X_{23}	X_{24}	X_{25}	D_2-D_1+G
		S_2	Y_3	X_{34}	X_{35}	D_3-D_2+G
			S_3	Y_4	X_{45}	D_4-D_3+G
				S_4	Y_5	G
Column sum	G	G	G	G	D_4+G	

TABLE II
COST COEFFICIENTS

0	c_{12}	c_{13}	c_{14}	c_{15}
0	0	c_{23}	c_{24}	c_{25}
	0	0	c_{34}	c_{35}
		0	0	c_{45}
			0	0

The blanks in Tables I and II refer to inadmissible activities. The Y_i are dummy variables and G may be defined as

$$G=0.5[D_1+\textstyle\sum_{i=2}^{i=4} |D_i-D_{i-1}|+D_4].$$

In general, the above transportation problem has $n+1$ row and $n+1$ column restrictions; using Proposition 1 below, we demonstrate that the dimensions may be reduced to $n\times n$.

We turn to a description of the properties of an optimal program. For a specific optimal program, let $Z_k{}^*$ be the amount of capacity available in period k, i.e.,

$$Z_k{}^*=\textstyle\sum_{r\leqq k} \sum_{t>k} X_{rt}^*. \qquad (k=1, 2, \cdots, n)$$

PROPOSITION 1. *There exists an optimal extreme point program*† *such that*‡

(*a*) *if* $Z_k{}^*>D_k$, *then* $X_{j,k+1}^*=0$ *for all* $j<k+1$ *and* $X_{kt}^*=0$ *for all* $t>k$;

(*b*) *if* $D_k=0$, *then* $X_{j,k+1}^*=0$ *for all* $j<k+1$ *and* $X_{kt}^*=0$ *for all* $t>k$;

† We say that X is an extreme point program if it is an extreme point of the convex set of programs satisfying (2). The general theory of linear programming ensures that at least one extreme point program is optimal when a feasible and finite optimal program exists.

‡ This and subsequent propositions are proved in the Appendix.

(c) *if* $Z_k^* > D_k$, *then there are integers* $r<k$ *and* $t>k$ *for which* $D_r = Z_r^* \geqq Z_{r+1}^*$, $Z_j^* = Z_{r+1}^*$, *for* $j=r+2, r+3, \cdots, t-1$, *and* $Z_{t-1}^* \leqq Z_t^* = D_t$;

(d) $Z_k^* \leqq \min(\max_{i \leqq k} D_i, \max_{i \geqq k} D_i) \equiv U_k$ *for every* k; *the function* U_k *is unimodal*;†

(e) *if* $k=1$, $k=n$, $D_k \geqq \max_{i<k} D_i$, or $D_k \geqq \max_{i>k} D_i$, *then* $U_k = D_k$, $S_k^* = 0$, *and* $Z_k^* = D_k$;

(f) *if* D_k *is a unimodal function of* k, *then* $S_k^* = 0$ *and* $Z_k^* = D_k$ *for all* k.

We now discuss the meaning of the proposition. Part (a) implies that we may restrict our attention to policies wherein no capacity is bought at the beginning or sold (released) at the end of a period k in which the total available capacity exceeds the total required. Part (b) states that if $D_k = 0$ for some k and if all period k activities are eliminated from the problem, an optimal solution for the revised $(n-1)$-period problem is also optimal for the original problem. More generally, if say, p of the D_k are zero, the problem may be reduced to an $(n-p)$ period model in which every requirement is positive. The conclusions of Part (a) are refined in Parts (c), (d), and (e), which assert respectively that available capacity coincides with the requirements schedule except over certain intervals where the available capacity remains constant; available capacity is never larger than the smaller of the preceding and subsequent peak requirements; and idle capacity never exists at the first or last period, or when the current requirement is at least as large as either the preceding or the subsequent peak requirement. Part (f) states that if the requirement pattern is unimodal, then we may seek a schedule in which no idle capacity occurs. These properties have been observed by other authors in a related context.[1,9]

One use of Part (e) is to eliminate certain S_k activities from the programming problem. In particular, since $S_1^* = S_n^* = 0$, we note from Table I that $Y_1^* = Y_{n+1}^* = G$; therefore, we may eliminate the first column restraint and the $(n+1)$-st row restraint. The new table of restraints is identical to Table I except for the deletion of the indicated row and column and the replacement of the first row sum, $D_1 + G$, by the quantity D_1 and the $(n+1)$-st column sum, $D_n + G$, by the quantity D_n.

The RTP Method

In order to describe the RTP[12] method for the transshipment problem, we give a network representation to the restrictions (8). The network is

† A function f_k is unimodal if there exists a number t such that f_k is nondecreasing for $k \leqq t$ and is nonincreasing for $k \geqq t$.

comprised of nodes, one for each restriction in (8), and of directed arcs connecting the nodes, one arc for each activity variable in (8). In Fig. 1 we illustrate this particular example. In the network representation, X_{ij} is interpreted as the amount of flow along arc (i, j), i.e., from node i to node j; c_{ij} is the cost per unit of flow along the arc. Similarly S_i is the flow from node $i+1$ to node i; here the cost per unit flow $c_{i+1,i}=0$.

The quantity associated with each node, e.g., D_i-D_{i-1} with node i, represents the amount by which the flow out of the node exceeds the total flow into the node. With this identification, our capacity scheduling problem is equivalent to the problem of finding flows in a network satisfying the nodal requirements at a minimum cost. From the point of view of

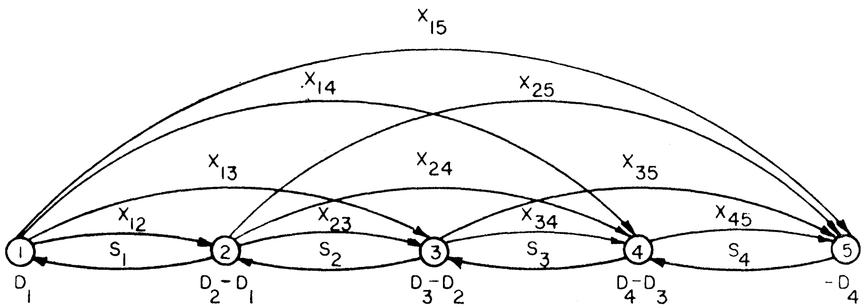

Figure 1

constructing a feasible set of flows, we observe that any positive node i (say), i.e., where $D_i>D_{i-1}$, may possibly supply the deficit at any negative node j (say), i.e., where $D_j<D_{j-1}$. Clearly, for any optimal program, the flow between a positive node i and a negative node j would travel over a minimal cost path, i.e., along an admissible set of arcs of the form (i,t_1), (t_1,t_2), (t_2,t_3), $\cdots$, (t_k,j) such that the sum of the associated unit costs is a minimum; we denote the minimal cost by f_{ij}. The RTP method is to solve a $w\times s$ ordinary transportation problem in which the 'surpluses' at the *w positive nodes* are used to satisfy 'shortages' at the *s negative nodes* and where f_{ij} is the unit cost of flow between positive node i and negative node j. The RTP approach necessitates solving first for the minimal cost paths to obtain the f_{ij}, and then for the optimal transportation schedule. Having such an optimal solution, we apply the obvious correspondences with the original X_{kt} variables to produce the desired capacity schedule.

The f_{ij} required for the RTP method may be determined in the following way. First, one readily concludes from Fig. 1 that $f_{ij}=0$ for $i>j$. Upon putting $f_{jj}=0$ for notational convenience, the following result, proved in reference 15† (c.f., reference 6 and 13), may be used to compute the f_{ij} for $i<j$ recursively.

PROPOSITION 2

(*a*) $f_{ij} \geqq f_{kt}$ *for every* $i \leqq k < t \leqq j$;

(*b*) *for every* $i<j$, $f_{ij}=\min_{i<k\leqq j}(c_{ik}+f_{kj})$;

(*c*) *for every* $i<j$, $f_{ij}=\min_{i\leqq k<j}(f_{ik}+c_{kj})$.

In order to use the 'backward equations' given in Part (*b*), we first observe that $f_{j-1,j}=c_{j-1,j}$. Then $f_{j-2,j}, f_{j-3,j}, \cdots, f_{ij}$ may be calculated

TABLE III

RELATIVE EFFICIENCY OF TP AND RTP METHODS[a]

	Total additions and subtractions	Total comparisons
TP-method	$2Kn^3+O(n^2)$	$Kn^3+O(n^2)$
RTP-method	$\left(7/48+\frac{K}{2}\right)n^3+O(n^2)$	$\left(7/48+\frac{K}{4}\right)n^3+O(n^2)$

[a] The symbol $O(x)$ stands for any function for which $\lim_{x\to\infty} O(x)/x$ is bounded.

The entries corresponding to the RTP method are upper bounds and are based on the assumption that the backward equations are used to compute all the f_{ij} when $s<w$ and that the forward equations are used otherwise (*see* footnote at the end of the preceding sub-section). When w or s is small in relation to $0.5n$, the advantage of the RTP method is even more pronounced.

recursively in the order given, since each term in the sequence depends only upon the terms preceding it and upon certain of the known cost coefficients. In the usual manner of dynamic programming,[3] the minimal cost computations provide the associated optimal path. Alternatively the 'forward equations' in Part (c) may be employed in an entirely analogous way.‡

† The proof consists essentially in exploiting (3) to show that there is an optimal path from i to j which does not involve 'back-tracking,' i.e., a path in which no arcs of the form $(k, k-1)$ appear.

‡ It is not necessary to determine every f_{ij} for the RTP method, but only those corresponding to a possible flow from a positive node to a negative node. There is a simple procedure that takes advantage of this fact. If $s<w$, use the backward equations to obtain $\{f_{i^*j}\}$ where i^* is the first (lowest numbered) positive node and j ranges over the negative nodes; otherwise use the forward equations in a similar manner. It is shown in reference 15 that this rule for choosing between exclusive use of the backward and forward equations respectively always selects the procedure requiring the smallest number of calculations when $\max(w, s) \geqq 0.6n+2$.

It is also shown in reference 15 that the computation of the f_{ij} may be simplified in a different way when the c_{ij} can be written in the form $c_{ij}=a^i m(j-i)+\sum_{k=i}^{j-1} h_k$ for all $i<j$. In particular $f_{ij}=a^i f(j-i)+\sum_{k=i}^{j-1} h_k$ for $i<j$ where $f(0)=0$ and $f(k) =\min_{0\leqq t<k}\{a^t m(k-t)+f(t)\}$ for $k=1, 2, \cdots, n$.

Comparison of the TP and RTP Methods

Our comparison of the TP and RTP approaches to capacity scheduling models is based on the assumption that the transportation simplex method is employed. We also postulate that the ratio of the expected number of iterations to the number of linear restrictions is a constant K.† Under these assumptions, it is shown in reference 15 that the amount of computation required for the TP and RTP methods is as shown in Table III.

Observe that if $K > 7/72$ then the RTP method is more efficient in terms of additions and subtractions for sufficiently large n. Similarly, the RTP method requires fewer comparisons if $K > 7/36$. Available evidence reference 8, pp. 119–20, points to the conclusion that K is close to 1, and therefore the RTP method appears superior to the TP method for large n.

MONOTONE AND PULSATING REQUIREMENTS SCHEDULES

WHEN THE D_k form a nondecreasing sequence for $k=1, 2, \cdots, n$, the RTP method yields a simple solution. Since $s=1$ and $j_s=n+1$, we obtain the minimal cost paths from each positive node to the $(n+1)$-st node by means of the backward equations. The RTP approach yields a w by 1 transportation problem with the evident solution: schedule the capacity for requirement level D_1 according to the minimal cost path associated with $f_{1,n+1}$; schedule the capacity for the increment D_2-D_1 according to the minimal cost path associated with $f_{2,n+1}$; etc. It is clear that in each period k, $X^*_{kt}=0$ for all $t\neq j$ where (k,j) is the first arc in the minimal cost path from node k to node $n+1$. In addition, $X^*_{kj}=D_k-\sum_{i<k,\,t>k} X^*_{it}$. Thus, in order to schedule optimally in period k, it is not necessary to know the exact values of $D_{k+1}, D_{k+2}, \cdots, D_n$, but only that this sequence is nondecreasing and that $D_{k+1}\geqq D_k$.

When the D_k form a nonincreasing sequence, we employ the forward equations to obtain the minimal cost paths from the first node to each negative node. The RTP approach yields a $1\times s$ transportation problem with the solution: Schedule the capacity requirement D_n according to the minimal cost path associated with $f_{1,n+1}$; schedule the capacity for the increment $D_{n-1}-D_n$ according to the minimal cost path associated with $f_{1,n}$; etc. Here it is necessary to know, in general, the exact values of $D_{k+1}, D_{k+2}, \cdots, D_n$ in order to determine correctly X^*_{kj}, $k<j$.

In the special case where $D_k=d$ for all k, we may employ the approach for either procedure above. It follows that only the activities associated with $f_{1,n+1}$ are employed, each at the level d, and that there is exactly d units of capacity available in each period.

† There is some evidence supporting this hypothesis in reference 8, pp. 119–20. Although our analysis was developed under the assumption that K is the same for both the *TP* and *RTP* methods, an alternative postulate that K differs for the two approaches would necessitate only minor changes.

As we shall observe in Part II, an important type of requirement scheduling problem stems from a pulsating pattern of minimal capacities, i.e., a pattern in which each element of $\{D_k\}$ equals either d or 0. In order to characterize certain properties of optimal extreme point programs for this problem, it is convenient to introduce some definitions. We say that a program X has *overlapping* components if for some integers $i<j<k<t$, both $X_{ik}>0$ and $X_{jt}>0$. A program X is said to have *nested* components if for some integers $i\leqq j<k\leqq t(t-i>k-j)$, both $X_{it}>0$ and $X_{jk}>0$.

By exploiting the fact that every extreme point of the set of feasible programs for a transportation problem involves only integral variables, it follows readily that every extreme point program X for the case of pulsating requirements has the following properties: (a) each X_{ij} equals zero or d; (b) X contains no nested components; (c) at most two activities may be employed in any period; and (d) if $X_{ij}=d$, then $D_k=d$ for at least some $i\leqq k<j$.

When (3) holds, it follows from Part (d) of Proposition 1 that there exists an optimum extreme point program for which $Z_k^*\leqq d$ for all k.† In addition, since this extreme point must satisfy property (a) in the preceding paragraph, it can have no overlapping or nested components. Hence, new capacity is acquired only after old capacity terminates.

An optimum extreme point for the case of pulsating requirements can be found by applying Part (b) of Proposition 1, i.e., by dropping periods with zero requirements, to form a new problem with identical requirements in every period. The new problem is readily solved by the method indicated in this section.

EXTENSIONS AND APPLICATIONS

Equipment Replacement and Overhaul

A NUMBER of important models in operations research have been recognized as yielding to both linear programming and dynamic programming methods. So far as we know, recognizing that equipment replacement models[4,5,10,11] are equivalent to a transportation model (actually, to an assignment model) appears a new observation. The replacement models are equivalent [when (3) holds] to the special case of our model in which $D_k=1$ for all k.

The assignment model representation is given by Table I (making the deletion of the first column and the $n+1$-st row, as indicated at the end of the sub-section "The TP Method") with each row and column sum entry being 1. If instead a dynamic programming approach is to be used, then

† This property also holds if (3) is relaxed to require merely that $c_{ij}\leqq c_{i,j+1}$ for all i and j or that $c_{ij}\leqq c_{i-1,j}$ for all i and j.

we propose solving the problem by computing $f_{1,n+1}$ employing either of the recursions in Proposition 2. An alternative recursion proposed in reference 4 for solving the problem requires twice the number of additions and the same number of comparisons as our method.

Since the c_{ij} may depend upon the length of the interval $[i,j]$ as well as the calendar dates i and j, our formulation easily encompasses situations in which technological progress, as reflected in economic terms, is present, and situations in which it is possible to purchase a used machine.[5]

Finally, we mention the possibility of including the decision to overhaul the equipment.[5] In particular, we permit (but do not require) a machine to be overhauled at the beginning of any period after the initial procurement and prior to the sale of the machine. We further assume that the length of time required to overhaul a machine is negligible in comparison with the length of a period. Thus, a machine that is overhauled at the beginning of a period is still available for use during the period. We suppose also that if a machine is bought at the beginning of period i, overhauled at the beginning of periods $t_1<t_2< \cdots <t_p(i<t_1)$, and sold at the beginning of period $j(j>t_p)$, the total net cost incurred can be written in the form:

$$o_i(i,t_1)+o_i(t_1,t_2)+ \cdots +o_i(t_{p-1},t_p)+b_{ij}(t_p). \tag{9}$$

We interpret $o_i(k,v)$ $(i\leqq k<v)$ as the cost of operating a machine during the interval $[k,v]$ and then overhauling it at the beginning of period v, given that the machine was initially purchased in period i and was last overhauled in period k (or bought, if $k=i$). The term $b_{ij}(t_p)$ represents the net total cost resulting from acquiring the machine in period i, from relinquishing the machine in period j, and from operating the machine during the interval $[t_p,j]$ between final overhaul and sale. The cost of providing a machine during the interval $[i,j]$ without overhauling is denoted by $b_{ij}(i)$.

Denote by O_{ij} the total cost of operating and overhauling a machine purchased in period i and overhauled in period $j(j>i)$ when an optimal overhaul policy is followed within the interval $[i, j]$. In view of (9), the O_{ij} satisfy the forward equations:†

$$O_{ij}=\min_{i\leqq t<j}\{O_{it}+o_i(t,j)\} \quad \text{for all } 1\leqq i<j\leqq n \tag{10}$$

where we put $O_{ii}=0$ for all i.

The first step in our approach is to employ (10) to calculate the O_{ij}; specifically, for each i we compute $O_{i,i+1}, O_{i,i+2}, \cdots, O_{in}$. Next we use these results to calculate the cost c_{ij} resulting from buying a machine in period i, selling it in period j, and overhauling it in an optimal manner

† It is inefficient to use backward equations here.

during the interval $[i, j]$. In view of (9), the c_{ij} may be obtained from

$$c_{ij}=\min_{i\leqq t<j}\{O_{it}+b_{ij}(t)\} \quad \text{for all } 1\leqq i<j\leqq n+1. \tag{11}$$

The final step is to employ Proposition 2 to calculate the optimal replacement policy. This method of solving the replacement-overhaul problem requires about two-thirds of the number of additions and the same number of comparisons as the proposal in reference 5.

The technique given above can be extended to the case where several types of machines and/or used machines can be acquired for certain intervals. For this case we find the optimal overhaul policies for each activity using the above procedure. We then determine the cheapest activity available for each interval and solve the resulting replacement problem.

Automatic Checkout, Repair, and Replacement

In the models of the preceding section, equipment replacement is justified solely in terms of economic factors, and the possibility of machine breakdown is taken into account implicitly, if at all, in the computation of the c_{ij}. The models in this section differ mainly in their focusing attention on the failure problem. Our purpose here is to indicate the basic idea permitting the use of the optimizing model in this paper for the solution of two stochastic failure problems. The model is also applicable to a number of generalizations of these problems; however, we postpone further extensions to another publication.

Problem 1: It is desired that a system be available for possible emergency service over n periods. However, a certain critical part in the system is subject to chance failure, which results in the system being unavailable for service with attendant penalty costs. Since it is impractical (for economic or technological reasons) to determine the serviceability of the part by means of checkouts, it is necessary to replace the part when it has been in service for a suitable interval in order to avoid extended system unavailability. We therefore seek a replacement policy that minimizes the combined (expected) costs of part replacements and system unavailability. The problem is formally equivalent to the replacement problem of the preceding section under the following circumstances: replacements are permissible only at the beginning of each period; a part placed in service for the interval $[i,j]$ has a known time to failure distribution $F_{ij}(t)$ with $F_{ij}(0-0)=0$ and $F_{ij}(j-i)=1$;† and the total cost $c(i,j,t)$ incurred during $[i,j]$ when the part placed in service for that interval fails at age $t(0\leqq t\leqq j-i)$ depends only on i,j, and t. The c_{ij}, which are defined by

† It is convenient for expository purposes to view removal of the part as a failure.

$$c_{ij} = \int_0^{j-i} c(i, j, t)\, dF_{ij}(t) \quad \text{for} \quad i<j,$$

here represent expected costs.†

Problem 2: A system, which is operational at the beginning of period one (i.e., at time zero), has a known time-to-failure distribution $F(t)$ with $F(0)=0$. If the system remains operational until the end of period n, it is removed from service at that point. Hence we assume for convenience that $F(n)=1$. It is possible to check out the system at the beginning of each period, employing a test that determines with certainty whether the system is or is not operational. If the system is found to have failed, no further checkouts are performed. The optimizing problem is to find a checkout schedule determining the specific periods in which the system should be tested in order to balance properly the costs of making frequent checks and of having the system in an undetected nonoperational state.

This problem is also mathematically equivalent to the replacement problem of the preceding section provided that the total cost (checkout and downtime) $c(i,j,t)$ incurred in the interval $[i,j]$ when successive checkouts are performed at the beginnings of periods i and j and when the system fails at time $t(\geqq i-1)$ depends only upon i,j, and t.‡ The c_{ij} are obtained from

$$c_{ij} = \int_{i-1}^{n} c(i, j, t)\, dF(t).$$

A continuous time version of this problem was formulated and studied extensively in reference 2.

Economic Lot Size, Product Assortment, and Deterministic Batch Queuing Models

The problem in the dynamic economic lot size model,[16,17] is to find a minimal cost inventory policy under the following circumstances: deterministic requirements R_t are to be filled in each period t; the cost of producing or ordering P_t items, $g_t(P_t)$, is a concave nondecreasing function, thereby permitting situations having set-up costs and quantity discounts that may possibly depend on the period t; inventory may be stored from one period to the next at a cost of h_t per unit on hand at the end of period t; and there is no inventory on hand initially.§

† We assume that $c(i,j,t)$ is nonnegative, bounded, and continuous in t. The latter assumption, which is imposed to ensure existence of the integral, can be relaxed.

‡ Here again we assume the $c(i,j,t)$ are nonnegative, bounded, and continuous in t.

§ If there is inventory on hand initially, we may assume without loss of optimality that it is used to satisfy the requirements consecutively in periods 1, 2, $\cdots$ for as long as it lasts. We then consider the revised problem in which the satisfied requirements are eliminated and in which there is no initial inventory.

In the assortment problem,[14,16] decisions must be made concerning the production of items possessing a property of "one-way technological feasibility." For example, consider selecting the assortment of structural steel beams, where a stronger beam can be technologically employed to meet the demand for a weaker beam, with an attendant economic loss. It was demonstrated in reference 16 that the assortment problem is mathematically equivalent to the lot-size model.

A third interpretation is a queuing system in which R_t persons arrive in period t and can be served in that period or subsequent periods. In deciding on an optimal batch size to serve in period t, an economic balance must be made between having persons wait and providing service capacity at a cost described by a concave function of capacity.

It was demonstrated in references 16 and 17 that an optimal policy exists for the above problems wherein for some integers $1=u_1<u_2<\cdots<u_p=n+1$, $P_k=\sum_{j=k}^{j=t} R_j$ for $k=u_i$ and $t=u_{i+1}-1$ and $i=1, 2, \cdots, p-1$, and $P_k=0$ otherwise. This result permits us to formulate the indicated problems in terms of the capacity scheduling problem (1), (2), by letting c_{ij} be the total cost associated with $P_i=\sum_{k=i}^{j-1} R_k$, i.e.,

$$c_{ij}=g_i(\textstyle\sum_{k=i}^{j-1} R_k)+\sum_{t=i}^{j-2} h_t[\sum_{k=t+1}^{j-1} R_k] \quad \text{for} \quad 1\leqq i<j\leqq n+1;$$

by letting $D_t=1$ for $1\leqq t\leqq n$; and by requiring that $S_t=0$ for $1\leqq t\leqq n$. Adopting this procedure, we have the result, just as for the equipment replacement models above, that the dynamic economic lot-size problem is equivalent to an assignment problem. The dynamic programming computational approach of Proposition 2 is the same as the technique in references 16 and 17.

Labor Force Planning

A large number of papers have appeared on the scheduling of labor force size, given a time sequence of minimal requirements D_k for manpower, e.g., *see* references 1 and 9. The formulations differ considerably,† but many of them attempt to isolate the costs resulting from hiring, firing, and employing. Our model is applicable to such situations where it is meaningful to let X_{ij} be the number of men hired in period i and fired at the beginning of j, where c_{ij} is the cost per man employed in this interval. Thus we are able to introduce explicitly the cost factors associated both with the length and with the calendar period of employment.

APPENDIX: PROOF OF PROPOSITION 1

We first prove that Part (a) holds for every optimal program under the assumption that the cost coefficients are strictly positive and that (3) is always satisfied

† Here again it has been long recognized that both linear and dynamic programming may be applied to some of these models.

with strict inequality. We recall that the primal problem is to find X_{ij}^* that

$$\text{minimize } \sum_{i,j} c_{ij} X_{ij} \tag{12}$$

subject to

$$\sum_{i\leqq k, j>k} X_{ij} \geqq D_k; \; (k=1, 2, \cdots, n; \qquad X_{ij} \geqq 0) \tag{13}$$

the dual problem is to find $u_k{}^*$ that

$$\text{maximize } \sum_{k=1}^{k=n} D_k u_k, \tag{14}$$

subject to

$$\sum_{k=i}^{j-1} u_k \leqq c_{ij}. \; (1 \leqq i < j \leqq n+1, \qquad u_k \geqq 0) \tag{15}$$

It is well known that every pair of optimal solutions to the primal and dual problems satisfy the relations

$$\text{if } \sum_{i\leqq k, j>k} X_{ij}^* > D_k, \quad \text{then } u_k{}^* = 0, \text{ and} \tag{16}$$

$$\text{if } \sum_{k=i}^{j-1} u_k{}^* < c_{ij}, \quad \text{then } X_{ij}^* = 0. \tag{17}$$

Employing these relations and the assumptions about the c_{ij}, we conclude that if $Z_k{}^* > D_k$, then

$$u_k{}^* = 0 < c_{k,k+1}, \text{ so that } X_{k,k+1}^* = 0;$$

$$\sum_{t=k}^{j-1} u_t{}^* = \sum_{t=k+1}^{j-1} u_t{}^* \leqq c_{k+1,j} < c_{kj}, \text{ so that } X_{kj}^* = 0 \text{ for } j > k+1;$$

$$\sum_{t=i}^{t=k} u_t{}^* = \sum_{t=i}^{k-1} u_t{}^* \leqq c_{ik} < c_{i,k+1}, \text{ so that } X_{i,k+1}^* = 0 \text{ for } i < k,$$

which proves Part (a) where the c_{ij} are all positive and satisfy (3) with strict inequality.†

The remaining assertions follow immediately from Part (a).

REFERENCES

1. Arrow, K. J. and S. Karlin, "Production Planning without Storage," in *Studies in the Mathematical Theory of Inventory and Production* (K. J. Arrow, S. Karlin, and H. Scarf, eds.), Stanford University Press, Stanford, 1958.
2. Barlow, R. E., L. C. Hunter, and F. Proschan, "Optimum Monitoring Policies," Mathematical Note No. 240, Boeing Scientific Research Labs., Seattle, April 1961.
3. Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, 1958.
4. ———, "Equipment Replacement Policy," *J. SIAM* **3**, 133–136 (1955).
5. Dreyfus, S., "A Generalized Equipment Replacement Study," *J. SIAM* **8**, 425–435 (1960).
6. Eastman, W. L., "A Note on the Multi-Commodity Warehouse Problem," *Management Sci.* **5**, 327–331 (1959).
7. Fetter, R. B., "A Linear Programming Model for Long-Range Capacity Planning," *Management Sci.* **7**, 372–378 (1961).
8. Gass, S. I., *Linear Programming*, McGraw-Hill, New York, 1958.
9. Karush, W., "On a Class of Minimum-Cost Problems," *Management Sci.* **2**, 136–153 (1958).

† The proof is extended to the case where some of the c_{ij} are zero and/or where (3) does not hold with strict inequality in reference 15.

10. KLEIN, M. AND L. ROSENBERG, "Deterioration of Inventory and Equipment," *Naval Res. Log. Quart.* **7**, 49–62 (1960).
11. LUTZ, F. AND V. LUTZ, *The Theory of Investment of the Firm*, Princeton University Press, Princeton, 1951.
12. ORDEN, A., "The Transhipment Problem," *Management Sci.* **2**, 276–285 (1956).
13. POLLACK, M. AND W. WIEBENSON, "Solutions of the Shortest-Route Problem—A Review," *Opns. Res.* **8**, 224–230 (1960).
14. SADOWSKI, W., "A Few Remarks on the Assortment Problem," *Management Sci.* **6**, 13–24 (1959).
15. VEINOTT, JR., A. F., AND H. M. WAGNER, "Optimal Capacity Scheduling," RM-3021, The Rand Corporation, Santa Monica, February 1962.
16. WAGNER, H. M., AND T. M. WHITIN, "Dynamic Version of the Economic Lot Size Model," *Management Sci.* **5**, 89–96 (1958).
17. ———, "A Postscript to 'Dynamic Problems in the Theory of the Firm,'" *Naval Res. Log. Quart.* **7**, 7–12 (1960).