COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 529

Note: Cowles Foundation Discussion Papers are preliminary
materials circulated to stimulate discussion and
critical comment. Requests for single copies of a
Paper will be filled by the Cowles Foundation within
the limits of the supply. References in publications
to Discussion Papers (other than mere acknowledgment
by a writer that he has access to such unpublished
material) should be cleared with the author to protect
the tentative character of these papers.

THE DESIGN OF DISTRIBUTED SERVICE SYSTEMS

David R. Strip and Richard Engelbrecht-Wiggans

July, 1979

THE DESIGN OF DISTRIBUTED SERVICE SYSTEMS[*][+]

by

David R. Strip[1]

Richard Engelbrecht-Wiggans[2]

Abstract

A variety of practical situations involve supplying a particular commodity by some locations to satisfy the demand at others. If the demands and the costs of producing varying amounts of commodity at each location are known, then the question is how much commodity should be supplied by each location in order to minimize the total system cost. Under some relatively general conditions, there will be an optimal solution with the property that the vector of amounts supplied by the various locations is one of a distinguished set of points. In the case of star networks, this combinatorial nature may be exploited to give a very efficient algorithm for finding an optimal solution. A numerical example illustrates the results.

Introduction

A variety of practical problems involve supplying a particular commodity by some locations to satisfy the demand at others. If the cost of supplying various amounts at each location, the cost of transporting the commodity from one location to another, and the total demand at each location are known, then one may ask how much of the commodity should each location supply to each other location so as to minimize the total cost. In general, it will not be optimal for each location to supply enough commodity to satisfy its own demand; likewise, it will, in general, not be optimal for a single location to be the sole supplier.

One example of such a supply-demand problem is that associated with computer networks. Each location has some demand for computer services. Each location could, at some cost, install an in-house computing facility sufficiently large to handle all local demand for services. Alternatively, the demand at one location could be satisfied by one, or possibly several, larger central processing centers; such an option becomes desirable if the pooling of demands results in a larger center sufficiently efficient to offset the cost of communications and other overhead. This example motivates the use of "distributed service systems" to describe the above supply-demand problem.

There are many other examples of distributed service system problems. These include designing electrical power networks, trash recycling programs and regional medical care systems, and locating manufacturing plants or regional centers for bulk mailings. In each of these examples, the designer has the option of having each demand for a commodity or service satisfied locally. Alternatively, all or part of the demand may be satisfied at some other location; this alternative may involve transportation costs and other overhead but may still result in lower overall costs due to, for example, economies of scale.

Distributed Service Model

There is a set $N = \{1, 2, \ldots, n\}$ of locations which demand or can supply a commodity. It is assumed that each location has a non-negative demand and that, at some price, each location can supply any amount of the commodity. The amount of commodity demanded at location j is assumed to be a fixed known amount d_j. If $x_{i,j}$ ($i \neq j$) denotes the amount of the commodity demanded by location j which is supplied by location i, then let $S(\underline{x})$ be the total cost of solution $\underline{x}$. The amount $d_j - \Sigma_{i \neq j} x_{i,j}$ must be supplied locally; this appears implicitly in the function S. (For concreteness, the reader may refer to the subsequent numerical example.)

The total cost of the solution $\underline{x}$ includes the cost of supplying the commodity, the cost of transporting the commodity, any overhead incurred because of remote processing, and any costs associated with time delays resulting from transportation and remote processing. In general, the cost of shipping $x_{i,j}$ units from i to j may depend on how much location i supplies to each other location, how many other locations are supplying j, and how much of the commodity is being shipped on each possible route. Indeed, there may be several different ways (e.g., alternative shipping routes) for a particular solution $\underline{x}$ to arise; if there is more than one way to realize a particular solution, then it is assumed that $S(\underline{x})$ corresponds to the least expensive implementation of the solution $\underline{x}$. (If there are infinitely many such implementations, then we assume the problem is sufficiently regular so that a minimum cost implementation exists.) The optimization problem may be summarized as follows.

Distributed Service System Problem:

Minimize $S(\underline{x})$ subject to: $x_{i,j} \geq 0$ ($\forall i \neq j$) .

(Implicit in the above problem, and in subsequent problems, is the condition that $\Sigma_{i \neq j} x_{i,j} \leq d_j$ $\forall j$. This condition is not stated explicitly since it will always be satisfied if there is a positive cost associated with supplying excess units; alternatively, an appropriate choice of the function S will assure that the above condition is satisfied even if negative costs are allowed.) While in come special cases the above problem may be solved using network flow algorithms [3], this will, in general, not be possible. Since the costs of transportation need not be linear in each $x_{i,j}$ or additive across $x_{i,j}$ and since, in general, the cost of supplying a commodity at a particular location is non-linear due to the start-up cost of supplying the first unit, the cost function will, in general, be sufficiently non-linear to make optimization difficult.

It will be shown that if $S(\underline{x})$ is lower semi-continuous and piecewise concave, then the above problem may be transformed into a combinatorial problem. Since almost any function can be approximated arbitrarily well by a lower semi-continuous piecewise linear function, this result appears quite general. Although the combinatorial problem is typically quite difficult to solve, a class of star network problems exists for which an efficient algorithm is developed. Since distributed service systems in which each location has only the options of satisfying its demand locally or at a single central processing facility (e.g., the components of trash may be separated at the location generating the trash or at a regional recycling facility, but typically not anywhere else) correspond to star networks, the efficient algorithm developed below is of potential importance to practical problems.

Relation to Previous Work

Existing technical literature related to the design of distributed service systems can be divided into two major categories; one dealing with the computer science aspects and the other dealing with the operations research aspects. The former field includes, for the purpose of this paper, such topics as network operating systems, access methods, systems protocols, packet switching techniques, and the like. While these topics are important considerations in network design, they are unrelated to the directions of our paper and will therefore not be dealt with here.

The operations research area has several major concerns. The first is the data base distribution problem: finding an allocation of data bases among users in a network to optimize the costs of maintaining the data base [1, 2, 8, 9, 14] . Generally, these papers consider the problem of optimally allocating copies of files among nodes in a network for which the topology, job assignment to computers, and frequencies are all known in advance. The models range from deterministic linear models to more complicated stochastic models; even at these extremes, they still have the common assumption of known network structure.

The second area of widespread attention in the operations research category is network design. Typically, network design has been interpreted as specifying network topology for a system in which message traffic patterns (source-node to destination-node) are known [4, 5, 6, 12] . Most of these papers deal with a waiting time objective and a cost constraint. They usually make the exponential and Poisson assumptions on the known message traffic patterns adopted by Kleinrock [7] in developing analytic expressions for calculating waiting time measures in message routing networks. Due to the difficulty of the problem, most authors restricted their design to tree-structured networks,

and adopted a variety of "greedy"-type heuristics to find locally optimal solutions. No analytic evaluation of the heuristics is given in any of the papers, and few give any comparison to calculated optima. Pye and Arozullah [11] and McGregor and Boorstyne [10] attempt to deal with the question of load distribution in the network. One model [11] is fairly elementary and provides no insight which may be used here. The other model [10] is basically a queueing model with assumptions adopted to satisfy Kleinrock's conditions. Despite this shortcoming, the model is fairly reasonable and an optimal gradient following algorithm is presented which calculates the proportions of the tasks performed by each server (computer).

A deficiency of all the existing models is the assumption that the capacity of the computers in the network is known and predetermined. In the network design papers this assumption is necessary to be able to determine the node-node message intensities. In addition, these papers assume that the assignment of tasks to processors is known. The last pair of papers attempts to determine the assignment of tasks to processors, but assumes not only the computer location, necessary to determine the processing rate for each node, but further assumes that the network topology is known and the routing pattern between each pair of nodes is known. Thus, while several important aspects of the problem have been dealt with, the interaction of task assignment and network topology has been ignored. More significantly, the capacity assignment aspects of the problem have not been considered at all. Since capacity assignment is the design variable which probably has greatest impact on cost, neglecting this variable, as well as its interaction with the two factors already considered, creates a serious deficiency in the set of tools available to a designer who is attempting to develop a system where none currently exists. Our distributed service model rectifies this situation.

Lower Semi-Continuous Piecewise Concave Model

The search for an optimal solution is simplified when $S(\underline{x})$ satisfies some relatively general conditions. For each pair of i and j $(i \neq j)$, define a set $A_{i,j}$ of points $a_{k_{i,j},i,j}$ such that $0 = a_{0,i,j} < a_{1,i,j} < \ldots < a_{m_{i,j},i,j} = d_j$. Assume that

1. $S(\underline{x})$ is lower semi-continuous on the set $\Pi_{j=1}^{j=n}[0, d_j]^{n-1}$;

2. $S(\underline{x})$ is concave with respect to the boundary of each rectangular

   set $\Pi_{\substack{j=1,i=1 \\ i \neq j}}^{j=n,i=n}[a_{(k_{i,j}-1),i,j}, a_{k_{i,j},i,j}]$, $k_{i,j} = 1,2,\ldots,m_{i,j}$, $i \neq j$;

   specifically, $S(\underline{x})$ satisfies the following condition, for each i

   and j, and for all $\underline{x}$ in each of the corresponding rectangular sets:

$$S(\underline{x}) \geq \frac{x_{i,j} - a_{(k_{i,j}-1),i,j}}{a_{k_{i,j},i,j} - a_{(k_{i,j}-1),i,j}} S(\underline{x}^{i,j-}) + \frac{a_{k_{i,j},i,j} - x_{i,j}}{a_{k_{i,j},i,j} - a_{(k_{i,j}-1),i,j}} S(\underline{x}^{i,j+})$$

where $\underline{x}^{i,j-}$ and $\underline{x}^{i,j+}$ denote the vectors obtained by changing the $(i,j)^{th}$ component of $\underline{x}$ to $a_{(k_{i,j}-1),i,j}$ and $a_{k_{i,j},i,j}$, respectively.

(The second condition is similar to piecewise concavity, but more general since it only requires concavity with respect to the boundaries of the sets.) A function which satisfies the above two conditions will be called lower semi-continuous and piecewise concave with respect to the sets $A_{i,j}$. Finally, a point $\underline{x}$ is said to be a corner of the sets $A_{i,j}$ if $x_{i,j} \in A_{i,j}$ $\forall i \neq j$.

Theorem 1: If $S(\underline{x})$ is lower semi-continuous and piecewise concave with respect to the sets $A_{i,j}$, then there is at least one optimal solution $\underline{x}^*$ to the distributed service problem such that $\underline{x}^*$ is a corner of the sets $A_{i,j}$.

Proof: The existence of at least one optimal solution follows from the lower semi-continuity, on a compact set, of the function S. If $\underline{x}'$ is an optimal solution not at a corner of the sets $A_{i,j}$, then there must be values

of i, j and k such that $a_{(k-1),i,j} < x_{i,j} < a_{k,i,j}$. Then, because

of the piecewise concavity of S , either increasing $x_{i,j}$ to $a_{k,i,j}$ or

decreasing $x_{i,j}$ to $a_{(k-1),i,j}$ (but not necessarily both) results in

another optimal solution. Iterate this procedure until the new optimal solution

is at a corner.

Notice that if for some j, $A_{i,j} = \{0, d_j\}$ $\forall i \neq j$, then there is at

least one optimal solution $\underline{x}$ with $x_{i,j} = 0$ for all i (i≠j) except at

most one, and $x_{i,j} = d_j$ for the one excepted i (i≠j) if any exists. In

other words, the demand at each j is satisfied entirely by a single supplier,

possibly j itself. (Note that if the hypothesis holds for several j , it is

not necessary that a single i supplies all the demand for all of the j's.)

The above theorem allows the distributed service system problem to be

transformed into a combinatorial problem. In the next section, we identify a

class of problems such that the combinatorial problem may be easily solved.

Star Networks

A star network consists of n + 1 locations; n of the locations are

demand points while the remaining location is a central supply facility. A

demand point may satisfy any or all of its demand itself. The demand not

satisfied locally must be supplied by the central facility.

Each of the previously listed examples of distributed service systems

may, as a special case, be a star network. For example, in the case of re-

cycling, it may be reasonable that trash is separated into components either

at the location where the trash is generated or at a central location, but no-

where else. Similar cases exist for power networks (a single utility providing any power which a residence cannot generate itself through solar collectors, etc.) , medical services (a single hospital providing services for a region served by several doctors), manufacturing problems (a single manufacturing plant to assemble parts from several feeder plants), and mailing systems (a single regional mailing center from which to mail a magazine).

We will consider problems in which each component of the cost can be associated with some node and the total cost is the sum of the total costs associated with each node. In particular, consider problems of the following form.

Star Network Problem:

$$\text{Minimize} \quad S(\underline{w}) = \Sigma_{j=1}^{j=n} s_j(w_j) + s_0 \left(\Sigma_{j=1}^{j=n} w_j \right) \quad \text{subject to} \quad w_j \geq 0 \, \forall \, j \; .$$

The various quantities have the following interpretations:

w_j = amount supplied by center to demand point j ;

$s_j(w_j)$, $j=1$, 2, ..., n = cost incurred by demand point j when the center supplies it w_j units;

$s_0(w)$ = cost of center supplying w units.

Notice that the w_j may include any overhead (costs of accounting/billing, checking for transmission errors, etc.) associated with remotely satisfied demand; the fact that similar overhead is not incurred for demand satisfied locally, is reflected in an appropriate choice of the functions s_j . The function s_j includes both the cost of locally satisfying any demand not satisfied when the center supplies w_j units and other costs such as the transportation cost resulting from the center supplying w_j units.

As in the previous problem, define a set A_j of points $a_{k,j}$ such that

$0 = a_{0,j} < a_{1,j} < a_{2,j} < \ldots < a_{m_j,j} = d_j$, where d_j is the amount (including

overhead) which the central facility must supply for demand point j if all of

that demand point's demand were satisfied by the center. It is assumed that

each s_j (j = 1, 2, ..., n) is concave with respect to the end points of each

interval $[a_{(k-1,j)}, a_{k,j}]$, that s_0 is concave on the interval $[0, \sum_{j=1}^{j=n} d_j]$,

and that all these functions are lower semi-continuous over the appropriate range.

It follows that $S(\underline{w})$ is lower semi-continuous and piecewise concave with respcet

to the sets A_j ; thus, the search for an optimal solution to the star network

problem may be restricted to corners of the sets A_j .

Minimally Piecewise Concave Functions

There are certain corners which never need be considered in the search

for an optimal solution.

Lemma: If for some j (j=1, 2, ..., n) , there is a $k_1 < k$ and a $k_2 > k$

such that $s_j(a_{k,j}) > \dfrac{a_{k,j} - a_{k_1,j}}{a_{k_2,j} - a_{k_1,j}} s_j(a_{k_1,j}) + \dfrac{a_{k_2,j} - a_{k,j}}{a_{k_2,j} - a_{k_1,j}} s_j(a_{k_2,j})$

then w_j will not be equal to $a_{k,j}$ in any optimal solution.

Proof: Consider any solution $\underline{w}$ with $w_j = a_{k,j}$ such that the hypo-

thesis is satisfied. Then, using the concavity of s_0 , it is easy to verify

that one of the following two changes in w_j results in a strictly better solu-

tion: either $w_j = a_{k_1,j}$ or $w_j = a_{k_2,j}$.

Construct the set A_j^* by deleting all the superfluous points (as defined by the above lemma) from A_j, and renumbering the elements $a_{k,j}^*$ of A_j^* so that

$$0 = a_{0,j}^* < a_{1,j}^* < a_{2,j}^* < \ldots < a_{m_j^*,j}^* = d_j \quad .$$ Note that the points in A_j^* are those of A_j which are on the lower boundary of the convex hull of the points $(a_{k,j}, s_j(a_{k,j}))$, $k=1, 2, \ldots, m_j$; $j=1, 2, \ldots, n$; thus, the set A_j^* is uniquely defined. Finally, define the function $s_j(w_j)^*$ $(j=1, 2, \ldots, n)$ as the piecewise linear continuous function obtained by connecting the points $(a_{k,j}^*, s_j(a_{k,j}^*))$ $k=1, 2, \ldots, m_j^*$. Note that $s_j(w_j) = s_j(w_j)^*$ for all $w_j \in A_j^*$ and that, although s_j^* is a convex function over the entire range, s_j^* is piecewise linear and continuous and, therefore, also lower semi-continuous and piecewise concave. Consider the following transformation of the star network problem.

Minimal Star Network Problem:

Minimize $S(\underline{w})^* = \Sigma_{j=1}^{j=n} s_j(w_j)^* + s_0(\Sigma_{j=1}^{j=n} w_j)$ subject to $w_j \geq 0$ $\forall j$.

In light of the above observations, Theorem 1 states that there is a corner of the sets A_j^* which is an optimal solution to the minimal star problem. This fact, the above lemma, and the fact that $S(\underline{w}) = S(\underline{w})^*$ for all corners of the A_j^* together imply that if a corner $\underline{z}$ of the sets A_j^* is an optimal solution to the minimal star network problem, it is also an optimal solution to the star network problem.

Define the gross unit cost (excluding central facility costs) $Q_{k,j}$ for units in the interval $[a_{(k-1),j}^*, a_{k,j}^*]$ $(k=1, 2, \ldots, m_j^*$; $j=1, 2, \ldots, n)$ as $Q_{k,j} = (s_j(a_{k,j}^*) - s_j(a_{(k-1),j}^*))/(a_{k,j}^* - a_{(k-1),j}^*)$; $Q_{k,j}$ is negative infinity for $k \leq 0$, and positive infinity for $k > m_j$. (Note that since

s_j^* is convex, $Q_{0,j}$, $Q_{1,j}$, $Q_{2,j}$, ... is a non-decreasing sequence.) The following theorem proves that there is an optimal solution to the minimal star network problem in which the central facility supplies precisely those units for which the gross unit cost is not greater than some threshhold Q .

Lemma: There exists an optimal solution $\underline{z}$ to the minimal star network problem of the following form: There are integers k_1, k_2, ..., k_n and threshold Q such that $z_j = a_{k_j,j}^* \ \forall j$; $Q_{k_j,j} \le Q \ \forall j$; and $Q_{(k_j+1),j} \ge Q \ \forall j$.

Proof: Since there is at least one optimal solution at a corner of the sets A_j^* and since $\underline{z}$ of the above form are corners, it need only be shown that any corner not of the above form is not an optimal solution. Consider any corner $\underline{v}$ not of the above form; thus, for some Q, j_1, and j_2 :
$v_{j_1} < a_{k_{j_1},j}^*$ and $Q_{k_{j_1},j} < Q$; and $v_{j_2} > a_{k_{j_2},j}^*$ and $Q_{k_{j_2},j} > Q$. Let

$D = $ minimum $\{(a_{k_{j_1},j}^* - v_{j_1})$, $(v_{j_2} - a_{k_{j_2},j}^*)\}$ and define $\underline{v}'$ as being

equal to $\underline{v}$ except $v_{j_1}' = v_{j_1} + D$ and $v_{j_2}' = v_{j_2} - D$. Notice that

$\Sigma_{j=1}^{j=n} v_j = \Sigma_{j=1}^{j=n} v_j'$, and thus the only difference in the value of the objective

function is $(s_{j_1}(v_{j_1}')^* - s_{j_1}(v_{j_1})^*) + (s_{j_2}(v_{j_2}')^* - s_{j_2}(v_{j_2})^*)$. However,

since $Q_{k_{j_1},j} < Q_{k_{j_2},j}$, the difference in objective functions is negative and

$\underline{v}'$ is a strictly better solution than $\underline{v}$; $\underline{v}$ is not an optimal solution.

Theorem 2: There exists an optimal solution $\underline{z}$ to the minimal star network problem of the following form: There are integers k_1, k_2, ..., k_n and threshold Q such that $z_j = a_{k_j,j}^* \ \forall j$; $Q_{k_j,j} \le Q \ \forall j$; and $Q_{(k_j+1),j} > Q \ \forall j$.

Proof: Since s_0 is concave, it follows that if there is a positive

amount of demand with gross unit cost Q supplied by the central facility in

an optimal solution, then all of the demand with gross unit cost Q is supplied

by the central facility. Thus, either, or both, of the inequalities in the

lemma may be replaced by a strict inequality.

This theorem suggests an efficient algorithm for finding an optimal

solution to a star network problem. First identify the sets A_j^* . Next calculate

all the $Q_{k,j}$. Finally, for each Q equal to $Q_{k,j} < \infty$ calculate $S(\underline{w})$ where

$w_j = \max_k a_{k,j}^*$ s.t. $Q_{k,j} \leq 0$. There will be a total of at most $M = 1 + \sum_{j=1}^{j=n} m_j^*$

such $\underline{w}$'s ; the one (or more) with a minimal $S(\underline{w})$ is an optimal solution to

the minimal star network problem, and therefore an optimal solution to the star

network problem. (If the computations are intelligently organized, the al-

gorithm's complexity is a low order polynomial of M .) Thus, there exists

an efficient algorithm for solving star network problems.

Numerical Example:

Consider a star network of two computer users tied into a central facility

and each user needs to prepare 1000 mailing labels. The first user has two

alternatives. The first, brute force, alternative is to have some, or all,

labels formated and processed at the central facility; the associated demands on

resources for labels printed at the central facility are 100 CPU units/label

for formatting and printing, 250 bytes/label for data transmission, and 20 CPU

units/label of overhead for remote processing. Under the second alternative,

the user compresses the data before transmission and trades some transmission

costs for some local processing costs; the associated demands on resources are

25,000 CPU units to compress all the data, 125 bytes/label for data transmission,

15 CPU units/label for expanding compressed data at the central facility, 100

CPU units/label for formatting and printing, and 10 CPU units/label of overhead for remote processing. Labels may be formatted and printed locally with a total resource demand of 100 CPU units/label. Although remote processing requires greater demands on resources, the various costs of the resources may be such that remote processing is desirable.

The cost $C(x)$ of x CPU units at the first user location is 0 for $x \leq 0$, .008x for $0 \leq x < 50,000$, and .008x $-$ 200 for $50,000 \leq x$. The discontinuity of $C(x)$ at 50,000 may result from a different computer being used if the load is at least 50,000 units. The cost $T(x)$ of transmitting x bytes of data is 0 for $x \leq 0$, 100 for $0 < x \leq 25,000$, and $25 + .003x$ for $25,000 \leq x$. The different unit communication costs may be caused, for example, by the fact that single telephone line available at a flat rate of 100 can handle up to 25,000 bytes.

The cost $H_1(x)$ incurred by the first user under the first option if the central facility provides x CPU units is 600 for $x = 0$, $C(100,000 - x/1.2) + T(2.5x/1.2)$ for $0 < w \leq 120,000$, and 775 for $120,000 \leq x$. Similarly, the cost $H_2(x)$ under the second option is 600 for $x = 0$, $C(125,000 - x/1.25) + T(x)$ for $0 < x \leq 125,000$ and 600 if $125,000 \leq x$. Since the user may be assumed to use the less expensive alternative, $s_1(x)$ is defined as the pointwise minimum of $H_1(x)$ and $H_2(x)$. The functions $H_1(x)$, $H_2(x)$, $s_1(x)$ and $s_1^*(x)$ are plotted in Figure 1, where $A_1 = \{0, 12500, 60000, 93750, 125000\}$ and $A_1^* = \{0, 93750, 125000\}$.

Without developing a similar scenario for the second user, define $A_2^* = \{0, 50000, 75000, 125000\}$ and define $s_2^*(x)$ to be the piecewise linear continuous function through the following points $(x, s_2^*(x))$: $(0,500)$, $(50000,400)$, $(75000,400)$, and $(125000,540)$; the function $s_2^*(x)$ is plotted in Figure 1. Finally, let $s_0(x)$ be any concave lower semi-continuous function

through the following points $(x, s_0(x))$: $(0,0)$, $(50000,90)$, $(143750,170)$, $(168750,185)$, $(218750,210)$, and $(250000,218)$. The total cost $S(\underline{x})$ of the solution $\underline{x} = (x_1, x_2)$ is given by $S(\underline{x}) = s_0(x_1 + x_2) + s_1(x_1) + s_2(x_2)$.

In order to find an optimal solution, we first calculate the $Q_{i,j}$, and obtain that $Q_{0,1} = Q_{0,2} = -\infty$, $Q_{1,2} = -.002$, $Q_{1,1} = -.001$, $Q_{2,2} = 0$, $Q_{3,2} = .0028$, $Q_{2,1} = .003$, and $Q_{4,2} = Q_{3,1} = +\infty$. By considering setting the threshold Q equal to each of the six possible values of $Q_{i,j}$ above, we obtain the solutions and associated costs tabulated in Table 1. Notice that $Q = -.001$ results in an optimal solution with $\underline{x} = (93750,50000)$ and a cost of 1076.25. Finally, it should be noted that although the costs in Table 1 form a unimodal function, this is not, in general, true (for example, if $s_0(0)$ were -1000 rather than 0, then $\underline{x} = (0,0)$ would be the optimal solution with a cost of 100, while all the cost associated with the other solutions would remain as in Table 1); in particular, the cost of the solution must be calculated for each threshold $Q_{i,j} < \infty$.

Conclusion

Under quite general conditions, a distributed service system problem has at least one optimal solution at a corner of the associated cost functions. The property reduces the optimization problem to a combinatorial problem. In the case of a general class starr network, the combinatorial nature results in an efficient algorithm for generating an optimal solution. The theory and algorithm are illustrated through a numerical example.

TABLE 1

Solutions and Costs Associated with Various Thresholds

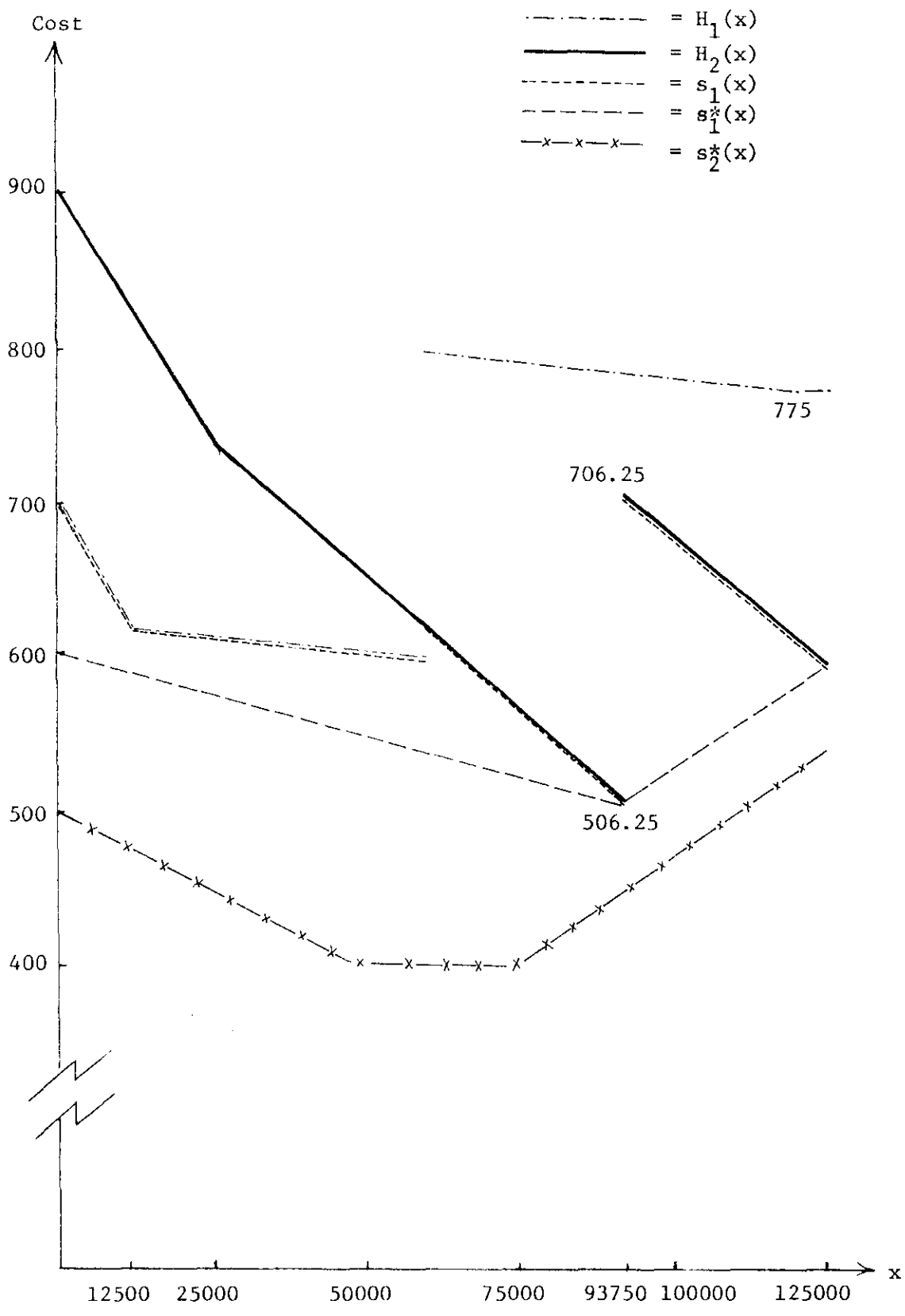| Q	x_1	x_2	$S(x_1, x_2)$	$(= s_0(x_1 + x_2) + s_1(x_1) + s_2(x_2)$
$-\infty$	0	0	1100	$(= 0 + 600 + 500)$
$-.002$	0	50000	1080	$(= 80 + 600 + 400)$
$-.001$	93750	50000	1076.25	$(= 506.25 + 400 + 170)$
0	93750	75000	1091.25	$(= 506.25 + 400 + 185)$
.0028	93750	125000	1256.25	$(= 506.25 + 540 + 210)$
.003	125000	125000	1358	$(= 600 + 540 + 218)$

FIGURE 1. Cost Functions of Numerical Example

REFERENCES

1. Casey, R. G., "Design of Tree Networks for Distributed Data," Proceedings of the Spring Joint Computer Conference, 1972, pp. 251-257.

2. Chu, W. W., "Optimal File Allocation in a Multiple Computer System," I.E.E.E. Transactions on Computers, Vol. C-18, No. 10, (Oct. 1969), p. 835.

3. Ford, L. R., Fulkerson, D. R., Flows in Networks, Princeton University Press, Princeton, N.J., 1962.

4. Frank, H., Frisch, I.T., Chou, W., Van Slyke, R., "Optimal Design of Centralized Computer Networks," Proceedings of the I.E.E.E. 1970 International Conference on Communications, Vol. 1, p. 19-1.

5. Frank, H., Frisch, I. T., Chou, W., "Topological Considerations in the Design of the ARPA Network," Proceedings of the Spring Joint Computer Conference, 1979, pp. 581-587.

6. Karnaugh, M., "A New Class of Algorithms for Multipoint Network Optimization," Proceedings of the I.E.E.E. 1970 Communications Conference, Vol. 3, p. 33-7.

7. Kleinrock, L., "Survey of Analytic Methods in Queuing Networks," Computer Networks, Rusin, R., ed., Prentice Hall, Engelwood Cliffs, N.J., 1972, pp. 185-205.

8. Levin, K.D., "Organizing Data Bases in Computer Networks," Ph.D. Thesis, University of Pennsylvania, Philadelphia, Pa., 1974.

9. Mahmoud, S., Riordan, J. S., "Optimal Allocation of Resources in Distributed Information Networks," ACM Transactions on Data Bases, Vol. 1, No. 1, (1976), pp. 68-78.

10. McGregor, P., Boorstyn, R., "Optimal Load Sharing in a Computer Network," Proceedings of the I.E.E.E. 1975 Communications Conference, Vol. 3, p. 41-14.

11. Pye, W. T., Arozullah, M., "Load Distribution in Enquiry Response Systems with Multiple Processors," Proceedings of the I.E.E.E. 1975 Communications Conference, Vol. 3, p. 33-22.

12. Sharma, R. L., El-Bardai, M. T., "Sub-optimal Communications Network Synthesis," Proceedings of the I.E.E.E. 1970 International Conference on Communications, Vol. 1, p. 19-11.

13. Strip, David R., "Design of Distributed Service Systems," Ph.D. Thesis, Cornell University, Ithaca, N.Y., 1978.

14. Whitney, V.K.M., "A Study of Optimal File Assignment and Communications Network Configuration in Remote Access Computer Message Processing and Communications Systems," Ph.D. thesis, University of Michigan, Ann Arbor, Michigan, 1979.