The Generalized *P*-Forest Problem on a Tree Network

Arie Tamir
Department of Statistics
Tel Aviv University
Ramat-Aviv, Tel Aviv 66978
Israel

Timothy J. Lowe
College of Business Administration
University of Iowa
Iowa City, Iowa 52242

In two recent papers, similar multifacility problems on tree networks were studied where facilities are subtrees of the tree. The complexity of the problems was left as open. In this paper, we formulate a general model that includes the above-cited problems as special cases and give a polynomial algorithm to solve the general model.

1. INTRODUCTION

Most of the literature dealing with optimally locating facilities on networks treat the facilities as points of the network. There has been significant interest in such problems due to their theoretical properties as well as to their practical applications in both the private and public sectors. Survey articles by Brandeau and Chiu [1], Dearing [10], Francis et al. [12], Halpern and Maimon [15], Krarup and Pruzan [27], McGinnis [29], and Tansel et al. [37] give an indication of the scope and nature of such problems. In addition, location problems on networks are discussed in texts by Christofides [2], Handler and Mirchandani [16], Larson and Odoni [28], Minieka [31], and Mirchandani and Francis [32].

Recently, there has been interest in extending the notion of a facility as a point on the network to that of a path or tree-shaped subgraph of the network. Most of this work has dealt with the case of locating a single such facility. Generally, the facility provides some form of service to demand points located at vertices of the network. The cost of service to a vertex is taken to be a function of the shortest-path distance on the network from the vertex to the closest point of the facility. In addition, it is generally assumed that the cost of the facility is some increasing function (often linear) of the total length of the

NETWORKS, Vol. 22, (1992) 217–230
 CCC 0028-3045/92/030217-14$04.00

facility. Examples of various versions of this problem can be found in Morgan and Slater [33], Slater [35], Minieka [30], Kim et al. [20–22], Kincaid et al. [23], and Richey [34]. These papers approach the solution to the facility (a path or subtree) location problem by exploiting network structure and thus assume that the network has a special structure e.g., it is a tree or a series-parallel graph, in order to obtain polynomial time algorithms.

In contrast, the papers by Hutson and ReVelle [17, 18], Current et al. [6–9], Current [5], and Church and Current [3] approach the facility location problem via integer programming models.

In this paper, we consider a very general model for locating p (≥ 1) tree-shaped facilities on a tree network and provide a polynomial dynamic programming algorithm for solving the general model. As we will show, the model captures generalizations (facilities as subtrees) of several instances of the facility location problem that have been considered in the literature.

Our interest in this problem was stimulated by recent working papers by Hakimi et al. [14] and Church and Current [4]. The former paper studies the complexity of 64 versions of the facility location problem. The versions are derived by considering such elements as locating one or $p > 1$ facilities, whether the facilities are paths or tree-shaped, whether the underlying network is a tree or a general network, and the objective function of the problem.

The latter paper considers the location of $p > 1$ tree-shaped facilities on a tree network. A multiobjective problem is considered with the objectives of minimizing the total length of all facilities located and maximizing the total number of vertices "covered" by the facilities. The solution approach is via integer programming.

Hakimi et al. [14] have shown that several of the problems that they consider can be solved in polynomial time if the underlying network is a tree, but are NP-hard if the network is arbitrary. That trees give rise to efficient solution methods to some of these problems is not totally surprising. When the facilities are considered to be points, certain convexity properties (see Dearing et al. [11]) are present. In the case of a single facility that is tree-shaped, submodularity properties (see Tamir [36]) are present.

Our main motivation for this paper was to resolve the complexity issue of the problem studied in Church and Current [4] and two of the problems, cited as open, in Hakimi et al. [14]. Our general model, which we solve in polynomial time, includes these problems as subcases.

We now give a brief review of the paper. In the next section, we formulate the general model that we consider and draw the relationship between the model and other facility location problems that have been considered in the literature. In Section 3, we give the dynamic programming recursive equations, which lead to an $O(n^3p^2)$ algorithm to solve the general model. Concluding remarks are given in Section 4.

2. THE MODEL

Let $T = (V,E)$ be an undirected tree with node set $V = \{1,2, \ldots, n\}$ and edge set E. A subtree $\tau = (V',E')$ of T is a connected subgraph of T, with V' a subset of V and E' a subset of E. We are concerned with finding a p-forest of T (p disjoint subtrees [facilities] of T, $1 \leq p < |V|$) in order to minimize the sum of the following cost terms:

1. The cost of the length of all of the facilities.
2. The cost of the edges of T that connect the components of the p-forest.
3. The travel cost from each node of T to a closest facility.
4. The cost of communication between facilities that occur on the edges of the facilities.
5. The cost of communication between facilities that occur on the edges of T that connect the facilities.
6. The cost of selecting exactly one vertex, a "server," in each facility.

To formulate the model, we make use of parameters defined on the edges and nodes of T. Each node $i \in V$ is associated with a nonnegative weight, $w(i)$, the cost of selecting i as a server, and a nonnegative nondecreasing transportation cost function $c_i(\cdot)$, whose argument is the total distance of travel to a nearest facility. Without loss of generality, we assume that $c_i(0) = 0$. Each edge $e \in E$ is associated with five nonnegative parameters:

$a_1(e)$—the contribution of e to the cost of the length of the p-forest if e is in a component of the p-forest.

$a_2(e)$—the contribution of e to the cost of the forest of T that connects the components of the p-forest, when e is in the *connecting* forest.

$a_3(e)$—the contribution of e to the travel distance from a node of T to a facility if e is in the path to the facility. We assume that $a_3(e) > 0$.

$a_4(e)$—the contribution of e to the cost of communication between components of the p-forest if e is in a path between components and e is itself in a component.

$a_5(e)$—the contribution of e to the cost of communication between components of the p-forest if e is in a path between components, but e is *not* in a component (is in the connecting forest).

The use of the above parameters in the cost terms 1–6 is as follows: the sum of the $a_1(e)$ over all edges in the components of the p-forest is term 1. The $a_2(e)$ are used in term 2. The parameters $a_3(e)$ are used to determine a closest facility to node i of T, and the length of a shortest path to a facility is the argument of $c_i(\cdot)$. The sum of these functions over the nodes of T is term 3. The parameters $a_4(e)$ and $a_5(e)$ are used in terms 4 and 5, respectively. Finally, the parameters $w(i)$ are used for term 6.

To more rigorously specify the role of the parameters in our general model, we need some additional notation. A p-forest, F_p, of T is a collection $\{\tau_1, \tau_2, \ldots, \tau_p\}$ of p mutually disjoint subtrees (facilities) of T. By disjoint we mean sharing no common nodes or edges. Each subtree τ_j is composed of edges of T. Given an F_p, we denote by S' the minimal forest of T whose components connect (span) the components of F_p. When $p > 1$, we note that S' consists of at least one, but no more than $p - 1$ disjoint components.

The cost associated with the "length" [via $a_1(\cdot)$] of F_p is determined as follows: For each $\tau_j, j = 1, \ldots, p$, let

$$L_1(\tau_j) = \sum \{a_1(e): e \in \tau_j\}.$$

If τ_j is a single point (node of T), $L_1(\tau_j)$ is zero.

The cost [via $a_2(\cdot)$] associated with the edges that connect the components of F_p, i.e., the edges of S', is denoted by

$$L_2(S') = \sum \{a_2(e): e \in S'\}.$$

For each $i \in V$, the unique path $\sigma(i,\tau_j)$ between i and the first-encountered node of T in facility j either has no edges, i.e., $i \in \tau_j$, or has one or more edges of T. We denote by $d_3(i,F_p)$ the "length of the shortest path" between i and F_p, which is defined by

$$d_3(i,F_p) \equiv \min \left\{\sum \{a_3(e): e \in \sigma(i,\tau_j)\} : j = 1, \ldots, p\right\}.$$

When $i \in F_p$, $d_3(i,F_p)$ is defined to be zero.

Two types of "communication" costs *between* facilities are considered: those costs measured across the edges of F_p and those costs measured across the edges of S'. Consider first the former case. Let $e \in \tau_j$ for some $j = 1, \ldots, p$. The removal of e from T disconnects T into two disjoint components $T_\alpha(e)$ and $T_\beta(e)$. Letting $\alpha(e)$ ($\beta(e)$) denote the total number of facilities that satisfy $\tau_k \subset T_\alpha(e)$ ($\tau_k \subset T_\beta(e)$), then the communication cost across edge e is $\alpha(e) \times \beta(e) \times a_4(e)$. The communication cost for facility τ_j is defined as

$$M_4(\tau_j) \equiv \sum \{\alpha(e) \times \beta(e) \times a_4(e): e \in \tau_j\}.$$

We note that $\tau_j \not\subset T_\alpha(e)$ and $\tau_j \not\subset T_\beta(e)$, but that for an edge $e \in \tau_j$, $\alpha(e) + \beta(e) = p - 1$. Each edge e of an "extreme" facility has either $\alpha(e)$ or $\beta(e)$ equal to zero. Also, if a facility ρ_j is a single point, $M_4(\tau_j)$ is zero.

Consider the communication cost across an edge of S', the forest spanning the p-forest F_p. The removal of e from T disconnects T into $T_\alpha(e)$ and $T_\beta(e)$. But in this case, since $e \in S'$, with $\alpha(e)$ and $\beta(e)$ as defined above, $\alpha(e) + \beta(e) = p$, and $\alpha(e), \beta(e) \geq 1$. The cost of communication for edge e is taken to be $\alpha(e) \times \beta(e) \times a_5(e)$, and the cost of communication on S' is defined as

$$M_5(S') \equiv \sum \{\alpha(e) \times \beta(e) \times a_5(e): e \in S'\}.$$

With the above definitions and descriptions of the cost terms, our objective is to find a p-forest $F_p = \{\tau_1, \ldots, \tau_p\}$ to minimize

$$\begin{aligned} f(F_p) = \sum \{L_1(\tau_j) : j = 1, \ldots, p\} \\ + L_2(S') + \sum \{c_i(d_3(i,F_p)) : i \in V\} \\ + \sum \{M_4(\tau_j) : j = 1, \ldots, p\} + M_5(S') \\ + \sum \{\min\{w(i) : i \in \tau_j\} : j = 1, \ldots, p\}. \end{aligned} \tag{2.1}$$

The terms of (2.1) correspond, respectively, to cost terms 1–6 discussed earlier.

Expression (2.1) unifies several location models discussed in the literature. For example, if $w(i)$ is constant and $c_i(y) = \mu y$, for all $i \in V$, where $\mu > 0$ is sufficiently small, $a_1(e)$ is sufficiently large for all e (so that the p facilities are points of T), and if $a_2(e) = a_4(e) = a_5(e) = 0$ for all $e \in E$, the problem reduces to the classical p-median problem on T (see, e.g., Kariv and Hakimi [19].

A second example is the simple plant location problem on a tree [25, 26], which is obtained from (2.1) by setting $a_2(e) = a_4(e) = a_5(e) = 0$, for all $e \in E$ and setting $a_1(e)$ sufficiently large for all $e \in E$ (so that, again, the p-facilities are points of T). We remark that although the simple plant location problem allows a variable number of facilities (p is a variable), a minor modification of our solution procedure will solve the above instance of the problem in polynomial time, since in an optimal solution, no more than n facilities will be chosen (no more than one at each vertex).

A third example is the minimal cost/maximal covering forest problem considered by Church and Current [4]. Their problem is obtained from (2.1) by setting, for $i \in V$,

$$c_i(y) = \begin{matrix} 0 \text{ if } y \le r_i \\ \theta_i \quad \text{otherwise,} \end{matrix}$$

where $r_i > 0$, and $\theta_i > 0$ for all i; setting $w(i)$ to a constant; setting $a_1(e) = \mu a_3(e)$ for some $\mu > 0$; and again setting $a_2(e) = a_4(e) = a_5(e) = 0$. Hakimi et al. [14] also considered this model with $\theta_i = +\infty$.

A fourth example is the fixed-cost spanning forest problem considered by Granot and Granot [13]. [This model and the second example have motivated the introduction of $w(i)$.] In their model there exists a set of demand points, V', a subset of V, and the problem is to select p sites for the service centers. Service centers can only be established at the vertices of T. The cost of opening a center at $i \in V$ is $w(i)$. The objective is to minimize the sum of the total cost of setting up centers and the length of the forest connecting the demand set V' to the established centers. The model is obtained from (2.1) by setting, for $i \in V'$,

$$c_i(y) = \begin{matrix} 0 \text{ if } y = 0 \\ +\infty \quad \text{otherwise,} \end{matrix}$$

and for $i \in V - V'$, $c_i(y) = 0$ for all y, and setting $a_1(e) = a_3(e)$, for $e \in E$, and $a_2(e) = a_4(e) = a_5(e) = 0$, for $e \in E$. We remark that Granot and Granot allow a variable number of centers (p is a variable) [13]. However, our comment with respect to the second example above applies here as well.

Although the analogy is not exact, our rationale for cost terms 4 and 5 (the communication cost between facilities) is partially due to the p-median problem with mutual communication considered by Kolen [25] among others. Kolen's problem considers the sum of weighted distances between pairs of facilities (which are points) as well as the sum of weighted distances between nodes of T and the facilities. In Kolen's problem, each facility is distinctly identified and the weight corresponding to a node–facility pair is prespecified. Thus, the travel cost for a node is *not* a function of the distance to a *closest* facility, but is the sum of weighted distances to all facilities.

3. THE DYNAMIC PROGRAMMING PROCEDURE

We (arbitrarily) root the tree T at some node $i_0 \in V$. For each $i \in V$, let $D(i)$ be the set of *all* nodes j having i on the unique path connecting them to i_0. $D(i)$ is called the set of descendants of i. Let $S(i)$ be the set of children of i, where a member of $S(i)$ is a descendant of i and is connected to i via an edge of T. We note that $S(i) \subset D(i)$ and that i is in $D(i)$, but not in $S(i)$. If $S(i)$ is empty, i is called a tip (or leaf) of T.

To solve (2.1) on T, we recursively solve a sequence of problems defined on certain subtrees of T, starting with the leaves of T. To define these subtrees, consider a node $i \in V$. Suppose that $S(i) = \{i(1), i(2), \ldots, i(s(i))\}$, where $s(i) = |S(i)|$. For any $t = 1, \ldots, s(i)$, let $T(i,t)$ denote the subtree induced by the nodes in $\{i\} \cup D(i(1)) \cup \cdots \cup D(i(t))$. In what follows, we let

$$N(i,t) = D(i(1)) \cup D(i(2)) \cup \cdots \cup D(i(t)),$$

so that the sets $N(i,t-1)$, $D(i(t))$, and $\{i\}$ are disjoint, but their union is the set of nodes of $T(i,t)$. Figure 1 depicts an example of these definitions.

Given a node $i \in V$ and $t = 1, \ldots, s(i)$, we define a sequence of restricted problems. Let j and k be nodes in $T(i,t)$ and $T - T(i,t)$, respectively. Consider the problem of selecting a minimum-cost q-forest $F_q = \{\tau_1, \ldots, \tau_q\}$ of $T(i,t)$, where $q \le p$ and j is a node in F_q that satisfies $d_3(i,j) = d_3(i,F_q)$.

We denote by $f(i,t,j,k|q)$ the minimum objective function value of

$$\begin{aligned}
&\sum \{L_1(\tau_r) : r = 1, \ldots, q\} \\
&\quad + L_2(S'(T(i,t))) \\
&\quad + \sum \{c_h(\min\{d_3(h,F_q), d_3(h,k)\}) : h \in T(i,t)\} \\
&\quad + \sum \{M_4(\tau_r) : r = 1, \ldots, q\} \\
&\quad + \{M_5(S'(T(i,t))) + \sum \{\min\{w(h) : h \in \tau_r\} : r = 1, \ldots, q\}. \qquad (3.1)
\end{aligned}$$

The optimization is over all q forests F_q in $T(i,t)$ that contain node j and node j is the closest node (in the sense of $d_3(\cdot)$) of the q forest to node i. In (3.1), if $q =$

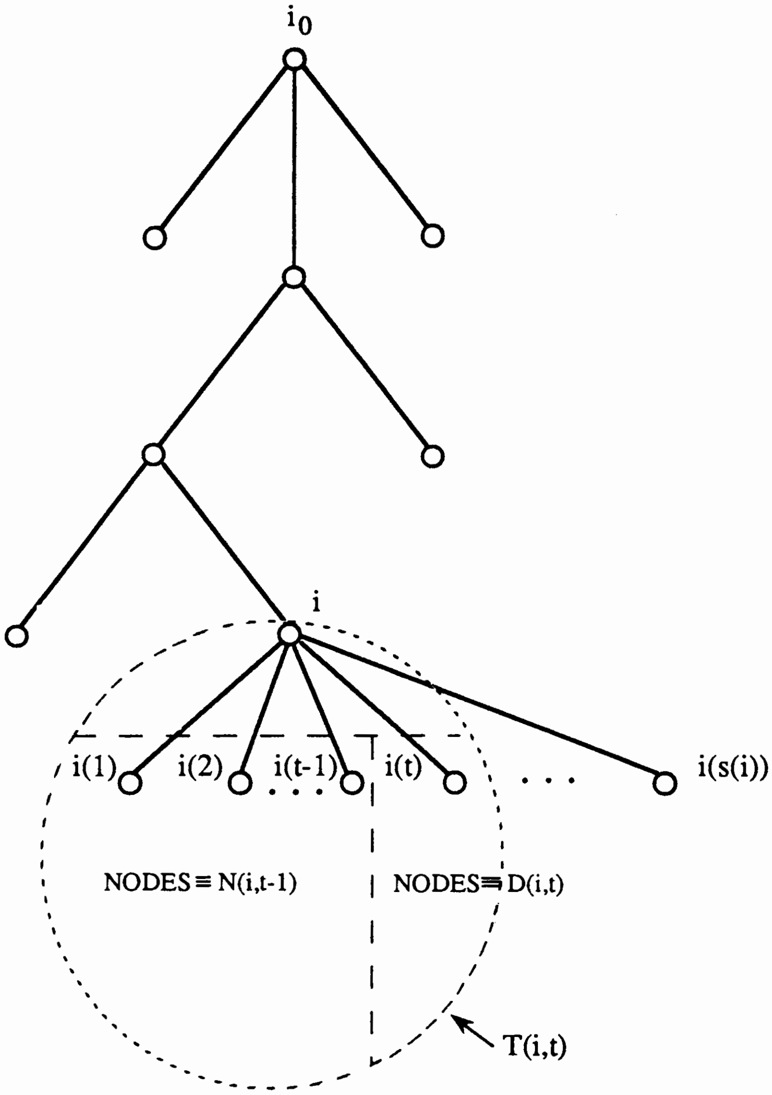

FIG. 1. Illustration of definitions.

p, or if $1 < q < p$ and i is in a component of the q-forest, then $S'(T(i,t))$ is defined to be the forest of $T(i,t)$ spanning the q-forest. If $q = 0$, $S'(T(i,t)) = \phi$. If $1 \leq q < p$ and i is not in a component of the q-forest, then $S'(T(i,t))$ is the forest of $T(i,t)$ spanning the q-forest and node i. Although $S'(T(i,t))$ clearly depends upon F_q, for ease of exposition, we delete the notation indicating dependence of $S'(\cdot)$ on F_q. Note that in (3.1), with respect to the node travel costs $c_h(\cdot)$, we are concerned only with nodes in $T(i,t)$. The travel cost for node $h \in T(i,t)$ is taken to be the minimum of the travel cost to the closest point in the q-forest and the point k in $T - T(i,t)$. The case with $q = 0$ (i.e., a 0-forest [the empty set]) is denoted by $f(i,t,-,k|0)$ and its value is $\Sigma\ \{c_h(d_3(h,k)) : h \in T(i,t)\}$. If no node k is selected in $T - T(i,t)$, the respective objective value is denoted by $f(i,t,j,-|q)$, and the minimization in the third term of (3.1) is replaced by $d_3(h,F_q)$. As a boundary condition, we define $f(i,t,-,-|0)$ as $+\infty$.

In the case that $i = j$ (i.e., i is restricted to be in the q-forest), it is necessary to define $f^-(i,t,i,k|q)$, which is the minimum objective value of (3.1), over all q-forests F_q of $T(i,t)$, with $i \in F_q$, but *not* including the cost of selecting a node in the component of F_q that contains node i; that is, $f^-(i,t,i,k|q)$ does not consider that element in the last term of (3.1) that counts the cost of a minimum cost node in the facility containing node i.

When i is a leaf node of T, we note that $S(i) = \phi$. In this case, we have the boundary conditions

$$f(i,-,i,k|1) = w(i) \qquad (\text{since } d_3(i,i) = 0),$$

$$f^-(i,-,i,k|1) = 0, \text{ and}$$

$$f(i,-,-,k|0) = c_i(d_3(i,k)). \tag{3.2}$$

We now give the recursive equations for computing $f(i,t,j,k|q)$ and $f^-(i,t,i,k|q)$. We will set $f(i,t,j,k|q) = f^-(i,t,i,k|q) = +\infty$ if it is not possible to select a q-forest in $T(i,t)$. For example, $f(i,-,i,k|q) = f^-(i,-,i,k|q) = +\infty$, for all $q \geq 2$. We begin by computing, via (3.2), $f(i,-,i,k|1)$, $f^-(i,-,i,k|1)$, and $f(i,-,-,k|0)$, $k \in T - \{i\}$, for all leaf nodes i of T. By induction, assume that $f(i,t-1,j,k|q)$ and $f^-(i,t-1,i,k|q)$ have been computed for all j,k, and $q \leq p$ and that $f(i(v), s(i(v)),j,k|q)$ and $f^-(i(v),s(i(v)),i(v),k|q)$, for all $v = 1, \ldots, s(i),j,k$, and q have been computed as well.

The recursive equations depend upon the parameters i,t,j,k, and q, but can be categorized into 10 subcases (not counting the boundary conditions for leaf nodes). These 10 subcases are best explained through the use of a "condition tree" (Fig. 2). The "tips" of the figure represent the 10 subcases. In the figure, there are two main branches labeled I and II. In branch I, j and k are such that $d_3(i,j) \leq d_3(i,k)$, and in branch II, $d_3(i,j) > d_3(i,k)$. Thus, for branch I, since $k \in T - T(i,t)$, for any $h \in T(i,t)$, $d_3(h,k) = d_3(h,i) + d_3(i,k) \geq d_3(h,i) + d_3(i,j) \geq d_3(h,j)$ so that k is irrelevant [see the third term in (3.1)]. Thus, $f(i,t,j,-|q) = f(i,t,j,k|q)$ on branch I. On branch II, since $d_3(i,j) > d_3(i,k) > 0$, the case $i = j$ is not possible, and so $f^-(\cdot)$ is only computed for cases IB1a, IB1b, IBta, and IBtb. Table I gives the recursive equations for each of the subcases. The reader will find it convenient to use both the figure and the table to verify the recursive equations. In the table, the notation $\pi(q)$ and $\delta(q)$ is used. $\pi(q)$ and $\delta(q)$ are defined as follows:

For any q, $0 \leq q \leq p$, the *indicator* function $\delta(q)$ is

$$\delta(q) = \begin{cases} 1 & \text{if } 0 < q < p \\ 0 & \text{otherwise} \end{cases}$$

and the *product* function $\pi(q)$ is

$$\pi(q) = q(p - q).$$

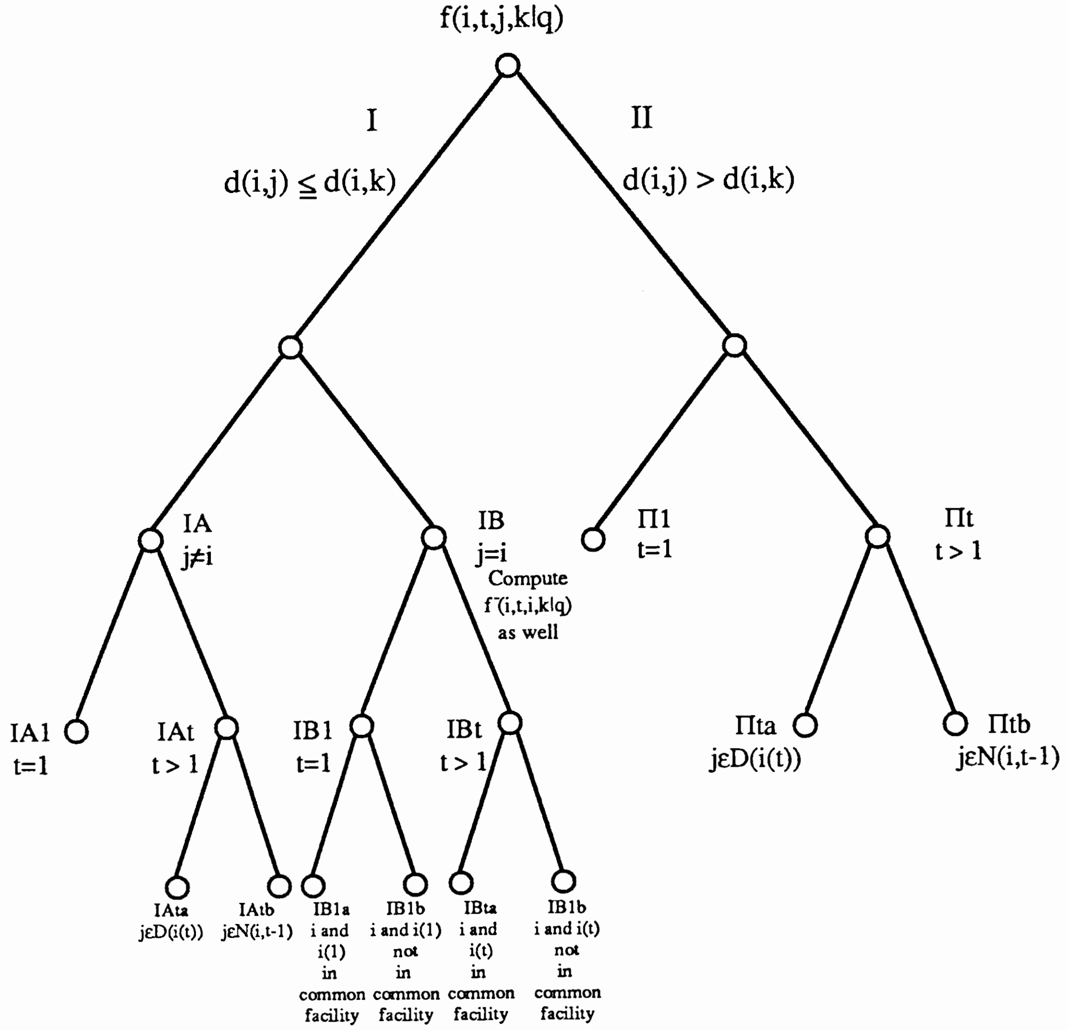

FIG. 2. Condition tree.

Note that $\delta(q)$ is used in conjunction with $a_2(e)$, where we include only the cost of edge e in a "connecting" forest, when edge e is on the path between at least two members of F_q. Also, $\pi(q)$ is used in conjunction with $a_5(e)$ to correctly count the communication cost across edge e when it is in a connecting forest. Since a total of p facilities will ultimately be chosen, if q facilities are on "one side" of e, then $p - q$ facilities will eventually be located on the "other side" of e. When no facilities ($q = 0$) are on one side of e, $\pi(q) = 0$. Thus, $\pi(q)$ accurately counts the communication traffic between facilities across edge e.

Finally, we note that with i_0 as the root of T, an optimal solution to (2.1) is found by solving

$$\min\{f(i_0, s(i_0), u, -|p) : u \in D(i_0)\}.$$

The verification of the complexity of the dynamic programming approach is straightforward. For every vertex i, we must consider $O(n^2)$ pairs $\{j,k\}$, with j

TABLE 1. Recursive Equations.

Case	Descriptor	Recursive equation
IA1	$t = 1, d_3(i,j) \le d_3(i,k), j \ne i$	$\delta(q) \times a_2(i,i(1)) + c_i(d_3(i,j)) + \pi(q) \times a_5(i,i(1)) + f(i(1),s(i(1)),j,k\|q)$
1Ata	$t > 1, d_3(i,j) \le d_3(i,k), j \in D(i(t))$	$\min\{\delta(q_1) \times a_2(i,i(t)) + \pi(q_1) \times a_5(i,i(t)) + f(i(t),s(i(t)),j,-\|q_1) + \min\{f(i,t-1,u,j\|q_2) : u \in N(i,t-1);\ d_3(u,i) \ge d_3(i,j)\} : q_1 + q_2 = q;\ q_1 \ge 1\}$
IAtb	$t > 1, d_3(i,j) \le d_3(i,k), j \in N(i,t-1)$	$\min\{\min\{f(i(t),s(i(t)),u,j\|q_1) : u \in D(i(t)), d_3(u,i) \ge d_3(i,j)\} + \delta(q_1) \times a_2(i,i(t)) + \pi(q_1) \times a_5(i,i(t)) + f(i,t-1,j,-\|q_2) : q_1 + q_2 = q;\ q_2 \ge 1\}$
IB1	$t = 1, j = i$	$f(i,1,i,k\|q)$ is the minimum of IB1a and IB1b, below:
IB1a	i and $i(1)$ are in a common facility ($q \ge 1$)	$a_1(i,i(1)) + (q-1)(p-q) \times a_4(i,i(1) + \min\{w(i) + f^-(i(1),s(i(1)),i(1),i\|q);\ f(i)1),s(i(1)),i(1),i\|q)\}$
IB1b	i and $i(1)$ are not in a common facility ($q \ge 1$)	$\delta(q-1) \times a_2(i,i(1)) + \pi(q-1) \times a_5(i,i(1)) + w(i) + \min\{f(i(1),s(i(1)),u,i\|q-1) : u \in D(i(1))\}$
		$f^-(i,1,i,k\|q)$ is the minimum of IB1a⁻ and IB1b⁻, below:
IB1a⁻	i and $i(1)$ are in a common facility ($q \ge 1$)	$a_1(i,i(1)) + (q-1)(p-q) \times a_4(i,i(1)) + f^-(i(1),s(i(1)),i(1),i\|q)$
IB1b⁻	i and $i(1)$ are not in a common facility ($q \ge 1$)	$\delta(q-1) \times a_2(i,i(1)) + \pi(q-1) \times a_5(i,i(1)) + \min\{f(i(1),s(i(1)),u,i\|q-1) : u \in D(i(1))\}$

IBt	$t > 1, j = i$	$f(i,t,i,k\|q)$ is the minimum of IBta and IBtb, below:
IBta	i and $i(t)$ are in a common facility	$a_1(i,i(t)) + \min\{(q_1 - 1)(p - q_1) \times a_4(i,i(t)) + \min\{f^-(i(t),s(i(t)),i(t),i\|q_1) + f(i,t - 1,i,k\|q_2 \in; f(i(t),s(i(t)),i(t),i\|q_1 + f^-(i,t - 1,i,k\|q_2)\} : q_1 + q_2 = q + 1; q_1, q_2 \geq 1\}$
IBtb	i and $i(t)$ are not in a common facility	$\min\{\delta(q_1) \times a_2(i,i(t)) + \pi(q_1) \times a_5(i,i(t)) + f(i,t - 1,i,k\|q_2) + \min\{f(i(t), s(i(t)),u,i\|q_1) : u \in D(i(t))\} : q_1 + q_2 = q; q_2 \geq 1\}$
		$f^-(i,t,i,k\|q)$ is the minimum of IBta$^-$ and IBtb$^-$, below:
IBta$^-$	i and $i(t)$ are in a common facility	$a_1(i,i(t)) + \min\{(q_1 - 1)(p - q_1) \times a_4(i,i(t)) + f^-(i(t),s(i(t)),i(t),i\|q_1) + f^-(i,t - 1,i,k\|q_2) : q_1 + q_2 = q + 1; q_1, q_2 \geq 1\}$
IBtb$^-$	i and $i(t)$ are not in a common facility	$\min\{\delta(q_1) \times a_2(i,i(t)) + \pi(q_1) \times a_5(i,i(t)) + f^-(i,t - 1,i,k\|q_2) + \min\{f(i(t)), s(i(t)),u,i\|q_1) : u \in D(i(t))\} : q_1 + q_2 = q; q_2 \geq 1\}$
II1	$t = 1, d_3(i,j) > d_3(i,k)$	$\delta(q) \times a_2(i,i(1)) + c_i(d_3(i,k)) + \pi(q) \times a_5(i,i(1)) + f(i(1),s(i(1)),j,k\|q)$
IIta	$t > 1, d_3(i,j) > d_3(i,k), j \in D(i(t))$	$\min\{\delta(q_1) \times a_2(i,i(t)) + \pi(q_1) \times a_5(i,i(t)) + f(i(t)),s(i(t)),j,k\|q_1) + \min\{f(i,t - 1,u,k\|q_2) : u \in N(i,t - 1); d_3(i,u) \geq d_3(i,j)\} : q_1 + q_2 = q; q_1 \geq 1\}$
IItb	$t > 1, d_3(i,j) > d_3(i,k), j \in N(i,t - 1)$	$\min\{\delta(q_1) \times a_2(i,i(t)) + \pi(q_1 \times a_5(i,i(t)) + f(i,t - 1,j,k\|q_2) + \min\{f(i(t)), s(i(t)),u,k\|q_1) : u \in D(i(t)); d_3(i,u) \geq d_3(i,j))\} : q_1 + q_2 = q; q_2 \geq 1\}$

$\in D(i)$ and $k \in V - D(i)$. For fixed i,t,j, and $k, O(p^2)$ total effort is required to compute $f(i,t,j,k|q)$ for all $q = 1, \ldots, p$. Since $t = 1, \ldots, s(i)$, the overall approach takes $O(n^2p^2(\Sigma\{s(i): i \in V\})) = O(n^3p^2)$.

4. CONCLUSIONS AND FINAL COMMENTS

In two recent papers, Hakimi et al. [14] and Church and Current [4], similar multifacility location models on a tree network were studied, where facilities are subtrees of the tree. The first paper called the model the center p-tree problem, while the second, considering a generalization, referred to it as the minimal cost/maximal covering forest problem. We have presented the generalized p-forest problem that extends and unifies the above models as well as several other multifacility location problems studied in the literature. We then provided an $O(n^3p^2)$ dynamic programming algorithm to solve our generalized model, thus resolving the open complexity issue.

Several comments are in order. First, we note that our algorithm can be used parametrically in p to solve the case where p is a variable, since p, the number of facilities, is bounded by the number of demand points (nodes). In fact, the case where p is a variable is considerably simpler when the cost of communication between the facilities is irrelevant, i.e., $a_4(e) = a_5(e) = 0$, for all $e \in E$. One can easily modify the recursive equations for this case to obtain an $O(n^3)$ algorithm.

Our next comment concerns the definition of a subtree. We have regarded a subtree as a subgraph of the underlying tree network, i.e., every edge of the subtree is also an edge of the tree. Hakimi et al. [14] considered this definition as well as a variant where the subtree is allowed to contain partial edges. Our algorithmic results can be extended to the partial edge case. Instead of dealing with the discrete entities $f(i,t,j,k|q)$ in (3.1), we will need a continuous version where j and k are "points" along edges of $T(i,t)$ and $T - T(i,t)$, respectively. Some models allowing partial edges are directly reducible to our model with "full" edges. To illustrate, consider the center p-tree problem with partial edges defined by Hakimi et al. [14]. To obtain this problem from our model, we set

$$c_i(y) = \begin{matrix} 0 \text{ if } y \leq r_i \\ \infty \quad \text{otherwise,} \end{matrix}$$

set $w(i) = 0$ for all i, $a_2(e) = a_4(e) = a_5(e) = 0$ for all $e \in E$, and $a_1(e) = \mu a_3(e)$ for all $e \in E$, where $\mu > 0$. Also, the "cost" of a partial edge is proportional to its length.

It is easy to see that there exists an optimal solution to the above problem where each boundary point of a facility, which is not at an original node of T, is at a distance of exactly r_i from some node i of T. If we add to the original set of nodes all points of the tree that are at a distance of exactly r_i from some node i, for all $i \in V$, the problem reduces to the p-tree problem with "full" edges. Also, if k is one of the added points, we set $c_k(y) = 0$ for all y. Since $O(n^2)$ points are added, the original problem can be solved in polynomial time.

Finally, we comment on the complexity bound of our algorithm. Our main goal has been to resolve the open complexity issues raised in the literature. Therefore, we focused in this paper on the relatively simple $O(n^3p^2)$ recursive procedure of Section 3 that requires minimal notation and machinery. We have also developed a more elaborate recursive approach that has an $O(n^2p^2)$ complexity. In this latter approach, the single function $f(i,t,j,k|q)$ of (3.1) is replaced by a two-function recursive system. The interested reader can obtain the details from the authors.

REFERENCES

[1] M. L. Brandeau and S. S. Chiu, An overview of representative problems in location research. *Management Sci.* **35** (1989) 645–673.

[2] N. Christofides, *Graph Theory: An Algorithmic Approach.* Academic Press, New York (1975).

[3] R. Church and J. Current, Maximal covering tree problems. Working paper no. 89-47, College of Business, Ohio State University, Columbus, OH (1989).

[4] R. Church and J. Current, The minimal cost/maximal covering forest problem on a tree. Fifth International Symposium on Locational Decisions, Lake Arrowhead, CA (1990).

[5] J. Current, The design of a hierarchical transportation network with transshipment facilities. *Transport. Sci.* **22** (1988) 270–277.

[6] J. Current, C. ReVelle, and J. Cohon, The shortest covering path problem: An application of locational constraints to network design. *J. Reg. Sci.* **24** (1984) 161–185.

[7] J. Current, C. ReVelle, and J. Cohon, The maximum covering/shortest path problem: A multiobjective network design and routing formulation. *Eur. J. Oper. Res.* **21** (1985) 189–199.

[8] J. Current, C. ReVelle, and J. Cohon, The hierarchical network design problem. *Eur. J. Oper. Res.* **27** (1986) 57–66.

[9] J. Current, C. ReVelle, and J. Cohon, The median shortest path problem: A multiobjective approach to analyze cost vs. accessibility in the design of transport networks. *Transportation Sci.* **21** (1987) 188–197.

[10] P. M. Dearing, Location problems. *Operations Res. Lett.* **4** (1985) 95–98.

[11] P. M. Dearing, R. Francis, and T. Lowe, Convex location problems on tree networks. *Operations Res.* **24** (1976) 628–642.

[12] R. L. Francis, L. F. McGinnis, and J. A. White, Locational analysis. *Eur. J. Oper. Res.* **12** (1983) 220–252.

[13] D. Granot and F. Granot, Computational complexity of a cost allocation approach to a fixed cost spanning forest problem. Working paper, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, BC, Canada (1990).

[14] S. L. Hakimi, E. Schmeichel, and M. Labbe, On locating path- or tree-shaped facilities on networks. Research report, University of California-Davis, Davis, CA (1990).

[15] J. Halpern and O. Maimon, Algorithms for the m-center problem: A survey. *Eur. J. Oper. Res.* **10** (1982) 90–99.

[16] G. Y. Handler and P. B. Mirchandani, *Location on Networks.* The M.I.T. Press, Cambridge, MA (1979).

[17] V. Hutson and C. ReVelle, Maximal direct covering tree problems. *Transport. Sci.* **23** (1989) 288–299.
[18] V. Hutson and C. ReVelle, Indirect covering tree problems on spanning tree networks. Research report, Johns Hopkins University, Baltimore, MD (1990).
[19] O. Kariv and S. L. Hakimi, An algorithmic approach to network location problems, II: The p-medians. *SIAM J. Appl. Math.* **37** (1979) 539–560.
[20] T. U. Kim, T. J. Lowe, and J. E. Ward, The minimum cost partial covering subtree of a tree. Working paper, Krannert School of Management, Purdue University, West Lafayette, IN (1989).
[21] T. U. Kim, T. J. Lowe, J. E. Ward, and R. L. Francis, A minimum length covering subgraph of a network. *Ann. Oper. Res.* **18** (1989) 245–260.
[22] T. U. Kim, T. J. Lowe, J. E. Ward, and R. L. Francis, A minimum length covering subtree of a tree. *Naval Res. Log.* **37** (1990) 309–326.
[23] R. K. Kincaid, T. J. Lowe, and T. L. Morin, The location of central structures in trees. *Comput. Oper. Res.* **15** (1989) 103–113.
[24] A. Kolen, *Location Problems on Trees and in the Rectilinear Plane*. Stitchting Mathematisch Centrum, Kruislaan 413, 1098 SJ, Amsterdam, The Netherlands (1982).
[25] A. Kolen, Solving covering problems and the uncapacitated plant location problem on trees. *Eur. J. Oper. Res.* **12** (1983) 266–278.
[26] A. Kolen and A. Tamir, Covering problems. *Discrete Location Theory* (P. Mirchandani and R. Francis, Eds) Wiley, New York (1990) Chap. 6.
[27] J. Krarup and P. Pruzan, Selected families of location problems. *Ann. Disc. Math.* **5** (1979) 327–387.
[28] R. C. Larson and A. R. Odoni, *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, NJ (1981).
[29] L. McGinnis, A survey of recent results for a class of facility location problems. *AIIE Trans.* **9** (1977) 11–18.
[30] E. Minieka, *Optimization Algorithms for Networks and Graphs*. Marcel Dekker, New York (1978).
[31] E. Minieka, The optimal location of a path or tree in a tree network. *Networks* **15** (1985) 309–321.
[32] P. B. Mirchandani and R. Francis, *Discrete Location Theory*. Wiley, New York (1990).
[33] C. Morgan and P. Slater, A linear time algorithm for the core of a tree. *J. Algor.* **1** (1980) 247–258.
[34] M. Richey, Optimal location of a path or a tree on a network with cycles. *Networks* **21** (1990) 391–407.
[35] P. Slater, Locating central paths in a graph. *Transport. Sci.* **16** (1982) 1–18.
[36] A. Tamir, A unifying location model on tree graphs based on submodularity properties. Research report, Tel Aviv University, Tel Aviv, Israel (1990).
[37] B. C. Tansel, R. L. Francis, and T. J. Lowe, Location on networks, A survey, Parts 1 and 2. *Management Sci.* **29** (1983) 482–511.

Received October 1990
Accepted August 1991