A new approach for a restricted concentrator location problem

Karina Marcus Sylvia Boyd

1 Introduction

In any network where a large number of widely dispersed "users" share a limited number of "resources", the strategy for access will play a large part in determining the cost and performance of the network. The users may include time-sharing terminals, terminals used for message transfers, remote automatic sensing devices (such as might be found in an environment-monitoring situation), manned sensing stations and several other. The resources may be as sophisticated as many heterogeneous computer tied together in a packet-switching high-level subnet, or as simple as a single computer processing data received from automatic remote sensing devices. An almost endless number of user and resource combinations, both covering and extending the range described above, appear possible. Effective, economical user access will depend on the development of hardware to facilitate access, prococols to ensure satisfactory operation, and topological design techniques to efficiently utilize the hardware.

In this paper we consider a restricted version of the problem of locating "access facilities", or concentration points, to permit economical connection of users to resources. This problem has been already studied [16]. Actually, we consider only one resource that does not have a important place in the optimization problem. This assumption leads to a strong relation with the problem of finding a minimum r-rooted 2-height spanning tree, giving as result conceptually easy heuristics, that are easilly implemented as well.

In the next section we describe in details the restricted concentrator location problem, and some related problems often found in the literature. In section ?? the integer formulation of the problem is given and in section ?? the r-rooted 2-height spanning tree problem is introduced. Two heuristcs deriving from this formulation of the problem are presented in section ??, and the results comparing lower bounds and approximate solutions are presented in section ??. We close the paper with some conclusions and directions for further research.

2 The restricted concentrator location problem

The special version of the concentrator location problem that is considered in this work is the following. We are given a network of n users that communicate one to another using existent links between them, and can be congested. To introduce this congestion factor into the network, to each existent link (i, j) between users i and j is associated a penalty $p(i, j)$ – a high penalty indicates that the link is often used and may easily be congested.

We want to place some concentrators in this network, concentrators that will treat information coming from the users and will send them to a center through some means of communication, for which the distance and capacity involved are not important, for instance, a wireless one. We are allowed to place a concentrator only within an existent user, and the cost of placing a concentrator within a user depends on the user. The users may route their messages to a concentrator through other users, using the existent links.

The objective is to assign concentrators to some users in the network in such a way to minimize the total cost of the concentrators, without congesting too much the network links, when using them intensively to send messages from users to concentrators. What done in order to avoid higher congestion in the final assignment is to transform the link penalties into link costs, applying a function expressing how much it would "cost" to allocate an often congested link to route messages to a concentrator (for instance, this cost can reflect the cost of duplicating the link or increasing its capacity). Thus we want to minimize the total cost of concentrators plus the cost of the selected links.

One should remark that the center does not play a role in the cost function to be minimized, but the cost of placing a concentrator on a user may reflect its distance to the centralconcentrator. This is a factor that shall be handled by the expert estimating the concentrator costs.

This special concentrator location problem is tightly linked to the network access planning, the simple plant location problem and the p-median problem.

3 Related problems

Many problems having a similar definition have already been well studied. In the following we describe four related problems, and compare them to the concentrator location problem, highlighting the differences and similarities of the descriptions.

3.1 Concentrator location problem in computer network design

[16] One of the most well known and important research questions in computer network design is the *concentrator location* problem. In this problem, we have a large number of users at known locations and one central site to which all users must be connected. We wish to design a minimal cost tree network which connects all user to that central site, either directly or via one or more intermediate devices called *concentrators*. (Note that we are using the word concentrator in a generic, rather than technical, sense. Any type of line-sharing device – e.g. multiplexor, cluster controller, minicomputer – will be referred to here as a concentrator.)

The teleprocessing network design problem consists of three main components: (1) selection the number and location of concentrators (*concentrator location*), (2) assigning each user to a concentrator (*user assignment*), and (3) determining how to connect every concentrator to its assigned users (*user layout*).

Thus, teleprocessing design methods apply to local access network planning when we treat users as distribution points and the central computer as the switching center. Most teleprocessing network design methods proposed in the literature first determine the concentrator location and user assignment decision using a single model, called the *Capacitated Concentrator Location Problem* (CCLP), that approximates the actual costs of connecting users to concentrators by (separable) assignment costs.

If we associate plants with concentrators and customers with users, the CCLP is structurally similar to the plant location model.

3.2 Local access network planning

[3] The lowest level of national telephone networks are trees that connect individual customers to the rest of the national network through special nodes known as *switching centers*, which route telephone calls to their final destination. Each local access network (tree) T has its own switching center. As demand for service increases, telephone companies have two basic options for increasing the capacity of a local access network: (1) they can install more copper cables on the arcs of the networks; or (2) they can install devices, called *multiplexors* (or *concentrators*), at the nodes.

The multiplexors compress calls so that they use less downstream cable capacity. We assume that once a call reaches a multiplexor, it requires negligible cable capacity to send it to the switching center. Every call must be routed through the tree T either to the switching center or to one of the multiplexors.

Constructing a concentrator at any node in the telephone network incurs a node-specific cost and assigning each customer to any concentrator incurs a "homing cost" that depends on the customer and the concentrator location.

We say that a node i homes on another node j if the traffic from distribution point i is processed at node j. Node i homes on the switching center (node 0) if the traffic is not processed at any intermediate node. We have the following assumptions:

1. The modes permits at most one level of traffic processing, and assumes a single service type. For simplicity, we assume that traffic can arrive at different frequencies at the switching center.
2. Contiguity assumption: The model assumes that if a node i homes on node j, then all nodes on the (unique) path from i to j also home on node j. We refer to this routing restriction as the contiguity assumption since the set of all nodes homing on a particular processor induces a single contiguous or connected subgraph of the original network.
3. The model does not permit bifurcated routing, i.e., all the traffic originating at a particular node must follow the same route to the switching center(i.e., must use the same links and undergo processing at the same node).
4. The model accounts for "homing costs" when node i homes on a processor located at node j. By selectively setting these homing costs to a high value, we can prohibit homing patters that violate restrictions as proximity and others.

We want to identify the optimal location of concentrators to service these customers (assume that we must assign each customer to one of the concentrators).

3.3 The simple plant location problem

Also known as the uncapacitated facility location problem, in the simple plant location problem facilities of unrestricted size are placed among m possible sites with the objective of minimizing the total cost for satisfying fixed demands specified at n locations. Cost include a fixed charge for opening each facility and a constant amount for each unit of location j's demand supplied from facility i.

Various formulation of this problem have been treated with numerous solution techniques [4, 6, 7, 8, 9, 10, 12, 17], and good surveys on this problem have already been done [5, 13, 18].

Contrarily to our problem, in this one there are demands associated to the nodes, and a location node must be assigned to a facility node. The assumption of having two different sets seems to be also a restriction, but we will see in Section **??** that the concentrator location problem may be reduced to the simple plant location problem; thus all the already developed techniques for this problem could be used to solve the concentrator location problem.

3.4 The p-median problem

We consider a connected undirected graph $G(V, E)$ with weights $w(v)$ associated with each one for its $|V| = n$ nodes, and lengths $l(e)$ associated to each of its $|E|$ edges. Let $X_p = \{x_1, x_2, \ldots, x_p\}$ be a set of p nodes on G. The distance $d(v, X_p)$ between a node of G and a set X_p on G is

$$d(v, X_p) = \min_{1 \le i \le p} \{d(v, x_i)\}$$

where $d(v, x_i)$ is the length of a shortest path in G between vertices v and x_i. For each set $X_p = \{x_1, x_2, \ldots, x_p\}$ on G we define

$$H(X_p) = \sum_{v \in V} w(v).d(v, X_p). \tag{1}$$

The p-median of G is a set X_p^* such that

$$H(X_p^*) = \min_{X_p \text{ on } G} \{H(X_p)\}.$$

This problem has also been well studied as a companion problem to the simple plant location problem [12, 14, 18]

Compared to the concentrator location problem, the p-median problem presents a cost function (1) that includes a proportionality factor $w(v)$ that depends on the node v, but no cost associated to a node chose to be in X_p is explicit. Moreover, this problem is usually given with the fixed parameter p, the number of chose nodes, which is not the case in the concentrator location problem.

4 Simplification of the model and IP formulation

In the definition of the problem we do not assume that all the links are present, that is, we do not have as hypothesis that the corresponding graph is complete. But usually the description of algorithms and even of problems is easier when every node is connected to all nodes.

The concentrator location problem can be modeled using a complete graph in the following way. For each pair of nodes i, j, one computes the value $c(i, j)$ of a minimum cost path using the existent links. We then build a complete graph, with edge costs $c(i, j)$.

In the remaining of the paper we will suppose that the graph is complete, obtained as described above. We should remark that in this context the problem is reduced to find a subset of nodes where to install the concentrators, and to each node that does not contain a concentrator we must assign a "homing node", defining the used link to connect them. Notice also that this link corresponds to a path in the original network, and by the definition of the costs in the complete graph, all the nodes on this path will have the same homing node, providing the equivalent to the contiguity assumption describe in subsection 3.2.

Using this model we can formulate the problem as a simple integer programming problem. We associate a decision variable y_i to each node i, where

$$y_i = \begin{cases} 1, & \text{if there is a concentrator on node } i, \\ 0, & \text{otherwise,} \end{cases}$$

and introduce decision variables x_{ij} such that

$$x_{ij} = \begin{cases} 1, & \text{if the link } (i,j) \text{ is used to route messages,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that this definition implies that $x_{ij} = x_{ji}$. We will uses indistinctively x_{ij} and x_{ji}, but only one is actually stored.

The objective function to be minimized is

$$\min \sum_{i=1}^{n} f_i y_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} c_{ij} x_{ij}$$

subjet to

$$\begin{array}{l} \sum_{i=1}^{n} x_{ij} + y_i \geq 1 \\ y_i + y_j \geq x_{ij} \\ y_i \in \{0,1\}, \quad x_{ij} \in \{0,1\} \end{array}$$

since each node must home a concentrator or be assigned to a concentrator, and the link (i,j) can be used if and only if there is a concentrator in one of its endpoints.

Remark

We can reduce the concentrator location problem to the simple plant location problem in the following way. Each node is subdivided into two, one facility node $i \in I$ and one demand node $i \in J$. The fixed costs are the costs of opening the concentrators, and the cost of serving a demand node j by a facility node i is $c(j,i)$. To make sure that a node that contains a concentrator should not be served by another node, we just set $c(i,i) = 0$, for all node i.

Thus there is also the following integer programming formulation obtained directly from the plant location problem:

$$\begin{array}{l} \min \sum_{i=1}^{n} f_i y_i + \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \\ \sum_{i=1}^{n} x_{ij} + y_i = 1 \\ y_i \geq x_{ij} \\ y_i \in \{0,1\}, \quad x_{ij} \in \{0,1\}. \end{array}$$

Notice that in this case the variables x_{ij} and x_{ji} are different, and represent an assignment, not the utilization of the link, as in the previous formulation.

5 Complexity of the problem

Let G be a graph. We will show that the problem of finding a minimum cardinality domination set reduces to a special case of the concentrator location problem.

Let G' be a complete graph on $|VG|$ nodes, $c : VG' \rightarrow \mathbb{Z}_+$ and $l : AG' \rightarrow \mathbb{Z}_+$ such that $c(v) = 1$, for all $v \in VG$, and

$$l(uv) = \begin{cases} 0, & \text{if } \{u, v\} \in AG, \\ M, & \text{if } \{u, v\} \in AG. \end{cases}$$

where $M >> |VG|$.

The solution for the concentrator location problem is a set S such that

1. S contains edges e such that $l(e) = 0$;
2. S minimizes the number of nodes being concentrators;
3. all node that is not a concentrator is adjacent to a node that is a concentrator by an edge e of cost $l(e) = 0$.

So the set of concentrators is a dominating set of minimum cardinality. In the other direction it is easy to see that a minimum domination set is a set of concentrators that solves this concentrator location problem.

Thus we have proved the following.

Proposition 5.1 *The concentrator location problem is NP-complete even if all the fixed node costs are equal to one and the edge costs assume only two values, namely 0 and M, where $M > |VG|$.* □

6 Lower bounds

6.1 Trivial lower bound

By definition, We know that each node has to contain a concentrator or to be linked to one that contains a concentrator. So the trivial lower bound lb_t is the following.

For each $v \in VG$, let

$$l(v) = \min\{c(uv), \text{ for all } u \text{ adjacent to } v;\ f(v)\}.$$

And $lb_t = \sum_v l(v)$.

6.2 Tree lower bound

Create a new node r and join r to all the other nodes, setting $c(rv) = f(v)$. Let G^r be the resulting graph. Find a minimum spanning tree T G^r. Then $lb_T = c(T)$.

Clearly, if the spanning tree contains a path with one end point in r and whose length is greater than 2, then it is not a feasible solution.

Proposition 6.1 $lb_t \leq lb_T$.

Proof.: Let S be the set of edges is G^r corresponding to a subset of nodes and edges that attains the lower bound lb_t. If S does not contain a cycle, then S is a spanning tree. Thus every spanning tree in G^r is a possible set for establishing lb_t, and so $lb_t \leq lb_T$. □

We next analyze the "performance" of this lower bound.

Proposition 6.2 *If z^* is the cost of a best solution for the concentrator location problem, then*

$$z^* < (n-1)lb_T - (n-2)\min_{v \in VG} c(r,v).$$

Proof.: The worst case occurs when the tree T is a path. We will build a feasible solution from this path just by attaching each node to the one already adjacent to r. Let the path be ordered as $r, 1, 2, \ldots, n$ and the edges in T are $\{r,1\}, \{1,2\}, \{2,3\}, \ldots, \{n-1,n\}$. We will construct a feasible solution S with the edges $S = \{\{r,1\}, \{1,2\}, \{1,3\}, \ldots, \{1,n\}\}$.

It is not difficult to see that $c(1,i) \leq c(1,2) + c(2,3) + \ldots + c(i-1,i)$, since the graph is constructed from the shortest paths distances. So

$$\begin{aligned}
c(S) &= c(r,1) + c(1,2) + \ldots + c(1,n) \\
&\leq c(r,1) + c(1,2) + c(1,2) + c(2,3) + \ldots + c(1,2) + \ldots + c(n-1,n) \\
&= c(r,1) + nc(1,2) + (n-2)c(2,3) + \ldots + (n-i)c(i,i+1) + \ldots + c(n-1,n) \\
&= c(r,1) + c(1,2) + \textstyle\sum_{i=1}^{n-1}(n-i)c(i,i+1),
\end{aligned}$$

and so $c(S) < (n-1)c(T) - (n-2)\min_{v \in VG} c(r,v)$.

One can see in Fig. 1 an example where the constructed solution S is in fact optimal, showing that this lower bound can be really underestimating the value of a solution. □

7 Polynomial cases - path,tree (p-median)

Proposition 7.1 *If for all v, $c(v) - c(v^*) > \sum_{v \in VG} c(v, v^*)$, then $S = \{v^*\}$ is a optimum solution for the concentrator location problem.*

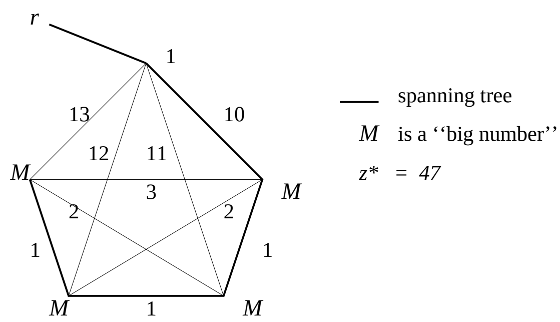

Figure 1: Example for an underestimating lower bound.

8 Relation to the 4-diameter minimum spanning tree

Using the idea of the tree lower bound, on can figure out a relation between the concentrator location problem an the minimum 4-diameter spanning tree, or more precisely, the minimum 2-height r-rooted spanning tree.

Taking the new node r as the root of a 2-height spanning tree, then every such spanning gree corresponds to a solution for the concentrator location problem, since the nodes at height 1 correspond to the nodes homing concentrators, and the nodes at height 2 are those who are linked to some other node containing a concentrator, and viceversa.

Thus, finding a minimum 2-height r-rooted spanning tree in the extended graph is equivalent to finding a best solution for the concentrator location problem.

Not surprisingly, the minimum 2-height r-rooted spanning tree is an NP-hard problem, as it was shown in [11] and [15]. The case when only the diameter of the tree is restricted to be less or equal to 3 is trivial.

Some work in terms of studying the restricted diameter spanning tree was done in [1] and [2], with integer programming formulations and a Branch and Bound concentratord heuristic. In [2] the minimum h-height r-rooted spanning tree was studied, and a mixed integer linear programming formulation is given for this problem, reducing it into a directed graph as follows.

The formulation involves transforming a given weighted graph $G = (V, E)$ with root vertex r into a weightted directed graph $G' = (V, A)$ as follows. Let $V \setminus \{r\} = \{1, 2, \ldots, n-1\}$. Then

$$A = \{(r, j) : j = 1, 2, \ldots, n-1\} \cup \{(i, j) : 1 \leq i \neq j \leq n-1\}.$$

The weight w_{ij} of the arc (i, j) is taken to be the wight of the corresponding edges (i, j) in G; if the edge is not in G, then the corresponding weight is ∞.

The MILP formulation is:

$$\min \sum_{(i,j)\in A} w_{ij} x_{ij}$$

subject to

$$x_{rj} + \sum_{i=1, i\neq j}^{n-1} x_{ij} = 1$$
$$x_{ij} \in \{0,1\},$$

and for all ordered h-subsets $\{i_1, i_2, \ldots, i_h\} of \{1, 2, \ldots, n-1\}$

$$\sum_{t=1}^{h-1} x_{i_t i_{t+1}} - x_{r i_1} \leq h-2$$
$$\sum_{j=1}^{t-1} x_{i_j i_{j+1}} - x_{i_t i_1} \leq t-1, \quad t = 2, 3, \ldots, h-1$$

9 Heuristics

Our approach is to use the relation of the concentrator location problem and the 2-height r-rooted spanning tree to suggest two heuristics.

10 Results

References

[1] N.R. Achuthan and L. Caccetta. Minimum weight spanning trees with bounded diameter. *Australasian Journal of Combinatorics*, 5:261–276, 1992.

[2] N.R. Achuthan, L. Caccetta, P. Caccetta, and J.F. Geelen. Computational methods for the diameter restricted minimum weight spanning tree problem. *Australasian Journal of Combinatorics*, 10:51–71, 1994.

[3] A. Balakrishnan, T.L. Magnanti, A. Shulman, and R.T. Wong. Models for planning capacity expansion in local access telecommunication networks. *Annals of Operations Research*, 33:239–284, 1991.

[4] O. Bilde and J. Krarup. Sharp lower bounds and efficient algorithms for the simple plant location problem. *Annals of Discrete Mathamtics*, 1:79–97, 1977.

[5] M.L. Brandeau and S.S. Chiu. An overview of representative problems in location research. *Management Science*, 35:645–674, 1989.

[6] P.M. Camerini, A. Colorni, and F. Maffioli. Some experience in applying a stochastic method to location problems. *Mathematical Programming Study*, 26:229–232, 1986.

[7] G. Cornuejols, M. Fisher, and G.L. Nemhauser. On the uncapacitated location problem. *Annals of Discrete Mathematics*, 1:163–177, 1977.

[8] G. Cornuejols, G.L. Nemhauser, and L.A. Wolsey. The uncapacited facility location problem. In R.L. Francis and P. Mirchandani, editors, *Discrete Location Theory*. Wiley Interscience, 1990.

[9] M.A. Efroymson and T.L. Ray. A branch-bound algorithm for plant location. *Operations Reseach*, 14:361–368, 1966.

[10] D. Erlenkotter. A dual-based procedure for uncapacitated facility location. *Operations Research*, 26:992–1009, 1978.

[11] M.R. Garey and D.S. Johnson. *Computers and Intractablility – A guide to the Theory of NP-completeness*. Freeman, 1979.

[12] O. Kariv and S.L. Hakimi. An algorithmic approach to network location problem. II: The p-medians. *SIAM Journal on Applied Mathematics*, 37:539–560, 1979.

[13] J. Krarup and P.M. Pruzan. The simple plant location problem: Survey and syntheses. *European Journal of Operations Research*, 12:36–81, 1983.

[14] G.L. Narula, U.I. Ogbu, and H.M. Samuelson. An algorithm for the p-mediam problem. *Operations Research*, 25:709–713, 1977.

[15] C.H. Papadimitriou and M. Yannakakis. The complexity of restricted spanning tree problems. *Journal of the ACM*, 29:285–309, 1982.

[16] G.M. Schneider and M.N. Zastrow. An algorithm for the design of multilevel concentrator networks. *Computer Networks*, 6:1–11, 1982.

[17] K. Spielberg. Algorithms for the simple plant-location problem with some side conditions. *Operations Research*, 17:85–111, 1969.

[18] B.C. Tansel, R.L. Francis, and T.J Lowe. Location on Networks: a survey. Part I: the p-center and p-mediam problems. *Management Science*, 29:482–511, 1983.