Efficient Solution Procedure to Develop Maximal Covering Location Problem Under Uncertainty (Using GA and Simulation)

K. Shahanaghi* & V.R. Ghezavati

K. Shahanaghi, Department of Industrial Engineering, Iran University Of Science and Technology, Tehran, Iran
V.R. Ghezavati, PhD student at the same Department, Tehran, Iran

Keywords
maximal covering;
location; genetic algorithm;
simulation;
uncertainty modeling

ABSTRACT
In this paper, we present the stochastic version of Maximal Covering Location Problem which optimizes both location and allocation decisions, concurrently. It’s assumed that traveling time between customers and distribution centers (DCs) is uncertain and described by normal distribution function and if this time is less than coverage time, the customer can be allocated to DC. In classical models, traveling time between customers and facilities is assumed to be in a deterministic way and a customer is assumed to be covered completely if located within the critical coverage of the facility and not covered at all outside of the critical coverage. Indeed, solutions obtained are so sensitive to the determined traveling time. Therefore, we consider covering or not covering for customers in a probabilistic way and not certain which yields more flexibility and practicability for results and model. Considering this assumption, we maximize the total expected demand which is covered. To solve such a stochastic nonlinear model efficiently, simulation and genetic algorithm are integrated to produce a hybrid intelligent algorithm. Finally, some numerical examples are presented to illustrate the effectiveness of the proposed algorithm.

© 2008 IUST Publication, All rights reserved. Vol. 19, No. 4

1. Introduction
This paper examines a class of maximal covering location problem which has wide real-world application. The maximal covering location problem (M.C.L.P) was originally developed to determine a set of facility locations which would maximize the total customers’ demand serviced by the facilities within a predetermined critical service criterion. Obviously, the model can be directly useful to most facility location planning such as warehouses, health-care centers, fire stations, recreation centers, emergency centers, etc. In addition, this model can be applied to data abstraction problems, to portfolio formation [1]. This model maximizes the number of demand points covered with a specified critical coverage or time by a fixed number of facilities. It does not need that all demand points be covered. The M.C.L.P and its many developments compose a significant class of problems in location literature.

Brandeau and Chiu [2] presented a survey of representative problems that have been studied in location research. They recognized more than 50 problem types and indicated how those problem types relate to one another. Galvão and ReVelle [4] proposed a LaGrange an and surrogate relaxations to solve maximal covering location problem. The first method had the integrality property and the surrogate relaxed problem they resolved was the linear programming relaxation of the original 0–1 knapsack problem. The performances of these two methods were compared with 331 test problems in the literature.

The error of these heuristics were very low and did not differ significantly among themselves. Galvão and ReVelle [4] proposed a LaGrange an heuristic for the maximal covering location planning. In this algorithm, the upper bound was specified by a vertex addition and
another heuristic and lower bounds were generated through a sub gradient optimization method. This algorithm was validated by solving 150 instances. For large sized problems a duality gap was computed at the end of the algorithm. Church and ReVelle [5] were the first who introduced the MCLP. In their model, the aim was to maximize the population covered within a desired service distance by optimally locating a fixed number of facilities. ReVelle et al. [6] used some met heuristics methods to solve large scale problems in maximal covering area. They believed that exact methods may be too unwieldy for real-world applications, and so heuristics can allocate faster solution times with sub-optimal results. In their instances exact algorithms are provided by linear programming and branch and bound. Karasakal et al. [7] developed maximal covering location problem as a partial coverage so that coverage can change from full covering to no covering. They also, introduced solution procedure based on LaGrange an relaxation method and compared it with two classical approaches. Richard L. Church et al. [8] developed a notation of a model based on the premise that reserve selection or site prioritization can be defined as a classical covering problem usually applied in many location problems. Specially, they developed a structure of the maximal covering location model to classify sets of sites which represent the maximum possible representation of specific species. Boffey and Narula [9] survey the multi objective covering and routing problems. Karasakal [10] formulated the MCLP in the presence of partial coverage, and developed a solution procedure based on LaGrange an relaxation and showed the result of the approach on the optimal solution by comparing it with the classical methods. Yupo Chan and M. Mahan [11] suggested a variant of the maximal covering location problem to locate up to p signal-receiving stations. In their study the “demands,” called geolocations, to be covered by these stations are distress signals and/or transmissions from any targets. Araz et al., [12] proposed a multi-objective MCLP based on an emergency vehicle location. They considered the maximization of the covered customers and minimization of the total travel distance from the emergency services. Younies and George [13] introduced a zero–one mixed integer formulation for a maximal covering problem where points were covered by inclined parallelograms in a plane. A version of maximal expected covering location problem was proposed by Daskin [14] as an extension of the maximal covering location problem (MCLP) formulated by Church and ReVelle [5], to account for the possibility of server unavailability due to a congested system. This model relates the problem of optimally locating servers so as to maximize the expected coverage of demand, considering the possibility of server unavailability when a is call received by the server. When formulating this problem the author makes 3 simplifying assumptions: servers operate independently, each server has the same busy probability and server busy probabilities are invariant with respect to their location. Luis Gonzalo Espejo et al. [15] proposed a 2-level hierarchical development of the MCLP and also a combined Lagrangean–surrogate (L–S) relaxation to solve the model. Their results were compared with exact results obtained using CPLEX software.

Berman and Krass [16] considered a generalization of the maximal cover location problem which allowed for partial coverage of customers, with the degree of coverage being a non-increasing step function of the distance to the nearest facility. John R. Current and David A. Schilling [17] proposed two bi-criterion routing problems including the median tour problem and the maximal covering tour problem. In both problems the tour must select only p of the n facility on the network. Indeed, both problems have as one of their objectives the minimization of total travel distance. The second objective in both problems is to maximize access to the tour for the facilities not directly on it.

The rest of this paper is as follows: Section 2 presents a new mathematical model for the given problem. Section 3 proposes the solution procedure based on simulation and genetic algorithms. The computational results are illustrated and discussed in Section 4. Finally, the remarking conclusion is given in Section 5.

2. Model Formulation

In this section, we describe in more detail the probability maximal location problem in which we are interested. There is a set of demand points \( N \), at which requests are generated, and a set of locations \( M \), where facilities may be opened. We assume that the requests at a demand point \( i \in D \) are generated independent of the processes at other demand points in \( N \).

We develop a formulation based on the classical \( p \)-median formulation. However, instead of minimizing the total distance, this model maximizes the expected coverage of the demand points by determining one of the selected facility sites, which ensures maximum coverage confidence for each demand point.

2-1 Parameters:

\( d_i \) : The demand for customer \( i \).

\( \theta_j \) : Coverage time for DC \( j \).

\( f_{ij} \) : Random variable denoting traveling time between customer \( i \) and DC \( j \) which has normal distribution function with the mean \( \mu_{ij} \) and variance \( \sigma_{ij}^2 \).

\( Cap_j \) : Maximum capacity for DC \( j \).
P: Total number of DCs to be located.

2-2. Decision Variables:

\[ X_{ij} = \begin{cases} 1 & \text{if demand of customer is satisfied by DC } j, \\ 0 & \text{Otherwise.} \end{cases} \]

\[ U_j = \begin{cases} 1 & \text{if a DC is located at site } j, \\ 0 & \text{Otherwise.} \end{cases} \]

\( p_i \): Total probability of coverage for customer i.

\( p_{ij} \): Probability of coverage customer i by DC j.

A notable point in this model is that all distribution centers can not service all customers because each DC has special coverage time and if a customer isn’t in this coverage, the DC can not service that customer. In this model we consider this parameter that leads to the model that will be more realistic.

We assume that this traveling time is probability and then “covering” or “not covering” a customer by a DC will be probabilistic. So, to compute the objective function, we measure expected value for it. (Amount of demand × Probability of coverage).

So, this constraint must be added to compute probability of coverage for each customer per each DC:

\[ P_j - \Pr(f_j \leq \theta_j) = 0 \Rightarrow \]

\[ P_j - \left( \int_0^{\theta_j} \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_j^2}} dx \right) = 0 \]

\[ P_i - \sum_j P_{ij} \cdot X_{ij} = 0 \Rightarrow P_i = \sum_j P_{ij} \cdot X_{ij} \]  

2-3. Model Description

Max \[ Z = \sum_i \sum_j d_{ij} \times P_i \]  

Subject to:

\[ X_{ij} \leq U_j \quad \forall i \in M, \forall j \in N \]  

\[ \sum_j X_{ij} \leq 1 \quad \forall i \in M \]  

\[ \sum_i X_{ij} \cdot d_{ij} \leq cap_j \quad \forall j \in N \]  

\[ \sum_j U_j = p \]  

\[ P_j - \left( \int_0^{\theta_j} \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_j^2}} dx \right) = 0 \quad \forall i \in M \]

\[ P_i = \sum_j P_{ij} \cdot X_{ij} = 0 \quad \forall i \in M \]  

\[ X_{ij} = (0,1), U_j = (0,1), P_i \geq 0, P_{ij} \geq 0 \]

The objective function maximizes total expected demands which are covered. Set constrain (4) says that a customer can allocate to DC j when that DC is opened. Set constrain (5) ensures that a customer cannot be allocated to more than one DC. Set constrain (6) forces that total demand assigned to DC j must be less that its capacity. Set constrain (7) guarantees that total number of DCs to be located should be equal to \( p \). Set constrain (8) computes probability of coverage customer i per each DC (Based on normal distribution). Set constrain (9) shows total probability of coverage for each customer by all DCs. Set constrains (10) determines type of variables.

3. Solution Procedure

Because this modeling is nonlinear and software such lingo gives us local optimum solution and capacitated location problems are NP-hard, we use hybrid met heuristics to solve this problem. So, in order to solve general stochastic programming models for capacitated network design problem, we use GA and simulation to produce a hybrid algorithm. Contrasting the general GA, this algorithm will reduce the computation greatly, which makes it possible to deal with problems of quite large size by the algorithm.

3-1. Simulation Process

Because of the complexity of the proposed model, we design a simulation model to compute uncertain functions covering or not covering customers by distribution centers. First, we experiment the simulation process to generate a 0-1 matrix, namely \( Z_{ij} \). Also, we define a lower bound for probability of coverage. If probability of coverage is greater than the lower bound, we assume that the customer can be covered by distribution center completely. For instance, if probability of coverage customer 5 by distribution center 8 is 0.78 and we define the lower bound as 0.70, we assume that \( Z_{5,8} = 1 \) and it means that Customer 5 can be covered by distribution center 8 [18].

\[ Z_{ij} = \begin{cases} 1 & \text{if } \Pr(f_{ij} \leq \theta_j) \geq r \\
0 & \text{otherwise} \end{cases} \]

Due to the complexity, we design some stochastic simulations to calculate uncertain functions shown in Figure 1. (Uncertain function is covering or not covering customers by DCs).

After the mentioned simulation process, we can obtain \( Z_{ij} \) matrix. Output of this simulation is input for genetic algorithm. (Output of simulation is \( Z_{ij} \) matrix).
After simulation $Z_{ij}$ can be defined and the model will be changed as follows:

$$Z_{ij} = \begin{cases} 1 & \text{If DC at site } j \text{ can cover customer } i. \\ 0 & \text{Otherwise.} \end{cases}$$

### 3-1-1. Model Linearization

$$\text{MaxZ} = \sum_i \sum_j X_{ij} \cdot d_{ij} \cdot Z_{ij} \quad (12)$$

Subject to:

$$X_{ij} \leq U_{ij} \cdot Z_{ij} \quad \forall i \in M, \forall j \in N \quad (13)$$

$$\sum_j X_{ij} \leq 1 \quad \forall i \in M \quad (14)$$

$$\sum_i X_{ij} \cdot d_{ij} \leq \text{cap}_j \quad \forall j \in N \quad (15)$$

Set constraint (13) says that customer $i$ can be allocated to DC $j$ if that DC is opened and the customer $i$ can be covered by DC $j$. The definitions of the other constraints have been said before.

### 3-2. Genetic Algorithm

GA is a stochastic search and heuristic optimization technique, which has been widely adopted by many researchers to solve various problems. This algorithm was first developed by Holland [19]. It mimics the mechanism of genetic evolution in the biological nature and consists of a population of chromosomes (strings or individuals) that are composed of genes. These genes represent a number of values, called alleles. Each chromosome (genotype) represents one potential solution (phenotype).

#### 3-1-2. Representation of Chromosome

In our GA-based approach, each chromosome or bit string (i.e., an example solution) consists of a matrix which is made of 0 and 1. The dimension of this matrix is equal to:

$$\sum_j U_j = p \quad (16)$$

$$X_{ij} = (0,1), U_j = (0,1) \quad (17)$$

GA does not rely on analytical properties of the function to be optimized (Goldberg [20]). In short, GA has two major processes: 1) GA is iteratively and randomly generating new solutions; and 2) these solutions are checked for the optimality according to predefined fitness functions. This becomes the most powerful principle of GA. It makes them widely suitable for finding an optimal solution in many complex problems, such as the traveling salesman problem (TSP) and any forms of scheduling problems. For more details figure 2 illustrated classical genetic algorithms.
The number of customers in the network \( \times \) number of DCs. The matrix represents the assignments of the customers to DCs and gets the value 1 in element \([ij]\) if customer \(i\) is assigned to DC \(j\). Sample of chromosome structure has shown in figure 3.

**Step 1.** Initialize pop_size chromosomes \( X^k = [x_{ij}^k] \), \( k = 1,2,\ldots,\text{Pop\_Size} \) from the potential region \( \{(X) \mid g_i(x_i) \leq 0, x_i = 1,2,\ldots,n \} \) randomly.

\[
[x_{ij}^k] =\begin{cases} 
1 & \text{If DC at site } j \text{ covers customer } i \text{ in chromosome } k. \\
0 & \text{Otherwise.}
\end{cases}
\]

At first \(p\) DCs which should be opened among all potential DCs randomly are selected and then a customer is selected randomly. In the second phase, for each customer the number of DCs which can cover it is determined. Then for this customer, a random DC from those DCs determined previously is selected randomly and the customer in chromosome \(k\) is allocated to it considering the capacity of DCs. (Because DCs have capacity, some customers cannot be allocated to DCs).

**Step 2.** Compute the fitness of all chromosomes \( V_k, k = 1,2,\ldots,\text{Pop\_Size} \). The rank-based evaluation function is defined as the objective function for chromosome \(k\).

**Step 3.** It is very important to create new chromosomes (i.e., offspring) from the selected chromosomes (called parents) with the current population. This process is carried out by the use of genetic operators, namely crossover and mutation.

Renew the chromosomes \( V'_k, k = 1,2,\ldots,\text{Pop\_Size} \) by crossover operation. In order to determine the parents for crossover operation, we repeat the following process from \(k = 1\) to \(\text{pop size}\): generating a random real number \(r\) from the interval \([0, 1]\), the chromosome \(V_k\) will be selected as a parent provided that \(r < P_c\); where the parameter \(P_c\) is the probability of crossover.

Then we group the selected parents \(V'_1, V'_2, V'_3,\ldots\) to the pairs, \((V'_1, V'_2), (V'_3, V'_4),\ldots\) without loss of generality; let us illustrate the crossover operator on each pair by \((V'_1, V'_2)\). At first, we make a matrix with \(i \times 2j\) dimension. \((i=\text{No. of customers} \text{ and } j=\text{No. of DCs})\). In other words, we accrete parent matrixes and make one matrix.

Then in the new matrix, in each row, if number of \(X_{ij}\) with value 1 is more than one, select one of them randomly and set it to be 1 and set the other to be 0. By this, a customer is not allocated to different DCs. In the new matrix we have at most \(2 \times p\) opened DCs. For
generating offspring we select \( p \) DC from opened DCs in the new matrix, randomly.
In fact we select \( p \) columns from the new matrix. (The note in this procedure is, if column \( r \) \((r \leq j)\) is selected then column \( r + j \) must not be selected and if column \( r \) \((r \geq j)\) is selected then column \( r - j \) must not be selected). By doing the mentioned procedure we can get offspring by crossover operator. An example of designed crossover is illustrated in figure 4.

**Step 4.** Update the chromosomes \( V_k, k=1,2,...,\text{Pop Size} \) by mutation operation. Similar to the proves of selecting parents for the crossover operation, we repeat the following steps from \( k =1 \) to Pop Size: generating a random real number \( r \) from the interval \([0, 1]\), then chromosome \( V_k \) will be selected as a parent provided that \( r < P_m \); where the parameter \( P_m \) is the probability of mutation. For each selected parent: \( X^k = [x^k_{ij}] \), we mutate it in the following way.
Select randomly from opened DCs and name it \( j \) and then select randomly from closed DCs and name it \( j' \). Then with below procedure close DC \( j \) and open DC \( j' \). First for customers allocated to DC \( j \), we set \( X^k = [x^k_{ij}] = 0 \). Then start to allocate unassigned customers to DC \( j' \) considering coverage radius and capacity constraint, randomly. Example of designed mutation is illustrated in figure 5.

**Step 5.** Compute objective function for all chromosomes where the chromosomes \( V_1, V_2, ..., V_{\text{Pop Size}} \) are assumed to have been rearranged from good to bad according to their objective values.

**Step 6.** Select the chromosomes for a new population. The selection process is based on selecting 50% from the best chromosomes and 50% randomly. Thus we obtain pop size copies of chromosomes, denoted also by \( V_k \).

**Step 7.** Repeat the second to seventh steps for a given number of cycles.

**Step 8.** Report the best chromosome \( X^* = [x^*_{ij}] \) as the optimal solution.

A notable point in the mentioned GA’s operators is to generate feasible offspring according to the capacity and the other constraints. So, in this process, it needn’t to repair or reject offspring since they are generated feasible.

**4. Computational Results**

**4-1. Validation Genetic Algorithm**

In order to illustrate the effectiveness of the genetic algorithm, we give some random numerical examples that are performed on a personal computer. All algorithms considered in this study were coded in Visual Basic 6 and run on a Pentium IV PC with 1.5 GHz CPU and 256MB RAM. For this purpose we solve some examples and compare our solutions with global solutions obtained from lingo 8 software. In this section, we just try to validate genetic algorithms alone without using simulation. In other words, in this section, genetic algorithm is applied to solve the second model and we assume that the \( Z_{ij} \) matrix is a deterministic parameter in order to measure just the effectiveness of the genetic algorithm. In the next section we will measure robustness of hybrid genetic.

Consider a manufacturer decides to locate new distribution centers. Assume that there are 25 customers. Indeed, suppose that a decision maker
needs to select distribution centers from 6 potential distribution centers to cover customers. In this study, we consider 30 population sizes and 200 generations for each test problem solved by the proposed algorithm. Customers demand has been generated uniformly from interval \([10, 20]\). Also, DC capacity has been generated uniformly from interval \([5.0, 3.0]\). 

\[ \sum_{i} d_{i} \]

where \( \sum d_{i} \) is the total demand of all customers.

In Table 1, we compare solutions when different problems are taken with the same generations as a stopping rule. It appears that all the objective function and global optimum differ little from each other. In order to account for it, we present a parameter, called the percent error, i.e. \((\text{global optimum} - \text{objective function})/\text{global optimum}\), where the global objective value is obtained from lingo 8 software. The last column named by ‘error’ in table 1 indicates this parameter. It follows from table 1 that the percent error does not exceed 2.31% when different problems are selected, which implies that the genetic algorithm is effective enough to solve model.

**4-2. Sensitivity Analysis for Simulation**

To illustrate the effectiveness of the hybrid genetic algorithm, we solve a special problem for eight times with numerous different runs for the simulation model. We show that this hybrid algorithm is also robust to the simulation settings. Also set parametering on GA and simulation was shown in Table 2.

Consider a manufacturer who wants to locate new distribution centers. Assume that there are 45 customers. Suppose that a decision maker needs to select 5 distribution centers from 12, potential distribution centers to serve customers. In Table 2, we compare solutions when different simulation runs are taken with the same generations as a stopping rule.

It appears that all the maximal covered demand differ little from each other. In order to account for it, we present a parameter, called the percent error, i.e. \((\text{the best objective value} - \text{objective value})/\text{the best objective value}\), where the best objective value is the
maximum of all the ten maximal covered demand obtained above. The last column named by ‘error’ in Table 2 is just this parameter. It follows from Table 2 that the percent error does not exceed 2.18% when different simulations are selected, which implies that the hybrid genetic algorithm is robust to the simulation runs. In other words, differences in simulation runs don’t have any major impact on the final solution. Thus, the hybrid method is robust to the simulation settings and therefore is also effective to solve the model.

<table>
<thead>
<tr>
<th>No. of Customers</th>
<th>P</th>
<th>No. of DCs</th>
<th>Pop_Size</th>
<th>Iteration</th>
<th>No. of Opened DC</th>
<th>Objective Function</th>
<th>The Best Objective</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>2,4,8,10</td>
<td>725</td>
<td>735</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>1,3,8,10</td>
<td>731</td>
<td>735</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>2,3,4,8</td>
<td>719</td>
<td>735</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>2,4,8,11</td>
<td>724</td>
<td>735</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>4,6,8,9</td>
<td>732</td>
<td>735</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>2,3,4,10</td>
<td>720</td>
<td>735</td>
</tr>
<tr>
<td>7</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>2,3,6,8</td>
<td>735</td>
<td>735</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>4,7,8,10</td>
<td>728</td>
<td>735</td>
</tr>
<tr>
<td>9</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>2,4,8,10</td>
<td>721</td>
<td>735</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>5</td>
<td>12</td>
<td>25</td>
<td>300</td>
<td>1,2,4,8</td>
<td>732</td>
<td>735</td>
</tr>
</tbody>
</table>

5. Conclusion
In this paper, we defined notation of location-allocation problem while we have stochastic traveling time between service centers and customers this time is stochastic with normal distribution function. In the proposed model, we assumed that service centers have coverage radius restriction. We formulated the problem as a non-linear integer programming. The advantages of the proposed study are as follows: considering probabilistic coverage radius which yields more flexibility for the model, presenting hybrid method including simulation for realizing uncertainty and genetic algorithm to solve the model. Performance of the solution procedure was verified by randomly generated test problems. These experiments show that the proposed algorithm is efficient to generate high quality solutions in a short period of time. We suggest some new areas to develop the mentioned model: By using this contribution in the integrated supply chain systems, it assumes that such inventory and transportation can be a suitable development as future research for the presented model. Also, considering service level constraint to cover each customer can be an interesting future studies. Because in this model we had traveling time between service centers and customers, considering the time window constraint for customers to get service is another suitable development.

References
[8] Richard, L., Church, David, M., Stoms, Frank, W., Davis; Reserve Selection as a Maximal Covering Location Problem, Biological Conservation, Vol. 76, Issue 2, 1996, pp. 105-112.


