

On optimal location of network equipment in large scale communication networks

Domagoj Mikac, Matija Mikac, Mladen Kos

University of Zagreb,
Faculty of Electrical Engineering and Computing, Zagreb
{domagoj, matija, kos}@tel.fer.hr

Abstract: *In the process of planning a large scale communication networks, much more parameters play an important role than in designing a metropolitan or local area networks. Large scale networks, such as mobile GSM, UMTS or fixed national SDH or WDM transport networks are usually spread over the large geographical area, so the distance and coverage factors are very important. Telecommunication equipment used in such networks is very expensive so the minimal number of network sites and its optimal location is very important issue in achieving a minimal cost network. In this paper a facility location problem and its integer formulation will be explained. AMPL/CPLEX model for solving explained location problem and greedy heuristic will be given and results obtained with both methods will be compared. At the end of the paper the results of example network will be shown, which were obtained by software tool developed at our department. Also there is a real case example in which 45 nodes network was tested on implemented heuristic algorithm.*

Keywords: *facility location, network design, optimization, linear programming, AMPL/CPLEX*

Introduction

In our work we consider the process of planning a large scale communication network. Distances in such networks are significant, and optimal placement of backbone nodes is very important issue. The number of backbone nodes is another parameter that should be minimized. Optimal number of nodes means that all the customers are satisfied and investment from the operator's point of view is minimal. There are numerous methods that address this problem [2],[4],[5],[6],[7]. In our paper we will give a classical formulation of facility location problem using linear/integer programming technique. Also it will be shown that for large scale networks, it is extremely hard to find an exact optimal solution, and the use of suboptimal heuristics algorithm is required. The paper is organized as follows: in the first section a classical integer formulation of facility location problem will be introduced. In our work we consider an uncapacitated facility location problem. We do not place a constraint on capacity of the nodes because in our case, demands were much smaller than currently available commercial network equipment can handle. Therefore, there is no need on placing a capacity constraint, even though it is possible with small changes in algorithm. In third section an AMPL/CPLEX implementation of integer formulation of the problem is given. Given implementation exactly solves the facility location problem. In the third section a greedy heuristic algorithm for solving a facility location problem is described. In fourth section a real case network example is given which justifies correctness of described heuristics.

Integer formulation of facility location problem

Facility location problem can be described as a problem of deciding which locations out of many to choose in order to place the backbone equipment. We are given a set of possible locations for locating facilities $J = \{1, \dots, n\}$, and the set of clients locations $I = \{1, \dots, m\}$. The locations in set J can be viewed as a locations where some operator has offices or buildings, while the locations from set I are the locations with significant concentration of potential customers. It is common sense that equipment should be placed closer to the locations with heavily populated areas, and areas with smaller population can be connected to the backbone nodes located further away. Further on there is a profit variable c_{ij} which is the profit made by satisfying a customer i from the location j . This profit variable

can vary with distance and the amount of bandwidth that user requires. For example if the customer i is far away from the location j and requires a lot of bandwidth than the operator must provide a lot of infrastructure in order to satisfy demands of distance customer. It means that connecting such distance customer to location j is not very profitable for network operator, and operator should consider setting up a backbone node closer to such high demanding customer. On the other side if customer does not require a large amount of bandwidth, then it might be cheaper not to set up a new backbone node but to connect such low demanding customer to the node located further away. Setting up a new node requires some investment from the operator, and for this reason we introduce a new constant assigned to each location f_j . Constant f_j represents a cost of setting up a new telecommunication node at the location j . For the small network the location and number of nodes can be determined manually, but for large scale network a specialized software tools are required. The complexity of the problem is too big for manual computation.

The problem is to choose a set of locations S ($S \subseteq J$) to set up a backbone communication nodes at chosen sites, and assign each client to exactly one site in such manner that the difference between variable profit and fixed installation cost is maximized. The number of optimal locations is not prespecified. In some works the number of required location is specified in advance, but in our case a number of locations is determined from an optimal solution.

Facility location problem can be written as a classical linear optimization problem as follows [1]:

$$\begin{aligned} & \max \sum_{i \in I} \sum_{j \in J} c_{ij} \cdot y_{ij} - \sum_j f_j x_j \\ \text{subject to: } & \sum_{j \in J} y_{ij} = 1 \quad \forall i \in I \end{aligned} \quad (1)$$

$$y_{ij} \leq x_j \quad \forall i \in I, \forall j \in J \quad (2)$$

$$x_j \in \{0,1\} \quad \forall j \in J \quad (3)$$

$$y_{ij} \in \{0,1\} \quad \forall i \in I, \forall j \in J \quad (4)$$

For each potential site $j \in J$, there is a zero-one variable x_j which can take the value 0 or 1. This variable tells us if the site is chosen or not. There is also a zero-one variable y_{ij} for each customer-node pair. This variable tells us weather the particular customer $i \in I$, is connected to $j \in J$. The meaning of the first constraint is that each customer is connected to exactly one backbone node. If we would like to change the constraint so that each customer can be connected to two different backbone nodes (for example for protection purposes), we would have to substitute the right side of the equivalence with number 2 instead of 1. Second constraint ensures that the customer can be connected only to the chosen site, because if the site $j \in J$ is chosen the variable x_j must be set to 1. Constraints (3) and (4) ensure that x and y variables can take only the values 0 or 1.

Solving the linear formulation gives us exact solution of the problem, but since the problem is in the class of NP hard problems, the calculation time significantly increases with number of nodes. The problem can be easily solved using AMPL/CPLEX mathematical modeling language, but for larger models a heuristic algorithms are more applicable.

AMPL/CPLEX implementation of facility location problem

Described integer formulation problem can be easily modeled in AMPL/CPLEX mathematical modeling language [3]. CPLEX solver efficiently solves linear/integer class of problems. For different classes of optimization problems (quadratic, nonlinear, stochastic, etc.) other solvers are available.

In Figure 1, an AMPL model of facility location problem is shown. Results obtained by AMPL/CPLEX, are optimal. At our department we have a student version of the software, which is limited to 300 variables. With 300 variables, large network configurations cannot be tested. We have tested a network made of 6 possible node locations and with 4 customers (each customer is one region), and it required 30 variables, with execution time of few seconds. When we have increased number of customers to 10 and number of potential sites to 10, 74 variables were required and execution time slightly increased. In our work we had a real network of 45 potential sites, and 545

customers, and it was impossible to solve it by using student version of AMPL/CPLEX. Even if we had a full version, we assume that execution time would be too long because such problem would have more than 20000 variables. For this reason we had to implement a software tool which can solve a problem for large number of variables in reasonable amount of time. In the next section a simple heuristic algorithm is described, which can be used for large network models.

```

option solver CPLEX;
set CLIENT;
set SITE;
param f{SITE} >= 0;
param cost {CLIENT, SITE} >= 0;
var x {SITE} binary;
var y {CLIENT, SITE} binary;

maximize UFL:
    sum{i in CLIENT, j in SITE}cost[i,j]*y[i,j]
    - sum{j in SITE}f[j]*x[j];

subject to OnlyOneSite{i in CLIENT}:
    sum{j in SITE}y[i,j] = 1;

subject to SiteMustBeOpen{i in CLIENT, j in
SITE}:
    y[i,j] <= x[j];

```

Figure 1: AMPL/CPLEX implementation of facility location problem

Heuristics for solving facility location problem

In this section a greedy heuristics for solving the uncapacitated facility location problem is represented.

Let set $S \subseteq J$ represents a set of chosen locations. At the beginning the set is empty. In each iteration the site that improves the solution the most is added in the set S . For an set S an objective value is given by:

$$z(s) = \sum_{i \in I} \max_{j \in S} \{c_{ij}\} - \sum_{j \in S} f_j$$

The change in objective value when a new site is included into solution is equal to the difference of the new objective value and the old objective value. This change will be denoted by p_j , where $j \in I \setminus S$ is a new site added to the solution. Further on, for the currently chosen sites, for each customer we can find the best site to connect to. It is the site where operator makes the biggest profit by connecting customer i to the site j , where $i \in I$ and $j \in S$. This profit is given by: $u_i(S) = \max_{j \in S} \{c_{ij}\}$. Specially when S is empty set, $u_i(S) = 0$. From here we can write the final equation for $z(S)$ and $p_j(S)$:

$$Z(S) = \sum_{i \in I} u_i(S) - \sum_{j \in J} f_j$$

$$p_j(S) = \sum_{i \in I} (c_{ij} - u_i(S))^+ - f_j$$

In each iteration of greedy heuristics, $P_j(S)$ is computed for a previously changed S .

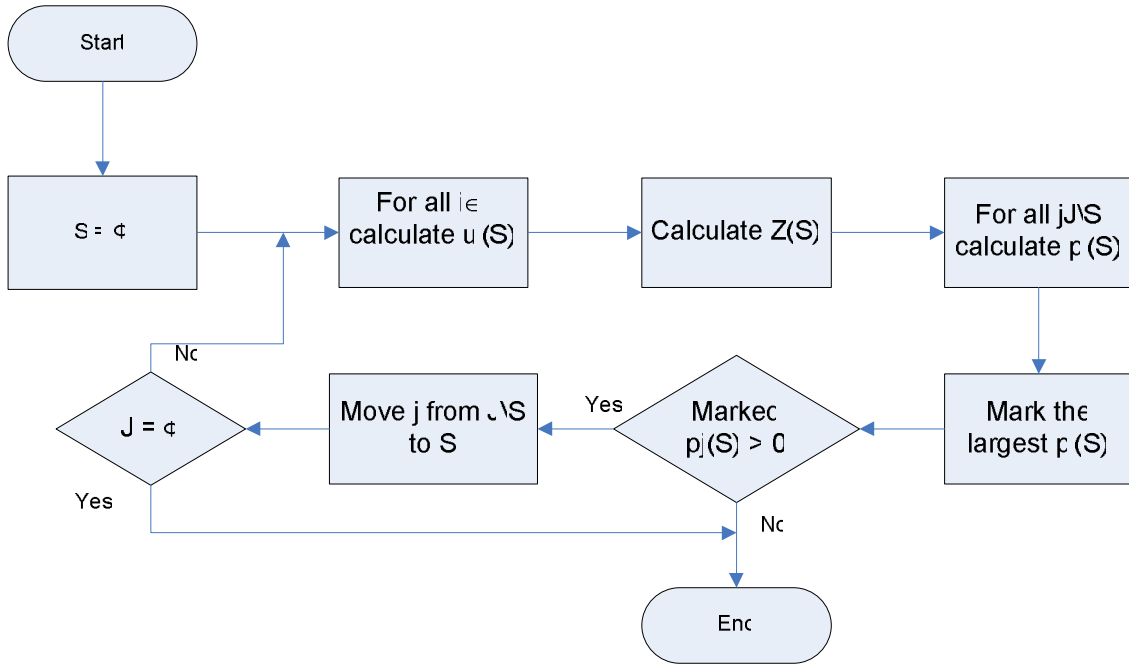


Figure 2: Flow diagram of the greedy algorithm

The algorithm is completed when either $J \setminus S$ is empty, or $p_j(S) \leq 0$ for each $j \in J \setminus S$. Flow diagram of the algorithm is shown in figure 2.

This is well known heuristics, recently described in [1]. The problem we have faced using this heuristic algorithm is that if values of f_j which are in our case the fixed costs of deploying new backbone equipment, are too high, the algorithm does not return any solution. On the other side, if the fixed costs of deploying new equipment is too low, the algorithms simply returns all the sites in the final solution. While the second case might be technically acceptable (network equipment is so cheap that operator can set up all sites), the first case can not be neither technically nor economically acceptable. Therefore, algorithm is not usable unless, exact and correct values for profit and costs are known. Costs of deploying a new backbone node are not only the cost of the equipment, but also the cost of building, power supply, air condition, taxes etc. which are very hard to include in calculation. Since, the network planner usually does not have an exact data, and all statistics, it is impossible to use this algorithm unless certain changes are made.

Let us consider a network with 6 potential sites. Each site covers the certain region, and produces certain amount of traffic demands towards all other sites. Therefore, each potential site is also a customer. After the optimization, some of the sites will become backbone sites, while the other will remain access sites just collecting traffic from the region and proceeded to some backbone node. The cost factors f and the profit factors c_{ij} are:

$$C = \begin{bmatrix} 1 & 6 & 4 & 3 & 0 & 6 \\ 6 & 2 & 3 & 4 & 6 & 6 \\ 5 & 8 & 9 & 5 & 3 & 0 \\ 4 & 5 & 6 & 2 & 4 & 4 \\ 3 & 4 & 6 & 3 & 3 & 4 \\ 2 & 3 & 4 & 5 & 3 & 4 \end{bmatrix} \quad f = [3 \quad 2 \quad 2 \quad 2 \quad 3 \quad 3]$$

For given example network algorithm gives as a solution in which operator has to deploy only two sites, $S = \{2 \ 5\}$, while the other sites $J \setminus S = \{0 \ 1 \ 3 \ 4\}$ will remain just access sites.

If for example price of deploying new equipment has been $f = \{300\ 200\ 200\ 200\ 300\ 300\}$, the algorithm would have returned $S = \{\}$. From this simple demonstration it can be seen how important it is to have very precise input parameters in the algorithm.

Real case network

In our project, we had to perform a network planning of future backbone network of new telecomm operator in Croatia. For this purpose, we had to collect data of potential customers, and potential location of backbone nodes. We were given a geographical coordinates of each location and belonging population. From this data we have produced a traffic matrix, which gave us a point-to-point demand for each potential site pair. It has been chosen that the transport network will be based on SDH technology, with possible point to point WDM connection between larger concentration points. We have used an above described heuristic algorithm to determine an optimal locations for WDM equipment. In figure 3 locations of potential sites are shown with larger circles, and belonging customers with smaller circles. In described algorithm an profit variable c_{ij} was introduced. In the planning process it is very hard to determine an actual value of c_{ij} variables. In our model we have used a traffic matrix as a C matrix, which basically means that higher traffic demands are equivalent with higher profit.

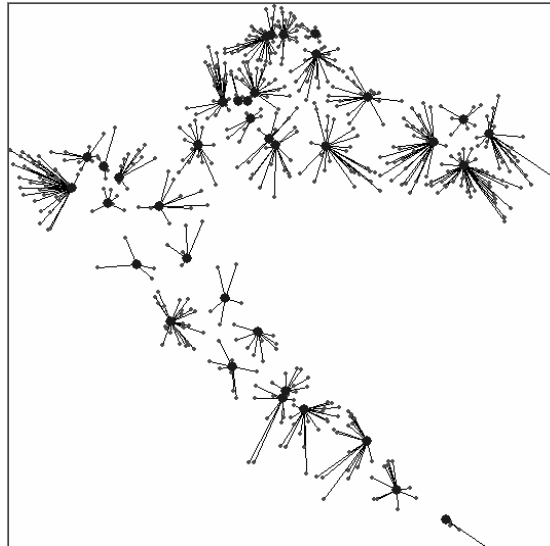


Figure 3: Location of potential sites, and customers

Technically it is correct interpretation, because network nodes with high traffic demands will produce more profit to the operator than the nodes with the low traffic demands. After we have specified a C matrix, the problem with f has aroused. Too small values in f matrix yield to the optimal solution where all potential sites are included into solution. It is obvious that such solution was not acceptable for the operator, because network with all nodes being WDM capable is simply too expensive. On the other side, to large values in f matrix result in solution with none locations included in optimal solution.

In figure 4 traffic load of potential sites is depicted. Larger circles symbolize a higher traffic load, while smaller circles symbolize smaller traffic load. One could easily point out that it is expected that sites with lot of traffic load are the most likely the candidates for WDM equipment.

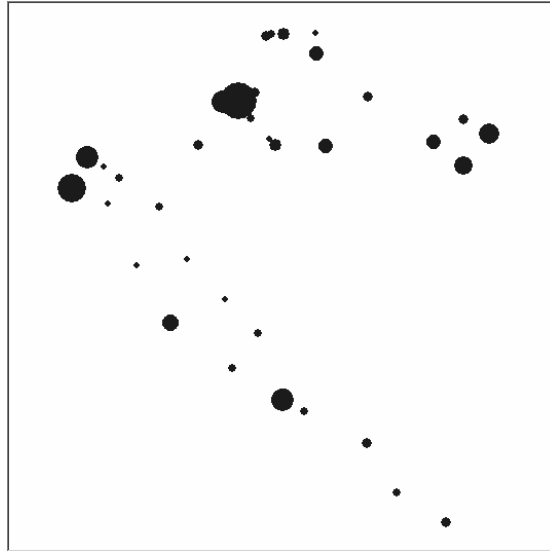


Figure 4: Traffic load at the potential sites

In figure 5 a visual representation of the results obtained by above described heuristics is shown. Large circles represent a five chosen location for WDM equipment, and small circles represent locations that are not included in the solution. This location will be used for SDH equipment. It can be seen from the pictures that sites included in the optimal solution are exactly the one with the most traffic load. It justifies our assumption, and proves that correct values in C and f matrix were chosen.

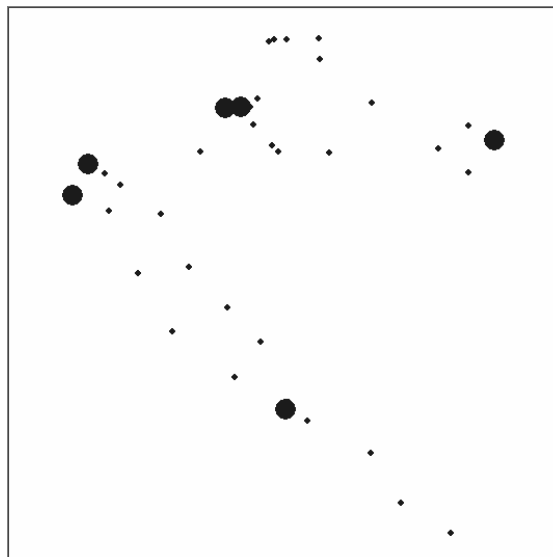


Figure 5: Optimal solution of facility location problem, for WDM equipment in Croatia

The chosen locations are also locations of the biggest cities in Croatia, which is logical solution since the majority of population and industry is located around large cities.

Conclusion

In this paper a problem of optimal location of backbone equipment is presented. In the planning process of communication networks that are spread over large geographical area, it is very important to place network equipment at optimal locations. Number of required equipment is also an issue that has to be carefully planned. In the paper a classical linear/integer mathematical formulation of the problem is given, as well as its AMPL/CPLEX implementation. It has been shown that exact solution

is hard to obtain for networks with large number of network nodes and large number of customers, therefore a approximation heuristic algorithms are required to address this problem. A Greedy heuristics algorithm is presented, as well as the justification of the algorithm on the real case network. Further work will include, verification and improvement of other facility location algorithms, as well as node interconnecting problem. Interconnecting problem is closely related to the facility location problem, since chosen sites have to be connected in such manner, that resulting topology is resistant of possible link or node failures.

REFERENCES

- [1] J. Cheriyan, R. Ravi, “*Aproximation algorithms for network problems*”, Lecture notes, Dept. of Combinatorics and Optimization, University of Waterloo, Canada, September 1998.
- [2] S.M. Daskin, “*Network and Discrete Location: Models, Algorithms and Applications*”, 1995, John Wiley and Sons, Inc. ISBN 0-471-01897-X
- [3] R. Fourer, D.M. Gay, B. W. Kerningham “*AMPL – A Modeling Language For Mathematical Programming*”, Scientific Press. 1993, ISBN: 0-78950-701-3
- [4] G. Y. Handler, P.B. Mirchandani, “*Location on Networks – Theory and Algorithms*”, The MIT Press, 1979, ISBN – 0-262-08090-7
- [5] Z. Liu, Y. Gu, D. Medhi, “*On optimal Location of Switches/Routers and Interconnection*”, Technical Report, Computer Science Telecommunications, University of Missouri-Kansas City, July 1998
- [6] R. F. Love, “*Facilities Location – Models & Methods*”, Elsevier Science Publishing Co., 1998, ISBN 0-444-01031-9
- [7] Resende M.G.C. “*Computing Approximate Solutions Of The Maximum Covering Problem With GRASP*”, Journal of Heuristics, Volume 4, 161-177 (1998)