

Facility Location and Capacity Acquisition: An Integrated Approach

Vedat Verter

Faculty of Business, University of Alberta, Edmonton, Alberta, T6G 2R6, Canada

M. Cemal Dincer

*Department of Industrial Engineering, Bilkent University, 06533,
Bilkent, Ankara, Turkey*

The facility location and capacity acquisition decisions are intertwined, especially within the international context where capacity acquisition costs are location dependent. A review of the relevant literature however, reveals that the facility location and the capacity acquisition problems have been dealt with separately. Thus, an integrated approach for simultaneous optimization of these strategic decisions is presented. Analytical properties of the arising model are investigated and an algorithm for solving the problem is devised. Encouraging computational results are reported. © 1995 John Wiley & Sons, Inc.

1. INTRODUCTION

A company might consider investing in the construction of new facilities for a variety of reasons, such as, increasing its production capacity of an existing product, or extending its product range by new product introduction, or entering new markets with the existing and/or new products. Here *facility* refers to the smallest productive entity that manufactures a single commodity (or, at most a single family of commodities). A *plant* however, refers to a collection of facilities in the same location, and hence in general will be producing multiple commodities. Construction of a new facility therefore, might mean expansion of an existing plant if it takes place at that site, or otherwise would require opening a new plant.

In many investment projects, decisions regarding the location and the size of a new facility to be established are interrelated since capacity acquisition costs are location dependent. A typical example is new facility investments in the international context, where subsidized financing as well as low tax rates are provided by the national governments to attract the multinational companies to locate production plants in their country. In this case, it is clear that not only the fixed costs that occur due to opening the new facility at a particular site but also the capacity acquisition costs that vary with the size of the new facility are location specific. In a recent review however, Verter and Dincer [20] pointed out that the facility location and the capacity acquisition problems have been dealt with separately in the literature. That is the capacity acquisition costs are not incorporated in the facility location models which implies an implicit assumption that they would be the same at all sites. Therefore, the information provided by a facility location model regarding the size of a new facility might be far from optimal when the above assumption is not valid.

Whereas, the capacity expansion models dwell on a given set of existing plants which may not necessarily contain the optimal site to locate the new facility. Thus, separate treatment of the facility location and capacity acquisition decisions could be justified only under quite stringent assumptions.

The aim of this study therefore, is to present an integrated approach for the problem of simultaneously deciding the optimal location and size of each new facility to be established. Thus, the remainder of this article is organized as follows: Section 2 provides the problem definition and a model formulation. The relevant literature is reviewed in Section 3. Analytical properties of the model for the uncapacitated version of the problem are explored in Section 4 which constitute the foundations of the algorithm presented in Section 5. Computational results obtained via the implementation of this algorithm are reported in Section 6. Section 7 justifies the applicability of the same algorithm (with a minor adjustment) for also solving the capacitated version of the problem. Finally, some concluding remarks are provided which suggest directions for further research.

2. THE FACILITY LOCATION AND CAPACITY ACQUISITION PROBLEM

Given a set of alternative facility locations and a set of markets to be served, the *facility location and capacity acquisition problem* involves simultaneously locating an undetermined number of new facilities, and deciding their size to minimize the total cost of serving the clients. For each alternative location the cost items are:

- Fixed setup cost of establishing a new facility at that site,
- Variable cost of capacity acquisition associated with the size of the new facility, and
- Variable costs of operation and transportation for serving the markets.

Fixed setup costs such as the acquisition of land and the construction of infrastructure will be incurred only if a new facility is opened. Capacity acquisition costs would normally represent economies of scale, and hence would be monotone increasing concave functions of the capacity to be built-in at each new facility. Although it is also possible that the operation and transportation costs represent economies of scale, we presume that this would not have a considerable effect on the strategic location and sizing decisions. Therefore, variable costs of operation and transportation are assumed to be linear.

The facility location and capacity acquisition problem is by definition concerned with a single commodity. Furthermore, the problem is deterministic, static, and has no transshipment points. Note that the problem boils down to the uncapacitated facility location problem (UFLP) when the capacity acquisition costs are ignored. UFLP is shown to be NP-complete by Krarup and Pruzan [14] which in consequence means that the facility location and capacity acquisition problem also belongs to the NP-complete class of problems.

Let n denote the number of markets (indexed by j) and m denote the number of alternative facility locations (indexed by i). The problem can be modeled as follows:

$$\text{Minimize} \quad z = \sum_{i \in I} \left[F_i Y_i + f_i \left(\sum_{j \in J} X_{ij} \right) + \sum_{j \in J} c_{ij} X_{ij} \right], \quad (1)$$

$$\text{subject to} \quad \sum_{i \in I} X_{ij} = D_j, \quad j = 1, \dots, n, \quad (2)$$

$$0 \leq X_{ij} \leq Y_i D_j, \quad i = 1, \dots, m, j = 1, \dots, n, \quad (3)$$

$$Y_i \in \{0, 1\}, \quad i = 1, \dots, m, \quad (4)$$

where

- I, J = the sets of alternative facility locations and markets respectively,
- F_i = the fixed setup cost of opening facility i ,
- $f_i(\cdot)$ = the total capacity acquisition cost at facility i ,
- c_{ij} = the unit cost of producing and shipping from facility i to market j ,
- D_j = the demand of market j ,

and the decision variables are

- X_{ij} = the quantity shipped from facility i to market j ,
- Y_i = 1 if facility i is opened, 0 otherwise.

It should be emphasized that F_i excludes any costs associated with the capacity of facility i , and obviously will depend on whether the new facility is established at an existing plant or it requires construction of a new plant. Thus, the total cost that is incurred due to the location and sizing of the new facilities as well as the facility-market allocations is minimized, while constraints (2) guarantee that each market's demand will be fully satisfied, and constraints (3) ensure that markets receive shipments only from open facilities.

The above mathematical program models the problem where there are no upper bounds on the capacity acquired at the new facilities, which we call the uncapacitated facility location and capacity acquisition problem (UFL&CAP). If the size of each new facility however, is constrained due to a variety of reasons e.g., availability of land, then the following constraints are appended to the model:

$$\sum_{j \in J} X_{ij} \leq CAP_i \quad i = 1, \dots, m \quad (5)$$

where, CAP_i is the maximum capacity that can be built-in at facility i . This version of the problem is called the capacitated facility location and capacity acquisition problem (CFL&CAP). It is evident that CFL&CAP boils down to the capacitated facility location problem (CFLP) when the capacity acquisition costs are ignored. Note that, it is possible to interpret CFLP as the problem of locating an undetermined number of new facilities where the capacity of each alternative facility is given, and hence capacity acquisition costs are incorporated in the fixed setup costs. It should be emphasized that however, for a new facility, such a predetermined size might be far from optimal.

3. REVIEW OF THE RELEVANT LITERATURE

In this section the relevant literature is reviewed in order to provide the reader with further insight about the problem we address here. Note that, the mathematical program

presented in Section 2 is also suitable for modeling the concave cost facility location problem. It is possible to show this by redefining $f_i(\cdot)$ as the total cost of operating facility i , and c_{ij} as the unit cost of transport from facility i to market j . If the parameters are defined as above then (1)–(5) constitutes the concave cost capacitated facility location model. Therefore, the techniques available for solving the concave cost facility location problem are also suitable for solving the facility location and capacity acquisition problem.

The literature on concave cost facility location however, is rather sparse. In their seminal work on UFLP, Efroymson and Ray [4] suggested that if the cost functions are piecewise linear then the concave cost uncapacitated facility location problem can be formulated as an UFLP by associating a separate “facility” with each segment. Khumawala and Kelly [13] suggested a heuristic approach to the uncapacitated case where operating costs are represented via power functions. The earliest algorithm that guarantees an optimal solution with no further assumptions regarding the form of the concave cost functions is due to Soland [18]. He devised a branch-and-bound algorithm, where the nodal problems can be solved by inspection in the uncapacitated case. They are however, in the form of the standard transportation problem when there are capacity constraints on the volume of production. At each node, linear underestimates (chords) are used to approximate the concave operating costs (including the fixed costs), and hence a lower bound on the problem solution is obtained by solving the associated linear program (LP). Branching involves partitioning the constraint space by the aid of the LP solution, and hence narrowing the relevant range of the concave cost functions in the subsequent nodal problems. Thus, the chord inscribed in a facility’s operating cost curve is replaced with two adjacent chords which results in a better approximation. Therefore, the algorithm generates progressively better lower bounds. Khumawala and Kelly [13] however, observed the efficiency of tangent line approximation in assigning the facilities to markets for a given set of open facilities. Whereas, in the presence of fixed setup costs of opening new facilities, Kelly and Khumawala [12] had to resort to a combination of the tangent line and chord approximations. Their algorithm for the capacitated problem seems to have less memory requirements than that of Soland [18] since only the previous solution needs to be retained at each iteration. Note that objective function approximation is a well-studied technique in the general area of global optimization. The reader is referred to Geoffrion [9] and Pardalos and Rosen [17] for the use of approximation methods in solving problems with nonlinear objective functions.

It is also possible to cast the facility location and capacity acquisition problem as a minimum concave cost network flow problem (MCNFP). Figure 1 depicts a network flow representation of the UFL&CAP. Here, facilities acquire the necessary capacity from a single source by paying the associated costs, to be able to produce and ship the market demand. CFL&CAP can also be represented by the same network by imposing upper bounds on the flows from the capacity source to the facilities. Therefore, the techniques available for solving MCNFP are also suitable for solving our problem.

Guisewite and Pardalos [10] provided an extensive review of the solution techniques for MCNFP. There are a variety of algorithms based on branch-and-bound, dynamic programming, and extreme point ranking. Falk and Soland [7] presented a branch-and-bound algorithm for minimizing a separable nonconvex function over a linear polyhedron, e.g., MCNFP. At each node, a linear underestimate of the nonconvex function is minimized over the associated partition of the constraint space. Note that, Soland [18] involves minimization of the associated linear underestimate over the entire constraint space at each

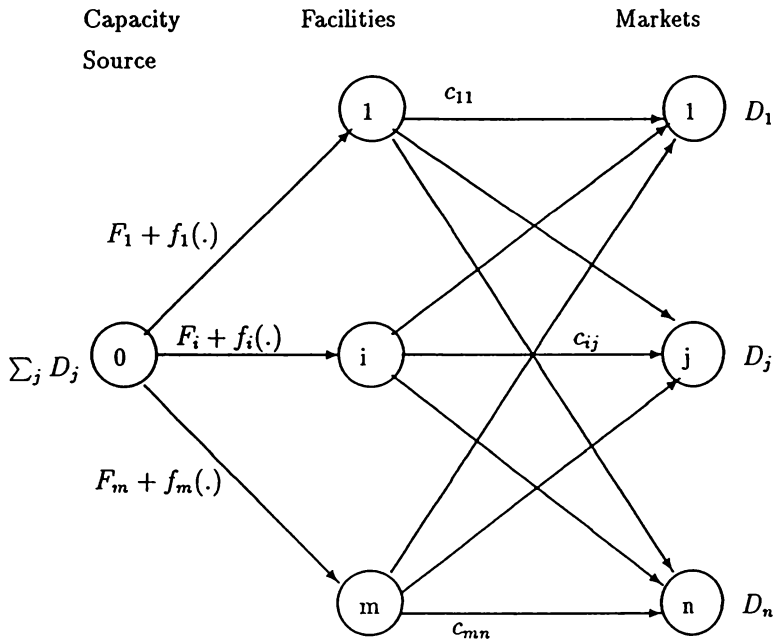


Figure 1. Network flow representation of UFL&CAP.

node, and hence constitutes a simplified version of Falk and Soland [7]. Florian and Robillard [8] however, suggested transforming the original network to an equivalent uncapacitated bipartite network using Wagner's [21] transformation. Thus, their branching involves forcing flow either on an arc in the original network, or on a "slack" arc. Nevertheless, their algorithm requires implicit enumeration of quite large branch-and-bound trees for modest problems, since the number of arcs and the number of nodes increase considerably due to the transformation.

Erickson, Monma, and Veinott [5] provided a dynamic programming approach for the single-source uncapacitated version of MCNFP that is called the send-and-split method. The algorithm involves recursively solving subproblems $l \rightarrow L$, where l denotes a node which is assumed to have a preflow that is sufficient for serving demand of the nodes in set L , in order to obtain a minimum cost solution to the original problem $0 \rightarrow J$. At any node l , *sending* corresponds to shipping all the demand of L to another node l' which is accessible from l , whereas *splitting* corresponds to the decision to serve two complementary subsets of L via separate descendants. The complexity of this method arises from the splitting operation. Nevertheless, this technique is shown to be polynomial in the number of (transshipment) nodes, and arcs, but exponential in the number of demand nodes.

Evidently, a local optimum is not necessarily a global optimum for MCNFP, due to concavity of the objective function. It is also well known however that if a finite optimum solution exists then there exists an extreme flow i.e., an extreme point of the polyhedron defined by the set of linear constraints, that is optimal. Extreme point ranking methods dwell on this property in searching for a local optimum. That is, based on an initial extreme solution, at each iteration, the adjacent vertices are enumerated for moving (if possible) to

the best adjacent vertex. Guisewite and Pardalos [11] provided a computational comparison of the various local search algorithms for the single source uncapacitated MCNFP.

The review of the relevant literature reveals that although there exist a variety of techniques applicable for solving the facility location and capacity acquisition problem, the ones that exploit the problem structure are not that many. Furthermore, since Soland's [18] computational experiments were confined to the fixed-charge cost structure, and Kelly and Khumawala [12] provided only a numerical example, the computational performance of the existing algorithms that recognize the problem structure remains to be investigated for problems of the type commonly encountered in practice.

4. ANALYTICAL PROPERTIES OF THE UFL&CAP

Given that each market is accessible by at least one alternative new facility, there always exists a feasible solution to the UFL&CAP since there are no constraints on the capability of the new facilities to serve the market demand. Further, since the associated MCNFP does not contain any negative cost cycles, the UFL&CAP always has a finite optimum solution.

Efroymsen and Ray [4] observed that for any given set of open facilities, the optimal allocation decisions for UFLP can be obtained by allowing each market to be supplied from the "closest" facility. Such a *dominant* facility has the least unit variable cost of serving the market among the open facilities. Existence of a dominant facility for each market leads to a significant increase in the computational efficiency of their branch-and-bound procedure. This is because dominance enables decomposition of a nodal problem into n easily solved subproblems. Note that, each market's dominant facility among the set of open facilities is decided independent of the quantity demanded. Linearity of the variable supply costs is a sufficient condition for this property to be present which is called *consistent dominance*.

It is possible to show that a different version of the dominance property holds for UFL&CAP where the effective variable costs of serving the markets are no longer linear due to the presence of concave capacity acquisition costs. That is, for a given set of open facilities, the optimal sizing and allocation decisions for UFL&CAP can still be obtained by allowing each market to be served by its dominant facility. In this case however, "closeness" of a facility to a market depends on the demand to be served in addition to the unit variable costs of production and transportation. Thus, the dominant facilities and hence the optimal facility-market allocations might differ due to variations in market demand while costs remain the same. We call this *conditional dominance*.

To formalize the conditional dominance property, let $\{1, 2, \dots, m'\}$ be the set of *open* facilities to serve a single market j with demand $D_j = D$. Further, let

$$g_i(X_i) = f_i(X_i) + c_{ij}X_i, \quad i = 1, 2, \dots, m' \quad (6)$$

represent the total cost of providing X_i units of the commodity to market j from facility i ($X_i \geq 0$). Note that, $\sum_{i=1}^{m'} X_i = D$, and $g_i(\cdot)$ are monotone increasing concave functions when $f_i(\cdot)$ and c_{ij} are defined as in Section 2.

THEOREM 1: *Conditional dominance in the UFL&CAP:*

At the optimum solution of the UFL&CAP, market j will be fully served by a dominant facility that varies with D_j , all other parameters remain the same.

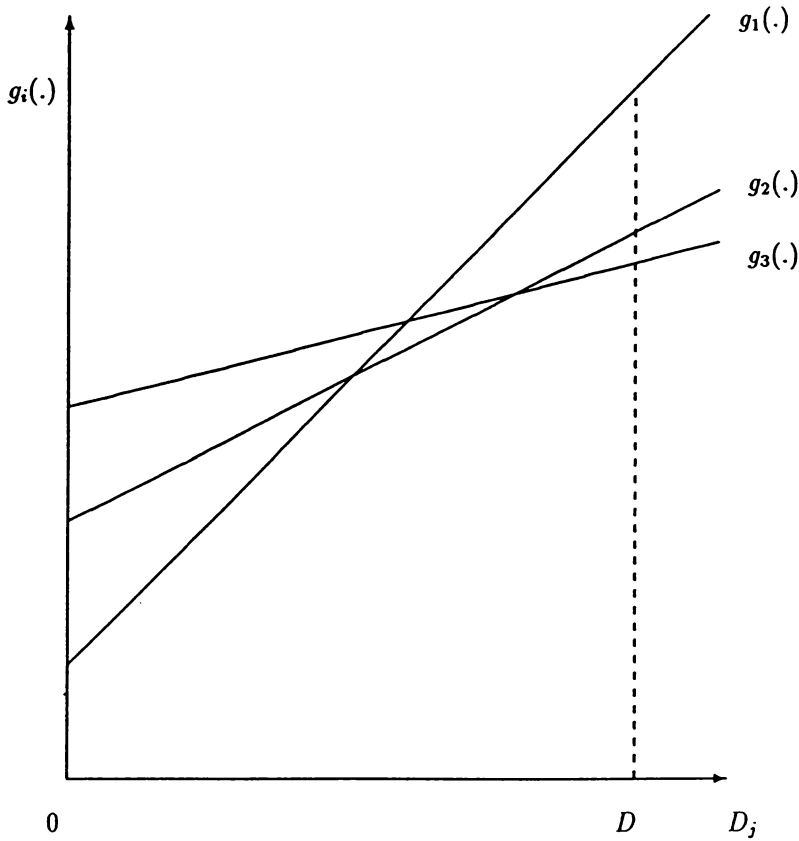


Figure 2. Conditional dominance for fixed-charge costs.

PROOF: It suffices to show that

$$\text{Min} \{ g_1(D), g_2(D), \dots, g_{m'}(D) \} \leq \sum_{i=1}^{m'} g_i(X_i). \quad (7)$$

Since the cost functions are concave,

$$g_i(X_i) \geq \frac{D - X_i}{D} g_i(0) + \frac{X_i}{D} g_i(D), \quad i = 1, 2, \dots, m'.$$

Summing over the open facilities, and realizing that $g_i(0) = 0$,

$$\sum_{i=1}^{m'} g_i(X_i) \geq \sum_{i=1}^{m'} \frac{X_i}{D} g_i(D) \geq \text{Min}_i g_i(D) \sum_{i=1}^{m'} \frac{X_i}{D} \quad (8)$$

This proves (7) since $\sum_{i=1}^{m'} X_i = D$ \square

Theorem 1 shows that market j will be fully served by the facility that has the minimum total cost of providing D_j among $\{1, 2, \dots, m'\}$. For the ease of exposition, Figure 2 depicts

the presence of conditional dominance for fixed-charge costs in the three-facility case. Although, facility 3 happens to be the dominant facility for serving market j when $D_j = D$, it should be realized that first, facility 2 and then facility 1 become dominant as D_j decreases. Clearly, in an UFL&CAP where there are more than one markets, each market's dominant facility depends on the total demand to be served rather than only the demand of that market, due to the scale economies in capacity acquisition costs.

The presence of conditional dominance is very useful in the characterization of the set of alternative sizes for a facility in the UFL&CAP. Since each market will be fully served by *one* facility at the optimum solution, each facility can fully serve all possible combinations of the markets. Thus, any feasible solution to the UFL&CAP that contains more than one facility serving a market qualifies to be nonoptimal irrespective of the costs incurred. Therefore, there are at most 2^n alternative sizes for any facility i , despite the fact that the acquired capacity is represented by a continuous variable (i.e., $\sum_{j \in J} X_{ij}$) in the mathematical model. Note that, this result is in parallel with Zangwill [22] stating that an extreme flow of the single-source uncapacitated MCNFP would include at most one arc (with a positive flow) entering each node. Zangwill's property however, does not provide any insight about the relation between the arcs included in an extreme flow and the market demand.

5. AN ALGORITHM FOR SOLVING THE UFL&CAP

In the UFL&CAP, the total capacity acquisition cost is a separable concave function. This enables utilization of an underestimate that is also separable in terms of the facilities. Thus, the cost of providing capacity at each facility, i.e., sum of the fixed setup and the capacity acquisition costs, is approximated by a piecewise linear concave function. At each iteration of the algorithm, a *pseudo-facility* is associated with each segment of the current linear underestimate for each facility. Hence, the UFL&CAP is transformed to an equivalent UFLP on the basis of the piecewise linear approximation. Therefore, the algorithm devised for solving the UFL&CAP involves solving a sequence of UFLPs. Optimum solution of an UFLP however, corresponds to an extreme flow of the network underlying the UFL&CAP (see Figure 1). This is because in an extreme flow of the network, there will be one facility serving each market, which is in parallel with the dominance property as discussed in the previous section. At each iteration, the underestimate is improved by the aid of the optimum solution to the associated UFLP, and hence progressively better lower bounds are provided. Thus, we call this *progressive piecewise linear underestimation technique*. The algorithm finds a global optimum of the UFL&CAP in a finite number of iterations since the number of the extreme flows of the underlying network is finite.

At an iteration of the algorithm, let m_i denote the number of pseudo-facilities (indexed by k) associated with facility i . Further let,

- F_{ik} = the fixed setup cost of opening pseudo-facility k of facility i ,
- c_{ik} = the unit cost of capacity acquisition at pseudo-facility k of facility i ,
- R_{ik}, R_{ik+1} = the lower and upper bounds on the size of pseudo-facility k of facility i respectively.

Thus, m_i represents the number of linear segments in the current approximation of the cost of providing capacity at facility i . Observe that the pseudo-facilities represent size

ranges for a facility, and hence at most one pseudo-facility associated with the facility must be open in a feasible solution to the UFL&CAP. For pseudo-facility k of facility i , F_{ik} is the intercept obtained by extending the associated linear segment back to the y-axis whereas, R_{ik} and R_{ik+1} are the endpoints of the associated partition of the x-axis, and c_{ik} is the slope of the linear segment. On the basis of the current approximation the UFL&CAP can be modeled by the following mathematical program:

$$\text{Minimize } z_L = \sum_{i \in I} \sum_{k \in K_i} \left[F_{ik} Y_{ik} + \sum_{j \in J} c_{ijk} X_{ijk} \right], \quad (9)$$

$$\text{subject to } \sum_{i \in I} \sum_{k \in K_i} X_{ijk} = D_j, \quad j = 1, \dots, n, \quad (10)$$

$$0 \leq X_{ijk} \leq Y_{ik} D_j, \quad j = 1, \dots, n, i = 1, \dots, m, k = 1, \dots, m_i, \quad (11)$$

$$Y_{ik} \in \{0, 1\}, \quad i = 1, \dots, m, k = 1, \dots, m_i, \quad (12)$$

$$Y_{ik} R_{ik} \leq \sum_{j \in J} X_{ijk} \leq Y_{ik} R_{ik+1}, \quad i = 1, \dots, m, k = 1, \dots, m_i, \quad (13)$$

$$\sum_{k \in K_i} Y_{ik} \leq 1, \quad i = 1, \dots, m, \quad (14)$$

where

K_i = the set of pseudo-facilities associated with facility i ,
 c_{ijk} = the unit cost of serving market j from pseudo-facility k of facility i i.e.,
 $c_{ijk} = c_{ik} + c_{ij}$,

and the decision variables are

X_{ijk} = the quantity shipped from pseudo-facility k of facility i to market j ,
 Y_{ik} = 1 if pseudo-facility k of facility i is opened, 0 otherwise.

The optimal value of z_L constitutes a lower bound on the optimal solution value of the UFL&CAP. Constraints (10) ensure that each market's demand will be fully satisfied, constraints (11) guarantee that markets receive shipments only from open pseudo-facilities. Whereas, constraints (13) ensure that the total production of each open pseudo-facility is between its lower and upper bounds, and constraints (14) specify that at most one pseudo-facility can be open associated with each facility. Due to the concavity of the piecewise linear underestimates however, the constraints (13) and (14) are redundant in the above formulation. That is for each facility, cost minimization will automatically select the correct pseudo-facility which corresponds to the size range that contains the optimal size of the facility. Evidently, the remaining integer program constitutes a classical model for an-UFLP with $\sum_{i=1}^m m_i$ "facilities" and n markets. As pointed out by Verter and Dincer [20] however, there are very efficient techniques available for solving the UFLP such as the dual-based optimization procedure of Erlenkotter [6].

The Algorithm

Step 0. Initialize.

Input: For $i \in I, j \in J$, get $F_i, f_i(\cdot), c_{ij}, D_j$.

Set total demand, $TD = \sum_{j \in J} D_j$.

Initialize the arrays:

For $i \in I$, set

$m_i = 1, R_{i1} = 0, R_{i2} = TD, F_{i1} = F_i,$

$c_{i1} = (f_i(TD) - f_i(0))/TD,$

For $j \in J$, set $c_{ij1} = c_{i1} + c_{ij}$.

Label the current pseudo-facility as pseudo-facility 1 of facility i .

Step 1. Solve the UFLP.

For $i \in I$, For $j \in J$, For $k \in K_i$, obtain X_{ijk}^*, Y_{ik}^* , and the associated value of z_L^* that constitute the optimum solution of the UFLP corresponding to the current approximation.

For $i \in I$, For $j \in J$, set $X_{ij}^* = \sum_{k=1}^{m_i} X_{ijk}^*$.

Calculate the implied size for each facility:

For $i \in I$, set $SZ_i = \sum_{j=1}^n X_{ij}^*$.

Step 2. Improve the approximation.

For $i \in I$, For $k \in K_i$,

if $SZ_i \in (R_{ik}, R_{ik+1})$,

then partition this size range to generate two new pseudo-facilities k' and k'' to replace the pseudo-facility k . Set

$R_{ik'} = R_{ik}, R_{ik'+1} = R_{ik} = SZ_i, R_{ik''+1} = R_{ik+1},$

$c_{ik'} = (f_i(SZ_i) - f_i(R_{ik}))/ (SZ_i - R_{ik}),$

$c_{ik''} = (f_i(R_{ik+1}) - f_i(SZ_i))/ (R_{ik+1} - SZ_i),$

$F_{ik'} = F_i + f_i(R_{ik}) - c_{ik'} R_{ik},$

$F_{ik''} = F_i + f_i(SZ_i) - c_{ik''} SZ_i,$

set $m_i = m_i + 1,$

Relabel the pseudo-facility k' as k, k'' as $k + 1, k + 1$ as $k + 2$ and so forth.

Rearrange the arrays accordingly.

Step 3. Terminate.

if no new pseudo-facility is generated

then For $i \in I$ For $j \in J$,

Output: X_{ij}^*, Y_i^*, z_L^* as the global optimum of the UFL&CAP,

Terminate.

else For $i \in I$, For $k \in K_i$, set $c_{ijk} = c_{ik} + c_{ij}$,

Go to Step 1.

Figure 3 depicts an iteration of Step 2 for facility i . If at any iteration SZ_i equals to one of the interval endpoints (i.e., $R_{ik}, k = 1, \dots, m_i$) then the cost of providing capacity at facility i is exactly represented by the underestimate, and hence no new pseudo-facilities are generated. At each iteration, the value of the original objective function z^* for X_{ij}^*, Y_i^* constitutes an upper bound on the optimal solution value of the UFL&CAP. The availability of the lower and upper bounds provides an indication of the quality of the solution at each iteration. Thus, the algorithm can also be used as a heuristic for solving the

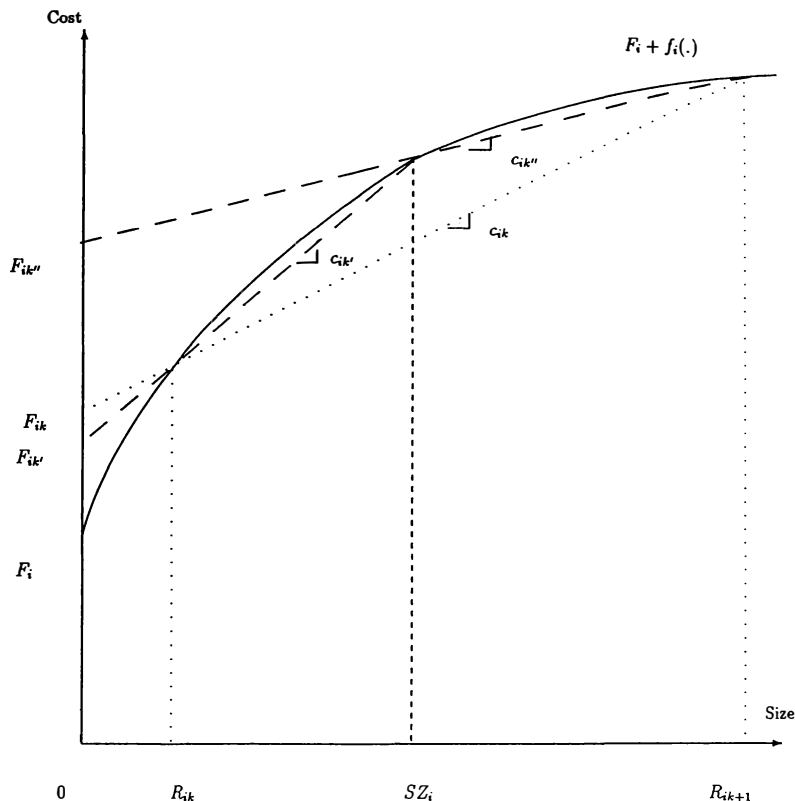


Figure 3. Step 2 of the Algorithm for facility i .

UFL&CAP. Let ϵ denote the maximum acceptable gap between the lower and upper bounds. Due to the separability of both the original objective function and its underestimate let

$$\epsilon = \sum_{i \in I} \epsilon_i, z^* = \sum_{i \in I} z_{Ui}^*, z_L^* = \sum_{i \in I} z_{Li}^*. \quad (15)$$

In a heuristic implementation, no new pseudo-facilities associated with facility i will be generated at Step 2 of the algorithm: if $(z_{Ui}^* - z_{Li}^*)/z_{Ui}^* \leq \epsilon_i$.

6. COMPUTATIONAL RESULTS

The algorithm devised for solving the UFL&CAP was programmed in Standard FORTRAN and tested on a number of problems generated on the basis of the UFLPs drawn from the literature. The current implementation contains the DUALOC code of Erlenkotter [6] for solving the UFLPs formulated during the operation of the algorithm. The computational experiments comprise two parts: First we investigated the average computational performance of the algorithm on Sun Microsystems. Then we used an IBM RISC/

Table 1. Computational results based on *cap71* for $\beta = 20$.

α	Total cost (\$)	# of open facilities	Change in the set of open fac.	# of pseudo-fac.	# of iter.	Total Sun time (sec)
1.00	2097975.8	11		16	1	0.23
0.95	1682095.1	9	{12, 9}	46	5	1.97
0.90	1411790.1	6	{1, 4, 6}	49	7	2.39
0.85	1242238.4	9	[1, 4, 6]	52	8	4.29
0.80	1132823.2	9		47	6	2.07
0.75	1062582.4	10	[9]	42	4	1.20
0.70	1016944.7	10		36	4	1.15
0.65	987540.7	10		34	4	1.11
0.60	968374.6	11	[12]	31	3	0.59

6000 to compare the computational performance of the algorithm with that of the Balakrishnan and Graves [1] algorithm for MCNFP. The demand, fixed setup cost, and variable operation and transportation cost data used in the computational experiments originates from the standard test problems of Kuehn and Hamburger [15] which are available via electronic mail from the OR-Library (see Beasley [3]). Concerning the capacity acquisition costs, Luss [16] pointed out the common use of power functions in the literature as well as the empirical evidence regarding the validity of these functions in modeling industrial problems. Verter and Dincer [20] stated that the fixed charge cost function and the piecewise linear concave cost function are also popular in modeling the capacity acquisition costs. Observe that the former constitutes a trivial case which can be solved in a single iteration by the algorithm. Whereas, in the latter case an equivalent UFLP can be formulated as described in the previous section that again can be solved in one iteration.

In the UFL&CAP the major trade off is between the cost of providing capacity and the operation and transportation costs at each facility. The UFLP named *cap71* in the OR-Library will be used as a basis for an illustrative example. The problem contains 16 alternative facility locations and 50 markets. The fixed setup costs of opening a facility at any location is given as \$7,500 except the already existing facility 11, i.e., $F_{11} = 0$. Further, let the capacity acquisition costs have the following form:

$$f_i \left(\sum_{j \in J} X_{ij} \right) = \beta_i \left(\sum_{j \in J} X_{ij} \right)^{\alpha_i}, \quad i = 1, \dots, m, \quad (16)$$

where $\alpha_i \in [0, 1]$ represents the economies of scale in capacity acquisition at facility i whereas, β_i is a positive scalar for scaling the capacity acquisition cost with respect to the fixed setup cost at that facility. Without loss of generality it will be assumed that $\alpha_i = \alpha$, $\beta_i = \beta$, for, $i = 1, \dots, m$ in investigating the impact of scale economies on the location and sizing decisions. This assumption ensures that the location and sizing decisions are not biased toward establishment of a larger facility at a certain location. For $\beta = 10$, $\alpha = 1$ the optimum solution to the *cap71* constitutes opening facilities 1–4, 6–9, and 11–13 for serving the 50 markets. When there is economies of scale in capacity acquisition and hence $\beta = 10$, $\alpha = 0.8$ however, facility 12 is closed in the optimum solution while its markets are allocated to facilities 4, 7, and 11. Although, the total operation and transportation cost of

Table 2. Computational results based on *cap71* for $\beta = 30$.

α	Total cost (\$)	# of open facilities	Change in the set of open fac.	# of pseudo-fac.	# of iter.	Total Sun time (sec)
1.00	2680655.2	11		16	1	0.25
0.95	2046686.0	6	{12, 9, 1, 4, 6}	43	6	1.70
0.90	1633869.5	5	{2}	43	6	1.90
0.85	1380066.5	5		41	6	1.45
0.80	1223853.8	5		43	6	1.91
0.78	1179137.6	6	[2]	46	6	1.71
0.75	1125328.2	9	[1, 4, 6]	48	7	3.55
0.70	1058199.8	9		42	4	1.20
0.65	1014372.8	10	[9]	36	4	1.20
0.60	985928.7	10		34	4	1.00
0.55	967491.9	11	[12]	33	3	0.78

serving facility 12's markets by facilities 4, 7, and 11 is higher, this is more than compensated by the economies achieved by increasing the size of each of the three facilities.

Table 1 presents the effects of varying the degree of scale economies in *cap71* for $\beta = 20$. Note that $\alpha = 1$ constitutes the case where the marginal cost of capacity acquisition is (a positive) constant, and the total cost decreases as the economies of scale in capacity acquisition increases. The fourth column in the table depicts the changes in the set of open facilities as α decreases. $\{\dots\}$ denotes the set of facilities that are closed due to a decrease in α whereas, $[\dots]$ denotes the set of facilities that are opened for the same reason. The total Sun time (including the input and output) relates to the total number of pseudo-facilities generated during the operation of the algorithm as well as the number of iterations in finding the optimum solution.

Table 2, presents the effects of increasing the weight of capacity acquisition costs with respect to the fixed setup costs in *cap71*. Note that $\beta = 30$ is a sufficiently large factor since for example facility 3 incurs a \$87,319 capacity acquisition cost (for acquiring 21,379 units of capacity to serve 37% of the total market demand for 58,268 units) compared to the \$7,500 fixed setup cost in the optimal solution for $\alpha = 0.8$. As the capacity acquisition costs increase due to the increase in the scaling factor β , the effects of scale economies are magnified. This leads to drastic changes in facility sizes. Note that as scale economies increases at an equal rate at all facilities, its effect in decreasing the number of open facilities gradually vanishes. Increasing the fixed setup costs of opening facilities, results in larger economies of scale in capacity acquisition as depicted by Table 3. Apparently, the demand, and the operation and transportation cost data is the same for all the problems referred in Table 3.

The computational performance of the algorithm is encouraging. All of the test problems mentioned above were solved in at most 8 iterations. Furthermore, for $\beta = 20$, $\alpha = 0.95$, the 25 facility locations problem *cap101* was solved to optimality in seven iterations generating 70 pseudo-facilities which took 6.44 seconds. Whereas, for the same values of α and β , finding an optimal solution to the 50 facility locations problem *cap131* required nine iterations involving the generation of 104 pseudo-facilities in 7.99 seconds. Note that the number of markets is 50 in all the test problems which would require 2^{50} iterations in the worst case. The above analysis also enabled us to observe the existence of a sequence in

Table 3. Computational results for increasing fixed setup costs for $\beta = 20$, $\alpha = 0.95$.

Problem name	F_i	Total cost (\$)	# of open facilities	Change in the set of open fac.	# of pseudo-fac.	# of iter.	Total Sun time (sec)
<i>cap71</i>	7500	1682095.1	9	{12, 9}	46	5	1.97
<i>cap72</i>	12500	1708123.6	5	{1, 4, 6, 2}	37	5	1.35
<i>cap73</i>	17500	1727685.5	4	{7}	35	6	1.05
<i>cap74</i>	25000	1745875.9	3	{8}	24	4	0.66

closing (and then opening) facilities as the scale economies increases. For example, facilities 5, 10, and 14–16 are closed in the optimum solution of *cap71*. Depending on the economies in building larger facilities, the following order can be observed in closing the facilities: 12, 9, 1, 4, 6, 2, 7, 8. Note that this example presumes the equality of scale economies at all the alternative facility locations. Nevertheless, such an analysis is also possible without the equality assumption and would provide the decision maker with valuable information about the priorities of the alternative facility locations. This is especially crucial when there is an upper bound on the number of facilities to be established due to a variety of reasons such as the scarcity of the financial resources or the strategic policies that discourage diversification of operations.

The 16 alternative facility locations and 50 markets UFL&CAP can be cast as a MCNFP with 67 nodes, 816 arcs, and 50 commodities where the required flow to each market from the capacity source is treated as a distinct commodity. In the equivalent MCNFP, the cost functions on the 16 arcs connecting the capacity source to the alternative facility locations are concave and the remaining cost functions are linear. (see Figure 1.) Balakrishnan and Graves [1] developed a composite algorithm for the MCNFP with piecewise linear concave costs that generates both heuristic solutions and good lower bounds. Their algorithm is based on Lagrangian relaxation of the flow conservation constraints in a mixed-integer programming formulation of the MCNFP. A combination of dual ascent and subgradient optimization is used to generate Lagrangian lower bounds, and a local improvement procedure is devised to construct locally optimal solutions from the Lagrangian solutions. The algorithm also contains a problem reduction stage for the improvement of the final lower bound.

In order to analyze the relative performance of our algorithm we generated six test problems by using *cap71* as a basis and varying the capacity acquisition cost functions. In each test problem, all piecewise linear concave cost functions have the same three cost ranges, and the total cost at the endpoint of each cost range is calculated using a power function in the form of (16). Table 4 depicts the range widths (in demand units) and the economies of scale parameters for the test problems. For example, in both *mcnfp1* and *mcnfp2* the range width to total demand ratios are 1/15, 5/15, and 9/15 for the three cost ranges. The optimal solutions to these UFL&CAPs were found by the use of our progressive piecewise linear underestimation algorithm (without an a priori transformation to UFLP). In using the composite algorithm for solving the equivalent MCNFPs, the maximum number of subgradient iterations was set at 100. Other parameters of the algorithm were set as described in Balakrishnan and Graves [1]. The current FORTRAN implementation of the composite algorithm however, assumes that the fixed cost for the first range would be zero for all arcs. Therefore, in order to incorporate the fixed costs of facility location in the

Table 4. Test problem specifications.

Problem name	Total demand	Width of range 1	Width of range 2	Width of range 3	β	α
<i>mcnfp1</i>	58268	3885	19423	34960	30	0.80
<i>mcnfp2</i>	58268	3885	19423	34960	30	0.75
<i>mcnfp3</i>	58268	3885	11653	42730	30	0.80
<i>mcnfp4</i>	58268	3885	11653	42730	30	0.75
<i>mcnfp5</i>	58268	1942	9711	46614	30	0.80
<i>mcnfp6</i>	58268	1942	9711	46614	30	0.75

equivalent MCNFPs, we divided range 1 into two subranges (and in effect, used four cost ranges) and fixed the width of the first subrange as one unit. Table 5 presents the results of our comparative experiments. The considerably higher computational effort required by the composite algorithm in providing acceptable heuristic solutions can be explained by the fact that it is designed for the more general MCNFP, and hence it does not exploit the structure of UFL&CAP.

7. ANALYTICAL PROPERTIES OF THE CFL&CAP

In the CFL&CAP there is an upper bound on the size of each new facility to be built at an alternative location. In terms of the underlying network (see Figure 1), these bounds impose constraints on the flows from the capacity source to the facilities. In consequence, the flow on each arc emanating from a facility is also constrained by the upper bound on the size of the facility. Thus, given that each market is accessible by at least one alternative new facility, the CFL&CAP has a feasible solution only if for each market, the set of facilities that have access to the market can acquire sufficient capacity for serving the demand. If a feasible solution to the problem exists however, then there exists a finite optimum solution.

The piecewise linear concave approximation to the total cost of providing capacity at the new facilities enables the transformation of the CFL&CAP to an equivalent CFLP. Therefore, the application of the progressive piecewise linear underestimation algorithm described in Section 5 for solving the CFL&CAP involves solving a sequence of CFLPs. Evidently, the UFLP solver used at every iteration of Step 1 needs to be replaced by a CFLP solver. Beasley [2] presented an efficient algorithm for solving sufficiently large CFLPs whereas, the cross decomposition method of Van Roy [19] remains to be the most efficient technique for solving moderate size problems.

Table 5. Results of Comparative Computational Experiments.

Problem name	Optimum sol. (\$)	Total IBM time (sec)	Heuristic sol. (\$)	Lower bound (\$)	Total IBM time (sec)
<i>mcnfp1</i>	1213422.9	0.42	1216472.5	1039411.6	81.09
<i>mcnfp2</i>	1112492.2	0.41	1112532.8	1102917.2	105.34
<i>mcnfp3</i>	1213813.1	0.38	1221458.0	1195663.9	109.45
<i>mcnfp4</i>	1117358.4	0.39	1118225.2	1106340.1	95.72
<i>mcnfp5</i>	1211816.9	0.39	1222351.4	1183563.4	111.83
<i>mcnfp6</i>	1117549.2	0.43	1122758.4	957347.3	87.94

Observe that the dominance property does not hold for CFLP, since more than one facility might serve a market in the optimal solution, due to presence of the capacity constraints. Nevertheless, optimum solution of a CFLP corresponds to an extreme flow of the capacitated network underlying the CFL&CAP. In a capacitated network however, a flow is extremal only if at most one of the arcs entering each node has a flow that is strictly between its bounds. The implication of this for the CFLP, and hence for the CFL&CAP is that in the optimal solution although a market might be served by multiple facilities, all of these facilities except at most one will be fully utilizing their capacity in serving the market. Investigation of properties of the extreme flows of the underlying capacitated network provides further insight about the analytical properties of the CFL&CAP. Since the number of extreme flows is finite, the algorithm will find an optimum solution to the CFL&CAP in a finite number of iterations. The following definitions facilitate the analysis:

DEFINITION 1: A facility that fully serves a market's demand is called a *full-server* of the market.

DEFINITION 2: A facility that partially serves a market's demand at its capacity limit is called a *partial-server* of the market.

DEFINITION 3: A facility that serves the remaining demand of a partially served market is called a *remainder-server* of the market.

In an extreme flow of the uncapacitated network associated with the UFLP, and hence the UFL&CAP, each market will have one full-server i.e., its dominant facility. This implies that each facility might be a full-server for all possible combinations of the markets. Thus, the acquired capacity at each facility can take 2^n distinct values. That is at a new facility the optimal size is essentially determined by making a "fully serve/do not serve" decision associated with each market. Note that the number of values that the acquired capacity can take constitutes a worst case bound on the number of iterations of the algorithm required for solving the UFL&CAP.

In an extreme flow of the capacitated network associated with the CFLP, and hence the CFL&CAP, each market will have one of the following:

- 1 full-server,
- 1 to $m-1$ partial-servers and 1 remainder-server,
- 2 to m partial-servers.

This implies that each facility might be one of the following:

- A full-server for some combination of the markets,
- A remainder-server for some combination of the markets,
- A full-server for some combination of the markets, and a remainder-server for a disjoint combination of the markets,
- A partial-server for one of the markets.

Facility i being a remainder-server for market j implies that market j is being partially served by a combination of the remaining $m-1$ facilities. Further, facility i being a full-server for market j is equivalent to facility i being a remainder-server of market j that is

partially served by none of the remaining $m-1$ facilities. Thus, there are actually 2^{m-1} remainder values that a facility can serve. Hence the following can be stated:

PROPOSITION 1: In the CFL&CAP the acquired capacity at a facility can take at most $(2^{m-1} + 1)^n + 1$ distinct values.

PROOF: For each of the n markets, a facility will either be a remainder-server providing one of the 2^{m-1} remainder values or not serve the market. Alternatively, the facility will be a partial-server for one of the markets, and hence its size will be CAP_i . The former implies the first term whereas the latter implies the second term in the expression that constitutes an upper bound on the cardinality of the set of alternative sizes for a facility. \square

Let p denote the number of alternative partial-servers for a market given that one of the facilities is qualified as a remainder-server, and t equal 1 if there are capacity constraints 0 otherwise. The following theorem enables the perception of the UFL&CAP as a special case of the CFL&CAP in terms of the set of alternative sizes for a facility:

THEOREM 2: In the facility location and capacity acquisition problem, the acquired capacity at each facility can take at most $(2^p + 1)^n + t$ distinct values.

PROOF: Observe that $t = 0$ implies $p = 0$ and $t = 1$ implies $p = m - 1$. \square

Note that however, Proposition 1 provides a weak upper bound on the cardinality of the set of alternative sizes in the CFL&CAP. To show this, let ψ denote the actual number of alternative sizes for a facility i.e., in the UFL&CAP $\psi = 2^n$.

PROPOSITION 2: In the CFL&CAP,

$$\psi < (2^{m-1} + 1)^n + 1. \quad (17)$$

PROOF: Let $R_j(S)$ denote the remainder of the demand of market j being partially served by the facilities in the set S . If facility i serves $R_j(S)$ and $R_{j'}(S')$ such that $j \neq j'$ then $S \cap S' = \Phi$ where Φ denotes the empty set. This is because if facility $i' \in S$, that is i' is a partial-server of market j , then $i' \notin S'$. Thus the strict inequality holds. \square

It is possible to provide closed form expressions for ψ in the CFL&CAP when n is small. Let $p = m - 1$, for $n = 2$, $p \geq 1$,

$$\psi = 2(2^p + 1) + \sum_{i=1}^p \binom{p}{i} (2^{p-i} + 1). \quad (18)$$

Whereas, for $n = 3, p \geq 2$,

$$\begin{aligned} \psi = & 4(2^p + 1) + 2 \sum_{i=1}^p \binom{p}{i} (2^{p-i} + 1) \\ & + \sum_{k=1}^{p-1} \binom{p}{k} \left[2(2^{p-k} + 1) + \sum_{i=1}^{p-k} \binom{p-k}{i} (2^{p-k-i} + 1) \right] + 4. \end{aligned} \quad (19)$$

The above expressions constitute enumeration of the possible sizes of a facility based on the properties stated in the proof of Proposition 2. Evidently, as the number of markets in the CFL&CAP increases, it becomes quite cumbersome to provide a closed form expression for ψ . Thus, for example in a CFL&CAP with four alternative facility locations three markets, the facilities can actually be of 170 different sizes compared to the upper bound of 729 provided by Proposition 1. As in the uncapacitated case ψ constitutes a worst case bound on the number of iterations of the algorithm required for solving CFL&CAP.

8. CONCLUDING REMARKS

This article presents an integrated approach for the facility location and capacity acquisition decisions. The arising model requires global minimization of a concave function over the constraint space which constitutes the set of feasible solutions to the problem. When the facility size is unconstrained, the algorithm devised for solving the problem involves solving a sequence of UFLPs. It is shown that the dominance property present in the UFLP also holds for the UFL&CAP, although in a conditional sense. The dominance property however, is instrumental in the characterization of the set of alternative sizes for a facility in the UFL&CAP. Note that the cardinality of this set provides a worst case bound on the number of iterations of the algorithm, required for solving the problem. The computational performance of the algorithm in solving the UFL&CAP is satisfactory. In the capacitated case however, the algorithm involves solving a sequence of CFLPs. Although the dominance property does not hold for the CFLP, the set of alternative sizes for a facility can still be characterized. This provides a framework for the perception of the UFL&CAP as a special case of the CFL&CAP in terms of the alternative sizes of a facility. The computational performance of the algorithm in solving the CFL&CAP is currently an open question.

The proposed algorithm can also be used as a heuristic for solving large size facility location and capacity acquisition problems, since both lower and upper bounds on the optimum solution value are available at each iteration. Further, if it is to be assumed that the capacity acquisition costs are piecewise linear concave functions (as assumed in Balakrishnan and Graves [1] for the arc flow costs in the MCNFP) then the set of alternative sizes provides valuable information for deciding the endpoints of the linear segments. At this stage, it is possible to relax the linearity assumption regarding the variable operation costs which was stated in Section 2 for the ease of exposition. Thus, the economies of scale in operation costs can also be incorporated in the model by redefining $f_i(\cdot)$ as the total capacity acquisition and operation cost at facility i , and c_{ij} as the unit cost of shipping from facility i to market j .

Facility location problems attracted the attention of many operations researchers since

the seminal work of Efroymson and Ray [4]. Despite abundance of the academic literature however, the papers describing real life applications of analytical approaches for locational decisions are not that many. In our view, this is primarily due to the strategic and complex nature of locational decisions, and hence the limitations of the available models in incorporating all of the relevant factors. Thus, the present model is in need of two major extensions to enhance its capability in assisting the strategic decision-making process. First, the model should be generalized for also dealing with the facility relocation and capacity expansion decisions via incorporation of the dynamic nature of the cost and demand parameters. No need to say however, the arising dynamic model will be much more difficult to handle in terms of its computational complexity compared to the static model provided in this paper. Second, the model should be generalized for dealing with multiple commodities which will enable simultaneous optimization of the plant location and sizing decisions. Note that however, there would normally be a set of alternative technologies for producing each family of commodities. Thus, such an extension constitutes a primary step in improving the model to also provide the optimal technology selection decisions. The more general model should be able to provide the optimal location of the plants, including their facility configuration (implying their product-mix) as well as the amount of capacity acquired in terms of each technology at each open plant. The authors' current research focuses on the incorporation of the technology selection decisions in the model presented in this paper.

ACKNOWLEDGMENTS

The authors acknowledge the comments and suggestions of two anonymous referees which improved the article. In particular, one of the referees suggested a shorter proof for Theorem 1. We are also grateful to Anant Balakrishnan and Steve Graves for enabling us to use their software for testing the relative performance of our algorithm.

REFERENCES

- [1] Balakrishnan, A., and Graves, S. C., "A Composite Algorithm for a Concave-Cost Network Flow Problem," *Networks*, **19**, 175–202 (1989).
- [2] Beasley, J. E., "An Algorithm for Solving Large Capacitated Warehouse Location Problems," *European Journal of Operational Research*, **33**, 314–325 (1988).
- [3] Beasley, J. E., "OR-Library: Distributing Test Problems by Electronic Mail," *Journal of Operational Research Society*, **41**, 1069–1072 (1990).
- [4] Efroymson, M. A., and Ray, T. L., "A Branch-and-Bound Algorithm for Plant Location," *Operations Research*, **14**, 361–368 (1966).
- [5] Erickson, R. E., Monma, C. L., and Veinott, A. F., "Send-and-Split Method for Minimum Concave Cost Network Flows," *Mathematics of Operations Research*, **12**, 634–664 (1987).
- [6] Erlenkotter, D., "A Dual Based Procedure for Uncapacitated Facility Location," *Operations Research*, **26**, 992–1009 (1978).
- [7] Falk, J. E., and Soland, R. M., "An Algorithm for Separable Nonconvex Programming Problems," *Management Science*, **15**, 550–569 (1969).
- [8] Florian, M., and Robillard, P., "An Implicit Enumeration Algorithm for the Concave Cost Network Problem," *Management Science*, **18**, 184–193 (1971).
- [9] Geoffrion, A. M., "Objective Function Approximation in Mathematical Programming," *Mathematical Programming*, **13**, 23–37 (1977).
- [10] Guisewite, G. M., and Pardalos, P. M., "Minimum Concave-Cost Network Flow Problems:

- Applications, Complexity and Algorithms,” *Annals of Operations Research*, **25**, 75–100 (1990).
- [11] Guisewite, G. M., and Pardalos, P. M., “Algorithms for the Single-Source Uncapacitated Minimum Concave-Cost Network Flow Problem,” *Journal of Global Optimization*, **1**, 245–265 (1991).
- [12] Kelly, D. L., and Khumawala, B. M., “Capacitated Warehouse Location With Concave Costs,” *Journal of Operational Research Society*, **33**, 817–826 (1982).
- [13] Khumawala, B. M., and Kelly, D. L., “Warehouse Location with Concave Costs,” *INFOR*, **12**, 55–65 (1974).
- [14] Krarup, J., and Pruzan, P. M., “The Simple Plant Location Problem: Survey and Synthesis,” *European Journal of Operational Research*, **12**, 36–81 (1983).
- [15] Kuehn, A. A., and Hamburger, M. J., “A Heuristic Program for Locating Warehouses,” *Management Science*, **9**, 643–666 (1963).
- [16] Luss, H., “Operations Research and Capacity Expansion Problems: A Survey,” *Operations Research*, **30**, 907–947 (1982).
- [17] Pardalos, P. M., and Rosen, J. B., “Constrained global optimization: Algorithms and Applications,” *Lecture Notes in Computer Science* 268, Springer-Verlag, New York, 1987.
- [18] Soland, R. F., “Optimal Facility Location with Concave Costs,” *Operations Research*, **22**, 373–385 (1974).
- [19] Van Roy, T. J., “A Cross Decomposition Algorithm for Capacitated Facility Location,” *Operations Research*, **34**, 145–163 (1986).
- [20] Verter, V., and Dincer, M. C., “An Integrated Evaluation of Facility Location, Capacity Acquisition and Technology Selection for Designing Global Manufacturing Strategies,” *European Journal of Operational Research*, **60**, 1–18 (1992).
- [21] Wagner, H. M., “On a Class of Capacitated Transportation Problems,” *Management Science*, **5**, 304–318 (1959).
- [22] Zangwill, W. I., “Minimum Concave Cost Flows in Certain Networks,” *Management Science*, **14**, 429–450 (1968).

Manuscript received September 28, 1992

Revised October 1994

Accepted October 29, 1994