ELSEVIER

European Journal of Operational Research 129 (2001) 481–495

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Theory and Methodology

Capacitated facility location/network design problems

Sanjay Melkote [a,*], Mark S. Daskin [b]

[a] Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA

[b] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA

Received 1 November 1998; accepted 1 November 1999

Abstract

We introduce a combined facility location/network design problem in which facilities have constraining capacities on the amount of demand they can serve. This model has a number of applications in regional planning, distribution, telecommunications, energy management, and other areas. Our model includes the classical capacitated facility location problem (CFLP) on a network as a special case. We present a mixed integer programming formulation of the problem, and several classes of valid inequalities are derived to strengthen its LP relaxation. Computational experience with problems with up to 40 nodes and 160 candidate links is reported, and a sensitivity analysis provides insight into the behavior of the model in response to changes in key problem parameters. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Location; Facilities

1. Introduction

The uncapacitated facility location/network design problem (UFLNDP) was introduced by Daskin et al. (1993). This problem was motivated by the simple observation that in some contexts, it may be more economical to change the configuration of the underlying *network* instead of locating new facilities. In general, combined facility location/network design problems are useful for modeling a number of situations in which tradeoffs between facility costs, network design costs, and operating costs must be made. Such situations arise in regional planning problems and in the design of less-than-truckload (LTL) distribution systems, telecommunications networks, airline networks, pipeline systems, and power transmission networks, to name just a few contexts. Melkote and Daskin (1997, 2000) solve instances of the UFLNDP with up to 40 nodes and 160 candidate links and identify a number of polynomially solvable cases of the problem.

The UFLNDP assumes that facilities may serve an infinite amount of demand. This assumption is valid in situations in which it is known in advance that the facilities will operate significantly below their capacity (e.g., fire stations in rural areas of Brazil and Indonesia). However, in many of the above applications, facilities have a constraining

* Corresponding author.

PII: S0377-2217(99)00464-6

upper limit on the amount of demand they can satisfy. For example, in telecommunications and power transmission networks, there are hard constraints on the capacities of concentrators, switches, and transformers. Hard capacities on pumps and storage facilities also exist in pipeline system design problems. Some production plants and warehouses also operate close to their capacity, as do certain types of public facilities (e.g. hospitals or fire stations) in densely populated urban areas. In this paper we introduce a variant of the UFLNDP in which facilities are capacitated. Since the problem may be viewed as a generalization of the classical capacitated facility location problem (CFLP) in which the network topology is determined endogenously, we call it the capacitated facility location/network design problem (CFLNDP).

Most research on the CFLP has focused on the development of efficient solution algorithms. These include branch-and-bound (Akinc and Khumawala, 1977), Lagrangian relaxation (Christofides and Beasley, 1983), Benders decomposition (Davis and Ray, 1969), a combination of the preceding two (Van Roy, 1986), dual-based methods (Guignard and Spielberg, 1979), and heuristics (Jacobsen, 1983). Magnanti and Wong (1990) provide an overview of solution techniques for the CFLP. Variants of the model are generally sparse and include problems in which the facility capacity is determined endogenously given a limited number of capacity options (Jacobsen, 1990; Mukundan and Daskin, 1991). Daskin and Jones (1993) examine a capacitated location model with single-sourcing constraints. Simchi-Levi (1991) studies the "capacitated traveling salesman location problem," in which a service station consisting of mobile servers with finite capacity is to be located. Recently a capacitated location model in which customer demands are stochastic has been investigated by Laporte et al. (1994).

Statement of contributions and outline

The contributions of this paper are, first of all, to introduce the capacitated facility location/network design problem, a model with a number of important applications, and to present a mixed integer programming formulation of the problem. Also, several classes of valid inequalities for the problem are derived, computational experience is presented, and insight into the behavior of the model is provided.

The rest of this paper is organized as follows. In the next section we give a mixed integer programming formulation of our model. We present several classes of valid inequalities in Section 3. In Section 4, we add the inequalities to the original formulation and solve the problem using branch-and-bound. Computational experience on a set of randomly generated test problems is presented. In Section 5, we explore the behavior of the model by performing sensitivity analysis on a problem instance. We draw conclusions in Section 6.

2. Formulation

Prior to formulating the CFLNDP, we make the following assumptions regarding the underlying network: (1) Each node represents a demand point, (2) facilities may only be located at the nodes, (3) only one facility may be located per node, and (4) the network is a customer-to-server system in which the demands themselves travel to the facilities to be served. Assumption 2 is directly from the definition of the CFLP (Magnanti and Wong, 1990), which our model is an extension of. This restriction is made as it is often the case in real life that the candidate locations are restricted to the population centers (for easy access to utilities, etc.), which are typically what the nodes represent. If the travel costs are symmetric, Assumption 4 is not restrictive and an optimal solution to our formulation provides an optimal solution to the corresponding server-to-customer system. To formulate the problem we define the following notation:

Inputs

$\mathscr{N}$	set of nodes in the network
$\mathscr{L}$	set of (undirected) links in the network
d_i	demand at node i
M	$\sum_{i\in\mathscr{N}} d_i$ (total network demand)

t_{ij} travel cost per unit flow on link (i,j)
f_i cost of constructing a facility at node i
K_i capacity of facility constructed at node i
c_{ij} cost of constructing link (i,j)

Decision variables

$$Z_i = \begin{cases} 1 & \text{if a facility is located at node } i, \\ 0 & \text{if not,} \end{cases}$$
$$X_{ij} = \begin{cases} 1 & \text{if link } (i,j) \text{ is constructed,} \\ 0 & \text{if not,} \end{cases}$$
$$Y_{ij} = \text{flow on link } (i,j),$$
$$W_i = \text{total demand served by a facility at node } i.$$

We assume all inputs are nonnegative. Then our formulation is:
(CFLNDP)

$$\text{minimize} \quad \sum_{(i,j)\in\mathscr{L}} t_{ij}Y_{ij} + \sum_{i\in\mathscr{N}} f_iZ_i + \sum_{(i,j)\in\mathscr{L}} c_{ij}X_{ij} \quad (1)$$

subject to

$$\sum_{j\in\mathscr{N}} Y_{ji} + d_i = \sum_{j\in\mathscr{N}} Y_{ij} + W_i, \quad i \in \mathscr{N}, \quad (2)$$

$$W_i \leqslant K_iZ_i, \quad i \in \mathscr{N}, \quad (3)$$

$$Y_{ij} \leqslant MX_{ij}, \quad (i,j) \in \mathscr{L}, \quad (4)$$

$$Y_{ij} \geqslant 0, \quad X_{ij} \in \{0,1\}, \ (i,j) \in \mathscr{L},$$

$$W_i \geqslant 0, \quad Z_i \in \{0,1\}, \ i \in \mathscr{N}.$$

Eq. (2) is a conservation of flow equation, where the inbound flow to a node consists of the total inbound demand plus the demand at the node, and the outbound flow is the total outbound demand plus the demand served at the node. (Recall our system is assumed to be customer-to-server.) Constraint (3) ensures that demand is only served at located facilities and that facility capacities are not exceeded. Eq. (4) states that flow is allowed only on constructed links. Note that this is a weak constraint, as a single link will typically carry far less than the total network demand. Similarly, Eq. (3) will be weak when the demand served at a facility is significantly less than its capacity.

If we set $K_i = M$ for all i (i.e., make all facilities uncapacitated), we have the UFLNDP. Since Melkote and Daskin (2000) have shown that the UFLNDP is NP-hard, we have the following result.

Property 1. *The CFLNDP is NP-hard.*

Also note that our model can easily be used with an existing network. This is accomplished by including in $\mathscr{L}$ any existing links (i,j) and setting their construction costs c_{ij} to zero.

3. Valid inequalities

In this section, we propose a number of valid inequalities to strengthen the LP relaxation of the CFLNDP. Prior to doing so, we first use a modeling device adopted from Magnanti and Wong (1984), Balakrishnan et al. (1989), and Melkote and Daskin (2000) that will aid our development. Specifically, we can define a "commodity" k to be the demand originating at node k. Then we can scale the size of all commodities to unity and define *commodity-specific* travel costs:

$$t_{ij}^k = \text{travel cost of commodity } k \text{ on link } (i,j)$$
$$= t_{ij}d_k.$$

We may then replace the Y_{ij} and W_i variables with the following: Y_{ij}^k is the fraction of commodity k that flows on link (i,j), and W_i^k is the fraction of commodity k served by a facility at node i.

We also henceforth assume that the X_{ij}s represent directed link variables. (Thus for each link $(i,j) \in \mathscr{L}$, we add the link (j,i) to $\mathscr{L}$. Both links have the same construction cost and travel cost as the original link (i,j).) The "direction" of a link simply specifies the direction of the flow carried on it, and does not mean that directed links are actually built. This assumption is used to exploit the structure of the problem.

First, note that (3) is now replaced by

$$\sum_{k\in\mathcal{N}} d_k W_i^k \leqslant K_i Z_i, \quad i \in \mathcal{N}. \tag{5}$$

Next, we use the above definitions to "split up" the conservation of flow constraint (2) for a node into two cases: (a) trans-shipped demand and (b) demand originating at that node.

Proposition 1. *The constraints*

$$\sum_{j\in\mathcal{N}} Y_{ji}^k = \sum_{j\in\mathcal{N}} Y_{ij}^k + W_i^k, \quad i,k \in \mathcal{N}\colon\ i \neq k, \tag{6}$$

$$W_i^i + \sum_{j\in\mathcal{N}} Y_{ij}^i = 1, \quad i \in \mathcal{N}\colon\ d_i > 0, \tag{7}$$

are valid for the CFLNDP.

Constraint (6) states that each demand entering node i is either shipped out or served at i, and constraint (7) states that demand at i must be served either by a facility at i or by another facility.

Proposition 2. *The constraint*

$$\sum_{i\in\mathcal{N}} W_i^k = 1, \quad k \in \mathcal{N}\colon\ d_k > 0 \tag{8}$$

is valid for the CFLNDP.

Eq. (8) also states that all demands must be served.

Proposition 3. *The constraints*

$$Y_{ij}^k \leqslant X_{ij}, \quad (i,j) \in \mathcal{L},\ k \in \mathcal{N}, \tag{9}$$

$$W_i^k \leqslant Z_i, \quad i,k \in \mathcal{N}, \tag{10}$$

are valid for the CFLNDP.

Eq. (9) replaces the weak forcing constraint (4), and (10) is included to improve the bound on Z_i in the LP relaxation of the CFLNDP. (The addition of a constraint similar to (10) to the CFLP has been shown to significantly strengthen its LP relaxation (Van Roy, 1986).)

Proposition 4. *The constraint*

$$X_{ij} + X_{ji} \leqslant 1, \quad (i,j) \in \mathcal{L} \tag{11}$$

is valid for the CFLNDP.

Proof. This constraint precludes oppositely directed flows on a link. See Melkote and Daskin (2000) for a short proof. □

Finally, if we wish to provide the user with control over the number of facilities to locate, we may add the following constraint:

$$\sum_{i\in\mathcal{N}} Z_i = p. \tag{12}$$

The addition of a similar constraint to the UFLP has been shown to strengthen its LP relaxation (Daskin et al., 1993). (In the remainder of this paper we assume (12) has *not* been added to the formulation.)

4. Computational experience

4.1. Test problems

We use the same test problems that Melkote and Daskin (2000) utilized for the UFLNDP. These networks vary in size from 20 nodes and 40 links to 40 nodes and 160 candidate links. The problems are similar in size to those used in the pure network design study of Balakrishnan et al. (1989), which was motivated by transportation, distribution, and telecommunication applications. The test networks were generated using GENRAND2 (Daskin, 1993), a software package designed to generate random networks that simulate real-life transportation networks. GENRAND2 uses the following procedure. First, the desired number of nodes (given in Table 1) are generated on a 100×100 grid, with their x- and y-coordinates randomly sampled from a Uniform$(0, 100)$ distribution. Then the desired number of candidate links (see Table 1) are randomly selected, with a bias toward shorter links to emulate transportation networks. We then calculate the Euclidean length of each link and round it to the nearest integer. The nodal demands are sampled from a Uniform$(10, 40)$ distribution, the fixed facility charges are sampled from a Uniform$(1200, 1500)$ distribution, and both are rounded to the nearest integer.

Table 1
Experimental design (72 problems in total)

Factor	Number of levels	Values of levels
$\lvert\mathcal{N}\rvert$	3	20, 30, 40
$\lvert\mathcal{L}\rvert$	3	$2\lvert\mathcal{N}\rvert,\ 3\lvert\mathcal{N}\rvert,\ 4\lvert\mathcal{N}\rvert$
u	2	2, 10
K	2	$K_1 = K_{\min} + 25,\ K_2 = K_{\min} + 75$
f	2	$f_1 = f,\ f_2 = 1.7f$

We vary the link costs as follows: Two values of a unit link construction cost u are used. In other words, the cost of each link is assumed to be directly proportional to its travel cost, which we may interpret as its length (i.e. $c_{ij} = ut_{ij}$). We assume the facility capacities are the same for all nodes so that K_i is replaced by K. We use two values of K as a function of $K_{\min}$, which we define as the minimum facility capacity needed to serve all demands ($= \sum_{i\in\mathcal{N}} d_i/\lvert\mathcal{N}\rvert$). Further, we weight the facility cost term (the second part of the objective (1)) by two different values. We henceforth refer to these weights as "levels": 100% and 170% of the "base" facility costs determined above. This gives a total of 72 problems. The complete experimental design is shown in Table 1.

4.2. Results

The formulation (CFLNDP) and the valid inequalities (except (12)) were modeled using the AMPL language of Fourer et al. (1993) and solved using the CPLEX MIP solver. We specified the LP relaxations to be solved using CPLEX's dual simplex algorithm, which was found to help avoid degeneracy for the UFLNDP (Melkote and Daskin, 2000). A time limit of 1200 s was imposed on the branch-and-bound algorithm, curtailing the search if a provably optimal solution had not been found within the time limit. Thus whenever the limit was reached, the procedure was essentially used as a heuristic with a guarantee of the maximum deviation from optimality for the best feasible solution it produced. All experiments were conducted on a Sun SPARCstation 10.

Tables 2 and 3 show the maximum deviation from optimality of the best feasible solutions found by the procedure for the two values of f. This percentage is computed as $(z - \mathrm{LP})/\mathrm{LP}\times 100$, where z is the objective value of the best feasible solution and LP is the value of the LP relaxation, and rounded to one place after the decimal point. Branch-and-bound nodes are given in Tables 4 and 5. For several of the problems in which the time limit was reached, we were unable to obtain the exact number of branch-and-bound nodes, as the search was interrupted in the middle of the solution of an LP subproblem at some node. For these problems, only a lower bound on the number of nodes is obtainable from CPLEX's output. (These figures are marked with a †.)

CPU times are presented in Tables 6 and 7. We see that provably optimal solutions were found for over one-third (27 out of 72) of the problems within the time limit, and 21 of them were solved in under 5 minutes. Overall, feasible solutions guaranteed to be within 10% of optimality were found for 68 of the 72 problems, and solutions within 5% of optimality were found for 43 problems.

LP/IP ratios for the formulation of Section 2 are given in Tables 8 and 9, and the ratios for Section 3's formulation are given in Tables 10 and 11. For problems marked with a *, IP is the objective value of the best integer (feasible) solution found. So for these problems the given LP/IP ratios are lower bounds on the actual ratios. All ratios are rounded to two places after the decimal point. We observe that Section 3's formulation provides significantly stronger LP relaxations in almost all cases compared to the formulation of Section 2.

Tables 12 and 13 show the number of multiply sourced demands for the best feasible solution found to each problem. The number of demands that are multiply sourced drops as capacity is increased, as we may expect. Notice that the capacity constraints are not binding in 11 of the problems, as the number of multiply sourced demands is zero in these cases. The number of multiply sourced demands increases with the facility charge level, as fewer facilities are located when the cost of capacity increases. When the unit link construction cost u increases, we observe that the number of demands that are multiply sourced drops. This is

Table 2
Maximum deviation from optimum for $f = f_1$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2 (%)	K_1 (%)	K_2 (%)	K_1 (%)	K_2 (%)
20	2	0	0	0	0	4.2	0
	10	6.4	0	8.7	0	7.5	0
30	2	6.4	0	4.2	0	9.9	0
	10	8.7	0	8.7	3.1	13.6	0
40	2	6.4	0	7.5	2.0	7.5	5.3
	10	8.7	3.1	9.9	7.5	11.1	6.4

Table 3
Maximum deviation from optimum for $f = f_2$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2 (%)	K_1 (%)	K_2 (%)	K_1 (%)	K_2 (%)
20	2	0	0	0	0	0	0
	10	3.1	0	0	0	3.1	0
30	2	8.7	0	2.0	0	2.0	0
	10	11.1	8.7	9.9	2.0	7.5	2.0
40	2	3.1	1.0	4.2	2.0	2.0	8.7
	10	9.9	7.5	7.5	13.6	7.5	9.9

Table 4
Branch-and-bound nodes for $f = f_1$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	1703	98	1153	17	2296*	9
	10	3062*†	313	1745*†	2	1378*†	26
30	2	1291*	34	879*	613	32*†	199
	10	1318*	38	919*	1116*	30*	402
40	2	31*†	147	326*	486*	42*†	101*†
	10	32*†	673*	39*†	413*	276*	315*

* Time limit reached.
† Lower bound.

because it becomes more costly to build transportation infrastructure to source demands at the higher link cost level. As a tradeoff, more facilities are located to serve the same demand. Hence the facilities are less fully utilized, and the capacity constraints (5) are less binding. The increased slack in these constraints is the probable cause of the weaker LP relaxations which result in the greater deviation from optimality for $u = 10$ compared to $u = 2$ (see Tables 2 and 3).

Table 5
Branch-and-bound nodes for $f = f_2$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	984	201	246	31	1401	6
	10	2542*†	458	2775	38	1365*	19
30	2	40*†	588	201*†	259	934*	1036
	10	31*†	14*†	38*†	107*†	4*†	623*
40	2	435*	95*†	51*†	166*	148*†	26*†
	10	223*	22*†	42*†	23*†	53*†	99*

* Time limit reached.
† Lower bound.

Table 6
CPU times (in seconds) for $f = f_1$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	244.30	19.33	254.77	5.12	*	4.42
	10	*	96.40	*	2.28	*	14.20
30	2	*	29.33	*	455.88	*	201.92
	10	*	31.40	*	*	*	372.18
40	2	*	160.48	*	*	*	*
	10	*	*	*	*	*	*

* Time limit reached.

Table 7
CPU times (in seconds) for $f = f_2$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	191.43	72.55	57.13	19.45	471.27	7.60
	10	*	225.62	912.08	23.87	*	15.18
30	2	*	516.92	*	295.47	*	833.72
	10	*	*	*	*	*	*
40	2	*	*	*	*	*	*
	10	*	*	*	*	*	*

* Time limit reached.

5. Sensitivity analysis

To gain further insight into the behavior of the model, in this section we conduct some sensitivity analysis on a particular test problem. This problem, shown in Fig. 1, is a widely used test network (Berman and Rahnama, 1983; Daskin, 1987; Hodgson and Rosing, 1992; Simchi-Levi

Table 8
LP/IP ratios, Section 2's formulation, for $f = f_1$

Nodes $\lvert\mathcal{N}\rvert$	u	Links					
		$2\lvert\mathcal{N}\rvert$		$3\lvert\mathcal{N}\rvert$		$4\lvert\mathcal{N}\rvert$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	0.81	0.56	0.82	0.61	0.73*	0.50
	10	0.81*	0.56	0.74*	0.49	0.76*	0.54
30	2	0.73*	0.50	0.75*	0.53	0.65*	0.43
	10	0.74*	0.50	0.66*	0.45*	0.68*	0.47
40	2	0.71*	0.45	0.76*	0.51*	0.68*	0.41*
	10	0.74*	0.50*	0.65*	0.40*	0.69*	0.44*

* IP = best integer solution found.

Table 9
LP/IP ratios, Section 2's formulation, for $f = f_2$

Nodes $\lvert\mathcal{N}\rvert$	u	Links					
		$2\lvert\mathcal{N}\rvert$		$3\lvert\mathcal{N}\rvert$		$4\lvert\mathcal{N}\rvert$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	0.93	0.78	0.94	0.78	0.93	0.82
	10	0.88*	0.72	0.90	0.73	0.90*	0.77
30	2	0.84*	0.71	0.89*	0.74	0.92*	0.76
	10	0.79*	0.61*	0.79*	0.67*	0.84*	0.70*
40	2	0.88*	0.67*	0.89*	0.72*	0.92*	0.71*
	10	0.79*	0.59*	0.84*	0.61*	0.85*	0.66*

* IP = best integer solution found.

Table 10
LP/IP ratios, Section 3's formulation, for $f = f_1$

Nodes $\lvert\mathcal{N}\rvert$	u	Links					
		$2\lvert\mathcal{N}\rvert$		$3\lvert\mathcal{N}\rvert$		$4\lvert\mathcal{N}\rvert$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	0.97	0.98	0.97	0.98	0.96*	1.00
	10	0.94*	0.96	0.92*	0.99	0.93*	0.99
30	2	0.94*	1.00	0.96*	0.98	0.91*	0.98
	10	0.92*	1.00	0.92*	0.97*	0.88*	0.98
40	2	0.94*	0.99	0.93*	0.98*	0.93*	0.95*
	10	0.92*	0.97*	0.91*	0.93*	0.90*	0.94*

* IP = best integer solution found.

and Berman, 1988). For this application, we have scaled the travel costs up by a factor of 10. The demands, which are shown in parentheses next to each node, are normalized so that they sum to 1000. The fixed facility construction costs are taken from a Uniform(500, 1500) distribution and normalized so that their mean is \$1000. These costs are assigned to the demand nodes in as-

Table 11
LP/IP ratios, Section 3's formulation, for $f = f_2$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	0.99	0.99	0.99	0.99	0.99	1.00
	10	0.97*	0.98	0.98	0.99	0.97*	1.00
30	2	0.92*	0.99	0.98*	0.99	0.98*	0.99
	10	0.90*	0.92*	0.91*	0.98*	0.93*	0.98*
40	2	0.97*	0.99*	0.96*	0.98*	0.98*	0.92*
	10	0.91*	0.93*	0.93*	0.88*	0.93*	0.91*

* IP = best integer solution found.

Table 12
Multiply sourced demands for $f = f_1$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	4	2	5	0	6	1
	10	3	0	2	0	4	0
30	2	7	2	8	3	7	2
	10	4	0	4	0	6	1
40	2	7	3	8	5	9	1
	10	6	3	7	0	7	0

Table 13
Multiply sourced demands for $f = f_2$

Nodes $\|\mathcal{N}\|$	u	Links					
		$2\|\mathcal{N}\|$		$3\|\mathcal{N}\|$		$4\|\mathcal{N}\|$	
	K:	K_1 (%)	K_2	K_1	K_2	K_1	K_2
20	2	5	1	4	0	4	1
	10	4	1	4	0	4	0
30	2	11	3	10	3	10	4
	10	7	2	8	3	8	2
40	2	12	9	12	7	12	7
	10	10	4	9	4	12	3

cending order, that is, they increase with the amount of demand. (This assignment is intended to approximate the phenomenon that land values tend to increase with the population density of towns or cities.) The travel cost t_{ij} of each candidate link may be interpreted as its length. We assume that all travel costs are symmetric so that $t_{ij} = t_{ji}$ for all (i, j). Also, all distances on this network satisfy the triangle inequality. We again assume that we have a unit link construction cost

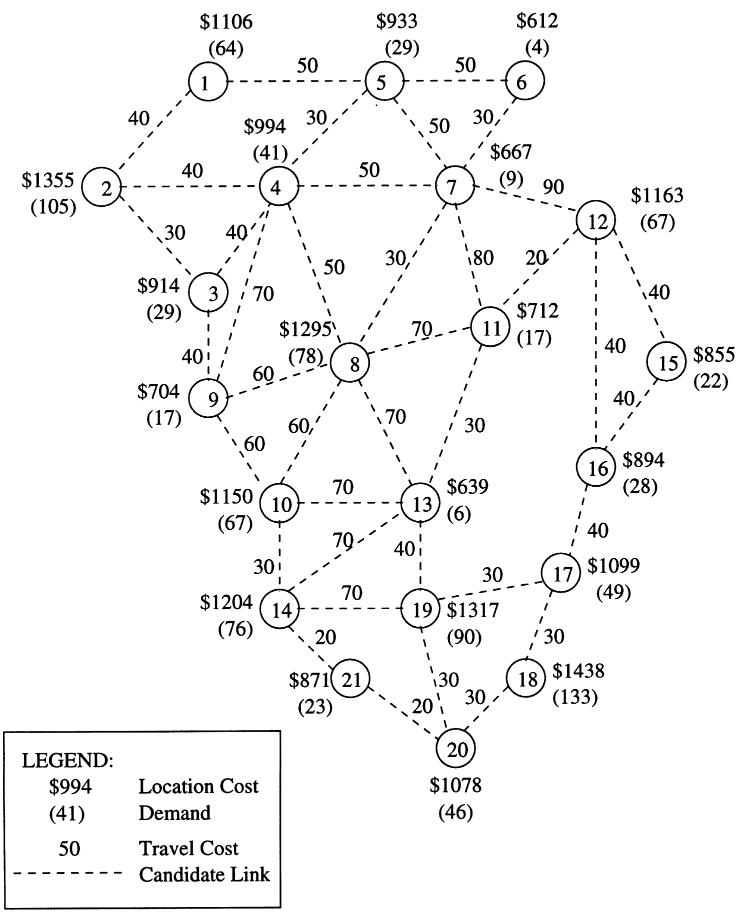

Fig. 1. 21-node network.

u. All facilities to be located are assumed to have the same capacity, so all K_i are replaced with simply K. This parameter is varied from 50 to 250 in increments of 10. Note that 50 is just slightly above $K_{\min}$, which we recall is the minimum facility capacity needed to serve all demands ($= \sum_{i \in \mathcal{N}} d_i / |\mathcal{N}|$). In addition, to assess sensitivity to the cost of this capacity we weight the facility charge term (the third part of the objective (1)) by four different values. We henceforth refer to these weights as levels: 100%, 120%, 144%, and 170% of the base facility charges shown in Fig. 1. These values were chosen so that each level is approximately 20% greater than the previous level (e.g. 144 is 20% greater than 120).

Figs. 2–7 contain plots of the model's output versus the capacity K for each level of facility charge, with the unit link construction cost u fixed at 10 in all cases. (The problem was solved to optimality in all cases.) From Fig. 2 we see that the objective appears roughly piecewise-linear convex in facility capacity. We henceforth denote the capacity after which further capacity increases do not diminish the objective (i.e., after which the problem is essentially uncapacitated), as $K_{\max}$. From the figure, we see that $K_{\max}$ is 140 units for the smaller three fixed charge levels and 150 for the 170% level. For a given capacity, total costs increase with the cost of capacity, as expected.

More interesting is the behavior of the individual components of the objective function.

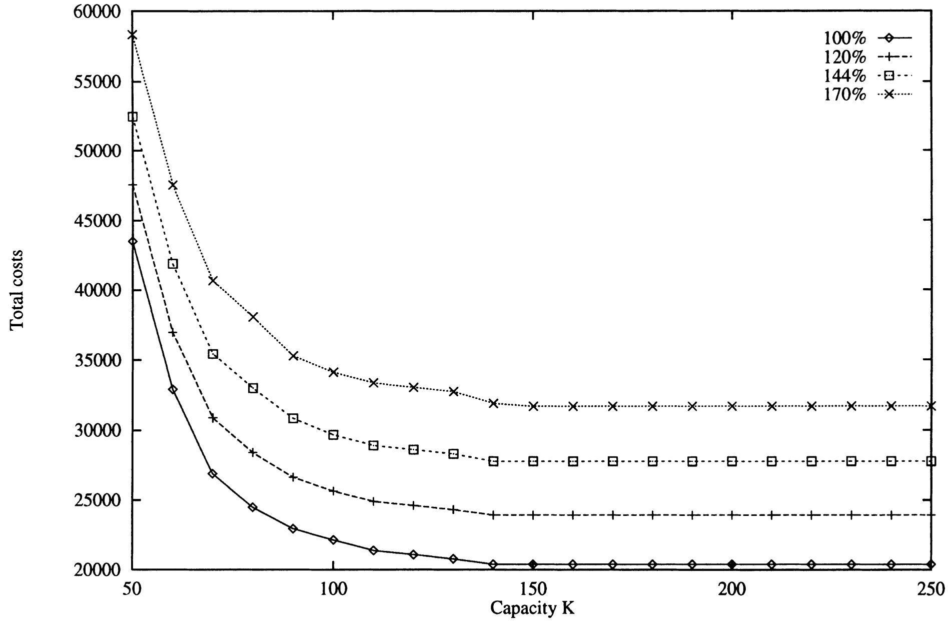

Fig. 2. Total costs vs. capacity by facility charge.

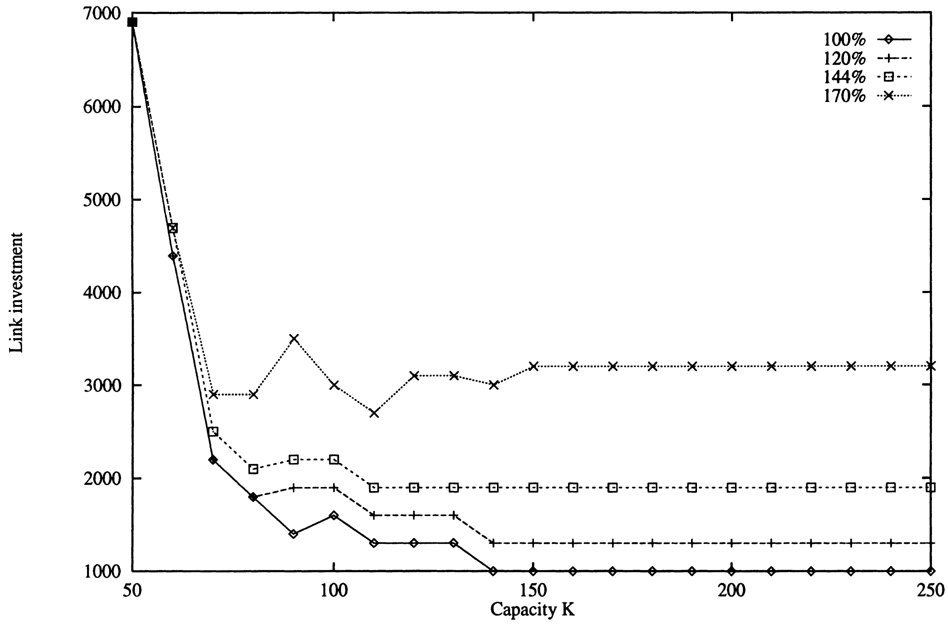

Fig. 3. Link investment vs. capacity by facility charge.

Figs. 3, 4, and 6 respectively display the link and facility investment and transport costs versus the capacity for each level of facility charge. In Fig. 3, the most surprising observation is that link investment does *not* monotonically decrease as capacity increases. Link investment first falls

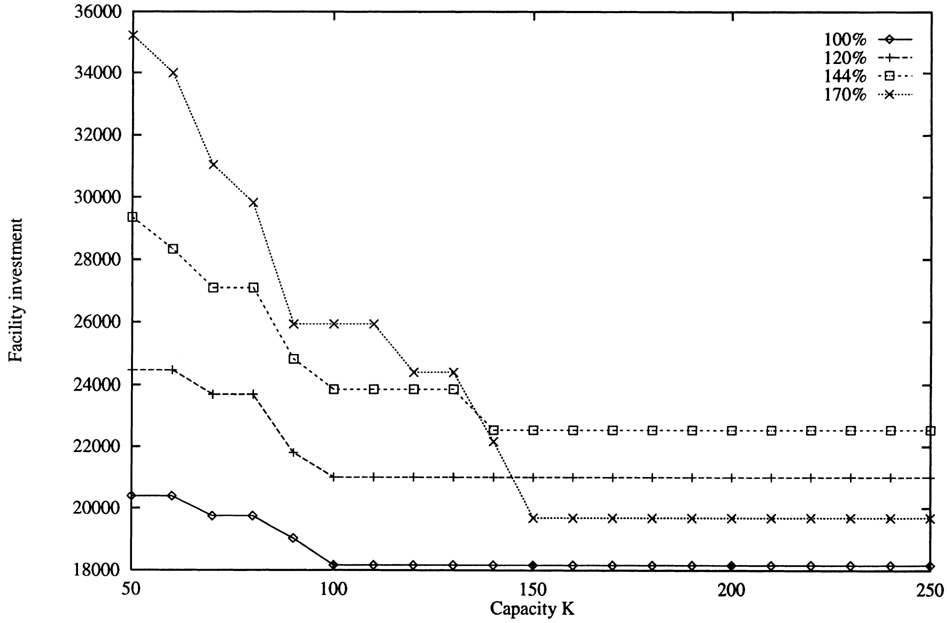

Fig. 4. Facility investment vs. capacity by facility charge.

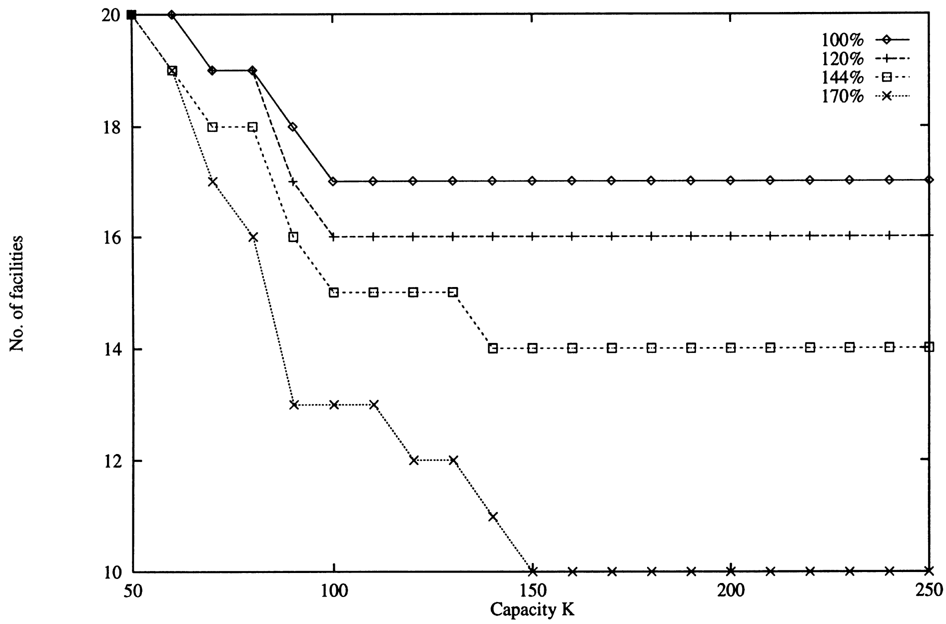

Fig. 5. Number of facilities vs. capacity by facility charge.

sharply as capacity is increased from $K_{\min}$, and then fluctuates significantly before reaching its uncapacitated value. The amount of fluctuation, in fact, appears to increase with the cost of capacity. The explanation for this behavior becomes clearer if we examine Figs. 4–7. When K is close

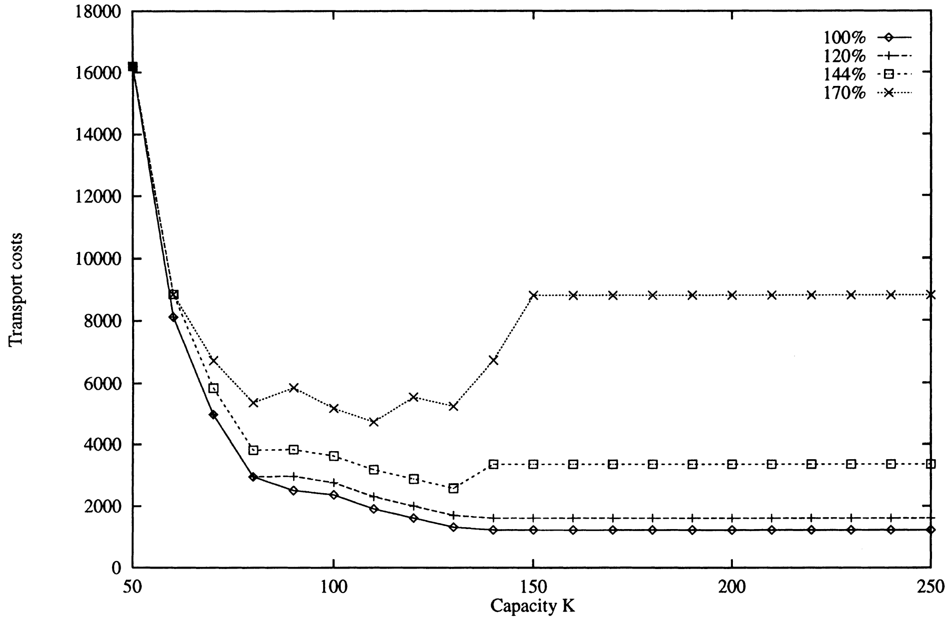

Fig. 6. Transport costs vs. capacity by facility charge.

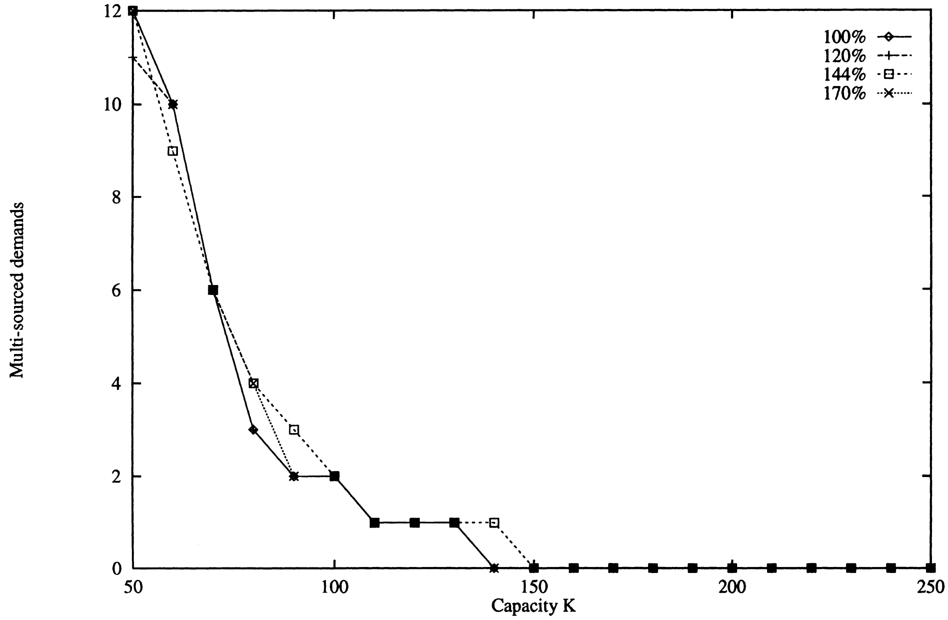

Fig. 7. Multiply sourced demands vs. capacity by facility charge.

to $K_{\min}$, many demands must be served by two or more facilities, which results in a high level of link investment to accommodate the multiple paths. This is evident from Fig. 7, which plots the number of multiply sourced demands as a function of capacity. Then when K is slowly increased, we see a large decrease in the number of demands that are multiply sourced. This allows us to substantially reduce link investment. After these large initial drops, the model significantly reduces investment in *facilities* instead, trading off higher link expenditures to serve the demands that were served at the former facility nodes. This is evident when we observe that for every increase in link investment in Fig. 3, there is a substantial corresponding drop in facility investment and/or the number of facilities located (Figs. 4 and 5). The capacity increase from 80 to 90 at the 170% facility charge level is a good example of this. Clearly, this type of tradeoff is more likely to occur at higher capacity cost levels.

Fig. 4 shows that facility investment monotonically decreases as capacity is increased, as expected, since more capacity allows us to invest in fewer facilities to serve the same demands. Also, facility investment is clearly more sensitive to larger capacity cost levels, as small increases in K can result in a large drop in facility investment. A good example of this is at the 170% facility charge level when capacity increases from 130 to 150. At this level, the model is so sensitive to the facility charge that it prefers a large decrease in these costs at the expense of an increase in both transport costs and link investment (Figs. 3 and 6) to reduce total costs, resulting in an uncapacitated facility investment level *below* that of two lower facility charge levels. (We return to this example shortly.)

In Fig. 5 we see that the number of facilities located also decreases monotonically as capacity is increased, as expected. For a given value of K, the number of facilities decreases when the facility charge level increases. When the cost of capacity is small, we may invest in a large number of facilities at relatively low cost. When the facility charge increases, we are forced to build fewer facilities at the expense of higher link investment and transport costs.

Fig. 6 contains the transportation costs plotted as a function of capacity. The plot is somewhat similar to the behavior of link costs. Most noticeable is that for the larger facility charge levels, transport costs drop sharply when capacity is first increased from $K_{\min}$ and continue to decrease before rising again to their uncapacitated value. This result appears somewhat counter-intuitive at first – increases in facility capacity lead to *increases* in transport costs. Alternatively, this implies that if we consider an UFLNDP and introduce capacities, the result may be a *decrease* in travel costs. The phenomenon is explained as follows. When capacities are imposed on an uncapacitated problem, typically more facilities must be located to serve all demands. As a result, demands are generally closer to facilities, reducing transport costs. As we keep reducing capacities, however, we come to a point at which each facility may serve so few demands that a significant number of demands must be multiply sourced (see Fig. 7). As a result, transport costs start to increase again.

Fig. 7 shows that, as expected, the number of multiply sourced demands is highest when $K = K_{\min}$. At this capacity, about half of the network's demand points are multiply served, and this number drops sharply and steadily as capacity is increased.

6. Conclusions

In this paper we have introduced an extension of the classical CFLP in which the topology of the underlying network is determined endogenously. This problem has a number of important applications in regional planning, distribution, telecommunications, energy management, and other areas. We presented a mixed integer programming formulation of the problem, and several classes of valid inequalities were derived. We observed that over a third of the problems we tested were solved to optimality in under 5 min, solutions within 5% of optimality were obtained for over half of the problems, and solutions within 10% of optimality were found for the vast majority of the problems. Finally, sensitivity analysis on a test problem was conducted to gain insight into the model's behavior. Compared to the uncapacitated model, when capacity constraints are imposed, we observed that (1) both link costs and transport costs may actually decrease, contrary to what we may expect, (2) the network is denser and hence more expensive, as expected, but only as capacity approaches its minimum allowable value, and (3) all three components of the objective function (link costs, facility costs, and transport costs) are more sensitive at higher capacity cost levels.

References

Akinc, U., Khumawala, B.M., 1977. An efficient branch and bound algorithm for the capacitated warehouse location problem. Management Science 23, 585–594.

Balakrishnan, A., Magnanti, T.L., Wong, R.T., 1989. A dual-ascent procedure for large-scale uncapacitated network design. Operations Research 37, 716–740.

Berman, O., Rahnama, M.R., 1983. Optimal path of a single service unit on a network to a "nonemergency" destination. Transportation Science 17, 218–232.

Christofides, N., Beasley, J.E., 1983. Extensions to a Lagrangian relaxation approach for the capacitated warehouse location problem. European Journal of Operational Research 12, 19–28.

Daskin, M.S., 1987. Location, dispatching, and routing models for emergency services with stochastic travel times. In: Ghosh, A., Rushton, G. (Eds.), Spatial Analysis and Location-Allocation Models. Van Nostrand Reinhold, New York, pp. 224–265.

Daskin, M.S., 1993. GENRAND2: A random network generator. Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA.

Daskin, M.S., Hurter, A.P., Van Buer, M.G., 1993. Toward an integrated model of facility location and transportation network design. Working Paper. The Transportation Center, Northwestern University, Evanston, IL, USA.

Daskin, M.S., Jones, P.C., 1993. A new approach to solving applied location/allocation problems. Microcomputers in Civil Engineering 8, 409–421.

Davis, P.S., Ray, T.L., 1969. A branch-bound algorithm for capacitated facilities location problem. Naval Research Logistics Quarterly 16, 331–344.

Fourer, R., Gay, D.M., Kernighan, B.W., 1993. AMPL: A Modeling Language for Mathematical Programming. The Scientific Press, South San Francisco, CA.

Guignard, M., Spielberg, K., 1979. A direct dual method for the mixed plant location problem with some side constraints. Mathematical Programming 17, 198–228.

Hodgson, M.J., Rosing, K.E., 1992. A network location-allocation model trading off flow capturing and p-median objectives. Annals of Operations Research 40, 247–260.

Jacobsen, S.K., 1983. Heuristics for the capacitated plant location model. European Journal of Operational Research 12, 253–261.

Jacobsen, S.K., 1990. Multiperiod capacitated location models. In: Mirchandani, P.B., Francis, R.L. (Eds.), Discrete Location Theory. Wiley, New York, pp. 173–208.

Laporte, G., Louveaux, F.V., Hamme, L., 1994. Exact solution of a stochastic location problem by an integer L-shaped algorithm. Transportation Science 28, 95–103.

Magnanti, T.L., Wong, R.T., 1984. Network design and transportation planning: Models and algorithms. Transportation Science 18, 1–55.

Magnanti, T.L., Wong, R.T., 1990. Decomposition methods for facility location problems. In: Mirchandani, P.B., Francis, R.L. (Eds.), Discrete Location Theory. Wiley, New York, pp. 209–262.

Melkote, S., Daskin, M.S., 1997. Polynomially solvable cases of combined facility location–network design problems. Networks, Under review.

Melkote, S., Daskin, M.S., 2000. An integrated model of facility location and transportation network design. Transportation Research, Forthcoming.

Mukundan, S., Daskin, M.S., 1991. Joint location/sizing maximum profit covering models. INFOR 29, 139–152.

Simchi-Levi, D., 1991. The capacitated traveling salesmen location problem. Transportation Science 25, 9–18.

Simchi-Levi, D., Berman, O., 1988. A heuristic algorithm for the travelling salesman location problem on networks. Operations Research 36, 478–484.

Roy, T.J., 1986. A cross decomposition algorithm for capacitated facility location. Operations Research 34, 145–163.