

Server Location in Mobile Computing

Geraldo Robson Mateus Antônio A.F. Loureiro
Ricardo Costa Rodrigues Olga Goussevskaia

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Caixa Postal 702 30123-970 Belo Horizonte, MG, Brasil
Email: {mateus,loureiro,rick,olga}@dcc.ufmg.br

Abstract

Mobile computing will continue to widen as the world of computing increasingly becomes the world of access to network services and resources. In this paper we study the problem of given a network comprised of users (in our case mobile) and a set of servers that provide a service, we are interested in finding an association between users and servers that minimizes the cost to access the servers in the network. We present the mathematical formulation of this problem and a user mobility model that is used to perform some experiments. This is a initial step towards studying this problem that will be very important in a mobile setting.

1 Introduction

Computing in the 1970s and 1980s meant using almost completely stationary machines that were designed to be situated in one location and rarely if ever moved. Those machines were heavy, awkward, and clumsy and physical movement was so impractical that software designers only considered and developed algorithms with the underlying assumption that location did not change for both clients and servers. The resulting software architectures were adequate for the time because those machines were not designed to move on any regular basis. Even the emerging "portable" machines at the end of the 1980s were not really portable. They could at best be described as "luggable," since putting a handle on a machine does not make it portable and, after all, it is not convenient to carry it.

During the last decade the hardware industry made rapid progress in chip and LCD technology and in general miniaturization as well. Computing devices that were truly portable and as powerful as a desktop computer started to emerge. A logical desire perceived was to benefit from the wireless infrastructure available in a world where people are increasingly mobile. In 1999, one third of the U.S. corporations provided their mobile workers wireless access to the corporate information and communication systems they needed to do their job. More and more, companies are providing sales forces, field service personnel, and traveling executives the tools they need to be as connected in the field as they would be at their desks. The adoption trend is growing at an accelerated pace. The number of mobile workers in the U.S. will grow from 43.5 million in 1996 to 62.4 million in 2001. There are expected to be 15.6 million wireless email subscribers by the year 2002.

T.P. x M.U.s ($Y(i,k), I=10, K=10$)

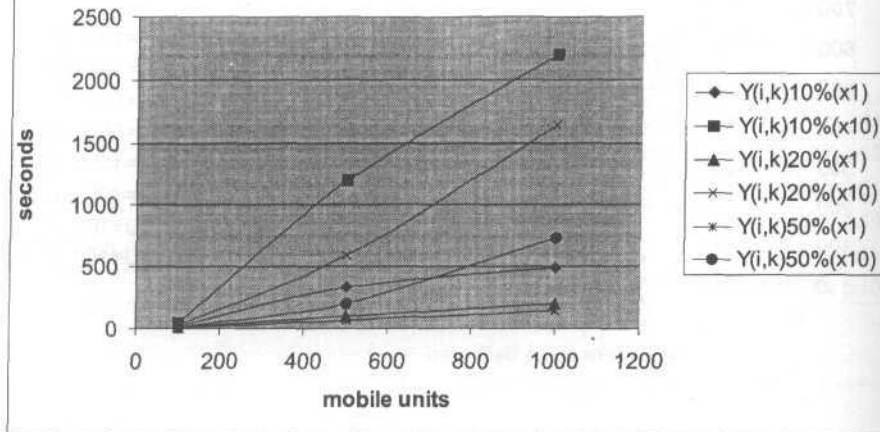


Figure 4: Second model where it can be chosen a different server for each period.

gets much more complex. As future work, we intend to explore variants of this problem as described above. We also plan to explore the design of new algorithms to solve the proposed problems efficiently.

References

- [1] K. Aardal, M. Labbé, J. Leung, and M. Queyranne. On the two-level uncapacitated facility location problem. *Journal on Computing*, 8(3):289-301, 1996.
- [2] American Express. Travel service location finder. <http://travel.americanexpress.com/travel/personal/resources/tso/>, 1999.
- [3] T.B. Boffey and J. Karkazis. P-median and multi-medians. *Journal of the Operational Research Society*, 35:57-70, 1984.
- [4] G. Cornuejols, M.L. Fisher, and G.L. Nemhauser. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science*, 23:789-810, 1977.
- [5] M.A. Efroymsen and T.L. Ray. A branch and bound algorithm for plant location. *Operations Research*, 14:361-368, 1966.
- [6] D. Erlenkotter. A dual-based procedure for uncapacitated facility location. *Operations Research*, 26(6):992-1009, 1978.

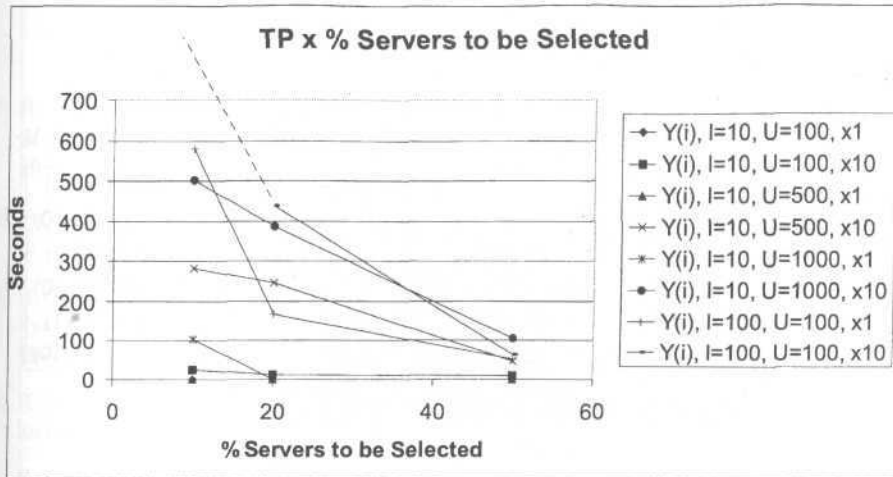


Figure 3: Behavior of the computational time with the minimal percent of candidate servers to be selected.

Figure 4 compares the computational time with the number of mobile users. Again, similar to the first model, the problem becomes more complex to be solved with lower percent of servers to be installed, higher fixed costs to install a server, higher number of mobile users and higher number of candidate servers.

Comparing Figures 2 and 4, we observe that the problems have similar behavior. Their curves follow the same pattern. It is important to emphasize, however, that the second problem, allowing the selection of different servers for each period, is always much more complex than the similar instance for the first problem.

7 Conclusions and Future Work

Mobile computing will continue to widen as the world of computing increasingly becomes the world of access to network services and resources. The problem studied in this paper is very important in this setting and can be adapted to a great variety of particular problems. Although it was treated as a problem similar to the Best Base Station Location Problem for Mobile Phone Systems, it can be viewed as the problem to install Redundant Web Servers in a wide range network. It can also be viewed as the problem of locating hot dog vendors around a city depending on the day and the events to occur in the town (for example, the finals of the Brazilian soccer championship).

The problem, as it is treated in this work, contributes to associate the user mobility to the location problem. However, this is a complex problem and further research needs to be done. As described in the paper, there is a large number of variables that modify the results and the computational time to solve the problem. For example, the user mobility itself is a very important problem and is subject to recent research. And, if we consider the possibility of having different servers active at different periods, the problem

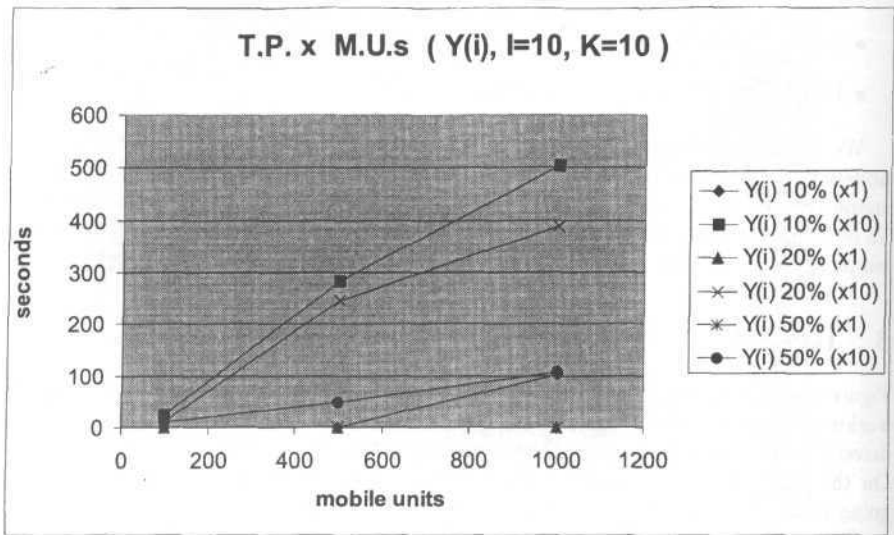


Figure 2: First model where a server is selected and active during all the time.

Lower percent of Servers and Higher fixed costs increases the necessity for choosing a very good location to install a server. A higher number of mobile users and candidate servers, on the other hand, increase the number of possible solutions to the problem and thus increases the computational time.

For example, for 100 candidate servers, 10 periods, 500 mobile users, 10% of candidate servers to be selected and a Fixed Cost factor of 10, the system did not finish the computation after 8000 seconds.

The second graph, depicted in 3, compares the behavior of the computational time with the minimal percent of candidate servers to be selected. The computational time decreases when we relax the problem allowing a higher number of candidate servers to be selected. On the other hand, choosing more servers increases the total cost to install the servers.

The results for the second model, where it can be chosen a different server for each period, are shown in Table 4.

I	K	U	10%(x1)	10%(x10)	20%(x1)	20%(x10)	50%(x1)	50%(x10)
10	10	100	26	45	16	38	10	11
10	10	500	332	1192	100	593	69	199
10	10	1000	493	2198	205	1637	148	737
100	10	100	560	2341	117	663	78	359
100	10	500	N/D	N/D	N/D	N/D	N/D	N/D
100	10	1000	N/D	N/D	N/D	N/D	N/D	N/D

Table 4: Second model where it can be chosen a different server for each period.

- Number of mobile users: 100, 500 and 1000
- Minimal percent of candidate server locations to be installed: 10%, 20% and 50%
- Fixed cost to install a chosen server: 1 and 10

We tested both models with each possible combination of the above variables. The models were implemented using the AMPL language [7] integrated to CPLEX [17]. The computer used to run the experiments is a dual-processor Sun UltraSPARC-II with 1.2 Gb of RAM Memory. As we will report in the next section, some problems became very complex that it was not possible to finish the computation.

6 Results

Figure 1 shows an example of user mobility behavior during five period times in a usual working morning. The user D, a 24-hour delivery boy, moves very frequently. A taxi driver (T) moves less frequently than the delivery boy. Both users move around the city. On the other hand, a common worker moves in a restricted area during the mornings going from home to the office and staying there.

Table 3 shows the results for the first model where a server is selected and active during all the periods. Figure 2 compares the processing time against the number of mobile users for a mobility simulation of ten periods according to different values of a Fixed Cost to install a server and the minimal percent of servers to be selected.

I	K	U	10%(x1)	10%(x10)	20%(x1)	20%(x10)	50%(x1)	50%(x10)
10	10	100	0	25	0	12	0	11
10	10	500	0	283	0	247	0	48
10	10	1000	103	504	0	388	0	106
100	10	100	579	2621	168	438	51	63
100	10	500	4468	> 8000	N/D	N/D	N/D	968
100	10	1000	N/D	N/D	N/D	N/D	N/D	N/D

Table 3: First model where a server is selected and active during all the time.

It is important to verify that for large instances of the problem, we did not get a solution due to computational resources limitation.

Figure 2 shows the graph for Table 3.

Analyzing Figure 2, we conclude that the problem gets more difficult to be solved according to four factors:

- Lower percent of Servers to be installed
- Higher fixed costs to install a server
- Higher number of mobile users
- Higher number of candidate servers

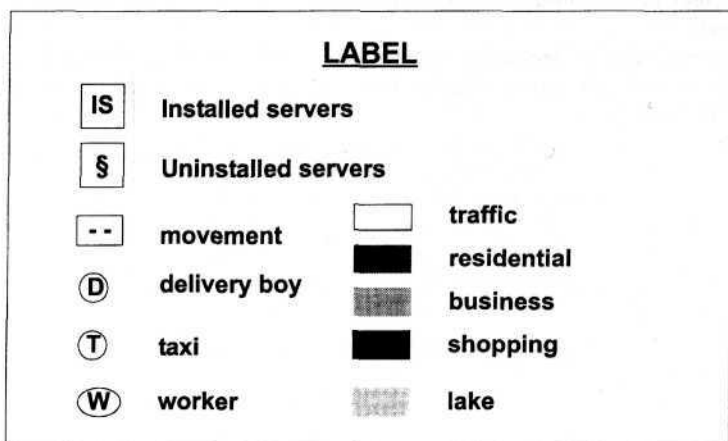
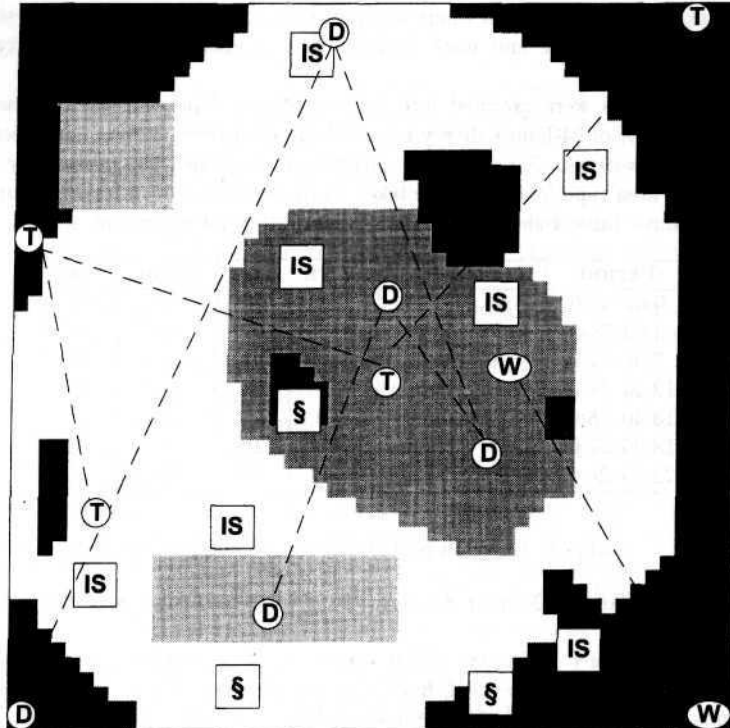


Figure 1: City areas.

- Number of periods: 5, 10 and 20

lake. The lake represents an area where any user cannot be. For each user, the simulator randomly defines its home and work location in a residential and business type area respectively.

The mobile users were grouped into four categories depending on their work and mobility profile: 3% of 24-hour delivery boys, 3% of taxi drivers, 31% of housekeepers and 62% of common workers. For each user category, it is defined the probability of being located in each area type (except in the lake), in its own home or in its work during each period of the day. Table 1 shows the location probability for a common worker.

Period	Home	Res.	Work	Bus.	Shop.	Traf.
0:00-4:00	98	00	00	00	00	02
4:00-7:00	100	00	00	00	00	00
7:00-12:00	00	00	100	00	00	00
12:00-13:30	00	00	75	20	05	00
13:30-18:00	00	00	100	00	00	00
18:00-22:00	55	05	00	10	30	00
22:00-24:00	80	20	00	00	00	00

Table 1: Location probability for a common worker

On the other hand, a 24-hour delivery boy is expected to be in traffic areas during almost all time.

We tested only part of the day of our simulated users because the main goal in this paper is not the user mobility itself, but the server location. Thus, the simulated period corresponds to only four hours of a common working day (between 6 and 10 am). We collected the user location every 714 seconds in order to obtain data for about 20 contiguous time periods.

The user velocity while moving around the city varies between 30 and 50 km/h. The probability of having a user moving on each time also varies according to the user type as shown in Table 2.

User Type	Movement Probability
24 hour delivery boy	90 %
Taxi driver	60 %
Housekeeper	30 %
Common Worker	50 %

Table 2: Probability of having a user moving

5 Experiments

We generated various scenarios with different configurations in order to analyze our model. We performed various experiments with different values for the following variables in the problem:

- Number of candidate server locations: 10, 50 and 100

$$x_{ij}^k \leq y_i^k, \quad \forall k \in K, \forall i \in I, \forall j \in U^k \quad (4)$$

$$x_{ij}^k \in \{0, 1\}, \quad \forall k \in K, \forall i \in I, \forall j \in U^k \quad (5)$$

(6)

The objective function (1) minimizes the distance between each mobile unit and its server in each period and the fixed cost to install each candidate server node in each period.

Constraints (2) ensure that each mobile unit is allocated to a server and constraints (3) that there are at least p servers to meet the demand for all mobile units. Constraints (4) ensure the fact that a mobile unit can be allocated to a server unless that server has already been pre-selected. Constraints (5) and (6) define the integrality of each variable.

The formulation below can be seen as special cases of the well-known formulations of the Uncapacitated Dynamic Location Problem (UDL) [16], and the p -median Dynamic Location Problem (PDL) [9].

The UDL Problem deals with the selection of facility sites from a set of candidate locations to meet the demand of a set of customers in each period. The objective function is to minimize the sum of fixed costs to install the facilities and the transportation costs to supply the demands. This problem was solved by different algorithms [1, 4, 5, 6, 10, 13].

The PDL formulation consists in locating p facilities (medians) on a network so as to minimize the sum of all distances from each node to its nearest facility [3, 16]. Successful attempts have been made to design algorithms for both problems. The algorithms are exact methods as branch and bound or heuristics [4, 15, 19, 9].

In this work we are interested in exploring different approaches. The fixed cost at the objective function can be used or not. We can work limiting the number of servers in period k in different ways: equal p , greater than or equal p and less than or equal p . We are exploring the mobility of servers as well, but we are also considering in the computational results all the servers static. Moreover, we can consider the boolean variable y_i independent of each period k . In this case the fixed cost to install each server is also independent of k and we have a fixed cost f_i to install server i for all periods. The set of constraints (3) is replaced by $\sum_i y_i > p$. Note that the working area must be served by a given number of servers which can vary depending on their cover areas in each period. We are exploring different cover area for each server.

Note that there are other variants as different server groups may depend on both the kind of service available, and groups of mobile units with similar demands and/or priorities.

In this paper, we solved the proposed formulation and we performed some computational experiments using the commercial optimization software CPLEX.

4 User Mobility Model

We tested the optimization models in a scenario with different kinds of mobile users in a hypothetical city. The user mobility behavior was assessed using a User Mobility Simulator [20, 21].

This simulator can represent a city as a 2-dimension grid of 100 x 100 units. The city contains several areas of five different types as shown in Figure 1: residential, business, shopping, lake and traffic. The user can be located in any of these areas except in the

In the telecommunications world there is a similar problem. In the future, telecommunications networks will not only route data, but they will also process it. The service processing will take place at special intelligent nodes in the network, possibly distributed over many nodes. In such an environment, the question arises as to how to distribute the processing of these services most efficiently. Note that with the integration of both computer and telecommunication networks, at some point this problem will have to be studied in a combined way.

In this work we follow a different track from the proposals described above in the sense that we allow the server to be either fixed or mobile and we also classify the different types of users present in the network.

3 Mathematical Formulation

The Server/Service Location Problem consists in selecting a set of servers to meet the demand of each mobile unit in each period of time. Each mobile unit can only access a given server in each period. This is not a strong restriction considering the kind of mobile device currently available or expected to be in the market in the next few years. We propose a combinatorial optimization formulation [18] and we consider alternative approaches.

The following notation is used in the formulation of the SSLP:

p : number of servers;

I : set of candidate server nodes;

K : set of periods;

U^k : set of mobile units for each period $k \in K$;

c_{ij}^k : distance from server i to mobile unit j in period k ;

f_i^k : non-negative location cost for the candidate server node i

x_{ij}^k : boolean variable which assumes the value 1 or 0 depending on whether or not the server i meet the mobile unit j in period k ;

y_i^k : boolean variable which is set to 1 or 0 depending on whether or not the server i is being selected to provide mobile units in period k .

The SSLP problem is formulated below, as a mixed integer programming model:

(M):

$$\min \sum \sum \sum c_{ij}^k x_{ij}^k + \sum \sum f_i^k y_i^k \quad (1)$$

s.t.:

$$\sum_{i \in I} x_{ij}^k = 1, \quad \forall j \in U^k, \forall k \in K \quad (2)$$

$$\sum_{i \in I} y_i^k \geq p, \quad \forall k \in K \quad (3)$$

In this new scenario it is expected that users (clients) will be able to connect to different servers for accessing different services such as Web servers, mail servers, databases, directories, fax machines and other variety of services coming our way. As the focus of interest shifts ever more onto the Internet and the Web, wireless network services will gain prominence in the minds of computer users everywhere. In the networked world of the future, interchangeable services will appear and disappear, and providing for the dynamic nature of their availability is an important accomplishment for the Server/Service Location Problem (SSLP).

The statement of this problem is very simple and will be presented in an informal way (the mathematical formulation will be given in Section 3): given a network comprised of users (in our case mobile) and a set of servers that provide one or more services, we are interested in *finding* an association between users and servers that minimizes the cost to access the servers in the network. From this statement we can derive several different variants such as the amount of servers is fixed/variable, each server has a fixed/variable location, each server serves one or more different services, each server has a limited/unlimited capacity, each server has a fixed/variable range for serving, each client can access one or more services simultaneously, clients may have different priorities, etc.

The main goal of this paper is to study the Server/Service Location Problem in a mobile environment typically supported by a cellular infrastructure. Currently there is a new generation of mobile phones and Portable Digital Assistants (PDAs) being deployed with multimedia capacities, simple Web browsers, and email facilities for the third generation of wireless networks. Those devices are expected to be cheaper and more popular than palmtops and laptops in the years to come. However, this will only come true if users can access different servers/services while mobile.

The rest of this paper is organized as follows. Section 2 presents some services available in the Internet that consider the location of a client. Section 3 presents our mathematical formulation of the Server/Service Location Problem. Section 4 discusses the user mobility model we used in the work to validate the mathematical formulation. Section 5 lists the experiments performed and Section 6 the results obtained. Finally, Section 7 presents our conclusions and future work.

2 Related Work

Currently in the Internet there are several Web servers that provide services that take into account the location of a client. For instance, the American Express Travel Service Location Finder [2] provides the location nearest a client or a destination, including a map and directions to reach the desired service location. Similar services are provided by Fujitsu [8] for client support, and commercial real estate in the U.S. [14].

The Internet Engineering Task Force has proposed a new standards-track protocol named Service Location Protocol [11, 12] designed to simplify network-based service discovery from a small, unadministered network up to an enterprise network. It is particularly well suited for client-server applications and establishing connections between peers that offer or consume generic services. A similar approach to the Service Location Protocol is the X.500 protocol proposed by ISO for service discovery in the OSI/ISO reference model. Yet another approach is the interface provided by CORBA (Common Object Request Broker Architecture) which can be seen as a "software bus" to access different kinds of services in a network.

- [7] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL - A Modeling Language for Mathematical Programming*. Duxbury Press / Brooks / Cole Publishing Company, 1993.
- [8] Fujitsu PC Corporation. Fujitsu service location. <http://support.fujitsu-pc.com/fidaspl/>, 1998.
- [9] R.D. Galvão and E.R. Santibanez Gonzales. A lagrangean heuristic for the p_k -median dynamic location problem. *European Journal of Operational Research*, 58:250-262, 1992.
- [10] R.D. Galvão and L.A. Raggi. A method for solving to optimality uncapacitated location problems. *Annals of Operations Research*, 18:225-244, 1989.
- [11] E. Guttman. Service location protocol: Automatic discovery of IP network services. *IEEE Internet Computing*, 3(4):71-80, July/August 1999.
- [12] E. Guttman et al. Service location protocol, version 2. <http://www.ietf.org/html.charters/svrloc-charter.html>, June 1999.
- [13] J. Krarup and P.M. Pruzan. The simple plant location problem: Survey and synthesis. *European Journal of Operational Research*, 12:36-81, 1983.
- [14] LoopNet, Inc. Commercial real estate links. <http://www.loopnet.com/directory/fiabci.html>, 1999.
- [15] G.R. Mateus, F.R.B. Cruz, and H.P.L. Luna. Algorithm for hierarchical network design. *Location Science*, 2:149-164, 1994.
- [16] P.B. Mirchandani and R.L. Francis. *Discrete Location Theory*. John Wiley & Sons, New York, 1990.
- [17] CPLEX Optimization. *Using CPLEX callable library and CPLEX mixed integer library. Version 5.0*. 1997.
- [18] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, New Jersey, 1982.
- [19] C.R. Reeves. *Modern Heuristic Techniques for Combinatorial Problems*. Blackwell, Londres, UK, 1993.
- [20] M.N. Rocha, G.R. Mateus, and S.L. Silva. Simulation of mobile unit tracking in wireless communication systems. *Proc. of the Wireless Personal Multimedia Communications (WPMC99) Conference*, September 1999.
- [21] M.N. Rocha, G.R. Mateus, and S.L. Silva. Traffic simulation and the location of mobile units in wireless communication systems. *Proc. of the 1st Latin American Network Operations and Management Symposium (LANOMS 99)*, pages 309-319, December 1999.