Optimal Two-Tier Cellular Network Design

Eylem Ekici
School of Electrical and Computer Eng.
Georgia Institute of Technology
Atlanta, GA, USA

Cem Ersoy
Computer Engineering Dept.
Bogazici University
Bebek, Istanbul, Turkey

***Abstract* - One way of improving the performance of cellular networks is to build a second layer of macrocells on top of the microcell level. The system performance can further be increased by using guard channels and allowing calls to overflow to the upper layer when needed. In this study, we used simulated annealing (SA) to determine the design parameters of two-tier cellular networks for which the cost is minimized. We experimented with the SA-based technique on different example problems and obtained promising results.**

I. INTRODUCTION

After the introduction of wireless communication systems a decade ago its growth has been rapid [1]. The number of wireless communication service users as well as the spectrum of the available services increased with an unexpected rate. The main reason for this growth was the newly introduced notion of terminal and user mobility. It is expected that the wireless communications will be the dominant mode of data access technology in the next century [2].

Cellular concept is the main idea on which majority of existing terrestrial wireless communication systems are based [3]. In this approach, the total area to be included in the system is divided into cells, where a subset of all available channels is used in each cell. The main idea is to use the channel subsets in cells that are far apart to ensure an interference-free communication scheme with potentially infinite coverage capability.

When a *Mobile Terminal* (MT) goes over to another cell while a call proceeds, the call is passed to the new base station. This procedure is called *handoff*. When the new calls are not serviced due to lack of channels it is said that the call is *blocked*. When a handoff call cannot find a free channel in the new cell, the call terminates forcefully. This situation is described as *call dropping*. But from the users' point of view, this situation is far less desirable than the first one [4]. Therefore in some newer systems the handoff calls are given higher priorities. Cellular networks in which handoff calls are prioritized are called *prioritized networks*. The non-prioritized systems are examined in [5]. One of the simplest ways to give priority to the handoff calls is to reserve some channels exclusively for them. These reserved channels are called *guard channels*. In [6], an efficient way to minimize the call dropping probability while staying below a given call blocking probability threshold is presented. It is also possible to design systems that have multiple layers of cells that cover the same area.

In [7], Hu and Rappaport describe a multi-tier system and give analytical results for a set of different parameters. The choice of parameters aims to present the performance of the system under different conditions. The efforts in [8] are aimed at the optimal spectrum partitioning once the system is set up. None of the mentioned works present how a system should be designed satisfying certain other constraints.

The focus of this work is chosen as designing a wireless cellular system with guard channels, deployment of two cellular layers on top of each other and overflow. The aim is to minimize the system cost while satisfying the maximum call blocking and forced termination probabilities. In order to solve the design problem, a well-known artificial intelligence technique, *Simulated Annealing* (SA) [9], is chosen.

The outline of this paper is as follows: In the second section, the design problem is formulated. Furthermore, the objective function that will be used in the solution techniques is described and the calculation methods are presented. The third section explains the solution technique extensively. In the fourth section, the results of the computational experiments are presented before the conclusion.

II. PROBLEM DEFINITION

The purpose of this work is to design a minimum cost multi-tier cellular network with call overflow and guard channels that satisfy certain performance constraints. We tried to determine the parameters that describe the multi-tier cellular network rather than giving the performance measures once these parameters are supplied. Here we assume two classes of MTs. The high mobility class represents the MTs that are used in cars and other vehicles. The low mobility class is made up of users that are primarily pedestrians. The speeds of MTs are exponentially distributed with means v_f and v_s respectively. Since the target deployment area is as large as a metropolitan area, the distribution of the members of both mobility classes can be considered as uniform and all mobile terminals are assumed to move in any direction equally likely. The call duration is also exponentially distributed with mean $1/U_t$ for both mobility classes. The time spent in a cell, which is called *dwell time* U_d, is calculated as in [8], where r is the radius of the cell and v is the sped of the mobile terminal, resulting in following identity:

$$\frac{1}{U_d} = \frac{r\pi}{2v} \tag{1}$$

The call arrivals to the cells follow a Poisson distribution for both mobility classes. The mean rate for the call arrivals is determined by the MT density and the mean call generation rate of the individual MT users. The antennas for both layers may be located at the same locations. Furthermore, the distinction between the mobility classes should be made using the mean of the mean mobility rates as the threshold.

The cells of the lower tier in the two-tier network are called microcells. Upper layer cells, macrocells, cover an integer number of microcells. Furthermore, the radius of the macrocells is constrained to be an odd integer multiple of the radius of the microcells. The total channel spectrum is divided by the cluster size into channel sets. Each channel set is then divided among one microcell and one macrocell. This means that if the microcells has one less channel, then that channel will be used by the macrocells. After the splitting of the channel sets some of the channels will be assigned as guard channels. The arriving calls are serviced as follows: The new calls of the slow MTs are primarily served by the appropriate microcells with an available non-guard channel. If only guard channels are available, then the new calls are overflowed to the macrocell that covers the microcell as *overflowed new call*. If handoff calls cannot be serviced in the microcells, then they are overflowed to macrocell.

At the macrocell level, all the calls of the high mobility terminals and the overflowed calls are serviced. The high mobility handoff calls and the overflowed handoff calls are treated equivalently. New calls of both classes may not use the guard channels upon their arrival. If no non-guard channel is available, new calls are blocked. The low mobility calls at macrocell level cannot return to microcell level even if channels become available in the microcells.

The metrics are chosen to be call blocking and dropping rates while the total system cost is minimized. Any solution is said to be feasible as long as it is in accordance with system description and the probability of call blocking and call dropping is below the specified thresholds. The values for the system parameters are in the order of 10^{-2} to 10^{-3} for the call blocking and for the call dropping 10^{-3} to 10^{-4}. The input parameters are used to describe the call and physical medium characteristics and to set the performance requirements of the designer. These are used, in turn, to produce the parameters that describe the cellular system.

The user supplied parameters are v_s and v_f (mean speed of low and high mobility users), SAm^2 (call arrival rate per second per m^2 for low mobility terminals), FAm^2 (call arrival rate per second per m^2 for high mobility terminals), $1/U_t$ (mean call duration), C_1 and C_2 (macrocell and microcell cost), A (total area), CS (cluster size), Ch_{total} (total number of available channels), p (Radius increase/decrease factor), cr (cooling rate), $P_{b,max}$ *and* $P_{d,max}$ (maximum allowable call blocking and dropping probability). The decision parameters are C (total system cost), Ch_1 and Ch_2 (number of channels reserved for each microcell and macrocell), G_1 and G_2 (number of guard channels reserved for each microcell and macrocell), R and r (radius of a macrocell and microcell).

The minimum cost two-tier cellular network design problem can be formulated as follows:

$$Minimize \quad C = C_1 \cdot N_1 + C_2 \cdot N_2 \tag{2}$$

subject to the constraints

$$P_b \le P_{b,\max} \tag{3}$$

$$P_d \le P_{d,\max} \tag{4}$$

$$\pi R^2 N_1 \ge Area \tag{5}$$

$$\pi r^2 N_2 \ge Area \tag{6}$$

$$\frac{R}{r} = 2n + 1, \qquad n \in Z \tag{7}$$

N_1 and N_2 are the number of the microcells and macrocells respectively. While minimizing the total system cost, the resulting probabilities for call blocking and call dropping should stay below the provided limits as expressed in (3) and (4). The calculation of the call blocking and dropping probabilities is explained in Section III.A.

III. SOLUTION TECHNIQUES

In order to solve the minimum cost two-tier cellular network design problem, a *Simulated Annealing* based technique is used. In this section, we describe the details of the algorithm used as well as the description of the cost and performance calculation procedures.

A. Calculation of Cost and Performance Measures

Given the total area, the number of microcells and macrocells that would cover the given total area is determined. If the resulting system should have N_1 microcells and N_2 macrocells with C_1 and C_2 as their respective unit costs, the total cost $C= C_1 N_1+ C_2 N_2$.

In order to evaluate the performance of a configuration, the system is divided into two parts. The first part corresponds to the microcell layer of the system. This part of the system is represented with a Markov Chain (M/M/s/s system) [10]. In this representation the state corresponds to the number of calls served by a microcell. The arrivals to the system are denoted as λ_{1s}. The dwell time of the low mobility users in the microcells U_{d1s} are calculated according to (1). A sample Markov chain for a microcell with five channels and two guard channels are presented in Fig. 1.

The steady state probabilities P_i can be calculated using Erlang-B formula [10] as shown in (8) – (10).

$$P_i = P_0 \cdot \left(\frac{a}{c}\right)^i \cdot \frac{1}{i!}, \ \ i \le Ch_1 - G_1 \tag{8}$$

$$P_i = P_0 \cdot \frac{a^{Ch_1 - G_1}}{c^i} \cdot \frac{b^{i - Ch_1 + G_1}}{n!}, \ \ Ch_1 \ge i > Ch_1 - G_1 \tag{9}$$

$$P_0 = \left[\sum_{i=0}^{Ch_1} P_i\right]^{-1} \tag{10}$$

The asymptotic handoff rate λ_{1sh} used in Erlang-B formula is calculated by iteration as described in [7], until Λ_{1sh} equals to λ_{1sh}.

$$\Lambda_{1sh} = \sum_{i=0}^{Ch_1} \left(i \cdot P_i \cdot U_{d1s}\right) \tag{11}$$

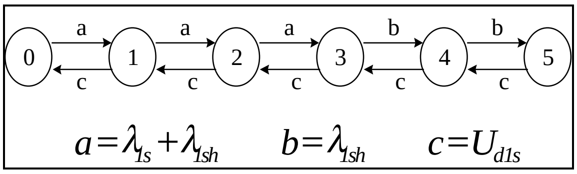

Fig. 1 Sample State Transition Diagram for Microcells (5/2)

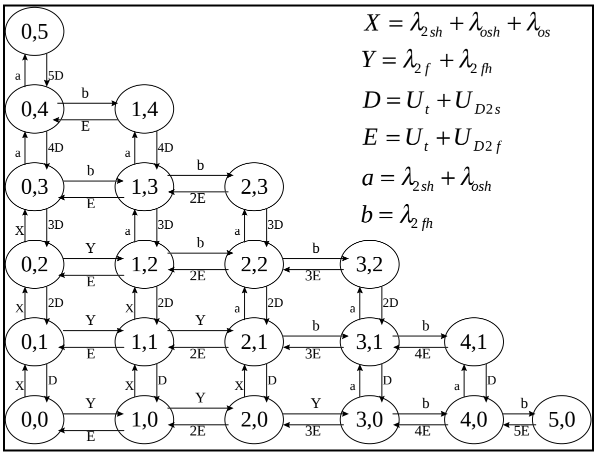

Fig. 2 Sample State Transition Diagram for Macrocells (5/2)

Having determined the probabilities for call blocking P_{b1} and call dropping P_{d1}, which correspond to the sum of the steady state probabilities for states $i=Ch_1-G_1$ to Ch_1 and P_{ch1} respectively, the total overflow new call and overflow handoff call traffic entering a macrocell (λ_{os} and λ_{osh}) covering N microcells can be calculated as follows:

$$\lambda_{os} = N \cdot \lambda_{1s} \cdot P_{b1} \tag{12}$$

$$\lambda_{osh} = N \cdot \lambda_{1sh} \cdot P_{d1} \tag{13}$$

The states in macrocell level correspond to the numbers of high and low mobility users (i and j) serviced by a macrocell. The state transition diagram for a macrocell with five ordinary and two guard channels is shown in Fig. 2.

The variables λ_{2f} and λ_{2fh} are the arrival rates of the new and handoff calls of high mobility users to a macrocell respectively. λ_{2sh} is the arrival rate of the handoff calls of the low mobility users once they entered the macrocells. $1/U_{d2s}$ and $1/U_{d2f}$ are the dwell times of low and high mobility users in the macrocells. The values for λ_{2fh} and λ_{2sh} are calculated using the same method that is described for the microcell handoff rate determination.

The system is solved for the steady state probabilities P_{ij} as described in [12]. For the real values of P_{ij} the handoff arrival rates should be calculated using an approach. The call blocking and call dropping probabilities for the macrocells (P_{b1} and P_{d1}) are calculated as follows:

$$P_{d1} = \sum_{i+j=Ch_2} P_{ij} \tag{14}$$

$$P_{b1} = \sum_{i+j \geq Ch_2 - G_2} P_{ij} \tag{15}$$

Finally, the overall call blocking P_b and call P_d dropping probabilities are calculated. The call blocking probability is calculated as in (16). The average call dropping rates for fast and slow mobility classes P_{ds} and P_{df}.are calculated based on the same idea.

$$P_b = \frac{\left(N \cdot \lambda_{1s} \cdot P_{b1} \cdot P_{b2}\right) + \left(\lambda_{2s} \cdot P_{b2}\right)}{N \cdot \lambda_{1s} + \lambda_{2s}} \tag{16}$$

B. Simulated Annealing Algorithm

The SA algorithm [9] is chosen because the design problem had multiple plateaus in the objective function. The algorithm starts with an initial feasible solution and makes random moves within the range of the neighbors that can be reached from the current solution. Each step in the algorithm corresponds to visiting a feasible neighbor. Fig. 3 shows the pseudo-procedure of the SA algorithm. The generation of the initial feasible solution is done by generating random radii for microcells and macrocells and by distributing the available channels among the tiers randomly. The available channels are split between macrocells and microcells randomly as well.

The neighbors of a system are determined by changing the decision parameters incrementally. The neighboring systems are either generated by transferring one channel from microcells to macrocells or vice versa, or by changing the microcell radius by a fixed percent value p supplied by the user, which effects also the macrocell radius or by changing the R/r ratio. The fixed ratio values constitute the plateau of the cost function. Those plateaus can be overcome by the nature of SA algorithm.

The SA algorithm stops in four different cases. If all the neighbors of initial feasible solution are infeasible, the algorithm does not proceed and tries to find another feasible solution. Second stopping criterion is the number of successive neighbors that are accepted due to their costs but do not confirm with the call blocking and dropping constraints. The third criterion is the number of moves that do

```
procedure SA
begin
reset the stopping criteria;
find initial feasible solution;
update the stopping criteria;
initialize temperature;
while termination criteria are not satisfied
begin
        generate neighbor;
        calculate cost;
        generate random number p;
        if  (exp. (-(New cost - Old cost)/(Temperature)) > p)   then
        if  (Neighbor is feasible)
                then  begin
                        accept the move;
                        reset the failure counter;
                        update the temperature;
                end;
                else begin
                        reject the move;
                        increase the failure counter;
                        update the stopping criteria with counter;
                end;
end;
end;
```

Fig. 3 Pseudo-Procedure of SA Algorithm

not change the cost of the system. Finally, as a precaution, the algorithm counts also the number of moves made since it started. The maximum allowable number of moves is set to be one thousand. The best value recorded is the output of the total algorithm.

C. Other Search Algorithms

In order to compare the results that are obtained by the SA algorithm, other neighborhood search algorithms are also implemented. These are the Greedy Search (GS) algorithm and Generate and Test (GAT) algorithm. The greedy search algorithm is most of the time used for problems that have a convex cost function. If the cost function is not convex, like in this case, then the GS algorithm is run starting from different point of the feasible solution space. When the cost cannot be improved any further, then the algorithm is restarted.

Setting the initial temperature in the SA algorithm equal to zero yields in the GS algorithm. In order to set the approximate running time of SA and GS algorithms, also the moves that are not accepted but for which the performance is calculated are counted. The mean value turned out to be around 1500 for the SA algorithm. Therefore, the stopping criterion for the GS algorithm is chosen as 1500 performance calculations. Then the number of feasible moves lies around 620.

The aim of GAT algorithm is to traverse the feasible solution space randomly. They can be used as benchmarks for heuristic algorithms. There is no rule that relates the generation of the successive points in the solution space. For the generation of the feasible solution the routine that calculates the initial feasible solution for the SA algorithm is used. The number of accepted solutions are bounded with the value 5000. After 5000 random feasible solutions are generated the algorithm terminates. On the average, approximately 25000 solutions are generated, but only 5000 of them are accepted for each run.

IV. COMPUTATIONAL EXPERIMENTS

The first group of experiments deals with the effect of the isolated parameters on the objective function. Secondly, the effects of the design decisions are questioned. Lastly, the selected solution technique is compared with other alternative solution techniques and the single-tier systems. The obtained results are presented graphically.

In order to prepare the problem sets for comparison, a sample problem is chosen as the base problem. This base problem reflects a typical case that can be faced when designing a cellular network. The experiment sets are prepared by changing the values of the given system. To demonstrate the effect of a parameter, all parameters except for the parameter in question are selected equal to the parameter values of the base problem. Parameters of the base problem are v_s=1 m/s, v_f=8 m/s, $SAm^2=8\text{x}10^{-8}$per sec. per m^2, $FAm^2=2\text{x}10^{-8}$per sec. per m^2, $1/U_t$=100 sec., C_1=10 Units, C_2=30 Units, A=5000 km^2, CS=7, Ch_{total}=150, $P_{b,max}$=0.01 and $P_{d,max}$=0.001.

The resulting solution that is obtained from the SA algorithm gives all the necessary decision parameters needed to implement the two-tier cellular network. For the base problem, resulting decision parameters are C=152060 Units, Ch_1=8, Ch_2=13, G_1=0, G_2=1, R=1275 m, R=425 m and R/r=3.

A. Effect of Some Parameters and Design Decisions

The first parameter to focus on is the mean speed of both mobility classes. The speed changes of both mobility classes are examined separately. Keeping the other parameters as in the base problem, the mean speed of the low mobility users are increased from 0.25 m/s to 2.5 m/s with a step size of 0.25 m/s. The resulting graph is shown in Fig. 4.

The cost of the system increases when the mean speed of the low mobility users increases. However this increase starts with a higher slope for small values of mean speed and becomes less and less when the mean speed increases. This behavior may be an indication of the fact that the system starts to benefit from the second tier of the network as a place to forward the excess calls instead of decreasing the cell radii and increasing the channel amount per m^2.

When the average speed of the high mobility users is increased, since the arriving high mobility calls cannot be overflowed to another tier, the system should react with incorporating more resources by paying more. In the previous case, the system could use the excess resources available in the upper tier. In this case, the excess traffic should be handled again within the same tier. Hence, the performance requirements can be met only if the number of available channels per unit area is increased. It can also be concluded that increases in the mobility do not effect the system setup cost too much, since the users change cells equally likely and the lack of resources are temporary.

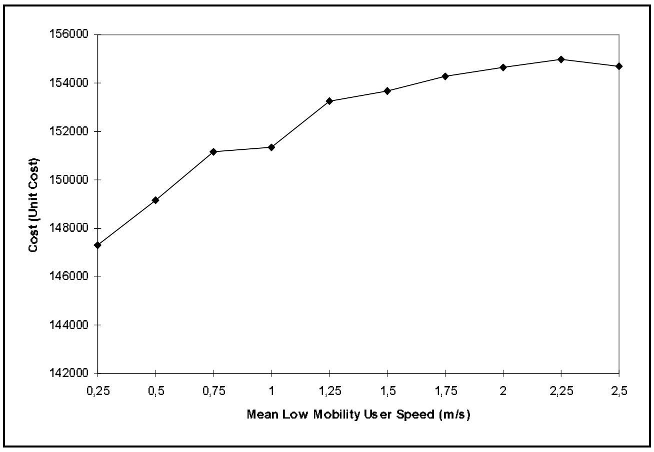

Fig. 4 Mean Low Mobility User Speed vs. Cost

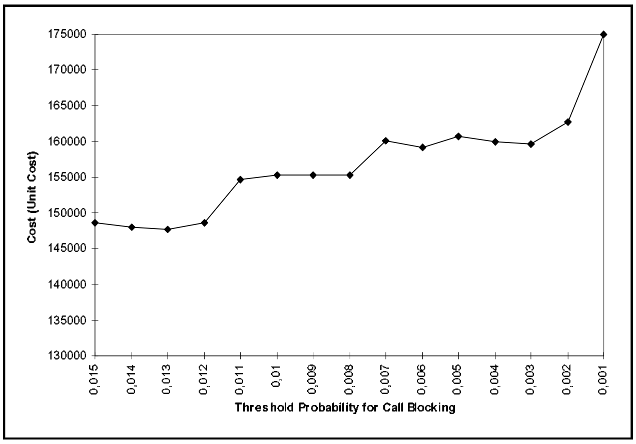

Fig. 5 Threshold Probability for Call Blocking vs. Cost

The maximum allowable call blocking probability is one of the constraints of the problem. In order to demonstrate the effect of maximum allowable call blocking probability, its value is decreased starting at 0.015 down to 0.001 with a step size of 0.001. The results of this experiment are shown graphically in Fig. 5. As the constraint gets stricter, the cost of the system goes over to different plateaus. It does not make any difference to choose a point belonging to a plateau. The value for maximum allowable call blocking threshold should be chosen such that it is as strict as possible within the limits of a plateau, since relaxing the constraints has no effect on cost as long as one stays on a plateau.

The next parameter to be inspected is the maximum allowable call dropping threshold. The value for this threshold is decreased from 0.01 to 0.001 with a step size of 0.001. The resulting values are shown in Fig. 6. For a given call blocking rate, the call dropping rate has no effect on the cost until it reaches a certain value. Only if it is set to be smaller than that value, the cost of the system starts to increase. The break point is set primarily by the call blocking rate. If that call blocking rate is satisfied, then the call dropping thresholds that are larger than the break point are satisfied anyway.

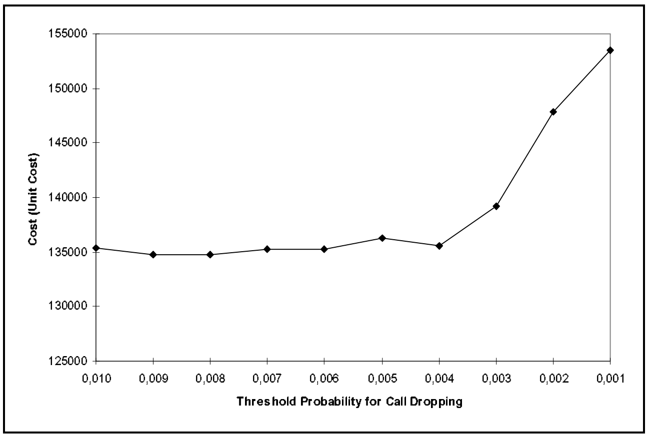

Fig. 6 Threshold Probability for Call Dropping vs. Cost

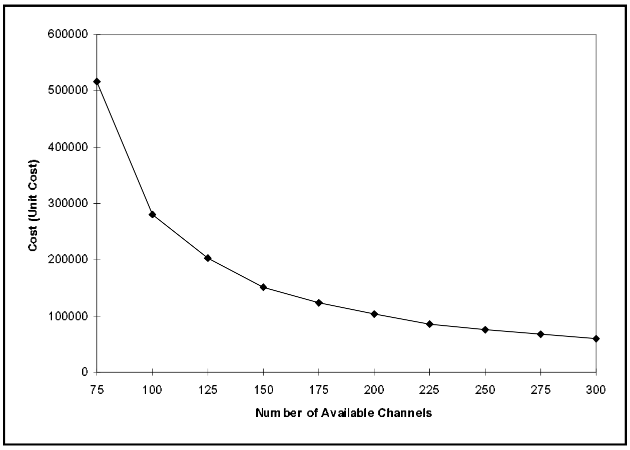

Fig. 7 Number of Available Channels vs. Cost

In order to observe the effect of additional resources on the system, the number of total available channels are increased from 75 to 300 with 25-channel. Fig. 7 summarizes the results. Increasing the number of available channels decreases the system setup cost. The gain per unit increase in the number of channels is high where the channel numbers are small. The increase decreases for larger numbers of available channels. The cost differences indicate the need for additional channels to relieve the system.

In addition to the effects of parameters and design decisions presented here, we also examined the effects of mean call arrival rates, call overflow and two-tier implementation decisions. The results can be obtained from [11].

B. *Assessment of SA Algorithm*

In order to demonstrate the behavior of the two-tier cellular network in question, several nominal and extreme parameter configurations are tested and solutions are compared with other solution techniques. All of these tests are based on the base problem defined in the introduction of Section IV. In order to compare the GS and SA algorithms fairly, the basic SA algorithm is run until 1000 accepted moves are reached. This may correspond to multiple runs of the basic SA algorithm. It is fair to run the basic SA algorithm more than once, since the results do not change very much with increasing number of runs. By doing this, the number of evaluated solutions is determined approximately. Then the number of performance evaluations in the GS algorithm is set equal to the mean number of performance evaluations in the SA algorithm. Hence, the running times of both algorithms are normalized. All the problems presented are variations of the base problem. The changed parameter is given in the description column as well as in the text. The obtained results are shown in Table I. More detailed explanations can be obtained from [11].

TABLE I

RESULTS OBTAINED WITH SA, GS AND GAT ALGORITHMS FOR SEVERAL PROBLEMS

PROB. NO	DESCRIPTION	SA	GS	GAT MIN	GAT AVG	GAT MAX
1	Base Problem	152060	155390	151890	2118037	10895180
2	Area=1000km^2	30990	30380	30570	429610	2193440
3	Area=50km^2	1550	1660	1550	20789	108860
4	Slow Low Mobility Speed=0.25m/s	147000	154050	148950	2099563	10901150
5	Fast Low Mobility Speed=3m/s	157200	161260	157360	2145840	10925040
6	Slow High Mobility Speed=5m/s	149900	154540	150460	2094752	10889240
7	Fast High Mobility Speed=20m/s	162440	165560	164150	2204543	10942990
8	Low Mobility Arrival Rate=2e-8	112590	118280	112300	2041311	10967040
9	Low Mobility Arrival Rate=1.5e-7	195400	195860	193880	2278864	10919060
10	High Mobility Arrival Rate=5e-9	76990	77380	92300	1325920	10913080
11	High Mobility Arrival Rate=8e-8	489980	774880	484280	3516816	10997170
12	Frequent Call Arrivals	646790	651960	645940	3765970	10979070
13	More High Mobility Users	218300	212540	213240	2529019	10948990
14	Short Duration=40sec	54580	56100	93250	1393697	10818160
15	Long Duration=200sec	356570	372160	338460	2952800	10954990
16	Loose Performance Limits	124360	123950	120970	1901178	10973050
17	Strict Performance Limits	161180	165680	161100	2177883	10979070
18	Very Strict Performance Limits	172760	180290	173660	2261686	10942990
19	Cluster Size=3	47690	60690	89760	1400909	10889240
20	Cluster Size=19	1054040	1069140	1052850	4531172	10991140
21	Same Unit Price	121960	123350	120960	1920096	8792900
22	Price Ratio=1:5	181300	188020	182160	2301842	13196610
23	No Macrocell Setup Cost	37500	38020	105990	1821174	7698010

V. CONCLUSION AND FUTURE WORK

In this work, we concentrated on the design of a two-tier cellular network. The goal was minimizing the cost while satisfying the performance constraints. The performance of the SA algorithm is compared with the performance of GS algorithm and GAT method. Computational experiments showed that the SA algorithm outperforms the GS algorithm in most of the cases. The quality of the solutions is compared with the results of the GAT method. The cost values obtained with the SA algorithm are either better than the solutions of GAT, or they are very close to each other. The effects of several parameters on the cost of the system are studied. Additionally, the effect of design decisions, having guard channels and allowing call overflow, are also considered. As a future work, we plan to investigate different neighborhood functions and alternative solution techniques. The time complexity of the algorithm will be studied in detail.

REFERENCES

[1] R. H. Katz, “Adaptation and Mobility in Wireless Information Systems,” *IEEE Personal Communications Magazine*, Vol. 1, No. 1, pp. 6-17, First Quarter 1994.

[2] M. Shafi, A. Hashimoto, M. Umehira, S. Ogose and T. Murase, “Wireless Communications in the Twenty-First Century: A Perspective,” *Proceedings of the IEEE*, Vol. 85, No. 10, pp. 1622-1639, October 1997.

[3] T. S. Rappaport, *Wireless Communications*, New Jersey: Prentice Hall, 1996.

[4] Ş. Tekinay and B. Jabbari, “A Measurement-Based Prioritization Scheme for Handovers in Mobile Cellular Networks,” *IEEE Journal on Selected Areas in Comm.*, Vol. 10, No. 8, pp. 1343-1350, October 1992.

[5] M. Sidi and D. Starobinski, “New Call Blocking versus Handoff Blocking in Cellular Networks,” *Proceedings of the Conference on Computer Comm.*, 1996.

[6] S. Oh and D. Tcha, “Prioritized Channel Assignment in a Cellular Radio Network,” *IEEE Transactions on Communications*, Vol. 40, No. 7, pp.1259-1269, July 1992.

[7] L. Hu and T. S. Rappaport, “Personal Communication Systems Using Multiple Hierarchical Cellular Overlays,” *IEEE Journal on Selected Areas in Comm.*, Vol. 13, No. 2, pp. 406-415, February 1995.

[8] L. Ortigoza and A. H. Aghvami, “On Optimal Spectrum Partitioning in a Microcell/Macrocell Layout with Overflow,” *Proc. of the Global Telecomm. Conference*, Arizona, 1997.

[9] S. P. Brooks and B. J. T. Morgan, “Optimization Using Simulated Annealing,” *The Statistician*, No. 44, pp. 241-257,1995.

[10] D. Bertsekas and R. Gallager, Data Networks, Second Edition, New Jersey: Prentice Hall, 1992.

[11] E. Ekici, “Optimal Two-Tier Cellular Network Design,” *Technical Report 98-104*, Bogazici University, 1998.