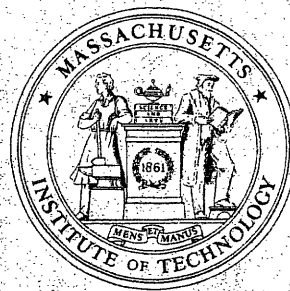# OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

THE CONGESTED MEDIAN PROBLEM

by

Oded Berman

and

Richard C. Larson

OR 076-78   July 1978

Table of Contents

·Abstract

The median problem has been generalized to include queueing-like congestion of facilities (which are assumed to have finite numbers of servers). In one statement of the problem, a closest <u>available</u> server is assumed to handle each service request. More general server assignment policies are allowed, however. The analysis requires keeping track of the states (available or unavailable) of all servers. Paralleling the standard deterministic median problem, the objective is to minimize the expected travel time associated with a random service request, weighted appropriately by the equilibrium state probabilities of the system. Under suitable conditions, it is shown that at least one set of optimal locations exists solely on the nodes of the network. This analysis ties together previously disparate efforts in network analysis and spatial queueing analysis.

Introduction

The problem of where to locate a set of facilities on a network so as to minimize the expected travel time to or from the facilities, for the population of their users, is one of the classic problems in location theory. This problem, known in the literature as the <u>median problem</u>, has been studied very thoroughly in the last two decades. The basic theoretical results in this area are due to Hakimi [3,4]. Subsequently, Goldman [2], Hakimi and Naheshwari [5], Levy [11] and Wendell and Hurter [12] have extended and generalized Hakimi's results.

When there are Q facilities to be located on a network G, the median problem is to find a set of Q points on G denoted as $Z^* = (Z_1^*, Z_2^*, \ldots, Z_Q^*)$, such that

$$\sum_{j=1}^{n} h_j d(Z^*,j) \leq \sum_{j=1}^{n} h_j d(Z,j) \; \forall \; Z \in G \qquad (1)$$

where h_j is the fraction of demand that is generated at node j ($\sum_{j=1}^{n} h_j = 1$), n is the number of demand points and $d(Z,j)$ is the shortest distance from node j to the closest point in the set Z. In [4] Hakimi proved that at least one set Z^* exists solely on the nodes of the network.

When considering the standard median problem for applications, four main assumptions are implied.

    (1)   Travel in the given area is restricted to take place solely along the links of the transportation network.

    (2)   Requests for service can occur only at a finite number of points – the nodes of the network.

(3) When the number of facilities is greater than one, a
service request from a particular location is always
handled by a server at a closest facility.

(4) There is always an available (free) server at the selected
(closest) facility.

The traveling associated with a service request could require the
"customer" (requester of service) to travel to a nearest facility or
a server at a nearest facility to travel to the customer. The former,
"customer-to-server" type system, includes outpatient clinics, "little
city halls," libraries, and even hamburger havens. The latter, "server-
to-customer" type system, includes emergency services (e.g., police,
fire, ambulance, emergency repair), special-order delivery services, and
certain home visitation medical services. In our work, we use the term
"travel time associated with a service request" to mean either the cus-
tomer-to-server or server-to-customer travel time.

The type of systems we consider are characterized by stochastically
generated requests for service (in time and space) and by nondeterminis-
tic service times for the service requests. [A service time is comprised
of travel time plus on-scene time.] In such an environment, it is often
likely that all servers at a nearest facility will be busy, thereby
yielding a congested network in which queues could form. Thus, assump-
tion (4) above often does not hold in practice. For these systems,
equation (1) is merely the problem of finding a set of points so as to
minimize the expected travel time for a random service request at very
special times, namely when servers are available at all facilities.

Since this is often not the case, it is the purpose of this work to

incorporate in the context of the median problem the possibility that

all servers at any subset of the Q facilities can be busy.

The objective function in this congested median problem is to

minimize the expected travel time associated with a random service

request weighted appropriately by the equilibrium state probabilities

of the system. Here "states" of the system are defined according to

the status of each of the facilities - at least one server available

at the facility or all servers busy. To avoid queue formation wherever

possible, we assume that the server that handles a service request is

a most preferred available server. Usually server preferences are

dependent solely on geographical proximity, but more general server

assignment policies are allowed. The basic result obtained is that

under fairly general assumptions at least one set of optimal locations

exists on the nodes of the network. This parallels the results of

Hakimi [3,4].

The analysis also ties together previously disparate research

efforts on network analysis and on spatial queueing analysis. In

particular we show that the hypercube model [8,9] and the algorithm of

Jarvis [7] on optimum locations can be useful to solve the congested

median problem for specific situations. In addition this work indicates

that the basic hypercube model does not suffer from a loss of general-

ity by considering only nodes (or atoms) for the locations of the

service units.

Notations and Assumptions

Let $G(N,L)$ be a network where N is the set of nodes with $|N| = n$, and L is the set of the links. Let X_Q be the set of all possible locations of Q facilities $(Q > 1)$, on the network G, i.e.,

$$\bar{X}_Q = \{X_Q = (i_1,\dots,i_Q) \; ; \; i_K \in G \quad K = 1,\dots,Q\}$$

Given any location $X_Q = (i_1,\dots,i_Q) \in \bar{X}_Q$, let $\bar{i}_K$ denote that the facility at i_K is not staffed with an available server (the facility is busy) and $\hat{i}_K$ that the facility at i_K does have an available server. Therefore, for any $X_Q \in \bar{X}_Q$ there are 2^Q combinations (states) of finding the network at any time, according to the status of the Q facilities. Let $Y_{X(Q)}$ be the set of all states for $X_Q \in \bar{X}_Q$ and let $Y_{X(Q)}$ (or for convenience y_Q) be a generic element of $Y_{X(Q)}$.

We assume that server assignment occurs according to a <u>fixed pre-ference</u> procedure. That is, for each demand point in the network there is a list of facilities that specifies the ordering of preferences for the assignment of servers (i.e., first preference for servers from facility i, second preference for servers from facility j, etc.) A most preferred <u>available</u> server is always assigned to a customer.* The goal

---

*When preferences depend directly on travel times, such a zero-look-ahead strategy is very unreasonable, but not always optimal in the sense of minimizing time-average mean travel time. An optimal policy occasionally requires assignment of other than the most preferred available server [7], in order to leave the system in a state which best anticipates future service requests. We do not consider such strategies in our formulation of the congested median problem.

of the optimization to be stated below is to minimize expected system
travel time under a given fixed preference procedure. The fixed prefer-
ence procedure itself need not be determined solely by relative travel
time, but can include characteristics of servers (e.g., bilingualness)
and needs of customers at the nodes.

Let $t(i,j)$ be the travel time on link (i,j), $(i,j) \in L$, and let
$d(y_Q,j)$ be the (minimum) travel time associated with a most preferred
available server to node j, when the system is in state y_Q.

As in the standard median problem we assume that service requests
are generated on the nodes of the network. However, in addition, we
assume that service requests occur according to a general renewal process,
with each request requiring a service time whose distribution is general
and not dependent on the identity of the server or the history of the
system. Thus variations in the service times that are due solely to
variations in travel times among potential servers are ignored. This
assumption is reasonable for systems having on-scene service times roughly
an order of magnitude greater than travel times.

Finally, we require that travel time is uniform over a link, i.e.,
the travel time over a fraction θ of some link (p,q) is $\theta t(p,q)$. This
assumption is not restrictive since the links and nodes can be defined in
such a way that this assumption holds to a specified degree of accuracy.

Model Formulation and Analysis

We will consider the steady state behavior of the system. For any possible set of locations $X_Q \varepsilon \overline{X}_Q$, let $P(y_Q)$ be the steady state probability that the network is in state $y_Q \varepsilon Y_{X(Q)}$. (We assume that the appropriate ergodicity conditions apply so that a unique steady state distribution exists.) Let y_Q^o be the state in which all the Q facilities are busy (i.e., $y_Q^o = (\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_Q)$ in our notation).

Conditioned on any state $y_Q \varepsilon Y_{X(Q)} - \{y_Q^o\}$, the expression

$$\sum_{j=1}^{n} h_j d(y_Q, j)$$

is the expected travel time associated with a random service request.

Suppose now that the network is in state y_Q^o. We will consider three policies regarding this state:

(a) Service requests that occur while all the service units are busy, are handled by a back-up service system (zero-line capacity case). Let R be the travel time cost of utilizing this special reserve server.

(b) Service requests that arrive while all the facilities are busy enter an infinite capacity queue that is depleted in a first-come, first-served manner; upon completion of service, the server is either assigned to the next request waiting in queue, or returns immediately home if none is waiting. Therefore,

$$\sum_{k=1}^{n} \sum_{j=1}^{n} h_k h_j d(i_k, j)$$

is the expected travel time of a random service request given that the network is in state y_Q^o.

(c)   Again service requests that arrive while all the facilities
      are busy enter a FCFS queue with infinite capacity, but now
      upon completion of service, the server always first returns
      to his/her home location. In this case the conditional expec-
      ted travel time of a random service request is

$$\sum_{k=1}^{n} \sum_{j=1}^{n} \frac{1}{Q} h_j d(i_k, j),$$

      given that the network is in state y_Q^o.

The appropriateness of any particular assumption depends of course on the

system being modeled. Assumption (a) often applies to ambulance systems,

in which emergency requests cannot be queued. Assumption (b) applies

frequently to police vehicles that may be dispatched back-to-back to suc-

cessive service requests. Assumption (c) applies to some ambulance and

fire services. The congested median problem is now stated:

$$\min_{X_Q \varepsilon \underline{X}_Q} F(X_Q) \tag{2}$$

with

$$F(X_Q) = \sum_{y_Q \varepsilon Y_{X(Q)} - \{y_Q^o\}} P(y_Q) \sum_{j=1}^{n} h_j d(y_Q, j) + P(y_Q^o) \sum_{j=1}^{n} h_j C(j)$$

where

$$C(j) \text{ is R or } \sum_{k=1}^{n} h_k d(k,j) \text{ or } \sum_{k=1}^{Q} \frac{1}{Q} d(i_k, j)$$

according respectively to (a), (b) or (c) above.

Obviously, the standard median problem is a special case of (2) aris-

ing when $P(y_Q) = 0, \forall \ y_Q \neq (\hat{i}_1, \ldots, \hat{i}_Q)$ - the state where all the units are

available $\underline{and}$ when $d(y_Q, j)$ is determined solely by geographic proximity

(i.e., minimizing travel time). The weights $P(y_Q)$ in (2) represent the fraction of time that the network is in each of the 2^Q possible states. Therefore, as noted before, we take into account that any subset of facilities can become depleted of servers.

Now the following important theorem can be proved.

<u>Theorem 1</u> For a <u>given</u> fixed preference server assignment procedure, at least one set of optimal solutions to (2) exists on the nodes of the network.

Proof: Let $X_Q^* = (i_1, i_2, \ldots, i_S, \ldots, i_Q)$ be the optimal solution to (2), and let $P(y_Q)$, $\forall\ y_Q\ \varepsilon\ Y_{X^*(Q)}$ be the corresponding steady state probabilities. Suppose that i_S is an interior point on the link (p,q). Then by the uniform speed assumption

$$\frac{t(p, i_S)}{t(p,q)} = Q \qquad 0 < Q < 1 \qquad (3)$$

The following proof is for the case $C(j) = \sum_{k=1}^{Q} \frac{1}{Q} d(i_k, j)$ in (2). The proofs for the other two cases are very similar and even slightly easier.

Let $Y_{i_S} \subset Y_{X^*(Q)} - \{y_Q^o\}$ be the set of all states in which the facility located at i_S is available. Then we can write $F(X_Q^*)$ as:

$$F(X_Q^*) = \sum_{y_Q \varepsilon Y_{i_S}} P(y_Q) \sum_{j=1}^{n} h_j d(y_Q, j) + P(y_Q^o) [\sum_{j=1}^{n} \frac{1}{Q} h_j d(i_S, j)] + A \qquad (4)$$

where the term A includes all server assignments that must exclude the facility located at i_S, i.e.,

$$A = \sum_{y_Q \in Y_{X^*(Q)} - Y_{i_S} - \{y_Q^o\}} P(y_Q) \sum_{j=1}^{n} h_j d(y_Q, j) + P(y_Q^o) [\sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq s}}^{Q} \frac{1}{Q} h_j d(i_k, j)] . \qquad (5)$$

Let $N_{y_Q}(i_S)$ be the set of all nodes that would be assigned to the server from node i_S, when the network is in state $y_Q \in Y_{i_S}$ and let $\overline{N}_{y_Q}(i_S) = N - N_{y_Q}(i_S)$. Therefore we can rewrite (4) as:

$$F(x^*) = \sum_{Q} \sum_{y_Q \in Y_{i_S}} P(y_Q) [\sum_{j \in N_{y_Q}(i_S)} h_j d(i_S, j)] + P(y_Q^o) [\sum_{j=1}^{n} \frac{1}{Q} h_j d(i_S, j)] + A + B \qquad (6)$$

where the term B corresponds to non-queued assignment of servers not located at i_S, even when the facility located at i_S is available, i.e.,

$$B = \sum_{y_Q \in Y_{i_S}} P(y_Q) \sum_{j \in \overline{N}_{y_Q}(i_S)} h_j d(y_Q, j) \qquad (7)$$

Recalling that i_S is assumed to be an interior point on the link (p,q), let $N_{y_Q}(i_S, p) \subset N_{y_Q}(i_S)$ be the set of all nodes that belong to the set $N_{y_Q}(i_S)$ and which communicate most efficiently with the facility at i_S via p, and let $N_{y_Q}(i_S, q) = N_{y_Q}(i_S) = N_{y_Q}(i_S, p)$. (The term "communicate" implies minimal travel time.) If a node communicates equally efficiently with i_S via nodes p or q for some y_Q, we can include that node in either $N_{y_Q}(i_S, p)$ or $N_{y_Q}(i_S, q)$, but not in both.

Let $N(i_S, p)$ be the set of all nodes which communicate most efficiently with the facility at i_S via node p and let $N(i_S, q) = N - N(i_S, p)$.

Therefore we can write (6) as

$$F(X_Q^*) = \sum_{y_Q \varepsilon Y_{i_S}} P(y_Q) \; [\sum_{j \varepsilon N_{y_Q}(i_S,p)} h_j(d(j,p) + t(p,i_S))$$

$$+ \sum_{j \varepsilon N_{y_Q}(i_S,q)} h_j(d(j,q) + t(q,i_S)) \;] + P(y_Q^o) \; [\sum_{\overline{Q} \; j \varepsilon N(i_S,p)} h_j(d(j,p)$$

$$+ t(p,i_S)) + \sum_{j \varepsilon N(d_S,q)} h_j(d(j,q) + t(q,i_S)) \;] + A + B. \tag{8}$$

Using (3) and rearranging terms we get

$$F(X_Q^*) = \theta[t(p,q) (\sum_{y_Q \varepsilon Y_{i_S}} P(y_Q) \sum_{j \varepsilon N_{y_Q}(i_S,p)} h_j + P(y_Q^o) \sum_{\overline{Q} \; j \varepsilon N(i_S,p)} h_j) \;]$$

$$+ (1-\theta) \; [t(p,q) (\sum_{y_Q \varepsilon Y_{i_S}} P(y_Q) \sum_{j \varepsilon N_{y_Q}(i_S,q)} h_j + P(y_Q^o) \sum_{\overline{Q} \; j \varepsilon N(i_S,y)} h_j) \;]$$

$$+ A + B + C \tag{9}$$

where the term C corresponds to "fixed components" of travel time to the

link (p,q), where

$$C = \sum_{y_Q \varepsilon Y_{i_S}} P(y_Q) \; [\sum_{j \varepsilon N_{y_Q}(i_s,p)} h_j d(j,p) + \sum_{j \varepsilon N_{y_Q}(i_s,q)} h_j d(j,q) \;]$$

$$+ \; P(y_Q^o) \; [\; \sum_{j \in N_{y_Q}(i_S,p)} h_j d(j,p) \; + \; \sum_{j \in N_{y_Q}(i_S,q)} h_j d(j,q) \;] \tag{10}$$

Assuming a fixed server assignment policy, once the "route-partitioning" sets $N_{y_Q}(i_S,p)$, $N(i_S,p)$, $N_{y_Q}(i_S,q)$, and $N(i_S,q)$ are specified, A, B, and C are independent of θ. Thus, $F(X_Q^*)$ is a linear function of θ implying its minimum occurs at an extreme point, either $\theta = 0$ or 1, corresponding to location at node p or q, respectively. Clearly the node p is optimal if the coefficient of θ in (9) is larger than the coefficient of $(1 - \theta)$; otherwise q is optimal or a tie exists, in which case either is optimal. Once the node p or q is reached, members of the route partitioning sets may have to be interchanged, corresponding to more efficient communication directly to the nodal location rather than through the entire link (p,q). This only improves matters, lowering the travel time below that achieved with the original route-partitioning sets. Moreover, the same proof with the new route-partitioning sets demonstrates the nonoptimality of moving away from the node. ▪

It is important to note that the fixed server assignment condition of the theorem does not imply that in practice the steady state probabilities are location independent. Server assignment preferences are usually heavily dependent on relative proximities of servers and hence state probabilities <u>are</u> affected by server locations. The theorem states that for any <u>given</u> set of server assignment preferences a set of optimal solutions exists on the nodes. As a result of this theorem the location problem has been reduced from optimization over an infinite set of points to an optimization over a finite set of nodes.

Notice also that if the expression (9) is concave in θ the same argument also holds. This can happen only if $P(y_Q)$ are all concave functions of θ. The meaning of this is not yet clear but can be of some interest in future research.

The Congested Median Problem and the Hypercube Model

"The hypercube model" is a spatially distributed queuing model

developed by Larson [8] to analyze analytically the performance of

urban emergency services. The model assumes a geographical region R

that is divided into n geographic areas of atoms. The fraction of

demand associated with each atom j is h_j ($\sum_{j=1}^{n} h_j = 1$) and the travel time

from atom i to atom j is d(i,j). Service requests over the entire

region are generated in a Poisson manner at a rate λ and at each atom

j independently in a Poisson manner with rate λj. ($\sum_j \lambda_j = \lambda$)

There are Q units to respond to the requests for service, located

at atoms i_1, i_2,...,i_Q. For Markov analysis, the service time for each

unit n is assumed to be exponential with mean μ_n^{-1}. Recent research has

shown that the assumption of exponentiality of the service time does not

markedly affect the predictive accuracy of the model when the mean of

a general distribution is entered into the exponential (Markov) model.

The mean service time is the sum of the travel time and on-scene time.

By the process of mean service time calibration [6,7,10], each server's mean

service time can be adjusted so that the model-computed mean travel times

(over the network) for each server are compatible with that server's total

mean service time μ_n^{-1}. For Theorem 1 to hold, we assume that μ_n^{-1} is

not affected by moving a server's home location along just one link. That

is, single link travel times are assumed to be negligible compared to total

mean service times.

States of the system are defined to be according to the status of

each service unit being busy or available. The model allows a zero-line

capacity queue, implying the existence of a special reserve unit, as well

as an infinite capacity queue. Given some dispatching policy, all the 2^Q steady state probabilities of the system can be obtained by solving 2^Q detailed balance equations [8]. In [9] Larson used a server sampling scheme adapted from the M/M/Q model to obtain fast approximate solutions for the required dispatch probabilities.

For a given set of single server locations at atoms $i,\ldots,i_Q$ the hypercube model computes several performance measures. Among them, the most important one is the mean region wide travel time, defined as

$$\sum_{j=1}^{n} \sum_{k=1}^{Q} \rho_{i_k,j} d(i_k,j) + P(\text{all units are busy}) \sum_{j=1}^{n} h_j r \qquad (10)$$

where $\rho_{i_k,j} \equiv$ fraction of all dispatches that send the unit from atom i_k to atom j; $k = 1,\ldots,Q$; $j = 1,\ldots,n$; $r \equiv$ travel time term arising from dispatches from queued service requests (infinite capacity case) or from service requests handled by a back-up service system (zero line capacity case). The $\rho_{i_k,j}$'s represent the _response patterns_ of units. They remain _fixed_ under a given set of dispatch preferences, even if the home locations of units change.

In [7] Jarvis developed an algorithm to find a set of "optimum" locations in the framework of the hypercube model where locations are constrained to atoms and each atom can contain not more than one facility. The key idea behind the Jarvis algorithm is to optimally locate the servers (facilities) for a given response pattern and then, given a new set of locations, to reassess the response patterns to determine if a new set of dispatch preferences (and thus response patterns) could improve

system performance further. This alternative iterative procedure is analogous to the "locate – allocate" scheme often used in deterministic location theory [1].

Jarvis' algorithm for the zero capacity case works as follows:

1. Initialize: Specify initial unit locations for units

    $1, 2, \ldots, Q$, corresponding to atoms $i_1, i_2, \ldots, i_Q$.

2. Allocate: Solve the hypercube model to obtain $\rho_{i_k, j}, k = 1, \ldots, Q;$

    $j = 1, \ldots, n$

3. Locate: Solve the following L.P problem:

$$\min \sum_{k=1}^{Q} \sum_{j=1}^{n} P(v, k) \; C(v, k)$$

$$\text{s.t.} \sum_{v=1}^{n} P(v, k) = 1 \qquad k = 1, \ldots, Q$$

$$P(v, k) \geq 0 \quad v = 1, \ldots, n; \quad k = 1, \ldots, Q$$

    where the decision variable $P(v, k)$ is the probability

    that server k is at node v when available $v = 1, \ldots, n;$

    $k = 1, \ldots, Q;$ and $C(v, k) = \sum_{j=1}^{n} p_{i_k, j} d(v, j) \quad v = 1, \ldots, n;$

    $k = 1, \ldots, Q.$

4. Test for Convergence: If the new Q locations are iden-

    tical to the old set of Q locations, stop. Otherwise

    go to step 2 with $i_1, \ldots, i_Q$ – new set of locations for

    units $1, \ldots, k,$ and reallocate.

Whenever the algorithm terminates, at least a local optimal solution is ensured. By taking several different initial sets of locations, the chances of getting closer to the optimal global solution are improved.

It is important to observe that step 3 of the algorithm is very simple because the problem can be reduced to Q independent trivial problems, each corresponding to a standard one-median problem with $\rho_{i_k,j}$ (j=1,2,...,n) being the nodal weight for the kth facility. To date in applications, the allocate step has been performed assuming that server preferences depend solely on proximity; however, more general (multi-attribute) procedures are allowed at this step.

The hypercube model can be applied in our congested median network context. The network G can represent the geographical region R, the nodes of the network being the atoms, and the links being the major streets connecting the atoms. We now demonstrate that if we take any Q points in the network to be the set of server locations, then $F(X_Q)$ - the cost function for the congested median problem (2) turns out to be identical to the mean region-wide travel time of the hypercube model (10). In terms of the congested median problem, the hypercube model disperses Q <u>single server</u> facilities over G.

Let $X_Q = (i_1, i_2, \ldots, i_Q)$ be a set of Q points in G. Then:

$$F(X_Q) = \sum_{y_Q \in Y_{X(Q)} - \{y_Q^o\}} P(y_Q) \sum_{j=1}^{n} h_j d(y_Q, j) + P(y_Q^o) \sum_{j=1}^{n} h_j C(j)$$

Let us consider now any $i_k \in X_Q$, k = 1,...,Q

Let $E_{i_k,j} = \{y_Q \in Y_{X(Q)} - \{y_Q^o\}$; the server at i_k is the most preferred available unit to node j }. Obviously

$$\sum_{y_Q \in Y_{X(Q)} - \{y_Q^o\}} P(y_Q) = \sum_{k=1}^{Q} \sum_{y_Q \in E_{i_k,j}} P(y_Q) \quad \forall \ j = 1, \ldots, n$$

Also $\forall\ y_Q\ \varepsilon\ E_{i_k,j},\quad d(y_Q,j\ ) = d(i_k,j)$

and hence by rearranging $F(X_Q)$ we get:

$$F(X_Q) = \sum_{k=1}^{Q} \sum_{j=1}^{n} d(i_k,j) \sum_{y_Q \varepsilon E_{i_k,j}} P(y_Q)\ h_j + P(y_Q^o) \sum_{j=1}^{n} h_j C(j)$$

Let us define $\rho_{i_k,j} = \sum_{y_Q \varepsilon E_{i_k,j}} P(y_Q)h_j$ which is the fraction of all dis-

patches that send the service out from i_k to j. Therefore,

$$F(X_Q) = \sum_{k=1}^{Q} \sum_{j=1}^{n} \rho_{i_k,j} d(i_k,j) + P(y_Q^o) \sum_{j=1}^{n} h_j C(j).$$

But $C(j)$ is the cost associated with a service request that occurs while all the servers are busy and hence $F(X_Q)$ is identical to (10) — the mean region wide travel time.

The conclusion of this discussion is that since the assumptions of Theorem 1 hold for the hypercube model (subject to our discussion of service times) both the hypercube model and Jarvis' algorithm do not suffer from a loss of generality by considering locations only on the atoms. In addition Jarvis' algorithm can be applied to the congested median problem whenever the hypercube model's assumptions are accepted. This result ties together two very different approaches in location theory, one which is purely deterministic as the median problem and another one which is stochastic as the hypercube model.

Example

The following example will illustrate some of our previous discussion. Suppose we want to locate three facilities on the simple network shown in Figure 1.
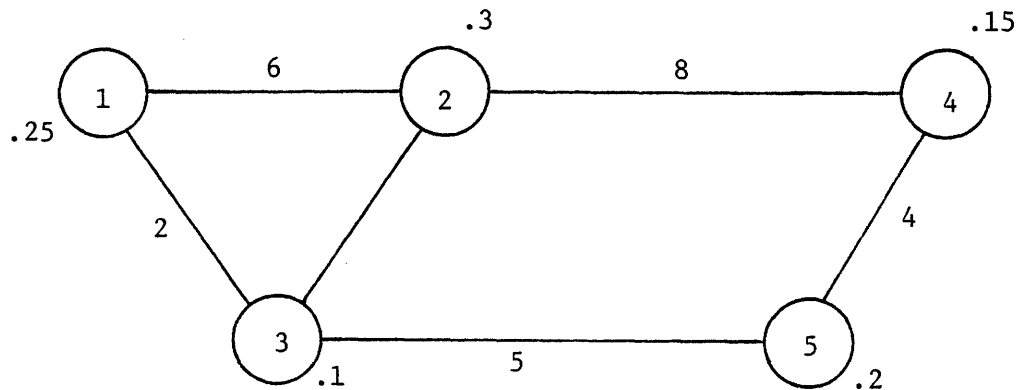

Figure 1

A Simple 5 Node Network

The numbers next to the nodes are the fractions of demands from each node λ_j; $j = 1,\ldots,5$ and the numbers next to the links are the travel times. There are $\binom{5}{2}$ possible distinct locations:

$$\{1,2,3\}\ ,\quad \{1,2,4\}\ ,\quad \{1,2,5\}\ ,\quad \{1,3,4\}\ ,\quad \{1,3,5\}\ ,\quad \{1,4,5\}\ ,$$
$$\{2,3,4\}\ ,\quad \{2,3,5\}\ ,\quad \{2,4,5\}\ ,\quad \{3,4,5\}\ .$$

The optimal location according to the standard 3-median problem is $\{1,2,5\}$, which can be obtained by hand. Suppose however that service requests occur in the network in a Poisson fashion with $\lambda = 4$, and the service time for each one of the three units is exponential with identical means $\mu^{-1} = 1$. Let us assume a zero capacity queue with R = 5 units of time - the cost

resulting when dispatching the reserve unit. We also assume that server preferences are determined solely by geographical proximity.

The Jarvis algorithm with an initial location at the absolute 3-median, i.e., {1,2,5}, converges after one iteration to the optimal solution at location {2,3,5}. The improvement achieved by moving from the location {1,2,5} to {2,3,5} is 3% in terms of the congested median problem. It is interesting to realize that the location {2,3,5} is among the weakest possible locations in terms of the standard median problem. This indicates that blind application of the absolute (deterministic) median problem can lead to erroneous results, even for such simple networks.

[1] Cooper, L., "Solutions of Generalized Locational-Equilibrium Problems," Journal of Regional Science, 7, 1967, pp. 1-18.

[2] Goldman, A.J., "Optimum Locations for Centers in Networks," Transportation Science, 3, No. 4, 1969, pp. 352-360.

[3] Hakimi, S.L., "Optimum Locations of Switching Centers and Absolute Centers and Medians of a Graph," Operations Research, 12, 1964, pp. 450-459.

[4] Hakimi, S.L., "Optimum Distribution of Switching Centers on a Communications Network and Some Related Graph Theoretic Problems," Operations Research, 13, 1965, pp. 462-475.

[5] Hakimi, S.L., and S.H. Hasheshwari, "Optimum Locations on Centers in Networks," Operations Research, 20, No. 5, 1972, pp. 967-977.

[6] Halpern, J., "The Accuracy of Estimates for the Performance Criteria on Certain Emergency Service Queueing Systems," Transportation Science, 11, No. 3, 1977, pp. 223-242.

[7] Jarvis, P.J., "Optimization in Stochastic Service Systems with Distinguishable Servers," TR-19-75, Innovative Resource Planning Project in Urban Public Systems, Operations Research Center, M.I.T., June, 1975.

[8] Larson, R.C., "A Hypercube Queueing Model for Facility Location and Redistricting in Urban Emergency Services," Computers and Operations Research, 1, March 1974, pp. 67-95.

[9] Larson, R.C., "Urban Emergency Services: An Approximate Method for Computing Operational Performance Measures," Operations Research, 23, No. 4, July-August 1975, pp. 845-868.

[10] Larson, R.C., "Computer Program for Calculating the Performance of Urban Emergency Service Systems: User's Manual," TR-14-75, Innovative Resource Planning Project in Urban Public Systems, Operations Research Center, M.I.T., March, 1975.

[11] Levy, J., "An Extended Theorem for Location on a Network," Operations Research Quarterly, 18, No. 4, 1967, pp. 433-442.

[12] Wendell, R.E., and A.P. Hurter, "Optimal Location on a Network," Transportation Science, 7, No. 1, February 1973, pp. 18-33.