8004387

CAPSHAW, GARY SCOTT

FACILITY LOCATION WITHIN A DISTRIBUTION SYSTEM

The University of Oklahoma PH.D. 1979

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FACILITY LOCATION WITHIN A DISTRIBUTION SYSTEM

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

GARY SCOTT CAPSHAW

Norman, Oklahoma

1979

FACILITY LOCATION WITHIN A DISTRIBUTION SYSTEM

APPROVED BY

Adel A. Aly, Chairman

Michael D. Devine

Bobbie L. Foote

David C. Kay

Hillel J. Kumin

ACKNOWLEDGEMENTS

Thanks go to my dissertation committee members whose support and signatures I appreciate more than I have ever said to them. The members were Drs. Adel Aly, Chairman, Michael Devine, Bobbie Foote, David Kay, and Hillel Kumin. Special thanks to Adel, whose help insured whatever quality this paper has and to Hillel whose support and friendship through my years in the department was warm and encouraging.

There are many friendships that I have formed during a time that will probably be the most memorable of my life. These people made any doctoral pain worth it: Robin, Billy, Becky, Bob, Abe and Roger.

I can only adequately express my feelings for the patience, support and love given to me by my parents and brother by always trying to be as good and wonderful as they are. I don't know if it is possible to pay back what they have given me.

And Amy.

ABSTRACT

In a typical facility location formulation a new facility is located with respect to a function involving each of the distances between the new facility and each existing facility. But in a distribution system, where the new facility to be located is a distribution point, service vehicles are many times dispatched to service more than one customer (existing facility) in a trip. Therefore the new facility should be located such that the length of these tours is minimized.

This paper examines this problem in many of its variations: discrete, continuous, deterministic, probabilistic, single-facility, multifacility, capacitated and uncapacitated. Some properties were developed and exact and heuristic solution procedures were found for some of the problems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation

The impetus for looking at this problem came from some work that was done for a solid waste collection system of a major city. While studying their solid waste disposal system the problem arose of where to locate disposal sites. During the course of the study the portion of the city under consideration had been divided into over 500 small regions each with some demand of garbage pickup services. So, in the language of facility location literature, the problem was to locate a number of new facilities (disposal sites) with respect to existing facilities (regions).

Assuming that the principal objective was to minimize the total distance traveled by the fleet of trucks, then the problem seems to be a typical facility location problem. Typical facility location formulations assume that each trip is made using the Eucleadian distance between a new facility and one existing facility. In this problem, and in many distribution problems, trips are made from the new facility to cover more than one existing facility. In fact in the garbage problem a truck would leave an established depot, service several regions, go to a disposal site, possibly service more regions, return to a disposal site and then return to the same depot. Therefore the objective would be to locate the disposal sites so that the total distance traveled would be minimized. Supposedly each of these trips would be a minimum-traveling-salesman tour* and would be in the order stated above.

But the problem is further complicated by the fact that each truck has a limited capacity, thus putting a limit on the number of regions that can be serviced in one trip. Also the number and types of disposal sites that are needed would have to be determined. This is complicated by the fact that different types of disposal facilities (landfills, transfer stations) have different disposal costs which vary further according to who owns and administers them.

The problem is quite similar when deliveries are made instead of pickups. In this case one may need to locate a warehouse or distribution center where trucks are dispatched to make deliveries to more than one customer in each trip.

The basic problem of locating a facility with the objective being the minimum-traveling-salesman tour through the existing facilities has been defined by Eilon, et al. (1971) and Burness and White (1976), each of which offered

---

*A minimum-traveling-salesman tour is a trip of minimum length through a number of points, each of which is visited once and only once.

solution methods to specific versions of the problem. Consideration of this problem also appeared in Griffiths (1968) and Webb (1968), who related tour distance to Euclidean distance, and Beattie (1973), who presented a case study of a problem of this type.

Since this type of formulation applies in many real-world distribution problems and since there has apparently been few published works on the subject, there seems to be a need for a formal look at this problem in its many variations. This research will attempt to investigate the problem more thoroughly, make some general observations of its properties, and propose methods of solution and test some of them. The disposal site selection problem mentioned above was not solved but did serve as motivation for this research.

1.2 Organization of This Paper

In this chapter a general introduction to the type of problems has been presented. Chapter II will contain a general literature review and a complete taxonomy of the traveling-salesman facility location problem along with a review of all related topics. Some properties were found and these are presented in Chapter III. The specific problems that were given attention were divided into two groups. Chapter IV discusses deterministic formulations and Chapter V contains work on the probabilistic formulations. Finally, Chapter VI includes a summary, conclusions and recommendations for further research.

CHAPTER II

PROBLEM FORMULATION AND RELATION TO PREVIOUS WORK

An excellent survey of location theory is Francis and White (1974). Much of the notation, formulation and facility location concepts used in this research parallel the ones found there. Other good surveys and bibliographies on location theory are Francis and Goldstein (1974), Lea (1973), Revelle, Marks and Liebman (1970), and Scott (1970).

2.1 The General Location Problem

In a very general sense the modeling of a facility location problem involves making a location decision based on a model similar to:

P2.1
$$\underset{X_j \varepsilon E_2}{\text{minimize}}\ f(X_1,\ldots,X_n) = \sum_{1 \le j < k \le n} v_{jk} d(X_j, X_k) + \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ji} d(X_j, P_i) + F(n)$$

The terms are

m = number of existing facilities,

n = number of new facilities,

$X_j = (x_j, y_j)$, the coordinate location of the j^{th} new facility,

P_i = (a_i, b_i), the coordinate location of the i^{th} existing facility,

d(U,V) = distance between points U and V (measured in any metric),

v_{jk} = cost per unit time per unit distance traveled between new facilities j and k,

w_{ji} = cost per unit time per unit distance traveled between existing facility i and new facility j,

F(n) = cost per unit time of constructing and operating n new facilities (this assumes that each facility has a similar cost structure).

Problem P2.1 can be reformulated to better fit the specific problem being solved: If n is known the term F(n) can be disregarded in the solution procedure since it is a function of n only. And, for the same reason, if n is a variable the problem can be solved for each value of n of interest and only then would F(n) be brought in to compare total costs of the alternatives.

The problem can be complicated by adding constraints: The distance norms must be defined and can then be limited to certain values; the possible locations of the new facilities can be constrained to certain discrete locations; or the structure of the w_{ji}'s may be amended in order to find the optimal allocation of new facilities to old facilities.

2.2 The Traveling-Salesman Facility Location Problem (TSLP)

The problem that this research effort deals with differs from the typical location problem in that the new facilities are to be located such that the traveling-salesman tours between each new facility and the existing facilities that it serves are minimized.

The objective function can be formulated as:

$$\text{P2.2} \quad \underset{X_j \varepsilon E_2}{\text{minimize}}\ f(X_1,\ldots,X_n) = \sum_{j=1}^{n} c(X_j,\{w_{ji}\},S) + \sum_{1\leq j<k\leq n} v_{jk} d(X_j,X_k)$$

In this case everything is the same as defined for P2.1 except

$\{w_{ji}\}$ = the set of all w_{ji}'s

S = the set $\{P_i,\ i=1,\ldots,m\}$, and

$c(X_j,\{w_{ji}\},S)$ = a function giving the cost of a minimum traveling-salesman tour through S (or subsets of S) and X_j based on $\{w_{ji}\}$.

Problem 2.2 will be simplified by assuming that $v_{jk} = 0$ for all j and k, i.e., there will be no interaction between new facilities. This is not an unrealistic assumption for many distribution problems. For example, in the solid waste problem each depot is an independent dispatching site--a truck leaves from and returns to the same depot. There are service vehicles that travel between depots, but this travel is insignificant compared to that of the garbage trucks.

Now the problem can be reformulated as

P2.3 $\underset{X_j \varepsilon E_2}{\text{minimize}}\ f(X_1,\ldots,X_n) = \sum_{j=1}^{n} c(X_j,\{w_{ji}\},S)$

The structure of the function c will depend on the specific problem to be solved.

2.2.1 The Traveling-Salesman Problem

Imbedded within the TSLP is, of course, one or more traveling-salesman problems. A traveling-salesman problem involves finding the minimum-cost tour between a group of n cities, existing facilities or demand points that each must be visited exactly once.

A complete review of the work on the traveling-salesman problem was not directly relevant to this research. But a fairly extensive review of the literature is presented in Appendix I.

It is assumed throughout that the distance between two points is the minimum distance. Bellmore and Nemhauser (1968) pointed out that this must be true to assure that the optimal traveling-salesman tour will visit each city once and only once.

2.2.2 Tour Length Estimation in the TSLP

Previous work has been done relating tour length in capacitated problems to Euclidean distances between the new and existing facilities. This was in an effort to use existing facility-location techniques in problems where vehicles make multiple stops. This could be simply done if a linear function could be found relating total length of all tours to the total of all Euclidean distances.

This was done using regression analysis by Griffiths (1968). The results were used in an actual location problem and the solution was deemed acceptable. But Webb (1968) continued investigations along these lines and concluded that a simple linear function could produce misleading results.

Eilon, et al. (1971) criticized the above work because they did not explicitly consider the average value of the maximum number of customers that can be supplied in one route. Eilon, et al. carried out experiments on randomly generated data and concluded that:

(1) There is a very well defined relationship between the sum of the Euclidean distances and the total of the tour lengths,

(2) The position of the depot does not affect this relationship appreciably,

(3) The standard deviation of a normally distributed customer demand affects the relationship in a well defined way.

This seems to indicate that one can get at least a very good initial solution to some TSLP formulations by solving the problem as a standard location problem and this idea was used in this research.

2.3 Taxonomy of the Problem

The general framework of the TSLP is shown in Figure 2.1. Typical variations of the problem will now be discussed.

2.3.1 Objective of the Cost Function

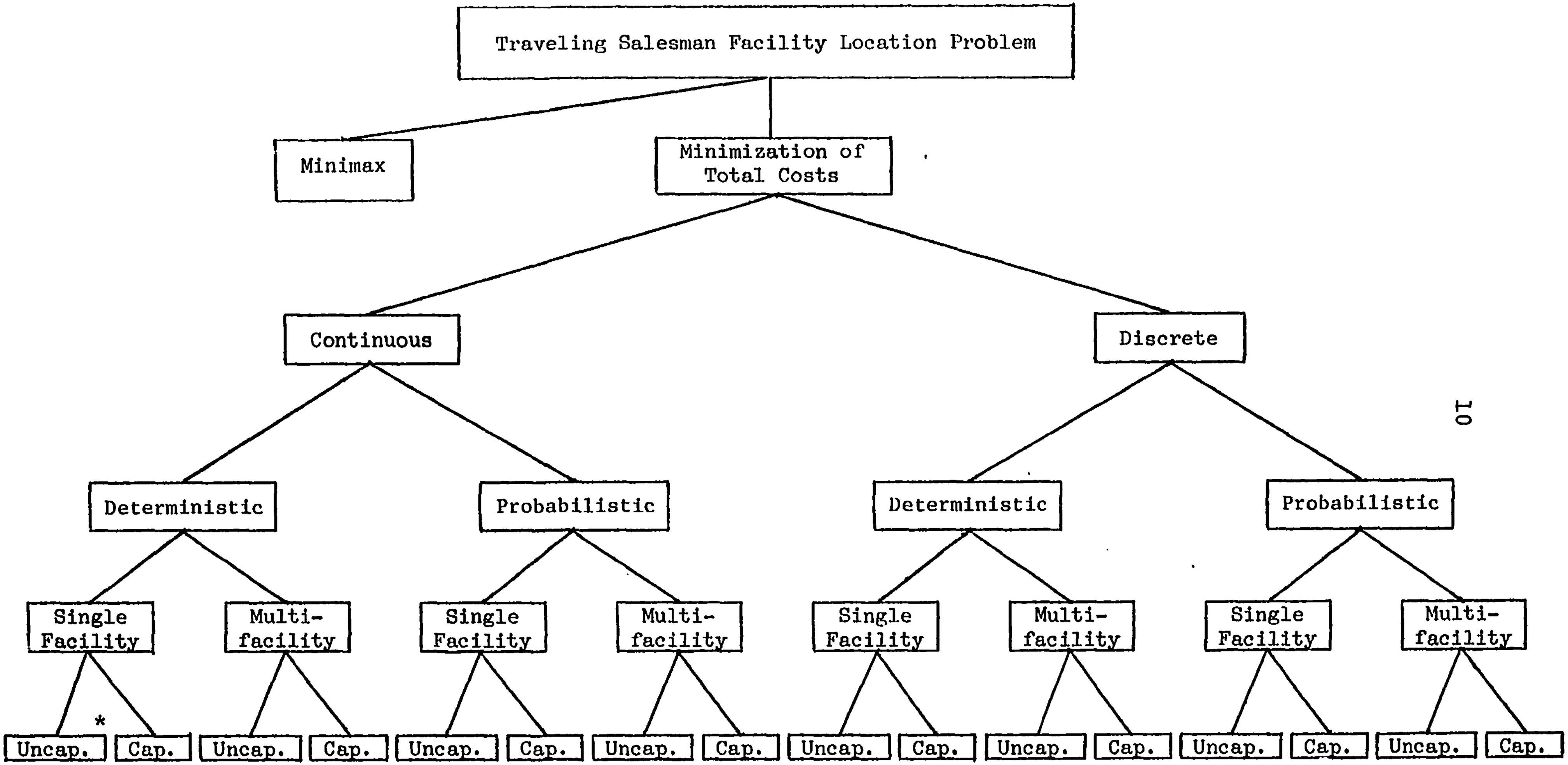
In location problems where the objective is directly tied to distance one is usually interested in minimizing total distance traveled (minisum) or in minimizing the maximum distance traveled (minimax). The minimax formulation is typically used in service or emergency facility location problems and its solution methods are different from those of the minimum distance problem.

To limit the scope of this investigation the minimax problem will not be discussed.

2.3.2 Continuous or Discrete Space

In some cases one may be interested in finding the sites for the new facilities anywhere in the plane that minimizes the costs and meets capacity constraints. In other cases the new facilities are restricted only to predetermined discrete sites (this is known as the plant or warehouse location problem if there are fixed costs on the new facilities or the p-median problem if there are no fixed costs).

There is also the problem of locating the new facility on the transportation network that joins all existing



Taxonomy of the Traveling Salesman Facility Location Problem

* Cap. - Capacitated
Uncap. - Uncapacitated

Figure 2.1

facilities. It has been found that an optimal solution to this problem can always be found on a node of that network (Hakimi, 1965), thus the problem is rendered a discrete problem. This result is extended to the TSLP problem on a network in the following chapter.

2.3.3 Deterministic and Probabilistic Formulations

In the general facility location problem P2.1 it is conceivable that either of the terms v_{jk} or P_i could vary probabilistically. But in this research these terms will always be assumed, without loss of generality, to be deterministic. Other probabilistic variations will be examined.

Let ρ_i be the probability that existing facility i will be visited in a given time period. If $\rho_i = 0$ then a tour from a new facility may not need to include i. Eilon, et al. (1971) present this formulation with the idea of using simulation to find the best new facility site.

Eilon, et al. (1971) state the following problem: Let h(r) be the probability that r existing facilities require service in a given time period. They then recommend a method of locating the new facility using their estimations of the length of traveling-salesman tours.

A third probabilistic formulation was examined by Burness and White (1976). In this work q_i is the probability that subset S_i is serviced in one tour, and the following function is to be minimized.

$$f(X) = \sum_{i=1}^{k} q_i d(X,S_i),$$

where

$$\bigcup_{i=1}^{k} S_i = S$$

k = number of subsets

$d(X,S_i)$ = distance of minimum traveling-salesman tour through X and existing facilities of S_i.

The probabilities in each of the above formulations are based on given time periods. More realism could be added to the problem by making these probabilities functions of time. In this case, when a truck is dispatched only a subset of the existing facilities would require service, but during the course of the trip others may very well develop a demand. These new demands could be handled at a later time or by another truck dispatched on a different route.

Aly (1975) has investigated the probabilistic variations in the interaction terms, w_{ji} and v_{jk}, in typical facility location problems and demonstrated two types of variation in w_{ji}. In one, the weight is a cost per distance traveled and the function to be minimized uses the expected value of w_{ji}.

The other possible probabilistic variation in weights comes about when they represent the frequency of trips. In typical facility location problems this involves a random sum of distances that is to be minimized. The complexities when one is dealing with traveling-salesman distances is obvious and it seems that this type of problem could better be modeled as above.

2.3.4 Single-Facility or Multifacility Formulation

It will be assumed that all new facilities to be located will provide equivalent services. This assumption can be extended to say that at the final solution each new facility will service its own subset of existing facilities and will have no interaction with facilities outside of this subset. Therefore a multifacility problem in this research will involve allocation of facilities and single facility location problems within subsets of existing facilities.

Location-allocation problems have been investigated in the literature by Cooper (1963, 1964, 1967) for the continuous case, and one can consider the typical plant or warehouse location models (for example, Ellwein and Gray, 1971) or the p-median problem (Narula, et al., 1977) as location-allocation in discrete space.

2.3.5 Capacitated Problem

If a problem is uncapacitated there is no limit on the number of existing facilities that can be visited in a tour. But in many cases the vehicle has a limited capacity of load that it can handle. Similarly there may be a limit on the amount of time that a tour can take.

These types of constraints necessitate multiple vehicles or at least multiple tours to service all points. A routing problem with these constraints (assuming locations of depots are known) has been called the transportation-routing problem, the vehicle-dispatch problem or the vehicle-scheduling problem. This problem is important to this research and a survey of the literature is given in Appendix II.

2.4 Heuristics in Distribution Problems

In the discussion of the traveling-salesman problem and the vehicle-dispatch problem in Appendices I and II it was mentioned that the most successful methods of solution have been the heuristic approaches. Both of these problems are np-complete (Karp, 1975), implying that the run-time of an optimal solution method can be expected to increase exponentially with the number of stops. The same is true in location theory with the location-allocation problem. Practical problems of large size cannot be solved optimally in a practical time.

Webb (1972) discusses the costs and benefits of finding the best possible solution to a vehicle routing problem. There are several things to consider:

(1) Practical problems can involve hundreds or sometimes thousands of stops. Therefore, because of run-times, an optimal solution is impossible.

(2) In a given application one should be convinced that the rewards of increased computation exceed the cost

of computer time and analyst time.

(3) It was found that the standard error of estimates of distances may be more than 10% and that of journey duration may be more than 25%. This implies that a small improvement in solution quality may be nonexistent in real terms.

(4) The objectives of the route planners and those doing the route work may not be the same. Also, their respective knowledge of the road network and its constraints may not be the same. It was found that the routes driven may not only be different from the routes that were planned but are usually considerably longer.

Considering these points when locating a facility with respect to at least "good" routes, one should remember that after the facility has been brought into service, the routes that are used may be very much different from those used in the location problem. This does not argue against getting the best possible solution, it only puts the problem in a realistic perspective.

2.5 Scope of this Work

The solution of all problems shown in Figure 2.1 is too great a task to be attempted in this one work--even with the simplifying assumptions made previously. So criteria for selecting the formulations to be researched will be those that seem to be most tractable, those that would seem to have the widest application and those that might have the most immediate application. Principally, discrete problems will be studied. In large area problems this is probably the most realistic since usually only certain sites can be considered. But observations in continuous problems were made and will be discussed.

In studying the problems they were most easily approached by dividing them between deterministic and probabilistic problems (see Figure 2.1). Of the deterministic problems, for the continuous case the uncapacitated problems were examined and for the discrete case the capacitated problems were examined. This included both single- and multifacility problems.

For the probabilistic cases, the single-facility problem was studied for both capacitated and uncapacitated cases. This was done for both continuous and discrete spaces.

CHAPTER III

DOMINANCE PROPERTIES

Certain general properties of the traveling-salesman facility location problem can be derived. When possible these will serve to reduce the size of the solution space.

3.1 Dominance in the General Location Problem

Wendell and Hurter (1973) proved the following property of the general location problem when the objective was to minimize problem P2.1.

Property 1

If 1. the location is to be on a plane,

2. all of the distance norms are the same, and

3. the costs are a nondecreasing function of distance (the greater the distance, the greater the cost),

then a solution at least as good as any other can be found in the convex hull formed by the existing facilities.

3.2 Properties of the TSLP

For the TSLP, Property 1 can easily be extended to

Property 2

Property 1 holds true for the traveling-salesman facility location problem.

Proof: One can consider the TSLP as a typical facility location problem with respect to the first and last existing facility visited in each tour. Thus Property 1 holds for the convex hull formed by these facilities. But this convex hull is contained in the convex hull of all existing facilities. The property is proven.

It will be assumed in any solution procedure presented in this paper that if there are enough feasible location possibilities within this convex hull, then nothing will be considered outside of it.

Property 3

For the unweighted, continuous, deterministic, single-facility, uncapacitated TSLP, the optimal solution is any point on the minimum-traveling-salesman tour through the existing facilities.

Proof: Burness and White (1976) stated this as being obvious. The simple argument is as follows: Since it was previously assumed that the distances between existing facilities was minimum, then if the new facility site was anywhere other than on one of the minimum tour links then the tour length would be increased.

Many times a location problem is constrained to a network whose nodes are the existing facilities and potential facility sites and the arcs are the transportation links between these points.

<u>Property 4</u>

If the solution of a continuous, deterministic TSLP problem is constrained to be on the transportation network joining existing facilities and potential facility sites, then an optimal solution can always be found on a subset of the nodes of that network.

<u>Proof</u>: For the uncapacitated case this follows immediately from Property 3. The capacitated case is proven in the following argument.

Assume that the optimal location X* of a new facility is somewhere on the link between nodes a and b which are on the same tour. Let a' and b' be two nodes on another tour between which it is determined that X* will lie on the transportation network. This can be visualized as in Figure 3.1.

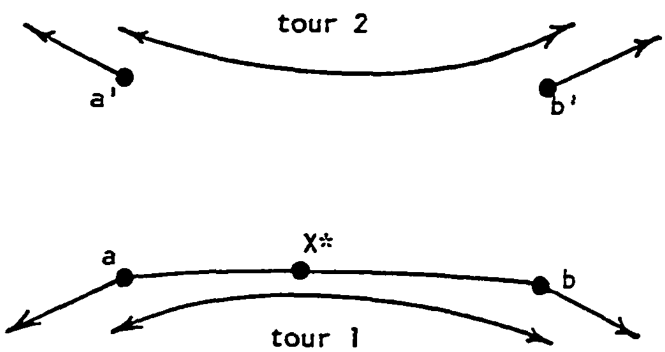


Figure 3.1. Location of new facility in a network.

Assume the most general case where all links (a',a), (a',b), (b',a) and (b',b) exist.

Note that tour 2 will include the links which determine the minimum of (d_1, d_2, d_3, d_4) where

$$d_1 = d(a',a) + d(a,X^*) + d(X^*,b) + d(b,b')$$

$$d_2 = d(a',a) + d(a,X^*) + d(X^*,a) + d(a,b')$$

$$d_3 = d(a',b) + d(b,X^*) + d(X^*,a) + d(a,b')$$

$$d_4 = d(a',b) + d(b,X^*) + d(X^*,b) + d(b,b')$$

and

d(U,V) = distance from U to V.

Now assume that X* lies on one of the nodes a or b. Then tour 2 will include the links which determine the minimum of (d_a, d_b) where

$$d_a = d(a',a) + d(a,b') \qquad (X^* = a)$$

$$d_b = d(a',b) + d(b,b') \qquad (X^* = b)$$

Since, by an earlier assumption, any link is the minimum distance between its end points then

$$\begin{aligned} d_a &\le d_1 \\ d_a &\le d_2 \\ d_b &\le d_3 \\ d_b &\le d_4 \end{aligned} \qquad (1)$$

There are two cases:

1) If $d_a \le d_b$ then from (1) $d_a \le \min(d_1, d_2, d_3, d_4)$ and X* = a.

2) If $d_b \le d_a$ then from (1) $d_b \le \min(d_1, d_2, d_3, d_4)$ and X* = b.

This argument can be successively applied to each tour in the system.

This is a result similar to that of Hakimi (1965) and Levy (1967) except that they were not concerned with traveling-salesman tours. The effect of this property is that continuous problems on a network can be solved as discrete problems if a discrete solution procedure is available.

CHAPTER IV

DETERMINISTIC FORMULATIONS

4.1 Introduction

This chapter will present deterministic formulations of the TSLP. Continuous space problems that are studied are both the single- and multifacility uncapacitated problems. The discrete problems investigated are the single- and multifacility capacitated problems.

4.2 Consideration of Existing Facility Weights

Property 3 solved the simplest TSLP, the continuous, deterministic, single-facility, uncapacitated problem. But no consideration was made of weights associated with each existing facility. Now assume that we have weights which represent costs per unit distance of travel from the existing facility to the new facility.

In the usual facility location problems each weight can easily be multiplied times the distance between the existing facility it is associated with and the new facility. But if the distances are traveling-salesman distances then problems arise. The trip from an existing facility back to the new facility may not be direct but through a series of other existing facilities. Therefore it seems that the weight per unit of travel should be ever increasing (or ever diminishing) as the tour visits each customer. In certain instances, e.g., when the facility changes the demands on a vehicle or when there is a cost associated with the time in transit of items being delivered, this may be realistic, but as will be shown below, the computational problems will be greatly compounded.

Assume that a route through the existing facilities is given. By Property 3, in the case of no weight, the optimal solution is anywhere on the route. With weights, the solution will be on this route but in a distinct location.

The route can be visualized as a circle with its circumference being the total distance of the route and all existing facilities are layed out along the circumference. One of the existing facilities is labeled as facility 1 and the remainder are numbered 2,3,...,m going in the direction of the route around the circle. Define the following

d_i = distance along the route from facility 1 to facility i.

x = distance along the route from facility 1 to the new facility.

D = total distance of the route.

$d_i(x)$ = distance from facility i to the new facility located at x along the direction of the route.

w_i = weight of facility i.

This model can be shown as in Figure 4.1.

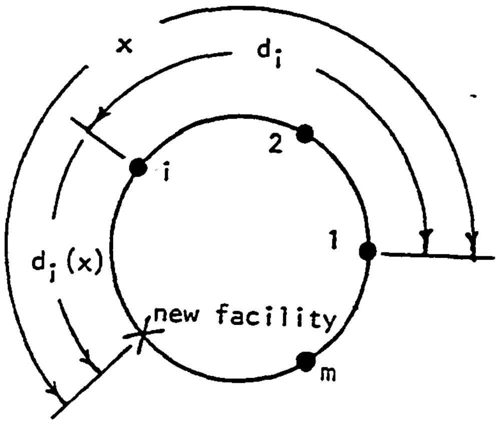


Figure 4.1. Visualization of an uncapacitated route.

Note that if facility i lies before the new facility in the direction of the route then $d_i(x) = x-d_i$. But if facility i lies ahead of the new facility then $d_i(x) = (D-d_i)+x$, where $(D-d_i)$ is the distance from the new facility to facility 1 (reference point). Or,

$$d_i(x) = \begin{cases} x-d_i, & d_i \leq x \\ (D-d_i)+x, & d_i > x \end{cases}$$

or,

$$d_i(x) = \begin{cases} x-d_i, & i = 1,\ldots,p \\ (D-d_i)+x, & i = p+1,\ldots,m \end{cases}$$

where p is the maximum value of i such that $d_i \leq x$.

The minimum must now be found of the cost,

$$z = \sum_{i=1}^{m} w_i \, d_i(x)$$

$$= \sum_{i=1}^{p} w_i(x-d_i) + \sum_{i=p+1}^{m} w_i(D-d_i+x)$$

$$= \sum_{i=1}^{m} w_i x - \sum_{i=1}^{m} w_i d_i + D \sum_{i=p+1}^{m} w_i$$

Letting $W = \sum_{i=1}^{m} w_i$ and noting that $\sum_{i=1}^{m} w_i d_i$ is constant,

$$z' = xW + D \sum_{i=p+1}^{m} w_i$$

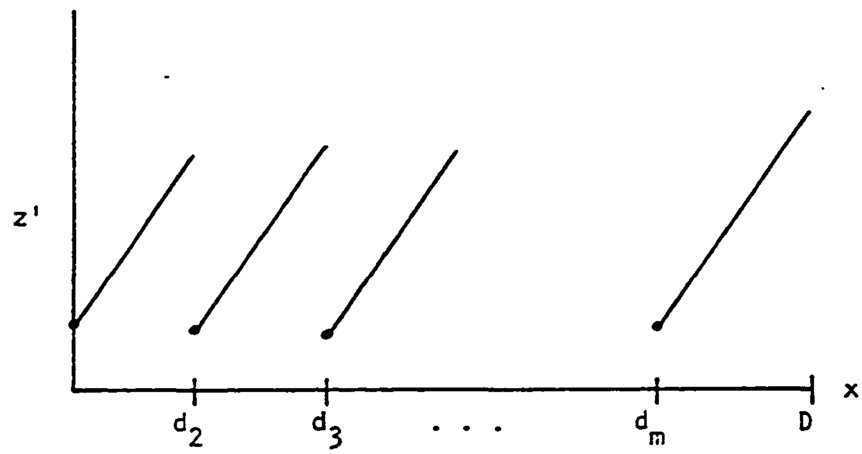


Figure 4.2. Graph of z'.

Figure 4.2 is a plot of z'. Notice that the minimum will obviously fall on one of the existing facilities, or, $x = d_i$ for some i. Therefore the minimum for a given direction of travel will be at facility j for

$$\min_{1 \le j \le m} \{d_j W + D \sum_{i=j+1}^{m} w_i\}$$

And for a route in the opposite direction:

$$\min_{1 \leq j \leq m} \{D \sum_{i=1}^{j-1} w_i - d_j W\}$$

This will give the optimal location on a given route. Perhaps this could be incorporated into an algorithm that would produce a location associated with a minimum weighted distance route.

This is the only case where weights will be considered explicitly. In the TSLP, the cumulative aspect of the weights increases the computational problems immensely.

These computational problems might be lessened at the expense of reality by assuming that the whole tour has one weight equal to the sum of all weights in the tour (which might be related to truck capacity).

A second interpretation might arise if the weights have no effect on the cost of travel. This might be the case if the vehicle is delivering or picking up demands based on the weight but its cost of travel is not affected and the stop is made no matter what the demand is. Here the weights would have no mathematical effect on the routing unless the vehicle has some capacity limit.

The weights might also be viewed as the frequency of demand or the number of trips required per unit of time. This would very much depend on the planning time horizon and on whether the facility required service as soon as it was demanded. Assuming that the times of the occurrences of the demands are known with certainty, the new facility might be located with respect to subtours through the facilities that required service in a given interval of time.

Webb (1972) has warned against using weighted distance, where the weight is the amount of goods transported, as a measure of cost in a distribution system. This approach can be very deceiving since many times the relationship between cost and weighted distance is a function of the type of vehicle used. Also a large part of the expenditures in a distribution system are on labor, thus decisions should be based on careful consideration of how labor should be used. This fact, along with the rising energy costs, make time on the road more important. Therefore, simple minimization of distance traveled would be a prime objective in distribution system design. One would probably be more interested in the "weights" as they related to capacity restrictions.

4.3 The Discrete, Single-Facility, Uncapacitated Problem

The objective will be to minimize the total distance of the tour connecting all of the existing facilities and one new facility. There are no capacity restrictions.

This problem might be rendered trivial in the special case shown in

Observation 4.1

If any of the potential new facility sites are the same as any of the existing facility sites, then distance of the tour would be minimized by locating on one of these sites.

Observation 4.2

If a potential site falls on one of the minimum tour links joining the existing facilities, then the distance would be minimized by locating there.

If neither of the above cases holds then other approaches must be tried. Let N be the number of the potential sites where the new facility can be located. For each of the N sites a traveling-salesman problem can be solved with one of the sites appended to the list of existing facilities. One then chooses the site which gave the minimum-length tour.

Another method will be proposed which will allow the solution of just one traveling-salesman problem and this solution will indicate the minimum distance site for the new facility. The network that is solved is made up of the m existing facilities plus 2N different nodes which are added as follows.

Let all of the existing facilities be numbered 1,2,...,m and the matrix of distances between these facilities be D which is made up of d(i,j) i,j = 1,...,m. The new matrix will be D' with elements d'(i,j) i,j = 1,...,m+2N. Let the proposed sites be numbered 1,...,N and $\underline{d}(i,j)$ be the distance from existing facility i to proposed site j or from proposed site j to existing facility i. This assumes that these distances are symmetric but this is not necessary.

These elements are used to make up the matrix D' shown in Figure 4.3 for N = 3. A traveling-salesman problem is then solved on the network represented by this matrix. Any of the methods reviewed in Appendix I can be used, except the r-optimal method which requires a symmetric distance matrix.

To facilitate discussion of the new network, rename the new nodes of Figures 4.3 and 4.4 as follows: nodes m+1, m+2 and m+3 are A, B and C; and nodes m+4, m+5 and m+6 are A', B' and C'.

Note from the figures that the only arcs in the appended network that have costs are those joining it to the network of existing facilities. The idea behind this is that if, for example, A is the minimum distance location, the tour will enter A and incur the appropriate cost. The tour will then visit all of the other nonprimed nodes and

	node	1	...	m	A m+1	B m+2	C m+3	A' m+4	B' m+5	C' m+6
	1	d(1,1)	...	d(1,m)	$\underline{d}(1,1)$	$\underline{d}(1,2)$	$\underline{d}(1,3)$	∞	∞	∞
	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	m	d(m,1)	...	d(m,m)	$\underline{d}(m,1)$	$\underline{d}(m,2)$	$\underline{d}(m,3)$	∞	∞	∞
A	m+1	∞	...	∞	∞	0	∞	0	∞	∞
B	m+2	∞	...	∞	∞	∞	0	∞	0	∞
C	m+3	∞	...	∞	0	∞	∞	∞	∞	0
A'	m+4	$\underline{d}(1,1)$	...	$\underline{d}(m,1)$	∞	∞	∞	∞	∞	0
B'	m+5	$\underline{d}(1,2)$	...	$\underline{d}(m,2)$	∞	∞	∞	0	∞	∞
C'	m+6	$\underline{d}(1,3)$	...	$\underline{d}(m,3)$	∞	∞	∞	∞	0	∞

Figure 4.3. The Matrix D* with N=3.

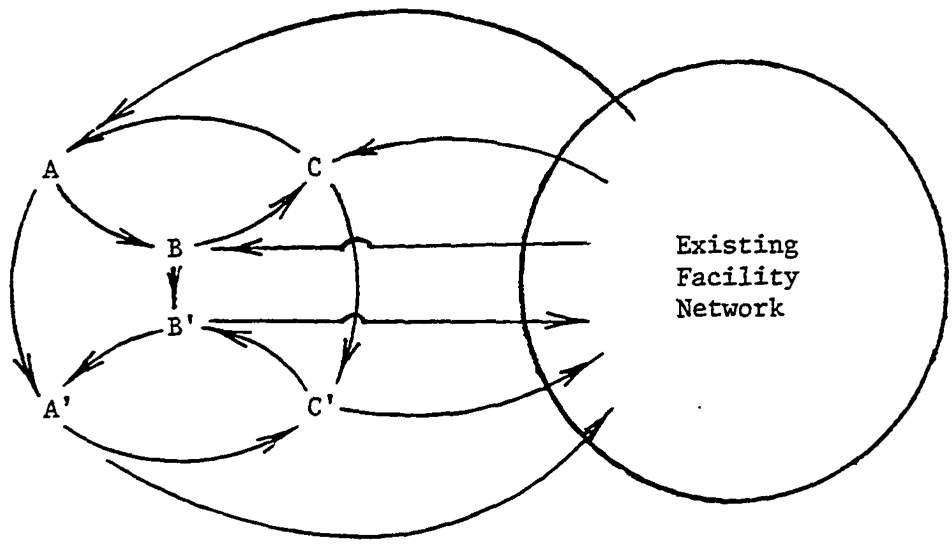


Figure 4.4. Representation of Network D* with N=3.

then all of the primed nodes at no cost and then leave through A', again incurring the appropriate cost.

This type of circuit is probable since in minimizing distance the route would tend to avoid those links with positive distances in favor of those with zero distances. Thus only one arc from the existing facility network to the appended network would be traversed as would only one arc going the opposite direction.

It would also be the case that the same number of primed nodes as nonprimed nodes would be visited in a tour through the appended network. If that were not the case a node would be "stranded" and could not be visited later. For example, if a tour went from an existing facility to A, B, B' and then back to an existing facility, then node A' could not be visited.

To assure that only one circuit be made through the appended network it might be necessary to add a constant, fictitious cost to each of the $\underline{d}(i,j)$'s. This would further force the tour to avoid those arcs. In fact, if all of the distances were scaled to be compatible with the fixed costs of opening facilities on each of the potential sites, then these fixed costs could be added to the $\underline{d}(i,j)$'s and the model would take this into account.

4.4 The Discrete, Single-Facility, Capacitated Problem

This problem involves finding the location for a single new facility among a set of predetermined, discrete sites. The objective is to minimize the distance of a set of tours that will serve each of the existing facilities. Each of these tours are limited by the capacity restrictions.

The obvious method of solving this is to solve vehicle routing problems using each of the new facility candidates in turn as depots. There are heuristic methods of solving the vehicle routing problem, e.g., the savings method (see Appendix II), that handle very large problems in just a few seconds. One then selects as the new facility the candidate giving the least distance tours.

4.4.1 Heuristic for the Discrete, Single-Facility, Capacitated Problem

Another method could use the relationships developed by Eilon et al. (1971). Here one could roughly find the point where tour distances were minimized by finding the point where the total of Euclidean distances between the candidate site and each existing facility was minimized.

To test this approach for the discrete problem 15 different customer sets were generated in a 100 by 100 square. The coordinates of the stops came from a uniform distribution with a range of 0 to 100 and the demands were generated from a normal distribution with mean 50 and a standard deviation of 15. There were 5 sets with 25 customers, 5 with 50 customers and 5 with 100 customers. Each of the sets were solved with four different capacities of 200, 300, 500 and 1000 units. This would mean an average number of stops per route of 4, 6, 10 and 20. For each of the 60 problems there were 10 candidate sites whose coordinates were generated from a uniform distribution of range 0 to 100.

For each candidate site in each problem a routing problem was solved. The method used was a variation of the savings method discussed in Appendix II. This method incorporated a list searching technique from Matthaus (1976) and a modified savings function introduced in Gaskell (1967). The algorithm was coded in FORTRAN IV on an IBM 370/158. and test results appear in Appendix II. The steps of the algorithm follow.

4.4.2 Algorithm for Single-Terminal Vehicle Routing

The data:

M = number of stops

D(I,J) = distance from stop I to stop J

DT(I) = distance from terminal to stop I

The algorithm variables:

A = list of terminal distances which is sorted and used to select linkages

BSTDIS = the best distance of all routes found so far

IA = list of stops associated with the values in A

J1,J2 = temporary variables for stops being considered for linking

K1,K2 = the best stops found for linking in one iteration

LINK = logical variable indicating whether any stops were linked on a route

PARA = route-shaper parameter in the modified savings function

S = savings

S0 = best savings found in an iteration

SPARA = the parameter associated with the best route found

TDIST = total distance of a route

The algorithm:

1. BSTDIS = a large number

   PARA = .2

2. Initialize data structure such that each stop is a single route.

3. A(I) = DT(I), IA(I) = I, for I = 1,...,M.

   Sort A and IA in descending order of A.

4. S0 = a very small number

   LINK = FALSE

   I = 1

5. J1 = IA(I), J1 is the stop associated with index I.

6. If A(I) + A(I+1) $\leq$ S0 then go to step 16, there cannot be an improvement.

7. J = I + 1

8. J2 = IA(J), J2 is the stop associated with index J.

9. If A(I) + A(J) $\leq$ S0 then go to step 15, there cannot be an improvement.

10. If J1 and J2 are on the same route, or if joining J1 and J2 will exceed capacity or distance constraints, then

go to step 14.

11. Calculate savings: S = DT(J1) + DT(J2) - PARA*D(J1,J2).
12. If S $\leq$ S0 then go to step 14.
13. Save values: S0 = S, K1 = J1, K2 = J2, LINK = TRUE.
14. If J = M then go to step 15,
    otherwise, J = J+1 and go to step 8.
15. If I = M-1 then go to 16,
    otherwise, I = I+1 and go to step 5.
16. If LINK = FALSE then go to step 20.
17. Update data structures such that the route containing K1 and K2 are linked at K1 and K2.
18. If K1 or K2 are no longer at the end of a route they are no longer considered for linking. Replace their entries in A with a very small number and resort A. Go to step 4.
19. Calculate total distance of the routes, TDIST.
20. If TDIST < BSTDIS then BSTDIS = TDIST and SPARA = PARA.
21. If PARA = 2 then go to step 22,
    otherwise, PARA = PARA + .2 and go to step 2.
22. The best route is the one associated with SPARA and the best distance is BSTDIS. <u>STOP</u>

4.5.3 Test Results

The results are shown in Table 4.1. The objective was to correlate the route distances with the sum of the Euclidean distances between the depots and the customers. A summary is given in Table 4.2.

TABLE 4.1

RESULTS OF EXPERIMENTS TO RELATE RADIAL DISTANCES TO ROUTE DISTANCES

Number of customers = 25

Problem set	Candidate site	Euclidean Distances		Route Distances Capacities 200		300		500		1000	
1	1	2505	(4)*	936	(4)	753	(4)	619	(4)	495	(3)
	2	3301	(9)	1149	(9)	912	(9)	711	(10)	530	(9)
	3	1964	(1)	791	(1)	669	(1)	526	(1)	470	(1)
	4	2089	(2)	883	(2)	674	(2)	569	(2)	478	(2)
	5	3016	(7)	1061	(7)	828	(7)	640	(6)	523	(6)
	6	3291	(8)	1146	(8)	898	(8)	710	(9)	529	(8)
	7	3362	(10)	1152	(10)	913	(10)	680	(8)	561	(10)
	8	2611	(5)	967	(5)	757	(5)	648	(7)	495	(4)
	9	2832	(6)	1009	(6)	808	(6)	632	(5)	525	(7)
	10	2361	(3)	917	(3)	742	(3)	605	(3)	498	(5)
	Correlations:			.9903		.9919		.9447		.9436	
2	1	1889	(1)	758	(1)	610	(1)	499	(1)	432	(2)
	2	2305	(7)	871	(7)	687	(7)	538	(7)	456	(6)
	3	2087	(4)	808	(3)	656	(4)	508	(2)	423	(1)
	4	3136	(9)	1139	(10)	832	(9)	643	(9)	514	(10)
	5	3174	(10)	1100	(9)	849	(10)	644	(10)	499	(9)
	6	2490	(8)	932	(8)	739	(8)	573	(8)	458	(7)
	7	2240	(6)	845	(6)	671	(6)	536	(6)	453	(5)
	8	1941	(2)	775	(2)	627	(2)	512	(3)	434	(3)
	9	2150	(5)	825	(5)	658	(5)	528	(5)	438	(4)
	10	2064	(3)	815	(4)	654	(3)	517	(4)	460	(8)
	Correlations:			.9950		.9959		.9936		.9320	
3	1	3054	(7)	1132	(7)	830	(7)	643	(7)	502	(6)
	2	3208	(10)	1159	(9)	871	(10)	668	(10)	515	(9)
	3	3016	(6)	1109	(5)	816	(6)	649	(8)	504	(7)
	4	2415	(4)	942	(3)	724	(4)	591	(4)	489	(2)
	5	1876	(1)	801	(1)	640	(1)	543	(2)	502	(5)
	6	3173	(9)	1162	(10)	850	(8)	640	(6)	513	(8)
	7	3097	(8)	1135	(8)	853	(9)	661	(9)	532	(10)
	8	3005	(5)	1114	(6)	794	(5)	624	(5)	490	(4)
	9	2310	(3)	943	(4)	698	(3)	564	(3)	464	(1)
	10	1933	(2)	813	(2)	654	(2)	534	(1)	490	(3)
	Correlations:			.9980		.9887		.9759		.5604	

* Numbers in parenthesis are rankings.

TABLE 4.1--Continued

Number of customer = 25

Problem set	Candidate site	Euclidean Distances		Route Distances: Capacities 200		300		500		1000	
4	1	2307	(5)	857	(4)	684	(5)	564	(6)	469	(6)
	2	2587	(7)	977	(7)	748	(7)	575	(7)	484	(7)
	3	2150	(3)	837	(3)	660	(3)	532	(3)	462	(4)
	4	2841	(8)	1038	(8)	776	(8)	616	(8)	487	(8)
	5	1783	(1)	732	(1)	590	(1)	493	(1)	438	(1)
	6	2979	(9)	1061	(9)	810	(9)	620	(9)	512	(9)
	7	2330	(6)	859	(5)	691	(6)	563	(5)	461	(3)
	8	3852	(10)	1323	(10)	986	(10)	729	(10)	557	(10)
	9	2262	(4)	870	(6)	673	(4)	552	(4)	466	(5)
	10	2109	(2)	792	(2)	643	(2)	528	(2)	461	(2)
	Correlations:			.9958		.9985		.9950		.9870	
5	1	2398	(5)	997	(5)	751	(3)	610	(6)	525	(7)
	2	3035	(9)	1175	(9)	904	(9)	668	(9)	556	(10)
	3	2942	(8)	1158	(7)	860	(8)	621	(7)	521	(5)
	4	1800	(1)	843	(1)	677	(2)	521	(1)	481	(1)
	5	2883	(7)	1161	(8)	835	(7)	651	(8)	547	(9)
	6	2282	(3)	981	(3)	768	(4)	596	(4)	495	(2)
	7	2345	(4)	979	(4)	775	(5)	588	(3)	504	(3)
	8	3102	(10)	1187	(10)	908	(10)	674	(10)	523	(6)
	9	2631	(6)	1067	(6)	776	(6)	603	(5)	535	(8)
	10	1931	(2)	881	(2)	661	(1)	533	(2)	503	(3)
	Correlations:			.9966		.9689		.9584		.8106	

TABLE 4.1--Continued

Number of customers = 50

Problem set	Candidate site	Euclidean Distances		Route Distances: Capacities 200		300		500		1000	
1	1	5263	(4)	1881	(4)	1357	(4)	1015	(4)	776	(4)
	2	6435	(9)	2200	(9)	1630	(9)	1133	(9)	876	(9)
	3	3969	(1)	1507	(1)	1123	(1)	883	(1)	682	(1)
	4	4147	(2)	1549	(2)	1173	(2)	896	(2)	728	(2)
	5	6104	(7)	2126	(7)	1509	(7)	1094	(7)	810	(7)
	6	6419	(8)	2195	(8)	1627	(8)	1132	(8)	876	(10)
	7	6856	(10)	2328	(10)	1647	(10)	1195	(10)	812	(8)
	8	5501	(5)	1946	(5)	1387	(5)	1083	(6)	770	(5)
	9	5743	(6)	2013	(6)	1451	(6)	1082	(5)	802	(6)
	10	4581	(3)	1693	(3)	1270	(3)	980	(3)	756	(3)
	Correlations:			.9990		.9908		.9852		.8919	
2	1	4189	(2)	1482	(2)	1147	(2)	837	(1)	656	(2)
	2	4336	(4)	1511	(3)	1176	(3)	870	(5)	673	(6)
	3	4809	(8)	1644	(8)	1261	(8)	910	(8)	682	(7)
	4	5878	(10)	1929	(10)	1457	(10)	1021	(10)	741	(10)
	5	5600	(9)	1857	(9)	1383	(9)	973	(9)	735	(9)
	6	4398	(5)	1542	(4)	1180	(5)	894	(7)	660	(3)
	7	4058	(1)	1452	(1)	1138	(1)	838	(2)	688	(8)
	8	4418	(6)	1536	(5)	1207	(7)	883	(6)	666	(4)
	9	4467	(7)	1557	(7)	1177	(4)	848	(3)	666	(5)
	10	4279	(3)	1553	(6)	1191	(6)	867	(4)	654	(1)
	Correlations:			.9953		.9911		.9685		.9043	
3	1	6346	(8)	2061	(8)	1492	(8)	1043	(8)	744	(4)
	2	6627	(10)	2132	(10)	1557	(10)	1084	(10)	790	(10)
	3	6282	(7)	2029	(7)	1472	(6)	1030	(6)	760	(7)
	4	5025	(4)	1714	(4)	1288	(4)	959	(4)	758	(6)
	5	3861	(1)	1398	(1)	1091	(1)	835	(1)	696	(1)
	6	6451	(9)	2096	(9)	1534	(9)	1052	(9)	763	(8)
	7	6110	(5)	1955	(5)	1489	(7)	1037	(7)	778	(9)
	8	6145	(6)	2003	(6)	1460	(5)	1024	(5)	747	(5)
	9	4790	(3)	1673	(3)	1280	(3)	920	(3)	719	(3)
	10	3979	(2)	1443	(2)	1123	(2)	850	(2)	705	(2)
	Correlations:			.9982		.9954		.9930		.8682	

TABLE 4.1--Continued

Number of customers = 50

Problem set	Candidate site	Euclidean Distances		Route Distances Capacities 200		300		500		1000	
4	1	4834	(5)	1654	(4)	1191	(6)	896	(6)	680	(5)
	2	5073	(7)	1806	(7)	1258	(7)	903	(7)	684	(7)
	3	4361	(3)	1593	(3)	1115	(3)	821	(3)	618	(2)
	4	6030	(9)	1971	(8)	1396	(9)	990	(9)	706	(9)
	5	3869	(1)	1393	(1)	998	(1)	772	(1)	594	(1)
	6	5787	(8)	2018	(9)	1388	(8)	971	(8)	704	(8)
	7	4848	(6)	1659	(5)	1171	(5)	875	(5)	684	(6)
	8	7286	(10)	2341	(10)	1624	(10)	1123	(10)	837	(10)
	9	4616	(4)	1668	(6)	1170	(4)	856	(4)	638	(3)
	10	4202	(2)	1465	(2)	1057	(2)	805	(2)	640	(4)
	Correlations:			.9874		.9951		.9969		.9522	
5	1	4563	(4)	1751	(4)	1260	(4)	946	(4)	755	(5)
	2	6555	(10)	2294	(10)	1639	(10)	1137	(10)	839	(10)
	3	5712	(9)	2095	(9)	1489	(9)	1058	(9)	813	(9)
	4	3535	(1)	1454	(1)	1085	(1)	836	(1)	681	(1)
	5	5440	(7)	1969	(6)	1406	(6)	1037	(7)	785	(8)
	6	4810	(5)	1809	(5)	1339	(5)	987	(5)	742	(4)
	7	4295	(3)	1649	(3)	1221	(3)	908	(3)	705	(2)
	8	5624	(8)	2031	(8)	1488	(8)	1043	(8)	773	(7)
	9	5391	(6)	1993	(7)	1459	(7)	1027	(6)	771	(6)
	10	3940	(2)	1576	(2)	1173	(2)	888	(2)	716	(3)
	Correlations:			.9982		.9944		.9960		.9632	

TABLE 4.1--Continued

Number of customers = 100

Problem set	Candidate site	Euclidean Distances		Route Distances: Capacities 200		300		500		1000	
1	1	10,542	(4)	3451	(4)	2467	(4)	1710	(4)	1194	(4)
	2	12,871	(9)	4105	(9)	2861	(8)	1964	(8)	1348	(9)
	3	7,963	(1)	2779	(1)	2008	(1)	1414	(1)	1080	(1)
	4	8,568	(2)	2933	(2)	2130	(2)	1525	(2)	1110	(2)
	5	11,807	(7)	3823	(7)	2710	(7)	1889	(7)	1274	(7)
	6	12,830	(8)	4104	(8)	2873	(9)	1965	(9)	1364	(10)
	7	13,127	(10)	4219	(10)	2940	(10)	1998	(10)	1333	(8)
	8	10,829	(5)	3566	(5)	2568	(6)	1749	(5)	1240	(5)
	9	10,957	(6)	3598	(6)	2556	(5)	1776	(6)	1258	(6)
	10	9,045	(3)	3071	(3)	2242	(3)	1565	(3)	1140	(3)
	Correlations:			.9996		.9981		.9968		.9879	
2	1	8,082	(1)	2719	(1)	1948	(1)	1400	(1)	1074	(4)
	2	8,815	(5)	2930	(6)	2106	(7)	1459	(4)	1093	(7)
	3	8,953	(7)	2925	(5)	2093	(6)	1505	(7)	1074	(5)
	4	12,777	(10)	3931	(10)	2732	(10)	1875	(10)	1253	(10)
	5	11,841	(9)	3691	(9)	2589	(9)	1795	(9)	1235	(9)
	6	9,669	(8)	3110	(8)	2190	(8)	1578	(8)	1125	(8)
	7	8,552	(3)	2848	(4)	2021	(2)	1430	(2)	1090	(6)
	8	8,501	(2)	2839	(2)	2034	(3)	1442	(3)	1036	(1)
	9	8,645	(4)	2845	(3)	2063	(4)	1462	(5)	1058	(3)
	10	8,940	(6)	2945	(7)	2088	(5)	1490	(6)	1037	(2)
	Correlations:			.9994		.9980		.9974		.9487	
3	1	13,111	(9)	4102	(8)	2856	(8)	1912	(8)	1260	(9)
	2	13,087	(8)	4112	(9)	2874	(9)	1925	(9)	1249	(8)
	3	12,647	(7)	4004	(7)	2782	(7)	1911	(7)	1222	(6)
	4	9,651	(4)	3171	(4)	2282	(4)	1558	(4)	1107	(4)
	5	8,315	(1)	2786	(1)	2007	(1)	1467	(1)	1055	(2)
	6	11,536	(6)	3711	(6)	2625	(6)	1763	(6)	1238	(7)
	7	13,484	(10)	4240	(10)	2900	(10)	1976	(10)	1288	(10)
	8	10,949	(5)	3541	(5)	2503	(5)	1721	(5)	1218	(5)
	9	8,789	(3)	2983	(3)	2115	(3)	1515	(3)	1072	(3)
	10	8,575	(2)	2880	(2)	2048	(2)	1487	(2)	1042	(1)
	Correlations:			.9994		.9982		.9966		.9659	

TABLE 4.1--Continued

Number of customers = 100

Problem set	Candidate site	Euclidean Distances	Route Distances: Capacities 200	300	500	1000
4	1	10,254 (6)	3323 (5)	2354 (6)	1637 (7)	1140 (7)
	2	9,866 (5)	3328 (6)	2319 (5)	1602 (5)	1134 (5)
	3	8,622 (2)	2975 (3)	2088 (3)	1479 (2)	1058 (3)
	4	12,695 (9)	3998 (9)	2819 (9)	1891 (9)	1263 (9)
	5	8,174 (1)	2810 (1)	2040 (1)	1444 (1)	1056 (2)
	6	11,137 (8)	3697 (8)	2559 (8)	1738 (8)	1192 (8)
	7	10,289 (7)	3336 (7)	2356 (7)	1624 (6)	1135 (6)
	8	14,631 (10)	4551 (10)	3184 (10)	2047 (10)	1349 (10)
	9	9,096 (4)	3136 (4)	2182 (4)	1534 (4)	1111 (4)
	10	8,895 (3)	2961 (2)	2077 (2)	1503 (3)	1044 (1)
	Correlations:		.9951	.9972	.9983	.9871
5	1	9,281 (4)	3286 (4)	2297 (4)	1545 (4)	1099 (4)
	2	13,544 (10)	4506 (10)	3118 (10)	2070 (10)	1340 (10)
	3	10.943 (7)	3624 (7)	2536 (7)	1760 (9)	1172 (7)
	4	7,345 (1)	2723 (1)	1904 (1)	1373 (1)	1013 (1)
	5	10,962 (8)	3736 (9)	2618 (9)	1732 (7)	1187 (9)
	6	10,035 (5)	3503 (5)	2414 (5)	1708 (6)	1166 (6)
	7	8,394 (3)	2924 (3)	2098 (3)	1486 (3)	1028 (2)
	8	11,031 (9)	3669 (8)	2568 (8)	1749 (8)	1155 (5)
	9	10,440 (6)	3535 (6)	2480 (6)	1695 (5)	1174 (8)
	10	7,774 (2)	2758 (2)	1991 (2)	1411 (2)	1052 (3)
	Correlations:		.9946	.9970	.9929	.9732

TABLE 4.2

AVERAGE CORRELATIONS FROM TABLE 4.1

Number of Customers	Capacities				Average Over All Capacities
	200	300	500	1000	
25	.9951(1)*	.9888(1.2)	.9735(1.2)	.8467(2)	.9510(1.35)
50	.9956(1)	.9934(1)	.9879(1.2)	.9160(1.4)	.9732(1.15)
100	.9976(1)	.9977(1)	.9964(1)	.9726(1.8)	.9911(1.2)
Average Over All Customer Sets	.9961(1)	.9933(1.07)	.9859(1.13)	.9117(1.73)	.9717(1.23)

*Starting from the location of minimum Euclidean distance, this is the average number of routing problems that must be solved to find the optimal location.

It is obvious that the correlations were extremely high for all cases except perhaps the problems with a capacity restriction of 1000 units. In general there were high correlations with low capacities and a large number of customers. The good correlation with low capacities makes sense because low capacity approaches the case of single customer routes where tour distance is equal to Euclidean distance.

This also holds true with the number of times that the location of minimum Euclidean distances corresponds with the location of minimum tour distances. With a capacity of 200 units the two minima correspond in every case. Other than that the correspondence held in almost every case. In the worst case routing problems for the four minimum Euclidean distance locations would have to be solved to find the minimum tour distance location. This occurred in a problem with 25 customers and a capacity of 1000 units--the case which had the lowest average correlation. In 56 out of 60 cases the solution of two routing problems would find the minimum distance location. The probability of this happening by chance is $\binom{60}{56} .2^{56} .8^{4}$, which is essentially zero.

This all suggests a heuristic: If it is felt that because of a lack of computing resource only K vehicle routing problems can be solved, then select the K candidates which give the minimum sum of distances between them and the existing facilities and solve a vehicle routing problem for each of these K candidates. The one giving the minimum tour distance is selected. If it is impossible to solve any routing problems then the location of minimum Euclidean distance would probably be near optimal, especially if there are several stops and a low capacity.

The results reported here might be made stronger by investigating how particular spacial arrangements of the customers affect the locations. For example, the effect of a certain percentage of the demand clustered in a certain area.

4.5 The Discrete, Multifacility Capacitated Problem

In this case there are m existing facilities or customers and n new facilities or depots to be located from a collection of N discrete possibilities. Krolak et al. (1972) found that their method could be applied to this problem if there was no constraint on the number of new facilities. A multiple-depot vehicle-routing problem is solved using all new facility possibilities and the ones not used in the final solution are discarded. It seems that any multiple-depot vehicle scheduling procedure could be used to this end.

But one may be interested in finding exactly n new facility sites. The same type of heuristic used with the single-facility problem can be suggested here. One can solve a discrete, location-allocation problem using Euclidean distances instead of tour distances. The n new facilities found with this procedure could then have the customers reallocated to then using a conventional multiple-depot routing algorithm.

4.5.1 Problem Testing

This idea was tested on problems with data generated the same way as with the previously discussed single-facility experiments. In this case there were two sets each of 25, 50 and 90 customers.

It was assumed all fixed costs of the facilities were equal, so the Euclidean distance problem that was solved was the p-median problem:

P4.1 minimize $\sum_{i=1}^{n} \sum_{j=1}^{n} a_i d_{ij} x_{ij}$

subject to $\sum_{j=1}^{n} x_{ij} = 1, \; i = 1,\ldots,n$

$\sum_{j=1}^{n} x_{jj} = p$

$x_{ij} \leq x_{jj}, \; x_{ij} = 0 \text{ or } 1, \; i,j = 1,\ldots,n$

where: x_{ij} = 1 if facility i is assigned to facility j,

a_i = demand of node i,

d_{ij} = distance from node i to node j.

This problem and its solution are reviewed in Appendix III. A method is described there which was used by Narula et al. (1977) and was coded in FORTRAN IV for use here. This method solves the Lagrangian dual of the problem by iteratively searching the solution space in the direction of the subgradient. The procedure is described as follows.

The Lagrangian relaxation of the problem with respect to the first constraint and for a given nonnegative n-vector is

$$L_d(\lambda) = \min_x \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i d_{ij} - \lambda_i) x_{ij} + \sum_{i=1}^{n} \lambda_i$$

subject to $\sum_{j=1}^{n} x_{jj} = p$

$x_{ij} \leq x_{jj},\ x_{ij} = 0 \text{ or } 1,\ i,j = 1,\ldots,n.$

The Lagrangian dual of this is

$$\max_{\lambda \geq 0} L_d(\lambda)$$

and subgradient optimization (Held et al., 1974) is used to solve this dual problem. The algorithm follows.

1. Select an initial point: $\lambda_i^0 = \min_{i \neq j}\{a_i d_{ij}\}\ \forall i.\quad k = 0$
2. $D_j^k = \sum_{i=1}^{n} \min\{0, a_i d_{ij} - \lambda_i^k\}\ \forall j$
3. Select the p smallest values from D_j^k. The sum of these values is S.
4. $L_d(\lambda^k) = S + \sum_{j=1}^{n} \lambda_j^k$
5. $d_i^k = \min_{j \varepsilon P}\{a_i d_{ij}\}\ \forall i$ where P is the locations with the p smallest D_j^k.
6. L_p^k = sum of n-p largest d_i^k.
7. $S_i^k = 1 - \sum_{j \varepsilon P} \bar{x}_{ij}$ where $\bar{x}_{ij} = \begin{cases} 1 \text{ if } a_i d_{ij} - \lambda_i^k < 0 \\ 0 \text{ otherwise.} \end{cases}$
8. Terminate if $S_i^k = 0\ \forall i$ or if $L_d(\lambda^k) = L_p^k$

9. Calculate t^k such that $0 < t^k < \pi \, (L_p{*}-L_d(\lambda^k))/|s^k|^2$.
where $L_p{*} = \min\,(L_p{}^1,\ldots,L_p{}^k)$

10. $\lambda^{k+1} = \lambda^k + t^k s^k$, $k = k+1$ and go to step 2.

For convenience all demand points were used as potential new facility sites--as the above algorithm implies. Any other configuration of candidate sites could easily be handled.

The p-median problem was solved for two new facilities and for five new facilities using each problem set. This gave the optimal Euclidean distance location-allocation for each data set. Then nine other sets of facility sites were selected and minimum distance allocations were made for them. These distances are shown under the "Euclidean distance" column in Table 4.3. The alternate sites for the five-new-facility problems were selected at random. Those for the two-new-facility problems were made up of the optimal sites in the five-facility problem.

Then multiple-depot routing problems were solved using each of the site sets as depots. These were solved with capacities of 200, 500 and 1000 units using the algorithm of Matthaus (1976, see Appendix II). Correlations were then calculated (Table 4.3).

4.5.2 Results

As with the single-facility experiments the correlations were highest with low capacity and a high number of

TABLE 4.3

RESULTS OF MULTIFACILITY EXPERIMENTS

Number of Customers = 25

Number of New Facilities = 2

Problem Set	Set of Candidate Sites	Euclidean Distances	Route Distances Capacities 200	500	1000
1	1*	692 (1)	673 (2)	512 (3)	470 (3)
	2	727 (3)	705 (3)	510 (2)	481 (5)
	3	1353(10)	1056(10)	609 (8)	490 (6)
	4	1047 (8)	869 (7)	541 (6)	423 (1)
	5	773 (4)	718 (4)	518 (5)	493 (7)
	6	1060 (9)	915 (9)	559 (7)	494 (8)
	7	905 (6)	842 (6)	512 (4)	469 (2)
	8	694 (2)	672 (1)	616 (9)	570(10)
	9	864 (5)	874 (8)	663(10)	473 (4)
	10	938 (7)	812 (5)	486 (1)	495 (9)
	Correlations:		.9609	.2338	-.2856
2	1	636 (1)	616 (1)	607 (9)	458 (5)
	2	1054 (9)	846 (8)	531 (2)	469 (7)
	3	852 (6)	736 (6)	525 (1)	445 (1)
	4	684 (4)	652 (4)	675(10)	491 (9)
	5	642 (2)	620 (2)	574 (4)	461 (6)
	6	890 (7)	776 (7)	575 (5)	543(10)
	7	677 (3)	651 (3)	598 (6)	457 (2)
	8	739 (5)	700 (5)	606 (8)	457 (3)
	9	1003 (8)	910 (9)	601 (7)	457 (4)
	10	1063(10)	910(10)	563 (3)	474 (8)
	Correlations:		.9750	-.5363	.1495

Given the demand distribution, the expected number of	200	500	1000
Customers per tour	4.000	10.000	20.000
Tours	6.250	2.500	1.250
Tours per facility	3.125	1.250	.625

*Site set number 1 is the optimal p-median location.

TABLE 4.3--Continued

Number of Customers = 25

Number of New Facilities = 5

Problem Set	Set of Candidate Sites	Euclidean Distances	Route Distances Capacities		
			200	500	1000
1	1	341 (1)	460 (1)	550 (7)	622 (7)
	2	426 (3)	588 (7)	519 (2)	604 (5)
	3	611 (9)	644 (9)	542 (5)	512 (1)
	4	483 (6)	633 (8)	586 (8)	566 (3)
	5	421 (2)	558 (3)	547 (6)	603 (4)
	6	450 (5)	571 (6)	528 (3)	620 (6)
	7	500 (7)	525 (2)	634 (9)	541 (2)
	8	448 (4)	559 (4)	639(10)	622 (8)
	9	666(10)	747(10)	443 (1)	737(10)
	10	531 (8)	568 (5)	531 (4)	641 (9)
	Correlations:		.8488	-.4191	.1805
2	1	364 (1)	499 (1)	680(10)	572 (5)
	2	500 (4)	593 (2)	672 (9)	690(10)
	3	487 (3)	626 (6)	642 (6)	631 (7)
	4	515 (6)	607 (3)	644 (7)	654 (8)
	5	517 (7)	645 (8)	558 (2)	546 (4)
	6	502 (5)	618 (5)	571 (4)	627 (6)
	7	605 (9)	643 (7)	528 (1)	505 (2)
	8	603 (8)	645 (9)	639 (5)	479 (1)
	9	440 (2)	608 (4)	656 (8)	687 (9)
	10	729(10)	768(10)	562 (3)	509 (3)
	Correlations:		.9241	-.6330	-.6049
Given the demand distribution, the expected number of					
Customers per tour			4.00	10.00	20.00
Tours			6.25	2.50	1.25
Tours per facility			1.25	.50	.25

TABLE 4.3--Continued

Problem Set	Set of Candidate Sites	Euclidean Distances	Route Distances Capacities 200	500	1000
			Number of Customers = 50; Number of New Facilities = 2		
1	1	1421 (1)	1218 (1)	815 (4)	694 (3)
	2	1831 (7)	1445 (7)	908 (7)	772 (5)
	3	1447 (3)	1323 (5)	783 (1)	685 (2)
	4	1894 (8)	1472 (8)	933 (8)	822(10)
	5	1505 (4)	1242 (3)	812 (3)	796 (7)
	6	1701 (6)	1383 (6)	817 (5)	811 (8)
	7	1555 (5)	1293 (4)	851 (6)	819 (9)
	8	2117 (9)	1619 (9)	934 (9)	782 (6)
	9	1426 (2)	1238 (2)	784 (2)	680 (1)
	10	2205(10)	1696(10)	956(10)	763 (4)
	Correlations:		.9811	.9343	.4639
2	1	1475 (1)	1279 (4)	872 (7)	784(10)
	2	1477 (2)	1266 (3)	780 (1)	656 (1)
	3	1662 (5)	1428 (7)	818 (3)	696 (5)
	4	1980 (9)	1511(10)	868 (6)	693 (3)
	5	1959 (8)	1455 (8)	878 (8)	734 (8)
	6	1744 (6)	1427 (6)	926(10)	689 (2)
	7	1999(10)	1463 (9)	846 (5)	699 (6)
	8	1892 (7)	1401 (5)	798 (2)	694 (4)
	9	1548 (3)	1233 (1)	830 (4)	712 (7)
	10	1565 (4)	1263 (2)	879 (9)	783 (9)
	Correlations:		.8941	.1861	-.2680
Given the demand distribution, the expected number of					
Customers per tour			4.00	10.00	20.00
Tours			12.50	5.00	2.50
Tours per facility			6.25	2.50	1.25

TABLE 4.3--Continued

Number of Customers = 50

Number of New Facilities = 5

Problem Set	Set of Candidate Sites	Euclidean Distances	Route Distances Capacities		
			200	500	1000
1	1	780 (1)	1030 (3)	819 (5)	763 (6)
	2	1077 (5)	1021 (2)	792 (4)	694 (1)
	3	1217 (9)	1085 (7)	890 (7)	809 (8)
	4	979 (2)	949 (1)	754 (1)	699 (2)
	5	1009 (3)	1034 (4)	900(10)	833 (9)
	6	1091 (6)	1037 (5)	791 (3)	759 (5)
	7	1099 (7)	1050 (6)	779 (2)	796 (7)
	8	1234(10)	1367(10)	879 (6)	719 (4)
	9	1060 (4)	1190 (9)	899 (9)	841(10)
	10	1111 (8)	1168 (8)	893 (8)	716 (3)
	Correlations:		.5658	.3362	-.0193
2	1	850 (1)	1066 (4)	786 (3)	674 (3)
	2	1198 (7)	1041 (1)	724 (1)	651 (1)
	3	1163 (3)	1075 (5)	828 (5)	759 (7)
	4	1182 (5)	1100 (6)	840 (6)	773 (8)
	5	1020 (2)	1108 (7)	895(10)	815 (9)
	6	1169 (4)	1046 (2)	869 (9)	887(10)
	7	1213 (8)	1056 (3)	783 (2)	651 (2)
	8	1279 (9)	1181 (9)	850 (7)	754 (6)
	9	1193 (6)	1117 (8)	851 (8)	691 (4)
	10	1570(10)	1306(10)	823 (4)	712 (5)
	Correlations:		.7316	.0	-.0536
Given the demand distribution, the expected number of					
Customers per tour			4.0	10.0	20.0
Tours			12.5	5.0	2.5
Tours per Facility			2.5	1.0	.5

TABLE 4.3--Continued

Number of Customers = 90

Number of New Facilities = 2

Problem Set	Set of Candidate Sites	Euclidean Distances	Route Distances Capacities 200	500	1000
1	1	2573 (1)	2005 (3)	1105 (1)	1003 (2)
	2	2607 (2)	1913 (1)	1122 (3)	927 (1)
	3	3290 (7)	2283 (7)	1315 (6)	1029 (5)
	4	4001 (9)	2641 (9)	1380 (8)	1086(10)
	5	3616 (8)	2485 (8)	1353 (7)	1026 (3)
	6	3144 (6)	2213 (6)	1389 (9)	1078 (7)
	7	4185(10)	2773(10)	1461(10)	1082 (8)
	8	3056 (5)	2167 (4)	1208 (4)	1044 (6)
	9	2677 (3)	1963 (2)	1114 (2)	1026 (4)
	10	2891 (4)	2180 (5)	1298 (5)	1084 (9)
	Correlations:		.9898	.8777	.6027
2	1	2458 (1)	1973 (1)	1119 (2)	979 (3)
	2	2817 (4)	2167 (4)	1234 (4)	983 (5)
	3	2820 (5)	2217 (5)	1399(10)	1089(10)
	4	2479 (2)	2055 (3)	1113 (1)	976 (4)
	5	3345 (6)	2511 (7)	1299 (5)	959 (1)
	6	3373 (7)	2469 (6)	1341 (6)	1034 (7)
	7	3569 (8)	2582 (8)	1398 (9)	1054 (8)
	8	3585 (9)	2641(10)	1396 (8)	1067 (9)
	9	3619(10)	2616 (9)	1362 (7)	997 (6)
	10	2596 (3)	2021 (2)	1155 (3)	970 (2)
	Correlations:		.9924	.8433	.4152
Given the demand distribution, the expected number of					
Customers per tour			4.00	10.00	20.00
Tours			22.50	9.00	4.50
Tours per facility			11.25	4.50	2.25

TABLE 4.3--Continued

Problem Set	Set of Candidate Sites	Euclidean Distances	Route Distances Capacities 200	500	1000
		Number of Customers = 90			
		Number of New Facilities = 5			
1	1	1415 (1)	1322 (1)	1015 (2)	929 (2)
	2	2769(10)	2110(10)	1228(10)	1026 (6)
	3	1799 (3)	1630 (4)	1164 (7)	1051 (7)
	4	2464 (8)	1830 (7)	1148 (5)	926 (1)
	5	2089 (5)	1711 (5)	1154 (6)	1054 (8)
	6	2564 (9)	2029 (9)	1146 (4)	1169(10)
	7	2147 (6)	1784 (6)	1029 (3)	1019 (5)
	8	2272 (7)	1868 (8)	1213 (9)	1143 (9)
	9	1676 (2)	1498 (2)	992 (1)	994 (4)
	10	1889 (4)	1579 (3)	1183 (8)	982 (3)
	Correlations:		.9792	.6382	.4289
2	1	1509 (1)	1510 (1)	1111 (4)	1101 (8)
	2	1696 (2)	1598 (2)	1118 (6)	931 (4)
	3	3066(10)	2310(10)	1328 (8)	1257(10)
	4	2674 (9)	2080 (9)	1336 (9)	1070 (6)
	5	2236 (7)	1843 (7)	1350(10)	1076 (7)
	6	1942 (4)	1686 (3)	1112 (5)	898 (1)
	7	1852 (3)	1686 (4)	1101 (3)	918 (3)
	8	1970 (5)	1768 (6)	1052 (1)	1060 (5)
	9	2031 (6)	1761 (5)	1073 (2)	1225 (9)
	10	2342 (8)	1971 (8)	1193 (7)	899 (2)
	Correlations:		.9944	.7769	.4408
Given the demand distribution, the expected number of					
Customers per tour			4.0	10.0	20.0
Tours			22.5	9.0	4.5
Tours per facility			4.5	1.8	.9

customers--probably because this comes closest to a Euclidean distance problem. Looking at it in other ways, few customers per tour or a large number of tours give high correlations as does a high number of tours per facility (Table 4.3).

A linear regression was done to see the effect of the expected number of tours on the correlations. With an intercept of .031 and a slope of .055 it was found that with 14 expected tours one might get a correlation of .8. With 16 expected tours a correlation of more than .9 was predicted. This all implies that the uncapacitated problem cannot be approached in this way since there would be very few tours.

The same trend was found in the number of times that the minimum Euclidean distance location matched the minimum route distance location. The correspondence rarely occurred with capacities of 500 and 1000 units, but with a capacity of 200 units the first or second lowest Euclidean distance location corresponded to the minimum route distance location in 10 out of 12 cases. The probability of this occurring by chance is $\binom{12}{10}.2^{10}.8^{2} = .0000043$.

Another observation can be made from the results of this experiment. Note that where there was poor correlation there were many site sets with route distances almost equal. In fact, since the routing algorithm was a heuristic it is very hard to make statements about the relative magnitude of the route distances when they are so close. This all implies that when there is a high capacity and few tours there is probably several acceptable configurations of new facilities.

CHAPTER V

PROBABILISTIC TRAVELING-SALESMAN FACILITY LOCATION FORMULATIONS

5.1 Introduction

This chapter will discuss certain formulations of the probabilistic traveling-salesman facility location problem (TSLP). Continuous, discrete, capacitated and uncapacitated aspects of the problem will be presented, but emphasis was placed on the single-facility case because this was found to be most tractable.

The different types of probabilistic variations in location problems were mentioned in Chapter II. This research examined the case where there is a probability on whether an existing facility will require service or not.

In preliminary studies the objective of this research on the probabilistic TSLP was to find an expression for the expected value of the length of a tour based on the location of the new facility. This term could then be minimized to find the new facility location. As will be seen this was not found to be practical. Instead, an approximation procedure was used for the continuous problem and simulation seemed to be the best approach for the discrete problem.

5.2 The Continuous Problem

The single-facility, uncapacitated case of this problem was examined by Burness and White (1976). As will be discussed, their approach was time consuming and did not guarantee an optimal solution. The attempt was to

$$\text{P5.1} \quad \underset{X \varepsilon E_2}{\text{minimize}}\ f(X) = \sum_{i=1}^{n} p_i d(X, S_i)$$

where X = new facility location

S_i = a subset of existing facilities

p_i = probability that S_i will demand service

$d(X,S_i)$ = distance of traveling-salesman tour through X and S_i.

The procedure is stated as follows. Given a starting solution,

1) Solve a traveling-salesman problem for every subset, each of which includes the current solution.
2) Solve a single-facility location problem with respect to the end-points of the tours where the weights are the p_i's.
3) Repeat steps 1 and 2 until there is no change in the solution.
4) Repeat steps 1 through 3 with different starting solutions.

This gave very good solutions but, as mentioned, there is no guarantee of optimality. More importantly, the procedure is very time consuming because of the number of traveling-salesman problems that must be solved. The biggest problems that they reported solving had ten customers and all possible subsets (1023). The average solution time with one starting solution was 6.5 minutes. As they pointed out, a realistic problem would not involve every possible subset, but the time problem lies not just in the number of subsets but in the size of the subsets. Therefore this approach is limited to small problems.

5.2.1 The Continuous, Single-Facility, Capacitated Problem

The previous procedure could be modified to handle capacitated problems. This could be done in step 1 by solving a vehicle routing problem instead of a traveling-salesman problem. In fact, since good vehicle routing heuristics usually require less solution time than traveling-salesman codes, then the capacitated problem would probably be solved much faster.

5.2.2 Expected Distance of a Tour

If one is given a set of probabilities or density functions describing the system under consideration, it seems that if a term could be found for the expected distance of a tour then minimization of this term would solve the problem. Such a term was developed in Appendix IV but it was found that, at least in this form, the use of this term was impractical. So the following procedure was developed.

5.2.3 Approximation Procedure for the Continuous Problem

Consider again Problem P5.1. The difficulty with the Burness and White (1976) procedure was the time involved in solving many traveling-salesman problems. Perhaps this can be overcome by using a result from Eilon et al. (1971). They found that for about ten or more cities the tour length, using Euclidean distances, can be estimated by

$$\alpha \sqrt{n} \sqrt{A} ,$$

where n = number of cities

A = area containing the cities

α = constant.

By simulation and regression Eilon et al. found the value of α to be .75.

Now define

$$\text{P5.2} \quad \underset{X \varepsilon E_2}{\text{minimize}}\ f(X) = \sum_{i=1}^{n} p_i d(X,S_i)$$

where all definitions are the same as P5.1 except

$$d(X,S_i) = \begin{cases} \text{distance of optimal traveling-salesman tour if } |S_i| < \ell. \\ .75\sqrt{|S_i|+1}\,\sqrt{A} \text{ if } |S_i| \geq \ell, \text{ where A is the area of the minimum rectangle containing } S_i \text{ and X.} \end{cases}$$

ℓ = the number of cities beyond which it is decided that the estimation is valid.

The problem P5.1 is a nonlinear function and any efficient unconstrained nonlinear programming method could be utilized.

5.2.4 Testing of the Approximation Procedure

The approximation procedure was tested using COMET NLP code (Staha, 1973). This is a general purpose NLP code and was used because of its availability. The traveling-salesman problems were solved using a code which utilized the 3-optimal method (Lin and Kernighan, 1973) discussed in Appendix I. Based on the results of Eilon et al. (1973) the value of ℓ was set at ten.

The first problems tested were two from Burness and White (1976). There were only four cities, so the estimation term was never used and exact traveling-salesman problems were solved. This allowed a comparison of solutions and they were the same with both methods.

Two five-city problems were uniformly generated in a 100 by 100 square and all possible subsets were used. Again the estimation term was not used so the solution of these problems could be compared with ones in Burness and White. They solved six-city problems in an average of 15.5 seconds, but this was with only one starting solution. The five-city test problems were solved in an average of about 29 seconds. The solutions in Burness and White were found in an average of four iterations but more starting solutions would be necessary to get a higher probability of finding the optimum. The NLP method took 14 iterations and found the optimum.

Eight other problems were similarly randomly generated with 100 cities and 40 subsets. The subsets were 10 each of average sizes of 10, 20, 30 and 40 cities. Because of storage restrictions the maximum sized subset had 50 cities. The average time to solve these problems was about 2.8 minutes. The largest problem that Burness and White reported had 10 cities and all possible subsets. Their average time was 6.5 minutes with one starting solution.

They solved problems with many more subsets than were solved in this study but their largest subset had 10 facilities plus the new one. As they pointed out, a realistic problem probably would not involve every possible subset. The new method is slowed by the number of subsets of size less than ℓ but not by the size of the subsets greater than ℓ.

5.3 The Discrete, Single-Facility Problem

This problem differs from that of section 5.2 because in this case there are only discrete candidate sites to be evaluated for the new facility or depot. Once again the idea of using a term for the expected length of a tour has intuitive appeal. Each alternative site could be evaluated using this term. But the term would be the same as the one developed in Appendix IV and the problems with its size would be the same as those discussed there.

5.3.1 Solution Procedure

Eilon et al. (1971) suggest solving these probabilistic formulations with simulation but they never reported any results. Given the complexity of the problem this may be the most efficient approach.

For this problem let p_i be the probability that customer i demands service in a given time period. Then for each possible depot site, successively simulate customer sets and, depending if there are capacity restrictions, solve a traveling-salesman problem or vehicle-routing problem on each of these sets. The sites with the least total tour length would be selected.

5.3.2 Simulation Results

The discrete, probabilistic, single-facility, capacitated problem was solved using the simulation approach mentioned above. This was tested on three customer sets of 25 customers and three of 50 customers. The coordinates were generated from a uniform distribution in a 100 by 100 square as were the coordinates of 10 new facility candidates for each customer set. Each set was used with capacities of 200, 500 and 1000 units. Thus there were 18 problems. The probabilities that each customer would require service was also generated from a uniform distribution with a range from 0 to 1. The demands of the customers were assumed to be normal with mean 50 and standard deviation of 15.

For the 25 customer problems 500 simulation trials were run and for the 50 customer problems there were 100 trials. The savings method without the route-shaper parameter (see Appendix II) was used for the routing. The simulation

TABLE 5.1

RESULTS OF EXPERIMENTS TO RELATE RADIAL DISTANCES AND DETERMINISTIC DISTANCES TO EXPECTED DISTANCE IN A PROBABILISTIC PROBLEM

Number of customers = 25

Number of simulation trials = 500

Cust. set	Site	Euclidean distance	Route distances — Capacities — 200: Deterministic	200: Probabilistic	500: Deterministic	500: Probabilistic	1000: Deterministic	1000: Probabilistic
1	1	3174 (9)*	1098 (10)	313,384 (4)	641 (8)	305,404 (7)	514 (6)	297,492 (8)
	2	2221 (3)	847 (2)	257,553 (2)	556 (3)	231,822 (2)	473 (1)	225,289 (1)
	3	2840 (5)	1003 (4)	343,363 (7)	637 (7)	280,572 (3)	535 (10)	260,399 (3)
	4	2928 (6)	1053 (8)	388,399 (9)	641 (9)	293,836 (5)	520 (7)	283,888 (5)
	5	1835 (1)	745 (1)	238,516 (1)	533 (1)	231,475 (1)	474 (2)	225,329 (2)
	6	2971 (8)	1042 (7)	300,372 (3)	615 (5)	290,400 (4)	500 (4)	284,255 (6)
	7	2120 (2)	851 (3)	326,504 (6)	555 (2)	301,257 (6)	491 (3)	286,419 (7)
	8	2641 (4)	1011 (5)	401,620 (10)	611 (4)	378,103 (10)	508 (5)	362,938 (10)
	9	3325 (10)	1140 (6)	324,121 (5)	657 (10)	315,600 (9)	525 (8)	307,142 (9)
	10	2965 (7)	1056 (9)	361,714 (8)	628 (6)	306,725 (8)	526 (9)	263,278 (4)
	Correlation with radial distances:			.5061		.5099		.4793
	Correlation with deterministic route distances:			.6095		.5515		.4328
2	1	2719 (7)	1049 (7)	358,321 (7)	591 (6)	317,432 (7)	488 (6)	304,702 (7)
	2	2795 (8)	1054 (8)	346,297 (6)	626 (10)	299,099 (5)	511 (10)	282,880 (5)
	3	2190 (4)	922 (4)	285,248 (1)	572 (5)	244,486 (1)	444 (2)	237,881 (1)
	4	2453 (5)	981 (5)	341,935 (5)	553 (4)	307,138 (6)	474 (5)	294,487 (6)
	5	2881 (9)	1077 (9)	400,829 (9)	611 (9)	349,235 (9)	493 (7)	334,566 (9)
	6	1981 (2)	855 (3)	287,512 (2)	542 (3)	262,698 (2)	446 (3)	258,049 (2)
	7	2656 (6)	1015 (6)	372,329 (8)	599 (8)	326,311 (8)	499 (8)	305,680 (8)
	8	2042 (3)	820 (1)	301,207 (4)	514 (1)	268,365 (3)	431 (1)	262,119 (3)
	9	1937 (1)	840 (2)	295,136 (3)	526 (2)	280,024 (4)	457 (4)	271,764 (4)
	10	3177 (10)	1148 (10)	428,830 (10)	595 (7)	370,006 (10)	509 (9)	354,842 (10)
	Correlation with radial distances:			.9468		.8891		.8684
	Correlation with deterministic route distances:			.9164		.6117		.7911
3	1	2631 (6)	929 (6)	358,190 (8)	562 (7)	325,414 (8)	446 (8)	310,430 (8)
	2	2580 (5)	919 (5)	375,494 (9)	533 (4)	336,621 (9)	439 (6)	320,179 (9)
	3	3517 (10)	1169 (10)	411,973 (10)	647 (10)	364,795 (10)	489 (10)	337,466 (10)
	4	2550 (4)	911 (4)	323,876 (5)	536 (5)	304,796 (6)	417 (4)	284,955 (6)
	5	2635 (7)	939 (7)	353,805 (7)	536 (6)	321,922 (7)	422 (5)	299,500 (7)
	6	2104 (2)	784 (2)	255,230 (2)	501 (2)	220,580 (2)	411 (2)	208,431 (2)
	7	3024 (8)	972 (8)	321,934 (4)	594 (8)	275,859 (4)	442 (7)	259,017 (5)
	8	1976 (1)	743 (1)	228,303 (1)	475 (1)	203,715 (1)	408 (1)	191,627 (1)
	9	3143 (9)	986 (9)	339,368 (6)	630 (9)	281,307 (5)	463 (9)	248,606 (4)
	10	2128 (3)	793 (3)	259,708 (3)	509 (3)	256,965 (3)	415 (3)	248,569 (3)
	Correlation with radial distances:			.8281		.7138		.6380
	Correlation with deterministic route distances:			.9006		.6225		.6211

* Numbers in parenthesis are rankings.

TABLE 5.1--Continued

Number of customers = 50

Number of simulation trials = 100

Cust. set	Site	Euclidean distance	Route distances, Capacities 200: Deterministic	200: Probabilistic	500: Deterministic	500: Probabilistic	1000: Deterministic	1000: Probabilistic
1	1	6921 (9)	2207 (9)	160,304 (9)	1143 (9)	128,152 (9)	781 (9)	121,867 (9)
	2	4989 (3)	1689 (3)	124,834 (3)	923 (3)	108,776 (3)	705 (3)	105,526 (3)
	3	5974 (7)	1999 (6)	143,098 (6)	1055 (7)	121,256 (5)	732 (5)	109,210 (4)
	4	5749 (5)	1972 (5)	130,624 (4)	1026 (5)	126,273 (6)	730 (4)	116,280 (8)
	5	4052 (1)	1450 (1)	108,930 (1)	825 (1)	90,872 (1)	678 (2)	95,031 (1)
	6	6533 (8)	2110 (8)	153,285 (8)	1091 (8)	127,108 (7)	760 (8)	115,373 (7)
	7	4135 (2)	1545 (2)	110,416 (2)	841 (2)	96,423 (2)	651 (1)	99,505 (2)
	8	5227 (4)	1816 (4)	134,173 (5)	982 (4)	116,175 (4)	749 (6)	112,863 (5)
	9	7225 (10)	2291 (10)	165,681 (10)	1178 (10)	127,489 (8)	797 (10)	115,173 (6)
	10	5969 (6)	2005 (7)	145,014 (7)	1036 (6)	146,799 (10)	755 (7)	140,127 (10)
		Correlation with radial distances:		.9860		.8036		.6449
		Correlation with deterministic route distances:		.9751		.8054		.6909
2	1	5493 (7)	1884 (7)	137,594 (7)	972 (6)	118,755 (10)	745 (8)	109,504 (10)
	2	5504 (8)	1896 (8)	147,324 (10)	996 (7)	115,555 (8)	758 (9)	99,090 (6)
	3	4611 (4)	1626 (4)	139,528 (8)	930 (4)	118,412 (9)	724 (5)	108,136 (9)
	4	5010 (5)	1757 (5)	127,540 (4)	952 (5)	107,091 (6)	730 (7)	97,984 (5)
	5	5885 (9)	1954 (9)	130,850 (5)	1043 (9)	104,331 (3)	694 (3)	93,292 (4)
	6	4061 (1)	1496 (1)	125,094 (3)	868 (2)	106,745 (4)	696 (4)	100,131 (8)
	7	5311 (6)	1840 (6)	142,843 (9)	1005 (8)	110,777 (7)	768 (10)	100,115 (7)
	8	4363 (3)	1545 (3)	119,763 (2)	873 (3)	99,267 (2)	690 (2)	90,625 (2)
	9	4066 (2)	1521 (2)	115,239 (1)	853 (1)	98,353 (1)	668 (1)	90,051 (1)
	10	6435 (10)	2105 (10)	137,000 (6)	1103 (10)	106,905 (5)	725 (6)	91,249 (3)
		Correlation with radial distances:		.6312		.3076		-.0164
		Correlation with deterministic route distances:		.6478		.3084		.5691
3	1	5263 (4)	1698 (4)	127,790 (5)	891 (4)	114,500 (6)	710 (5)	105,053 (6)
	2	5287 (5)	1702 (5)	123,510 (4)	928 (5)	106,368 (5)	709 (4)	107,998 (8)
	3	6928 (10)	2204 (10)	142,564 (8)	1044 (10)	122,790 (9)	788 (10)	114,742 (9)
	4	5487 (6)	1810 (7)	136,589 (7)	970 (7)	116,491 (7)	734 (7)	104,459 (5)
	5	5622 (7)	1808 (6)	156,898 (10)	963 (6)	133,900 (10)	712 (6)	131,280 (10)
	6	4271 (2)	1498 (2)	108,111 (1)	818 (1)	95,248 (2)	681 (2)	87,818 (1)
	7	5918 (8)	1928 (8)	142,928 (9)	1002 (8)	117,832 (8)	783 (9)	107,540 (7)
	8	3974 (1)	1401 (1)	110,936 (3)	827 (3)	100,347 (3)	677 (1)	93,616 (3)
	9	6090 (9)	1963 (9)	131,401 (6)	1014 (9)	106,217 (4)	750 (8)	103,501 (4)
	10	4316 (3)	1503 (3)	108,448 (2)	821 (2)	94,326 (1)	685 (3)	90,497 (2)
		Correlation with radial distances:		.8006		.7164		.6866
		Correlation with deterministic route distances:		.7763		.7232		.4829

of each depot candidate took about 10 seconds. The results are shown in Table 5.1.

Also, for each problem the sum of the Euclidean distances were calculated and a standard routing problem was solved disregarding the probabilities. This was done using the same savings algorithm described in section 4.4.1. Correlations were calculated and are also shown in Table 5.1. The trends were similar to those reported in Section 4.4 but the results were not quite as strong. Particularly, there was less correlation with higher capacity as found in Section 4.4 but overall the correlations were smaller. The average correlation with Euclidean distances was .6633 and the correlation with deterministic routes was .6686.

But in almost every case the best location in the probabilistic problem coincided with the best or second best location using both Euclidean distances and deterministic routing. This happened 13 out of 18 times using Euclidean distances and 16 out of 18 times using deterministic routing. This had probabilities of occurring by chance of just over zero for both cases.

These results seem to say that the simulation procedure need not be carried out on all candidate sites. The simulation analysis is very simple but can be time consuming with very large problems and with a large number of trials.

A more complex simulation analysis could be performed if the probabilities were time varying. In this case a given number of vehicles may be available and are dispatched as orders are accumulated. Even the working hours of the new facility could enter into the simulation.

Another area of further work might be to investigate the effect of different probability distribution assumptions and also the spatial arrangements of customers on the optimum locations and the corresponding correlations between Euclidean distances and traveling-salesman distances.

CHAPTER VI

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary and Conclusions

1. The traveling-salesman problem and the vehicle-routing problem are both np-complete--as is the discrete location-allocation problem. Thus nearly every formulation of the TSLP is np-complete. This implies that even when an optimal solution procedure is available the computational time might be excessive. On top of this it has been argued that optimal solutions are mostly unnecessary in distribution problems because of inaccuracies in data and realistic considerations of the solution's actual implementation. Therefore the principal thrust of TSLP research should not necessarily be toward demonstrating optimality.
2. Two dominance properties of the simple facility location models were extended to the TSLP: (1) An optimal solution can always be found in the convex hull formed by the existing facilities and (2) In locating on a network an optimal solution can always be found on the nodes.
3. Four deterministic problems can be solved to optimality

if their associated traveling-salesman problems can be solved to optimality. These are the continuous, uncapacitated problem for both the single- and multifacility cases, and the discrete, single-facility problem for both the capacitated and uncapacitated cases. Although it can be solved to optimality, a very accurate heuristic was found for the discrete, single-facility, capacitated problem.

4. The work with discrete, deterministic problems showed that there is a very high correlation between tour distances and Euclidean distances, thus for many TSLPs ordinary facility location techniques are all that are necessary. This affirms initial results of Eilon et al. (1973).

5. The discrete, deterministic, single-facility, uncapacitated problem was extended to consider weighted distances. This demonstrated the difficulties of using weighted distances in problems involving multiple-stop tours. This, along with the argument that weighted distance is not a good measure of distribution costs, tends to argue against this cost measure in the TSLP.

6. Burness and White (1976) gave an iterative, heuristic solution procedure for the continuous, probabilistic, single-facility problem which can be used on the capacitated and uncapacitated cases. Their method was not good for large problems because several traveling-salesman problems

must be solved. A general NLP procedure was utilized in this research for the uncapacitated problem. This approach utilized approximations for tour distances and proved to be efficient for large problems.

7. It was concluded that simulation is probably the best approach for the discrete, probabilistic, single-facility problem. Each candidate site is evaluated by successively generating customer sets according to some probability distribution and then solving routing problems on these sets.

8. Correlation experiments were done for the discrete, probabilistic, single-facility, capacitated problem, just as they were for the deterministic case. Once again it was found that Euclidean distance was a good indicator of expected distance in the probabilistic problem. An even better indicator was the distance of the deterministic routes.

6.2 Recommendations for Further Research

1. Only eight of the sixteen problems defined above were addressed in this paper. Obviously much work could be done on the remainder of the problems.

2. The correlation experiments of Sections 4.5, 4.6 and 5.5 were done using Euclidean distances. These should also be done using rectangular distances.

3. These correlation experiments should also be done to test different spatial arrangements of demand.

4. A Markovian approach to the probabilistic problems might be possible with cleverly conceived system states.
5. The simulation experiments of Section 5.3.2 for the discrete, probabilistic, single-facility, capacitated problem could be made more complex by allowing the probabilities to be time varying.
6. Different probability distributions of demands should be tried on all problems.

REFERENCES

Aly, A. A. Probabilistic Formulations of Some Facility Location Problems, unpublished dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1975.

Beattie, D. W. "Improving the Structure of a Distribution System," Operational Research Quarterly, Volume 24, Number 3, 1973, pp. 353-364.

Bellmore, Mandell and John C. Malone. "Pathology of Traveling-Salesman Subtour-Elimination Algorithms," Operations Research, Volume 19, Number 2, 1971, pp. 278-307.

Bellmore, Mandell and Saman Hong. "Transformation of Multisalesman Problem to the Standard Traveling-Salesman Problem," Journal of the Association for Computing Machinery, Volume 21, Number 3, 1974, pp. 500-504.

Bellmore, M. and G. L. Nemhauser. "The Traveling Salesman Problem: A Survey," Operations Research, Volume 16, Number 3, 1968, pp. 538-558.

Beltrami, E. J. and L. D. Bodin. "Networks and Vehicle Routing for Municipal Waste Collection," Networks, Volume 4, 1974, pp. 65-94.

Burness, Robert C. and John A. White. "The Traveling Salesman Location Problem," Transportation Science, Volume 10, Number 4, 1976, pp. 348-360.

Cassidy, P. J. and H. S. Bennett. "TRAMP--A Multi-depot Vehicle Scheduling System," Operational Research Quarterly, Volume 23, Number 2, 1972, pp. 151-163.

Christofides, Nicos and Samuel Eilon. "An Algorithm for the Vehicle-Dispatching Problem," Operational Research Quarterly, Volume 20, Number 3, 1969, pp. 309-318.

Christofides, Nicos and Samuel Eilon. "Algorithms for Large-Scale Traveling-Salesman Problem," Operational Research Quarterly, Volume 23, Number 4, 1972, pp. 511-518.

Christofides, Nicos. "The Vehicle Routing Problem," NATO Conference on Combinatorial Optimization, July, 1974, Paris.

Clarke, G. and J. W. Wright. "Scheduling of Vehicles from a Central Depot to a Number of Delivery Points," Operations Research, Volume 12, Number 4, 1964, pp. 568-581.

Cooper, Leon. "Location-Allocation Problems," Operations Research, Volume 11, Number 3, 1963, pp. 331-343.

Cooper, Leon. "Heuristic Methods for Location-Allocation Problems," SIAM Review, Volume 6, 1964, pp. 37-53.

Cooper, Leon. "Solution of Generalized Locational Equilibrium Problems," Journal of Regional Sciences, Volume 7, 1967, pp. 9-11.

Eilon, Samuel, G. D. T. Watson-Gandy and Nicos Christofides. Distribution Management: Mathematical Modeling and Practical Analysis, New York: Hafner Publishing Company, 1971.

Ellwein, Leon B. and Paul Gray. "Solving Fixed-Charge Location-Allocation Problems with Capacity and Configuration Constraints," AIIE Transactions, Volume 6, Number 4, 1971, pp. 290-297.

Francis, R. L. and J. M. Goldstein. "Location Theory: A Selective Bibliography," Operations Research, Volume 22, 1974, pp. 400-409.

Francis, R. L. and J. A. White. Facility Layout and Location: An Analytic Approach, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1974.

Garfinkel, R. S. and George L. Nemhauser. Integer Programming, New York: John Wiley and Sons, 1972.

Gaskell, T. J. "Bases for Vehicle Fleet Scheduling," Operational Research Quarterly, Volume 18, Number 3, 1967, pp. 281-295.

Gillett, Billy E. and Leland R. Miller. "A Heuristic Algorithm for the Vehicle Dispatch Problem," Operations Research, Volume 22, Number 2, 1974, pp. 340-349.

Gillett, Billy E. and Jerry G. Johnson. "Multi-Terminal Vehicle Dispatch Algorithm," Omega, Volume 4, Number 6, 1976, pp. 711-718.

Golden, Bruce L. "Evaluating a Sequential Vehicle Routing Algorithm," AIIE Transactions, Volume 9, Number 2, 1977, pp. 204-208. Note and Response: AIIE Transactions, Volume 9, Number 4, 1977, pp. 426-427.

Golden, B. L., T. L. Magnanti and H. Q. Nguyen. "Implementing Vehicle Routing Algorithms," Networks, Volume 7, 1977, pp. 113-148.

Griffiths, D. O. "The Use of Regression Analysis in a Depot Location Exercise," Applied Statistics, Volume 17, Number 1, 1968, pp. 57-63.

Hakimi, S. L. "Optimum Distribution of Switching Centers in a Communications Network," Operations Research, Volume 13, Number 3, 1965.

Held, R. L. and R. M. Karp. "A Dynamic Programming Approach to Sequencing Problems," SIAM Review, Volume 10, 1962, pp. 196-210.

Held, M. P., P. Wolfe, and H. Crowder. "Validation of Subgradient Optimization," Mathematical Programming, Volume 5, 1974, pp. 68-88.

Holmes, R. A. and R. G. Parker. "A Vehicle Scheduling Procedure Based upon Savings and a Solution Perturbaion Scheme," Operational Research Quarterly, Volume 27, Number 1, 1976, pp. 83-92.

Hong, Saman and Manfred W. Padberg. "A Note on the Symmetric Multiple Traveling-Salesman Problem with Fixed Charges," Operations Research, Volume 25, Number 5, 1977, pp. 871-874.

Isaac, Akkanad M. and Efraim Turban. "Some Comments on the Traveling-Salesman Problem," Operations Research, Volume 17, 1969, pp. 543-546.

Karp, Richard M. "On the Computational Complexity of Combinatorial Problems," Networks, Volume 5, Number 1, 1975, pp. 45-68.

Karp, Richard M. "Probabilistic Analysis of Partitioning Algorithms for the Traveling-Salesman Problem in a Plane," Mathematics of Operations Research, Volume 2, Number 3, 1977, pp. 209-224.

Knight, K. W. and J. P. Hofer, "Vehicle Scheduling with Timed and Connected Calls: A Case Study," Operational Research Quarterly, Volume 19, Number 3, pp. 299-310.

Krolak, P., Wayne Felts and James Nelson. "A Man-Machine Approach Toward Solving the Generalized Truck Dispatching Problem," Transportation Science, Volume 6, 1972, pp. 149-170.

Khumawala, Basheer M. "An Efficient Algorithm for the p-Median Problem with Maximum Distance Constraints," Geographical Analysis, Volume 5, Number 4, 1973, pp. 309-321.

Lam, Tenny N. "Comments on a Heuristic Algorithm for the Multiple Terminal Delivery Problem," Transportation Science, Volume 4, 1970, pp. 403-405.

Lea, A. C. "Location-Allocation Systems: An Annotated Bibliography," Discussion Paper number 13, Department of Geography, Toronto: University of Toronto, 1973.

Lentra, J. K. and A. H. G. Rinnooy Kan. "On the Expected Performance of Branch-and-Bound Algorithms," Operations Research, Volume 26, Number 2, 1978, pp. 347-349.

Levy, J. "An Extended Theorem for Location on a Network," Operational Research Quarterly, Volume 18, Number 4, 1967.

Lin, Shen. "Computer Solutions to the Traveling-Salesman Problem," Bell Systems Technical Journal, Volume 44, 1965, pp. 2245-2269.

Lin, S. and B. W. Kernighan. "An Effective Heuristic Algorithm for the Traveling-Salesman Problem," Operations Research, Volume 21, Number 2, 1973, pp. 498-516.

Little, J. D. C., K. G. Murty, D. W. Sweeney and C. Karel. "An Algorithm for the Traveling-Salesman Problem," Operations Research, Volume 11, 1963, pp. 979-989.

Matthaus, Fritz W. "Saving Approach to the Multiple Terminal Delivery Problem," Transportation Science, Volume 10, Number 2, 1976, pp. 216-221.

McDonald, J. J. "Vehicle Scheduling--A Case Study," Operational Research Quarterly, Volume 23, Number 4, 1972, pp. 433-444.

Miller, C. E., A. W. Tucker and R. A. Zemlin. "Integer Programming Formulations of Traveling-Salesman Problems," Journal of the ACM, Volume 7, 1960, pp. 326-329.

Mole, R. H. and S. R. Jameson. "A Sequential Route-building Algorithm Employing a Generalized Savings Criterion," Operational Research Quarterly, Volume 27, Number 2, 1976, pp. 503-511.

Narula, S. C., U. I. Ogbu and H. M. Samuelsson. "An Algorithm for the p-Median Problem," Operations Research, Volume 25, Number 4, 1977, pp. 709-713.

Norback, John P. and Robert F. Love. "Geometric Approaches to Solving the Traveling-Salesman Problem," Management Science, Volume 23, Number 11, 1977, pp. 1208-1223.

Papadimitriou, C. H. and K. Steiglitz. "Some Examples of Difficult Traveling Salesman Problems," Operations Research, Volume 26, Number 3, 1978, pp. 434-443.

ReVelle, Charles S. and Ralph W. Swain. "Central Facilities Location," Geographical Analysis, Volume 2, Number 1, 1970, pp. 30-42.

ReVelle, C., D. Marks and J. C. Liebman. "An Analysis of Private and Public Sector Location Models," Management Science, Volume 16, 1970, pp. 692-707.

Russell, Robert A. "An Effective Heuristic for the M-Tour Traveling-Salesman Problem with Some Side Conditions," Operations Research, Volume 25, Number 3, 1977, pp. 517-524.

Scott, Allen J. "Location-Allocation Systems: A Review," Geographical Analysis, Volume 2, Number 1, 1970, pp. 95-119.

Staha, R. L. "Constrained Optimization via Moving Exterior Truncations," Ph.D. Dissertation, University of Texas, Austin, Texas, May, 1973.

Svestka, Joseph A. and Vaughn E. Huckfeldt. "Computational Experience with an M-Salesman Traveling-Salesman Algorithm," Management Science, Volume 19, Number 7, 1973, pp. 790-799, Note and response: Management Science, Volume 22, Number 6, 1976, pp. 704-706.

Taha, Hamdy A. Operations Research, An Introduction, New York: The MacMillan Company, 1971.

Teitz, Michael B. and Polly Bert. "Heuristic Methods for Estimating the Generalized Vertex Median of a Weighted Graph," Operations Research, Volume 16, Number 5, 1968, pp. 955-961.

Tillman, Frank A. and Harold Cochran. "A Heuristic Approach for Solving the Delivery Problem," The Journal of Industrial Engineering, Volume 19, Number 7, 1968, pp. 354-358.

Tillman, Frank A. "The Multiple Terminal Delivery Problem with Probabilistic Demands," Transportation Science, Volume 3, Number 3, 1969, pp. 192-204. Note and Response: Transportation Science, Volume 4, 1970, pp. 226-237.

Tillman, Frank A. and Robert W. Hering. "A Study of a Look-Ahead Procedure for Solving the Multiterminal Delivery Problem," Transportation Research, Volume 5, Number 3, 1971, pp. 225-229.

Tillman, Frank A. and Thomas M. Cain. "An Upperbound Algorithm for the Single and Multiple Terminal Delivery Problem," Management Science, Volume 18, Number 11, 1972, pp. 664-682.

Turner, Wayne C., Pravhakar M. Ghare and Leslie R. Foulds. "Transportation Routing Problem--A Survey," AIIE Transactions, Volume 6, Number 4, 1974, pp. 288-301.

Webb, M. H. J. "Cost Functions in the Location of Depots for Multiple-Delivery Journeys," Operational Research Quarterly, Volume 19, Number 3, 1968, 311-320.

Webb, M. H. J. "Relative Performance of Some Sequential Methods of Planning Multiple Delivery Journeys," Operational Research Quarterly, Volume 23, Number 3, 1972, pp. 361-372. Note and response: Operational Research Quarterly, Volume 24, Number 2, 1973, pp. 305-309.

Webb, Michael. Transporting Goods by Road, London: London School of Economics and Political Science, Weidenfeld and Nicolson, 1972.

Wendell, Richard E. and Arthur P. Hurter. "Location Theory, Dominance, and Convexity," Operations Research, Volume 21, Number 1, 1973, pp. 314-320.

Wren, Anthony and Alan Holiday. "Computer Scheduling of Vehicles from One or More Depots to a Number of Delivery Points," Operational Research Quarterly, Volume 23, Number 3, 1972, pp. 333-344.

Yellow, P. C. "A Computational Modification of the Savings Method of Vehicle Scheduling," Operational Research Quarterly, Volume 21, Number 2, 1970, pp. 281-283.

APPENDIX I

THE TRAVELING SALESMAN PROBLEM

An attempt will not be made here to give a complete survey of the characteristics of the traveling salesman problem or of the successful methods used to solve it since it was found that this was not important to the main thrust of this research. A good survey of the literature up until 1968 is given in Bellmore and Nemhauser (1968), with the follow-up article of Akannad and Turban (1969), and a more detailed tutorial of the methods proposed up until 1971 appears in Eilon et al. (1971). This latter book also has a chapter on the estimation of traveling-salesman distances.

Mathematical formulations for the traveling-salesman problem can be found in Miller et al. (1960), Taha (1971) and Garfinkel and Nemhauser (1972). The most popular formulation is of the form

$$\text{minimize} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_{ij} = 1 \qquad j = 1,\ldots,n$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1,\ldots,n$$

$$x_{ij} = 0 \text{ or } 1$$

Plus constraints which insure that the solution is a Hamiltonian circuit through all of the points. The terms are

x_{ij} = 1 if the link between nodes i and j are in the solution and

c_{ij} = cost of link between nodes i and j (could be d_{ij}).

The constraints which insure a Hamiltonian circuit are quite numerous and render the problem unsolvable for very many nodes using standard integer programming techniques. Therefore different approaches have been proposed for the problem which do not consider these constraints explicitly.

Optimal Methods

The methods proposed so far that give optimal results are only good for problems no larger than about 40 or 50 nodes. Beyond this, computation time becomes impractically long. The most popular methods are those discussed in Held and Karp (1962), Little et al. (1963) and Bellmore and Malone (1971).

The Held and Karp method uses dynamic programming and suffers from the normal dynamic programming dimensionality problems.

Little et al. use a branch and bound strategy that builds tours by adding links according to a bounding strategy. The method is highly restricted in problem size but it has the advantages of being able to consider constraints other than those of the normal traveling-salesman problem and if the tree search must be stopped early a good solution will have probably already been found.

The other popular optimal method is the one surveyed and improved by Bellmore and Malone. This is also a branch and bound approach but in this case the bounding problem is the assignment problem that one can see is imbedded in the traveling-salesman formulation. If the assignment problem is solved and a complete Hamiltonian tour results, then the problem is solved. But usually the result is two or more subtours through subsets of the nodes. In this case a branch is formed for each link that can be deleted to block a subtour. This is done recursively until a complete tour is found whose length is less than any active lower bound.

This last approach is sometimes called the subtour-contraction or subtour-elimination technique. Bellmore and Malone reported the solution of an 80 node problem in just under three minutes using this method but they found that the run time increased at a polynomial rate making much larger problems impractical. This claim of polynomial run time came from impirical evidence but they also made theoretical arguments for it. Lenstra and Rinnooy Kan (1978) argue that this claim is not true and that a worst-case analysis would show an exponential run time.

Nonoptimal Methods

Since practical problems sometimes involve hundreds of nodes, and thus optimal methods are useless, much work has been directed toward efficient and effective heuristics which do not guarantee optimality. The most successful of these heuristics is one based on a work by Lin (1965) and improved by Lin and Kernighan (1973) and by Christofides and Eilon (1972). This algorithm essentially starts with a feasible tour (possibly randomly generated), makes feasible link exchanges which decrease the distance and stops when no more improvement can be made. This procedure can be repeated with different starting solutions as often as desired.

This method (called the r-optimal method because r, $2 \leq r \leq n$, links are exchanged) has produced solutions to 500 node problems in less than three minutes that were estimated to be 1.1 percent above the optimal. Other very good solutions were produced in just over three seconds for the same problems. To date this is the most successful approach especially since it many times gives optimal solutions to small problems. Papadimitriou and Steiglitz (1978) have shown a class of traveling-salesman problems which, using this r-optimal method, can only be solved to a local optimum with a very high cost. But they failed to state whether this type of problem was common in practice.

Recently Karp (1977) has shown that a certain class of nonoptimal traveling-salesman algorithms have a guaranteed upper limit on accuracy and run time. These algorithms divide the physical region of the problem into subregions and a traveling-salesman problem is solved in each subregion. Each of these subtours are then joined to make the final tour. He postulates that this method can be used to solve problems of many hundreds of nodes but no computational experience was given.

Another recent suggestion by Norback and Love (1977) which is not computationally superior does seem interesting. It assumes that all distances are straight-line, Euclidean distances between points. A "partial tour" is first made of the points forming the convex hull of all of the nodes. It is known that these points will be in the final tour in the order that they appear on the convex hull (Eilon et al., 1971). Interior points are then added to the tour according to rules that they give.

This method was used to solve a problem of 318 nodes but it took almost one-half hour of computer time. What is interesting about the approach is that small problems can be solved by hand giving very good solutions. Also, the approach does have an intuitive appeal.

APPENDIX II

THE VEHICLE ROUTING PROBLEM

The vehicle routing problem involves finding the routes for a set of vehicles from one or more depots or terminals to service a set of customers such that the total distance is minimized. Each customer has a given demand and each vehicle has a given capacity so it can only service a limited number of customers. The problem is essentially the same whether deliveries or pickups are being made.

Surveys of this type of problem can be found in Eilon et al. (1971), Turner et al. (1974), Golden et al. (1977). Different mathematical formulations of the problem can also be found in these references. These formulations minimize distance subject to constraints on the routing, capacity restrictions and demand requirements. But like the traveling-salesman formulation, the number of variables and constraints make problems of much size almost impossible to solve optimally. In fact, Christofides (1974) reported that the largest routing problem of any complexity that has been solved to known optimality had only 23 customers. Therefore only suboptimal heuristics have received extensive treatment in the literature. The successful of these methods will be reviewed.

M-Tour Traveling-Salesman Problem

A routing problem similar to this, yet uncapacitated, is known variously as the m-tour, multisalesman or multiple traveling-salesman problem. In this problem one has m salesmen or vehicles and it is desired to find m or fewer tours, each starting and ending at the same depot such that all customers are visited with minimum distance. Eilon et al. show that this problem is easily solved by transforming it to the standard traveling-salesman problem. This is done by replacing the depot in the distance matrix by m artificial depots all with the same physical location. Travel between artificial depots is prohibited by setting the distances between them to infinity.

Svestka and Huckfeldt (1973) found that the solution of the m-tour problem using a subtour-elimination approach (see Appendix II for the traveling-salesman problem) was easier if m was greater than one. Bellmore and Hong (1974) and Hong and Padberg (1977) have also investigated this problem and both papers present modified networks similar to the one mentioned above except that they include in certain cost elements the fixed charge associated with each additional salesman or vehicle. Their final solution not only gives the optimal tours but also the minimum number of tours less than or equal to m.

The m-tour approach has been successfully applied in solving the capacitated problem. This procedure essentially solves the uncapacitated m-tour problem and then if any capacity restrictions are violated, the violating tours are blocked and the problem is resolved. This continues until the stopping rules of the particular traveling-salesman algorithm are met.

As was pointed out in the traveling-salesman section of this paper, the r-optimal method has proven the most effective for solving the traveling-salesman problem. Thus it is not surprising that the most effective work in dealing with the capacitated m-tour problem has used the r-optimal method. This work was done by Christofides and Eilon (1969) and Russell (1977). Christofides and Eilon based their work on Lin (1965) and Russell based his on Lin and Kernighan (1973). For this reason Russell had better results but otherwise their approaches were quite similar. Russell's code, although slower than other methods to be described below, gave the best solution quality of any method he compared it to. The largest problem he solved had 159 customers and required just under nine minutes of computation time. He stated that the run time grew approximately as the number of customers raised to the 2.3 power.

Sweep Algorithm

Gillett and Miller (1974) have developed a novel and useful approach to the capacitated vehicle dispatch problem called the sweep algorithm. This was further generalized to more than one depot in Gillett and Johnson (1976).

For a single depot as the origin, the method first ranks the customers according to their polar coordinate angle. Then starting with one of the customers, new stops are added to a route according to the ranking as long as no capacity restrictions are violated. A traveling-salesman problem is then solved on each route. Later iterations switch customers between routes in attempts to reduce distances.

As long as the average number of stops remains constant, the execution time of the sweep algorithm increases linearly with the number of customers. But the computation time increases quadratically with the average number of stops per route. This happens because a traveling-salesman problem is solved on each route.

For this reason it was found that the sweep approach is a fast method for few stops per route but is not competitive with, for instance, Russell's code and the savings method for many stops per route. Sweep has solved a 250 customer problem with 25 routes in almost ten minutes. Also, the quality of sweep solutions are not quite that of the r-optimal, m-tour approach mentioned above.

Savings Approach

In 1964 Clarke and Wright proposed a heuristic for the vehicle dispatch problem called the savings method. The method, in its simplest form, starts with m single customer tours, where m is the number of customers. These tours each start at the depot, service the customer and then return to the depot. A tour with customer i is then joined with a tour containing customer j if the "savings" measure

$$s(i,j) = d(i,D) + d(D,j) - d(i,j)$$

is the maximum over all unassigned pairs. Here, D is the depot. This joining is of course only made if constraints on the tours are not violated.

There are variations of this referenced below and these references also give a more detailed description of the algorithm. But the general approach has received more attention and use than any other vehicle routing approach yet proposed. It has been used to solve the largest problems reported in the literature and the tour lengths are nearly always less than (many times much less than) 10 percent greater than the best known solutions to test problems.

Some variations and research on the approach will now be mentioned.

A criticism of the basic savings method is that once a customer is assigned to a route it is never considered for another route. One method to get around this is to check the effect of joining tours which do not have the maximum savings--to look ahead at tours with lower savings. This method has been investigated in Tillman and Cochran (1968), Tillman and Hering (1971), Tillman and Cain (1972) and Holmes and Parker (1976). These methods tend to get some improvement in solution quality but at the expense of much higher computation times.

Modifications of the savings function and of the solution approach were made in papers by Gaskell (1967), Yellow (1970), and Mole and Jameson (1976). These modifications have been well received in use and subsequent literature. Mole and Jameson solved a 225 customer problem in 160 seconds.

Webb (1972) tested the different savings functions suggested by Gaskell and found that none of them performed consistently as well as the original Clarke and Wright function. But McDonald (1972) found that intelligent use of the Gaskell functions as well as human interaction can produce exceptional results. Golden (1977) has shown that the basic savings approach can give very poor solutions in some cases. But he also pointed out that a poor solution is apparently a rarity. It should be mentioned that Webb solved problems of 1000 customers in 5 minutes using the original savings method. He found that the run time varied as the number of customers raised to the 1.6 power.

Two routing algorithms based on the savings method that were actually implemented were reported by Beltrami and Bodin (1974) and Golden et al. (1977). Beltrami and Bodin improved the savings solution by solving a traveling-salesman problem on each route. Golden et al. extended Beltrami and Bodin's work by including modifications suggested by Gaskell. But the work of Golden et al. was distinguished by the fact that two tours were only joined if they were close together and by the data structures used to store the basic data and the sorted savings list. This enabled them to solve very large problems very quickly. An actual 600 customer problem was solved in 20 seconds.

Further discussion of the problem of handling the savings list and other computational problems associated with this approach can be found in Webb (1972).

The savings approach has also been applied to the multiple terminal problem in Tillman (1969), Lam (1970), Matthaus (1976), Tillman and Cain (1972) and Golden et al. (1977).

Other Methods

There has been other published work on the vehicle dispatch problem that does not fit into the above categories. The three papers to be mentioned do have certain aspects in common, though. Two have routines that cluster customers according to proximity to each other and then try to form routes that meet capacity constraints from the clusters. All try to improve initial solutions by switching customers between routes by very simple procedures. And all handle multiple depot problems. None of these approaches offer much in speed of solution, but all give very acceptable solutions.

Krolak et al. (1972) have presented a man-machine approach to the problem that has the advantage of involving an experienced human controlling the solution. But it has the disadvantage of requiring complicated computer interaction and display apparatus.

Cassidy and Bennett's (1972) work is distinguished by the data structures used to represent route linkages. These linkages are easily modified to test new routes. Their method was successfully implemented in an actual system with 390 customers.

Wren and Holliday (1972) had very good results by just exhaustively switching customers between routes of an initial solution which is either supplied by the user or created by the program.

Computational Experience

Webb (1972) discussed and Matthaus (1976) implemented a clever way of finding the maximum savings which avoided exhaustive list searching in the savings method. Gaskell (1967) and later Golden et al. (1977) also improved the savings method by introducing a "route-shaper" parameter into the savings function. These modifications were incorporated into a code which was tested on 10 problems from several sources that have been published in Eilon et al. (1971). These have been very popular test problems so comparisons can be made among all of the popular codes, including the modified savings code. The results are in Table II.1.

The times are hard to compare because several different computers were used but the modified savings code was very fast. And since the new code is a quick heuristic the solution distances are within a tolerable range of the others. In fact it is equal to or better than the basic savings method and in some cases meets or beats the distances of more exhaustive methods.

Multiple-Depot Methods

The following papers were discussed above but are listed here as methods for handling more than one depot: Tillman (1969), Lam (1970), Tillman and Cain (1972), Cassidy and Bennett (1972), Krolak et al. (1972), Wren and Holliday (1972), Matthaus (1976), Gillett and Johnson (1976) and Golden et al. (1977).

Since the Matthaus savings algorithm proved efficient for the single-depot scheduling, it was coded for comparison with other published results of multiple-depot algorithms. Matthaus failed to make any such comparison. Krolak et al. and Gillett and Johnson published the data of some of their test problems and the results of the codes appear in Table II.2.

TABLE II.1

COMPUTATIONAL RESULTS OF FIVE SINGLE-TERMINAL VEHICLE SCHEDULING CODES

Problem	Number of Stops	Christofides & Eilon (1972). 3-optimal IBM 7090		Gillett and Miller (1974). SWEEP IBM 360/67		Russell (1977). M-tour with r-optimal IBM 370/168		Eilon, et al. (1971). Basic savings method IBM 7090		Modified Savings. IBM 370/158	
1	6	114*	.3	-	-	-	-	119	.1	119	.00075
2	13	290	.3	-	-	-	-	290	.1	290	.00318
3	21	585	1.8	591	.21	585	.063	598	.1	598	.01583
4	22	949	1.5	956	.17	956	.159	955	.1	956	.02293
5	29	875	2.4	875	.51	875	.255	963	.2	945	.03837
6	30	1414	2.5	-	-	1214	.310	1427	.2	1263	.02487
7	32	810	2.4	810	.62	810	.328	839	.2	831	.04903
8	50	556	6.0	546	2.0	524	.264	585	.6	583	.13347
9	75	876	12.0	865	1.23	854	4.098	900	1.3	893	.38865
10	100	863	30.0	862	6.0	833	1.690	887	2.5	886	.6836

*The first column is the best distance found for the route and the second column is computation time in minutes.

TABLE II.2

COMPUTATIONAL RESULTS OF THREE MULTIPLE-TERMINAL VEHICLE SCHEDULING CODES

Problem	Number of stops	Number of depots	Reported time (sec.)	Reported distance	Savings time (sec.) (IBM 370/158)	Savings distance
Krolak, *et al.** (XDS Sigma 7)						
26	91	9	148.44	27,771	18.59	32,446
28	91	9	110.46	26,188	22.42	32,898
Gillett and Johnson (IBM 370/168)						
8	249	2	457.61	4,832	69.71	4,606
9	249	3	444.33	4,220	107.24	4,251
10	249	4	344.04	3,822	153.56	4,072
11	249	5	315.99	3,754	216.60	3,972

*The distances were improved in the man-machine phase of the Krolak, *et al.* solution.

With different computers the times are hard to compare but the saving method is very fast. Some apparent trends can be seen: The savings code increases in time with more terminals while holding the number of customers constant. The opposite is true with the SWEEP algorithm. Also, it is very apparent that the quality of solution of the savings code degenerates as the number of terminals increases. But with 2 or 3 terminals the solution was very good.

This algorithm appears at the end of this appendix.

ALGORITHM FOR MULTIPLE-TERMINAL VEHICLE DISPATCHING

The data:

M = number of stops

NTER = number of terminals

D(I,J) = distance from stop I to step J.

DT(I,K) = distance from stop I to terminal K.

The algorithm variables:

A = list of quantities calculated in algorithm to speed comparisons of savings.

ASTER(I) = terminal assigned to stop I

IA = list of stops associated with the values in A

IK1,IK2 = the indices of A associated with stops K1 and K2

J1,J2 = temporary variables for stops being considered for linking

K1,K2 = the best stops found for linking in one iteration

K3 = best terminal found for linking in one iteration

KB = temporary variable for terminal being considered for linking

LINK = logical variable indicating whether any stops were linked on a route

NRTER(I) = the nearest terminal to stop I

OEND(I) = if I is the beginning or end stop of a route OEND is the stop on the opposite end of the route

S = savings

S0 = best savings found in an iteration

The algorithm:

1. Initialize data structures such that each stop is a single route.
2. Assign each stop to its closest terminal, thus initializing ASTER and NRTER.
3. A(I) = DT(I,ASTER(I)), IA(I) = I, for I = 1,...,M. Sort A and IA in descending order of A.
4. S0 = a very small number

   LINK = FALSE

   I = 1
5. J1 = IA(I), J1 is the stop associated with index I.
6. If A(I) + A(I+1) $\leq$ S0 then go to step 19, there cannot be an improvement.
7. J = I + 1
8. J2 = IA(J), J2 is the stop associated with index J.
9. If A(I) + A(J) $\leq$ S0 then go to step 18, there cannot be an improvement.
10. Only endpoints can be linked. If J2 is not an endpoint go to step 17.
11. If J1 and J2 are on the same route go to step 17, they cannot be linked.
12. Find KB which is the value of K which minimizes B = DT(OEND(J1),K) - DT(OEND(J1), NRTER(OEND(J1))) + DT(OEND(J2),K) - DT(OEND(J2), NRTER(OEND(J2))) for K = 1,...,NTER
13. If joining J1 and J2 to terminal KB will exceed distance constraint, go to step 17.

14. Calculate savings: S = A(I) + A(J) - B - D(J1,J2)

15. If S $\leq$ S0 then go to step 17.

16. Save values: S0 = S, K1 = J1, K2 = J2, K3 = KB, IK1 = I, IK2 = J, LINK = TRUE

17. If J = M then go to step 18, otherwise, J = J + 1 and go to step 8.

18. If I = M - 1 then go to step 19, otherwise, I = I + 1 and go to step 5.

19. If LINK = FALSE, then <u>STOP</u>.

20. Update data structures such that the route containing K1 and K2 are linked at K1 and K2 and the new route is assigned to terminal K3.

21. If step K1 is not the first or last stop of the new route, then it is no longer considered for linking and A(IK1) is set to a very small number. Similarly, if K2 is not at the end of the route, A(IK2) is set equal to a very small number.

22. Let END1 and END2 be the endpoints of the new route and J1 and J2 be the index of IA associated with END1 and END2. Then

```
A(J1) = DT(END1,ASTER(END1))
        + DT(OEND(END1), ASTER(OEND(END1)))
        + DT(OEND(END1), NRTER(OEND(END1)))
A(J2) = DT(END2,ASTER(END2))
        + DT(OEND(END2), ASTER(OEND(END2)))
        + DT(OEND(END2), NRTER(OEND(END2)))
```

Sort A and AI in descending order of A and go to step 4.

APPENDIX III

THE p-MEDIAN PROBLEM

Given a network containing n nodes (facility sites), it is desired to find p of those nodes which could service the remaining nodes with a minimum total distance traveled. This is the p-median problem which can be modeled as

$$\text{minimize} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} a_i d_{ij} x_{ij}$$

$$\text{subject to} \quad \sum_{j=1}^{n} x_{ij} = 1, \; i = 1,\ldots,n$$

$$\sum_{j=1}^{n} x_{jj} = p$$

$$x_{ij} \leq x_{jj}, \; x_{ij} = 0 \text{ or } 1, \; i,j = 1,\ldots,n$$

where: x_{ij} = 1 if node i is assigned to node j,

a_i = demand of node i,

d_{ij} = distance from node i to node j.

Revelle and Swain (1970) found that in almost every instance, if the 0-1 constraint is dropped and the problem is solved as a standard linear programming problem, the solution is integer. This makes the problem easily solved if its size does not exceed computer time and storage availability.

Teitz and Bart (1968) and Khumawala (1973) have both suggested heuristics for the p-median problem. The Teitz and Bart approach essentially interchanges nodes of a starting solution in an intelligent way until no improvement can be made. The Khumawala method constructs a solution by examining savings made by adding nodes to the solution. Both methods result in good solutions in reasonably low computer time, but neither guarantee optimality.

Narula et al. (1977) solved the problem using Lagrangian relaxation and subgradients. The procedure is simple, converges in an acceptable computation time, and was found to be optimal in every case that was tried. This solution method was used in this research.

The Lagrangian relaxation of the problem with respect to the first constraint and for a given nonnegative n-vector λ is

$$L_d(\lambda) = \min_x \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i d_{ij} - \lambda_i) x_{ij} + \sum_{i=1}^{n} \lambda_i$$

$$\text{subject to} \quad \sum_{j=1}^{n} x_{jj} = p$$

$$x_{ij} \le x_{jj},\ x_{ij} = 0 \text{ or } 1,\ i,j = 1,\ldots,n.$$

The Lagrangian dual of this is

$$\max_{\lambda \ge 0} L_d(\lambda)$$

and subgradient optimization (Held et al., 1974) is used to solve this dual problem.

APPENDIX IV

EXPECTED DISTANCE OF A TOUR

If one is given a set of probabilities or probability density functions describing the probable route of a service vehicle, it seems that a term could be found for the expected distance of a tour, given the new facility location. This term could then be minimized, with respect to the new facility location, to solve the problem. The following development of the term will show this particular approach to be impractical.

Define

$p_j(D,i_1,\ldots,i_r)$ = probability that customer j will be the first stop after the partial tour $(D,i_1,\ldots,i_r)$.

$I = \{1,2,\ldots,m\}$ where m is the number of customers.

Also assume that

$$\sum_{j\varepsilon I-\{i_1,\ldots,i_r\}} p_j(D,i_1,\ldots,i_r) = 1$$

for all admissible partial tours $(D,i_1,\ldots,i_r)$.

To simplify matters let m=4 and consider the network in Figure IV.1 as a visualization of how all possible tours might be found. One starts at the depot, D, and there are several possible paths that might be followed. The numbered nodes represent the customers. Each open, solid arrow represents a return to the depot. The dashed arrows represent parts of the network that were excluded to make room. (Capacity restrictions can be accounted for by shortening the network such that the maximum number of customers visited in a trip is not more than the maximum expected number of customers that could be serviced in one trip.)

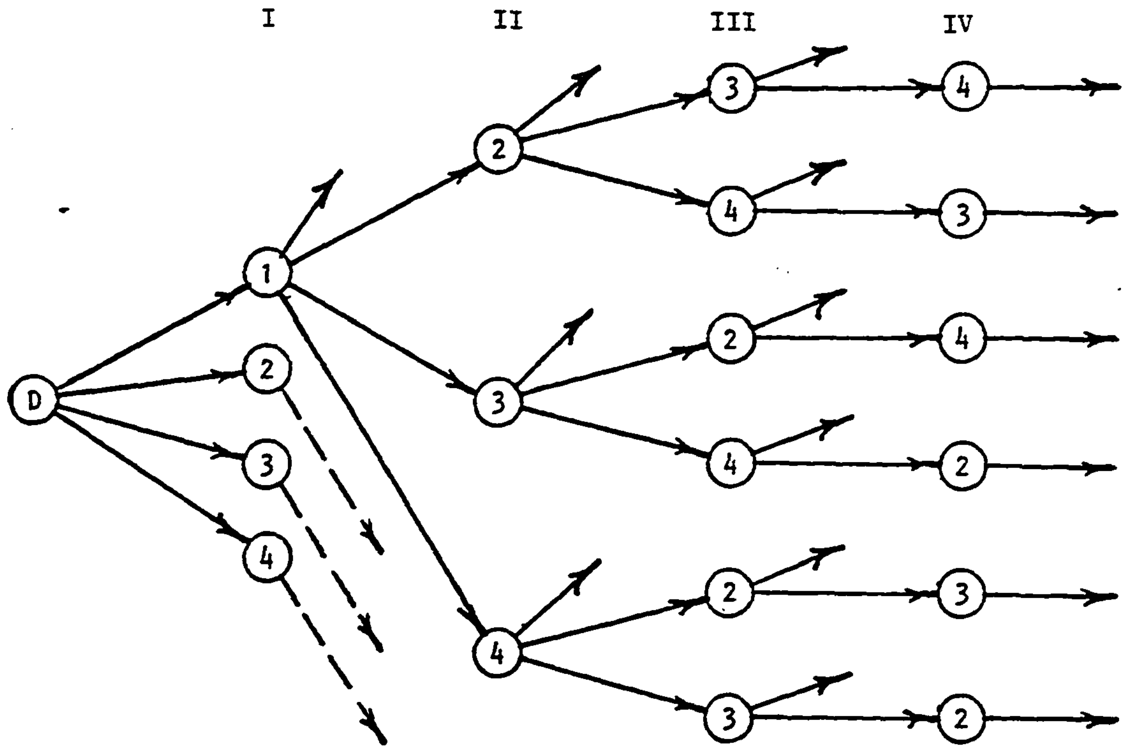


Figure IV.1. Network showing all possible tours through four customers.

To develop the expected distance of a tour, first note that (referring to Figure IV.1) the index notation is i_1 being the node visited at level I, i_2 the node visited at level II, etc.

Now start at level III of the network. At any node of this level there are two routes that can be taken: back to the depot or to the next (and last) level. The probabilities of these are $p_D(D,i_1,i_2,i_3)$ and $p_{I'}(D,i_1,i_2,i_3)$ where

$$I' = \{1,2,3,4\} - \{i_1,i_2,i_3\}$$
$$= \text{node visited at level IV.}$$

Thus the expected remaining distance at any node in level III is

$$p_{I'}(D,i_1,i_2,i_3)(d_{i_3I'} + d_{I'D}) + p_D(D,i_1,i_2,i_3)d_{i_3D}.$$

At any node i_2 of level II the argument is similar: The tour can either go to one of the nodes in level III or back to the depot. The expected remaining distance, using the term from level III, is

$$\sum_{i_3 \varepsilon I-\{i_1,i_2\}} p_{i_3}(D,i_1,i_2)\ [d_{i_2i_3}+p_{I'}(D,i_1,i_2,i_3)(d_{i_3I'}+d_{I'D})$$
$$+ p_D(D,i_1,i_2,i_3)d_{i_3D}] + p_D(D,i_1,i_2)d_{i_2D}$$

This analysis can be backed to the depot D in the network and the full expression is

$$ET = \sum_{i_1 \varepsilon I} p_{i_1}(D)[d_{Di_1} + \sum_{i_2 \varepsilon I-\{i_1\}} p_{i_2}(D,i_1)[d_{i_1i_2} +$$
$$\sum_{i_3 \varepsilon I-\{i_1,i_2\}} p_{i_3}(D,i_1,i_2)[d_{i_2i_3}+p_{I'}(D,i_1,i_2,i_3)(d_{i_3I'}+d_{I'D}) +$$
$$p_D(d,i_1,i_2,i_3)d_{i_3D}] + p_D(D,i_1,i_2)d_{i_2D}] + p_D(D,i_1)d_{i_1D}]$$

Since in minimization each term involving distances between customers is constant, then these terms can be left out. Rearranging the terms of ET, the expression ET' is

$$ET' = \sum_{i_1 \varepsilon I} p_{i_1}(D) d_{Di_1}$$

$$+ \sum_{i_1 \varepsilon I} \sum_{i_2 \varepsilon I-\{i_1\}} \sum_{i_3 \varepsilon I-\{i_1,i_2\}} p_{i_1}(D) p_{i_2}(D,i_1) p_{i_3}(D,i_1,i_2) p_{I'}(D,i_1,i_2,i_3) d_{I'D}$$

$$+ \sum_{i_1 \varepsilon I} \sum_{i_2 \varepsilon I-\{i_1\}} \sum_{i_3 \varepsilon I-\{i_1,i_2\}} p_{i_1}(D) p_{i_2}(D,i_1) p_{i_3}(D,i_1,i_2) p_D(D,i_1,i_2,i_3) d_{i_3 D}$$

$$+ \sum_{i_1 \varepsilon I} \sum_{i_2 \varepsilon I-\{i_1\}} p_{i_1}(D) p_{i_2}(D,i_1) p_D(D,i_1,i_2) d_{i_2 D}$$

$$+ \sum_{i_1 \varepsilon I} p_{i_1}(D) p_D(D,i_1) d_{i_1 D}$$

Note that each line ends with a distance between a customer and the depot. So it can be seen that for any given distance between the depot and a customer j the coefficient is the sum of:

1. The probability of going from the depot to customer j first.
2. The product of the probabilities of the links in each route serving 4 (m) customers, the last customer being j.
3. The product of the probabilities of the links in each route serving 3 customers, the last customer being j, and then returning to the depot.
4. The product of the probabilities of the links in each route serving 2 customers, the last being j, and then returning to the depot.

5. The product of the probability of serving customer j first and the probability of then returning to the depot.

From this it should be clear that the coefficient of any d_{jD} is the sum of the probability of visiting j first plus the sum of the products of the link probabilities of every possible tour that ends with customer j. The expression for this in general is very long and will be omitted here.

If each of these coefficients is interpreted as a weight, w(i), on the distance between a customer i and the depot, then this problem can be solved as a single-facility Weber problem (Francis and White, 1974).

The concept of this approach is very simple but the computation associated with calculating the product of probabilities of every possible route renders the method in its most general form impossible for a realistic problem. Most likely a large number of the routes would have zero probability and could be overlooked but a large amount of computation would probably remain.

There is still the problem of finding the probabilities. Each would be based on two things: the probability that a customer has a demand in a given time period and the probability of taking a certain route. This second probability might come from an idea mentioned by Knight and Hofer (1968). They said that since it appears that each successive stop in a traveling-salesman tour is usually the closest one, then it would probably be possible to get a distribution on the probability of visiting the kth closest customer each time a trip is made between stops.

But this brings up another problem. If one uses this approach there is no way of getting the probabilities of visiting the first city after leaving the depot if it is not known where the depot is located. This might be remedied by iteratively solving the problem with several different starting solutions. A starting solution would dictate some probabilities, then the problem could be solved to find a new location which would result in new probabilities, etc., until there was no change in location.

Again, even if a convenient method could be found for calculating the probabilities, there is still the problem of calculating all of the combinations of probabilities. A Markovian transition matrix might be considered with the states being the depot and customers. Then by finding all r-step transition matrices, $r = 1,\ldots,m$, the desire products of probabilities would be found. This is assuming that at a given stop it did not matter which customers had been serviced previously on the tour--which is a flaw to the Markovian approach since, except under special assumptions, one would not want to visit the same customer more than once in a tour.

Use of the expression for ET in solving this problem does seem to have intuitive appeal, but its practical application does not seem obvious.