A simulated annealing based location area optimization in next generation mobile networks

Vilmos Simon and Sándor Imre
Department of Telecommunications, Budapest University of Technology and Economics, Magyar tudósok krt. 2, H-1117 Budapest, Hungary
E-mail: {svilmos, imre}@hit.bme.hu

Abstract. Mobile networks have faced rapid increase in the number of mobile users and the solution for supporting the growing population is to reduce the cell sizes and to increase the bandwidth reuse. This will cause the number of location management operations and call deliveries to increase significantly, and result in high signaling overhead. We focus on minimizing this overhead, by efficient Location Area Planning (LAP). In this paper we seek to determine the location areas to achieve the minimization of the registration cost, constrained by the paging cost. For that we propose a simulated annealing algorithm, which is applied on a basic Location Area partition of cells formed by a greedy algorithm. We used our realistic mobile environment simulator to generate input (cell changing and incoming call statistics) for our algorithm, and by comparing the values of the registration cost function we recognized that significant reduction was achieved in the amount of the signaling traffic.

1. Introduction

In the next generation mobile communication networks the delivery of multimedia traffic will become an increasingly important and difficult issue as cell sizes shrink to accommodate continuously larger demand for high capacity. These networks should keep handover delay limited to reduce the impact on the perceived quality of time sensitive real time multimedia applications. Careful network planning is therefore indispensable to be able to guarantee the necessary QoS for mobile users making great demands on broadband mobile applications.

While IP is declared as the key technology of the future’s wired and mobile communication, the currently used version of IP, IPv4 itself is not suitable to be used in mobile scenarios. Therefore mobility management is an important issue in mobile communication systems (MCSs). Next generation mobile users require special support to provide connectivity, although they change their place of attachment to the network frequently. The task of mobility management is to provide this support, and the management consists of two parts: location management and handover management. The first one enables to originate and receive calls for the mobile terminals; the second is responsible for administering base station changes. In this paper we are concentrating on location management, and not on handover management.

Location management strategies are the methods of locating mobile terminals (MTs) by mobile systems when incoming calls arrive for MTs. The part of radio bandwidth allocated to each mobile system is used for location management. The problem is that the radio traffic generated by location management
is increasing rapidly due to the increases in the population and mobility of MTs. Since the bandwidth is limited, the location management strategies being employed in current mobile systems may not be able to handle the overload.

There are two tasks related to location management based on which component initiates the location management procedures. They are location updating (LU), which is initiated by the MT, and paging, which is initiated by the base station (BS). During the LU procedure the MT informs the mobile system of his current location by sending an LU message to the mobile system. In paging, the mobile system searches for the given MT by broadcasting paging messages through its BSs to the area where the MT probably resides. When the searched MT receives any of these paging messages, it sends a notification message to the BS to notify its current location.

2. Related work

The LA management is classified according to its use of time, distance, movement profile information in its paging and location update procedures. Location update strategies can be classified into two main categories regarding to the events that initiates the LU procedures: in time domain or spatial domain. In the time domain LU strategy the location update can be performed due to the time elapsed since the last registration process, which has the advantage of simple implementation, but a disadvantage of poor performance caused by unnecessary LUs performed by stationary MTs. Since the location information is in the spatial domain, the spatial domain LU strategies (or topology based strategies) are to find events in the same spatial domain to trigger LU procedures.

The three most significant spatial domain LU strategies are the distance-based [17], the movement-based [9] and the most common LU strategy the zone-based, which will be the scope of our paper. In the distance-based strategy the location update will be performed, when the MT moves a threshold number of cells away from the cell where the last registration process was carried out. This threshold distance can be adjusted for each MT dynamically. In the movement-based LU strategy the number of cell boundary crossings measured since the previous update is counted by the MT, and when this number exceeds a given threshold value, the MT performs a location update. Bar-Noy et al. [1] have compared time-, distance-, and movement based schemes in terms of location management cost, and they have shown that the distance-based one performs best. However, its implementation is hard since the distance of the mobile terminal has to be computed dynamically as it moves from cell to cell. The hybrid of distance-based and zone-based approach is studied by Casares-Giner et al. [4].

Other works in this area recommend dynamic mobility management schemes based on users profile information. In this category we find proposals like the personal LA approach, in which the LA sizes are defined on a per-user basis for reducing the signalling overhead of each individual user [8,11]. Sen et al. recommend that users can skip some update operations while they cross the LAs boundaries [13]. When a call comes in, the system uses the profile information to estimate the probability for each LA that it is the current LA of the user. While the profile-based schemes offer better utilization of network resources, they add significant complexity to the location management. It requires to keep track much more information per user and to get the mobility pattern of each mobile user. This complicates the incorporation of these schemes into the current systems, because they require the modification of the standards, as well as updating the handsets. Therefore we propose a zone-based LA solution, since they are used in the deployed mobile systems. In zone-based location updating the LA boundaries are fixed and the same boundaries are used for all registrations. Since all the existing cellular networks use the
zone-based scheme (the use of Location Area) in practice, our paper addresses the optimization of such zone-based schemes to minimize the signalling cost in next generation mobile networks.

In a zone-based LU scheme upon the arrival of a call, the system tries to find the mobile terminal (MT) by searching for it among a set of base stations. This domain of base stations is called location area (LA), and the search is called paging. The called user replies to this paging message and the system establishes a connection between the originator of the call and the called user. The cells of one or few LAs are associated with a single mobile switching center (MSC) that connects them to a fixed infrastructure. At each LA boundary crossing, the MTs register their new locations through signalling messages in order to update the location management database [18]. In this way the system is able to maintain the current location of each user. Signalling overhead will be produced only when we cross a LA border, but that is rarer than a cell handover, thus the traffic of signalling messages will be reduced.

3. The optimization problem

Therefore the question arises, what size the LA should be for reducing the cost of paging and LU signalling (or registration signalling) [12]. Both, increasing and decreasing the size have their own benefit. On the one hand if we join more and more cells into one LA, then the number of LA handovers will be smaller, so the number of location update messages sent to the upper levels will decrease. However in the case of large number of cells belonging to LA, an incoming call will cause lot of paging messages [20], since we must send one to every cell to find where is the mobile user inside that LA. That will increase the load of base stations. On the other hand if we decrease the number of cells, then we do not need to send so much paging messages (hereby we will load less links and the processing time will decrease, too), but then the number of LA changes will increase. Therefore the overall problem in LA planning comes from the tradeoff between the paging cost and the registration cost (see Fig. 1).

As we mentioned earlier the most significant problem is the tradeoff between the registration and paging cost, therefore most LA planning problems in the literature have a common objective: minimizing both costs. However, they do not consider the capacities of the cellular network, since the paging capacities
of base stations and switching centres should not be exceeded. In our paper we qualify the paging cost as a constraint; therefore the registration cost is left alone in the objective function. Hence we define and formulate a problem in which the final goal is the determination of optimum number of cells per LA for which the registration cost is minimum, with the paging cost as an inequality constraint function, which is the novelty of our research contribution. We propose a LA forming algorithm that contains two phases, first a greedy algorithm (GREAL) is adopted which forms a basic partition of cells into LAs, and then a simulated annealing based algorithm (SABLAF) is applied for getting the final partition. We developed a realistic mobile environment simulator for the generation of the algorithm input metrics (cell boundaries crossing and incoming call statistics).

This paper is organized as follows. Section II presents the problem formulation and the determination of optimum number of cells per LA, Section III describes the GREAL and SABLAF algorithms. In Section IV, we present the results of computation (comparing the registration costs of manual and our algorithmic LA forming in the function of input parameters). Finally, we conclude the paper in Section V with the scope for the future research.

4. LA planning problem formulation and solution techniques

4.1. Assumptions

The aim of employing LAs is to hide the cell boundary crossing inside the LA from the upper levels; therefore an administrative message for the registration of the new location of the MT will not be generated during the cell handover if it is an intra-domain movement. Hence for the purpose to make calculations about the movement of MTs among the LAs, the best is the fluid flow model [16]. The fluid flow model characterizes the aggregate mobility of the MTs in a given region (for example a LA) as a flow of liquid. It assumes that MTs are moving with an average speed \(v\), and their direction of movement is uniformly distributed in the region. Hence the rate of outflow from that region can be described by [7],

\[
R_{\text{out}} = \left( \frac{v \cdot \rho \cdot P}{\pi} \right)
\]

where \(v\) is the average speed of the MTs, \(\rho\) is the density of MTs in the region and \(P\) is the perimeter of the given region.

This model is very simple to analyze and to use for the definition of the registration cost function. We can define easily the density of the MTs in a LA:

\[
\rho = \frac{K}{N_k \cdot S}
\]

where \(K\) is the number of MTs in the \(k\)th LA, \(N_k\) is the number of cells in the \(k\)th LA, and \(S\) is the area of a cell.

4.2. Registration cost

Every time when a MT crosses a cell boundary which is a LA boundary also, a registration process is initiated, a Location Update message is sent to the upper level (home agent or gateway). Therefore the intra-LA boundary crossing cost is negligible, and this handoff cost should be not considered in the
registration cost. Hence we need to determine the number of cells located on the boundary of the $k$th LA, like a subset of $N_k$, and the proportion of the cell perimeter which contributes to the $k$th LA perimeter. Using this perimeter of the $k$th LA:

$$P_k = N_p \cdot \nu_p (N_k)$$

where $N_p$ is the number of boundary cells and $\nu_p$ is the proportion of the cell perimeter in the LA perimeter in the function of $N_k$.

The number of the boundary cells can be approximated as we done in our earlier work [14]

$$N_p = \kappa \cdot \sqrt{N_k}$$

The proportion of the cell perimeter which will be the part of the LA perimeter too can be expressed with an empirical relation [3]:

$$\nu_p (N_k) \approx \nu \cdot \left( a + b \cdot N_k^{\eta - 1} \right)$$

where $\nu$ is the perimeter of a cell and $a = 0.3333$, $b = 0.309$, $\eta = 0.574965$.

Substituting the values of $N_p$ and $\nu_p (N_k)$ in Eq. (3), the expression for the perimeter of the $k$th LA becomes:

$$P_k = \kappa \cdot \sqrt{N_k} \cdot \nu \cdot \left( a + b \cdot N_k^{\eta - 1} \right)$$

Therefore the number of crossing the $k$th LA boundary can be given by substituting the values of $\rho$ and $P_k$ in the outflow rate of the fluid flow model:

$$q_k = \frac{\nu \cdot K \cdot N_k \cdot N_k \cdot \kappa \cdot \sqrt{N_k} \cdot \nu \cdot \left( 0.333 + 0.309 \cdot N_k^{-0.425} \right)}{\pi}$$

As we mentioned earlier a registration process is initiated when the MT crosses a cell boundary which is a LA boundary too, hence the total registration cost will be:

$$C_{LU_k} = B_{LU} \cdot q_k$$

$$C_{LU_k} = B_{LU} \cdot \nu \cdot K \cdot \kappa \cdot \nu \cdot \left( \frac{0.333 \cdot N_k^{-0.5} + 0.309 \cdot N_k^{-0.925}}{\pi \cdot S} \right)$$

where $B_{LU}$ is the cost required for transmitting a location update message.

The final goal is determining the optimum number of cells per LA for which the registration cost is minimum, with the paging cost as an inequality constraint function.

4.3. Paging cost

The paging cost is a result of the arriving calls to MTs, because the called MT has to be searched within the LA. The goal is to decrease the registration cost by increasing the number of cells in one LA, hiding the cell crossings from the upper levels. However to have a feasible network, the paging capacities should not be exceeded, therefore we need to define a paging constraint per a LA.
The limited call processing capability of a mobile switching center (MSC) will cause a limit on the peak arrival rate (known as the busy hour call attempt), therefore we need to define an upper paging cost constraint for every LA. Another constrain should be defined, the maximum number of calls terminated in one cell \((\lambda_{\text{max}})\), because a base station has a finite number of transmitters, which defines the number of channels in that cell.

The paging cost for the \(k\)th LA should not exceed the paging cost constraint (the paging cost for the \(k\)th LA will be the sum of \(C_{P_i}\) over the \(N_k\) cells):

\[
C_{P_k} = \sum_{i=1}^{N_k} C_{P_i} = \sum_{i=1}^{N_k} B_P \cdot N_k \cdot K \cdot \lambda_i < C_k
\]

(10)

\[
C_{P_k} = B_P \cdot N_k \cdot K \cdot \sum_{i=1}^{N_k} \lambda_i < C_k
\]

(11)

where \(B_P\) is the cost required for transmitting a location update message and \(\lambda\) is the number of calls terminated to a MT.

If we assume that the mobile users have the same average number of terminated calls for all cells in the \(k\)th LA \((\lambda_i = \lambda)\), the paging cost reduces to

\[
C_{P_k} = B_P \cdot N_k^2 \cdot K \cdot \lambda < C_k
\]

(12)

4.4. Optimization of the registration cost

The problem is to find the optimum number of cells per LA for which the registration cost is minimum and the paging constraint \((C_k\) must not be exceeded) is satisfied.

If we know that the call arrivals \((\lambda)\) follow a Poisson process and the function of the registration cost Eq. (9) is a monotonically decreasing function, the paging constraint can be expressed in the following way:

\[
P(C_{P_k} < C_k) < 1 - e^{-\gamma}
\]

(13)

where \(\gamma = (10, 100)\), depending on the accuracy of the paging constraint.

The monotonically decreasing attribute of the registration cost function will mean, that we need to find the highest value of the \(N_k\) for which the Eq. (13) will be still satisfied.

Substituting the expression of the paging cost in Eq. (13):

\[
P(B_P \cdot N_k^2 \cdot K \cdot \lambda < C_k) < 1 - e^{-\gamma}
\]

(14)

Furthermore, if we know that the \(\lambda\) probability variable follows a Poisson process, then the maximum value \((N_{\text{max}})\) of \(N_k\) can be easily calculated:

\[
P(\lambda < \frac{C_k}{B_P \cdot N_k^2 \cdot K}) = 1 - e^{-\gamma}
\]

(15)

Substituting the calculated value of \(N_k\) in Eq. (9) will give us the minimum of the registration cost. We will use this calculated \(N_k\) as an input for our GREAL and SABLAF algorithm.
5. Location area planning algorithms

Our goal is to develop Location Area Planning algorithms, which consider the paging constraint, and take the available mobility pattern and cell perimeter information as input, and find an optimal or near optimal LA structure for which the registration cost will be minimal.

5.1. Greedy Algorithm (GREAL)

The registration cost is proportional to the number of handovers among different LAs ($q$), therefore the registration cost can be minimized by designing the LAs such that the cells belonging to one LA have the lowest boundary crossing rates among each other. We designed a realistic mobile environment simulator [15], which will generate this boundary crossing database, a handover rate for each cell pair, defined on the border of these cells. The incoming call statistic can be also generated for every cell; therefore the paging cost can be calculated at the same time for every LA. The GREAL algorithm starts with choosing the cell pair with the biggest handover rate in our cell structure ($q_{max}$), if multiple values are equal to this rate, then we choose one of them randomly and include the two cells into the LA set of cells. In the next step, we search for the second biggest handover rate (if there is more than one, we choose it in the same way as in the first step) among the cell pairs for which is true, that one of them belongs to the LA set of cells. We must check is inequality

$$N_k < N_{max}$$

satisfied, where $N_{max}$ is the maximum value of $N_k$ calculated from Eq. (15), namely the maximum number of cells in a LA which will give us the minimum of the registration cost. If the inequality is satisfied, the cell can be included into LA set of cells. If the inequality is not satisfied, this cell can not be included into this set, not exceeding the paging cost constraint Eq. (12). In this way we can join the cells which are in the same dominant moving directions, therefore the number of handovers among LAs can be decreased (highways, footpaths, etc.).

After the processing of all cell pairs in a sequential way, there will be cells that are not group of any set of cells. These cells will form another LA, which is not the best solution (greedy algorithm can only find the local minima of the solution space), but this will be only a basic LA partition which will serve as an input to the simulated annealing based LA forming algorithm.

5.2. The Simulated Annealing Based Location Area Forming Algorithm (SABLAF)

The optimal partition of cells into LAs is an NP-hard problem [2]. A general approximation algorithm that runs in polynomial-time needs to be used in order to solve this kind of problem. It is difficult to find such algorithm to obtain near optimal solution. Simulated annealing is considered an approximation algorithm which is applicable to various problems in general. In this paper we are focusing on the application of the simulated annealing (SA) approach to LA optimization problems. Metropolis algorithm [10] was the original idea behind the optimization technique of SA. Kirkpatrick et al. [6] has used Metropolis algorithm as a global optimizer. Thus, simulated annealing finds a suboptimal solution without searching the entire solution space.

The simulated annealing based algorithm here finds a basic LA partition formed by the GREAL algorithm, with the handover rate database among the cell pairs. The SABLAF algorithm starts with this initial solution, $s_0$. A neighbour to this solution $s_1$ is then generated as the next solution by simulated
annealing, and the change in the registration cost $\Delta C_{LU}(s_0, s_1)$ is evaluated. If a reduction in cost is found, the current solution is replaced by the generated neighbor, otherwise we decide with a certain probability set to $e^{-\Delta C_{LU}/T}$ (usually called the acceptance function) whether it remains the current solution, where $T$ is the control parameter (corresponds to temperature in the analogy with the physical annealing process). The SABLAF algorithm is started with a relatively high value of $T$, to have a better chance to avoid being prematurely trapped in a local minimum. The cooling schedule consists of three parameters, used like an input to SABLAF algorithm: the initial temperature ($T$), step of decreasing, and the stopping rule of the algorithm. The stopping rule is the maximal iteration step number or maximum number of steps when the $\Delta C_{LU}$ do not changes (see Fig. 2).

Another important input parameter is the calculated maximum number of cells in a LA ($N_{\text{max}}$). The performance of the SABLAF depends heavily on the cooling schedule and the initial partition, which should be carefully investigated and optimized to have the best results.

6. Results

We have tested the two LA forming algorithms for two types of networks (rural and urban mobile environment, already presented in [15]) and compared them with a manual solution where the cell partitions are made intuitively, the cells are grouped along the main moving directions of the MTs. Therefore the manual LAs can be considered as a planned partition, but not the optimal one. The input data, the boundary crossing and incoming call database, for the algorithms was generated by our realistic mobile environment simulator [15] (see Fig. 3.). We compared the performance of the GREAL and SABLAF algorithm and the manual solution, by using the two mobility environments.

As it is well-known, the rural mobility environment is rarely populated, but on the belonging highways a big number of mobile terminals are moving with high speeds, while the urban mobility environment is densely populated, with mobile terminals moving with smaller velocities. In the rural environment
the average cell size is larger then in the urban, accordingly there is a smaller number of cells. In this manner we designed the two mobility environments, the rural environment consisted of 42 cells, while in the urban system it was about 79 cells. We ran the simulation on these example networks and used the output of the simulation (boundary crossing and incoming call database) as the input parameter for our two-step LA forming program (we have coded the two algorithms). We examined how the registration cost changes by increasing the maximum number of cells in one LA. This way we could check if the registration cost function is correct, and if it reaches the minimum value when a LA consists of the calculated Eq. (15) maximum number of cells ($N_{\text{max}}$).

Figure 4 shows the registration cost in rural environment, where the horizontal axis represents the $N_{\text{max}}$, used in the Eq. (16) of the GREAL algorithm. As it can be seen the greedy algorithm can decrease the registration cost, but only the SA based solution finds the best solution for every value of $N_{\text{max}}$. For the calculated value of $N_{\text{max}} = 12$, the registration cost reaches the minimum value using the SA based technique.

Figure 5 shows the registration cost in the urban environment, where we have more cells, but the size of the cells is smaller. In the initial stages (where the value of $N_{\text{max}}$ is less than 5) SABLAF algorithm almost always accepts the partition formed by the GREAL algorithm or a very similar partition (which do not decreases the registration cost appreciably), but as the number of cells increases, the SABLAF is outperforming the other two solutions significantly. Using again Eq. (15), $N_{\text{max}} = 14$ (higher value than for the rural environment, because of the cell sizes and numbers), for that value the SABLAF algorithm gives again the minimum of the registration cost, therefore we can summarize, that for both mobility environments the SABLAF algorithm gives the best results, decreasing the registration cost almost for 50% more effective than the manual solution, and for approximately 20% more than the GREAL algorithm.
In order to design a feasible mobile network, we need optimized LA planning considering the constraints like paging handling capacity of a MSC. These networks should keep handover delay limited to improve the QoS parameters of the time sensitive real time multimedia applications. Optimized LA forming can help us attaining a significant reduction was in the signalling traffic that causes delay and delay variation.

In this paper, we presented a generalized formulation of the LA planning defining the registration cost function and the paging constraint, and a method for calculating the number of cells in a LA for which the registration cost is minimum. We also proposed a two-step LA forming solution, which consists of a greedy algorithm that gives the basic partition and a simulated annealing based algorithm which gives a suboptimal solution in reasonable running time. To evaluate the performance of our new scheme, we designed a rural and an urban environment in our mobility simulator, and with this database we run the LA forming algorithms. Then we compared them with a manually designed LA structure, examining the relation of the registration cost with an upper bound on the number of cells in a LA.

For the both mobility environments the SABLAF algorithm delivered the best performance, reducing the registration cost almost 50% more effective than the manual solution. This significant signalling load reduction can help us improving QoS parameters in real time multimedia applications.

7. Conclusion

Fig. 4. The registration cost in rural environment.

Fig. 5. The registration cost in urban environment.
Fig. 6. Future plan: The hierarchical entity forming algorithms.

An optimization algorithm based on the inter-LA movements need to be developed in the future, which will help us to plan a hierarchical mobile structure which results in the minimal signalling traffic. Therefore our future goal is to create a hierarchical network design algorithm based on the structure given by the LA forming algorithms, and an agent (MAP) router selection algorithm in Hierarchical Mobile IPv6 to optimize the handover management in IP based next generation mobile networks. The signaling cost is proportional to the number of handovers among different hierarchical entities; therefore the signaling cost can be minimized by designing the hierarchical structure such that the Location Areas belonging to one hierarchical entity have the lowest LA boundary crossing rates among each other (Fig. 6.). In this way the signaling messages will only be sent one level up in the hierarchy, and not to the top of the hierarchy. In the next step, the second level of the hierarchy entities will be grouped on the same principals.

Accordingly our future goal is to develop hierarchical entity forming algorithms, which consider the topology constraints, and take the available mobility pattern and LA boundary crossing information as input, and find an optimal or near optimal hierarchical structure for which the signaling cost will be minimum. Careful network planning is therefore indispensable to be able to guarantee the necessary QoS, to reduce the impact of handover delay on the perceived quality of time sensitive real time multimedia applications.

References


**Vilmos Simon** received the MSc degree in Telecommunication Engineering from the Budapest University of Technology and Economics in 2003. He is currently pursuing the PhD degree in the Department of Telecommunications at the Budapest University of Technology and Economics. His research areas include location management in next generation mobile networks, IP mobility, modelling of multimedia traffic and information spreading in biologically inspired networks.

**Sandor Imre** received the MSc degree in Telecommunication Engineering from the Budapest University of Technology and Economics in 1993. In 1999 he received the PhD degree in Electrical Engineering. He is an Associate Professor at the Budapest University of Technology and Economics and the R&D Director of the Mobile Innovation Center, Hungary. His research areas include IP mobility, routing, reliability; Wireless LANs; Software Defined Radio; Quantum Computing.