The Use of Prediction Areas to Improve Mobility Management Algorithms

Jens Biesterfeld, Klaus Jobmann

University of Hanover
Institute for Telecommunications
30167 Hannover, Germany
Tel: +49 511 7622831, email: biest@ant.uni-hannover.de

ABSTRACT

The paper discusses the two operations that are necessary to provide mobility service in telecommunication networks, the paging and the location updating messages. Several methods to optimise signalling traffic for mobility management purposes are discussed. They are classified into static and dynamic procedures. This paper proposes methods for the prediction of the location of users in order to minimise signalling traffic. One problem in using prediction methods is that they do not know the actual location of a subscriber. For that reason a solution is proposed to guarantee the localisation of a mobile user which is the introduction of prediction areas. Finally the proposed method is analytically compared with the classical method.

I. INTRODUCTION

In future telecommunication networks the number of subscribers and the user demand for higher transmission rates will increase significantly. Additionally the mobility service has to be introduced in all future networks that are intended to be competitive. Due to this changing needs new mobility management methods have to be developed.

A. Motivation

Existing mobility management methods do not take advantage of all information that are possible to get from the network and the users behaviour. E. g. the classical MM uses location updating every time a terminal leaves a location area and paging in the whole location area in case of mobile terminated call request. Other proposed methods use information that is obtainable from the users motion pattern to predict locations. These procedures are able to reduce signalling, but they cannot guarantee that the mobile terminal is found [1]. For that reason these methods have to be improved.

B. Paging-Location Update

In networks which offer a mobility service, two operations are necessary to track the users. The first task is the location update which serves to inform the network about the terminals. The second is paging which is used to inform the terminal about an incoming connection request. In case the terminal receives the paging message it answers the request in order to establish a signalling channel. Both operations require opposite area sizes, the location area should be large to minimise the location updating messages and the paging area should be small to minimise the paging traffic and vice versa. Existing solutions operate with a compromise [3].

Our proposal is to define different areas for the location update and the paging. The smallest area for the location update is called location area. It is used to determine points for sending update messages to inform the network about the location of the mobile terminal. For the paging operation paging areas are defined to determine the smallest area in which paging operation may be performed. The smallest area for both is one cell. A cell is defined in this paper as the coverage of one antenna sector.

Figure 1 shows one example of the relation between number of cells per location area and the cost for location updating and paging messages.

$$C_{total} = A \cdot N + B \cdot (\frac{1}{2} + \frac{1}{2 \cdot N}); \quad A = B = 1 \qquad (1)$$

It exists a minimum of the sum of both messages for a specific number of cells per location area [2].

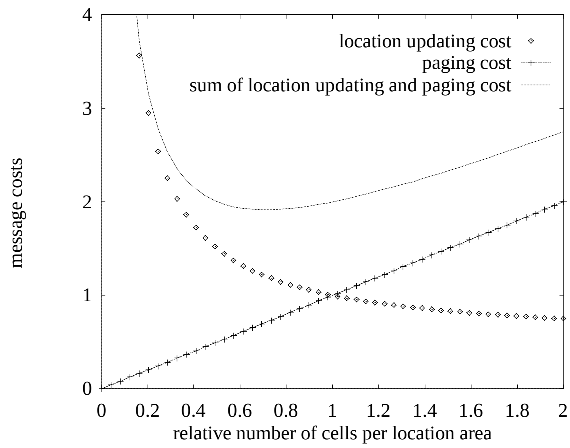

Figure 1: Location updating and paging traffic

The paper is organised as follows. In the second section the location update and the proposals to optimise it are introduced. The third section discusses the paging and the approaches to minimise the signalling traffic for it. The fourth section discusses the relation between paging and location areas. The fifth section introduces predictive mobility management methods and prediction areas. Finally a performance analysis closes the paper.

II. LOCATION UPDATE

A. Static Registration

The simplest strategy to send registration messages to the network is unconditional by space or time criteria. Both methods are static, because there is no change in the condition for updating the location area number.

The condition by space is often defined as the crossing of a location area border. In this case the network is permanently informed about the location of the terminal. The disadvantage of this method is that the information contained in the users motion pattern is not used to reduce the signalling amount.

The second condition, updating by time, leads to the fact that the network does not know the location of the subscriber at each time. When the user changes his location between to consecutive location updating times, his new location is not given to the network until the next updating time. The time difference between the registration messages can be adapted to the users motion pattern. This method has the same disadvantages as the space conditioned, for that reason the exclusive use of the time conditioned method is in general not sensible.

Today's mobile networks use a combination of both methods, though it should be noted that the time condition may be turned off, e. g. in the Global System for Mobile Communications (GSM).

B. Dynamic Registration

Dynamic registration is defined in this paper as the sending of location updating messages adapted to the users movements. Updating is carried out only if it is necessary, e. g. in case that the network is no longer able to predict the actual location or the paging cost would exceed the location updating cost.

For the dynamic registration several methods are proposed. Some of them use the information contained in the users motion pattern. All methods try to optimise the relation between update and paging messages. The first are called *predictive mobility management* methods. They try to predict the actual location of a user in case of incoming call. Registrations are only performed if the network is not able to reach the mobile terminal within appropriate time or without exceeding the limit for the number of paging operations.

A second method with dynamic registration as proposed in [4] calculates the sum of the paging and registration costs and performs a location update when the paging cost would exceed the registration cost. For the definition of the cost factors see [3]. The function to be minimised is the sum of the registration and paging costs

$$V_u(X(0)) = E\left\{\sum_{t=1}^{T} A \cdot u_t(X(t)) + f(X(T))\right\} \quad (2)$$

where A is a cost factor for registration messages, u_t is $\in \{0,1\}$ dependent on if the registration operation is performed or not and $f(X(T))$ is the cost of the paging messages at the time T. $X(0)$ defines the state in which the last call was terminated, that means the last connection between the terminal and the network was available and the location was known. This method leads to good results if the user is not very mobile and the ratio of paging and registration cost is high [4].

III. PAGING

As mentioned previously the paging is used to inform the mobile terminal of an incoming call request. The terminal answers with a signalling connection establishment request to the network. It is necessary since the status of the terminal and the exact location by means of radio sector are in general not known.

In most cases paging messages are sent to more than one cell. For that reason two different strategies are distinguished: parallel and sequential paging. Using *parallel paging* all messages are sent at the same time, which leads to a minimum delay in receiving the answer. If no answers arrives, e. g. if the terminal is switched off or the radio connection to the network is lost, a timer has to control that the maximum delay is kept. *Sequential paging* is a procedure which sends the messages in a given order at different times. The advantage is that if the answer from the mobile terminal arrives before all paging messages are sent no further paging is required. This method is able to save signalling capacity, but it increases the delay between call request and the answer to this request.

A. Static paging

If the amount of paging areas to reach a specific terminal is known in advance, the paging procedure shall be called static in this paper. The network is always sure about the capacity spend for paging messages. This method is often used with static registration methods. The disadvantage of this combination is again the disregard of the useful information contained in the users motion pattern. In current implementations, e. g. the GSM, the paging area is identically to the location area. This structure guarantees the localisation of the terminal if the mobile station is not switched off or the radio connection is not disturbed.

Static paging has the disadvantage that the worst case is performed every time, because paging messages are sent to all possible areas without any condition.

B. Dynamic paging

A paging method is called dynamic if the number of paging messages that are sent depends on the specific situation and is not set in advance from the network structure. This procedure can be used with static registration as well as in combination with dynamic registration methods.

In the case of dynamic paging in general the location area is not the same as the paging area. Since the exact location is not known in terms of paging area, several paging messages have to be sent to reach the mobile terminal.

If the amount of paging messages is not known in advance a limit of sequential paging attempts has to be determined. This limit influences quality of service, as the delay between call request and positive or negative answer should not exceed a defined value. If different service classes are defined this point may be a subject to choose for the subscriber to vary the cost of the service.

If dynamic paging is used with static registration the network knows the location area the user is located in, but the smaller paging area is not known exactly. That means the paging areas contained in the well known location area are paged according to a specific strategy. This method allows to improve existing networks which uses static registration and static paging. The following examples show the proposed algorithms. One example to optimise paging within static registration area is described [5], it pages sequential in the paging areas in one location area until the mobile terminal answers. Another strategy is the use of predictive methods which predicts the order of paging areas to page [1]. A third one is to provide a table containing the probabilities that the user is located in the different paging areas [6].

If dynamic paging is used with dynamic registration the signalling savings might be maximal. A predictive mobility management strategy witch uses sequential paging in the order of the location probability is a method witch uses dynamic registration and dynamic paging.

Another possible strategy is to page with rising geographical radius from the point the subscriber registered the last time [4].

IV. PAGING AND LOCATION AREAS

All the mentioned methods base on a defined logical structure of paging and location areas. The process of determining these areas is an optimisation process for its own. It is dependent on the network structure, the users behaviour and the resource allocation of the registration and paging messages. To find a good solution for the optimal structure heuristic methods [7] and genetic algorithms [8] constitutes the state of the art. The organisation of these areas may be done for the whole network or individual for each subscriber. The first method is used in most current network implementations, but the second one is necessary to optimise signalling, because users have different optimal structures.

Current networks use *static registration areas.* That means the network is divided into areas that are not changed dynamically and automatically during operation. These areas may be reconfigured when the traffic changes over time, but in a quasi static manner with the necessity of manual adaptations. In most existing networks registration area and paging area are the same.

In future networks *dynamic registrations areas* will be introduced. The generation of the area structure should be carried out automatically in dependence of the above mentioned parameters. To enhance the previously discussed mobility management methods it is useful to define user dependent areas. That means each user has his own pattern of areas in the network. Since processing power and memory cost decrease, it is better to save signalling capacity than to save computational equipment in the network. The area pattern is stored and computed at the same location as the data for the registration mechanism at the user data base. For that reason each database needs an additional control unit with powerful processing equipment.

V. PREDICTION METHODS

Predictive mobility management methods estimate the actual location of a user in case of incoming call requests. They use the information obtainable from the historical user movements, that are recorded in an observation phase, during which no prediction is carried out. If enough information on the users motion pattern is collected, the prediction method is trained. That means that a table is build up [6], [9] or a neural network is trained [1] with the movement data. After that, prediction can be used to save signalling capacity in the network, especially on the air-interface. Figure 2 shows a flowchart of the different phases that are necessary when using predictive mobility management methods.

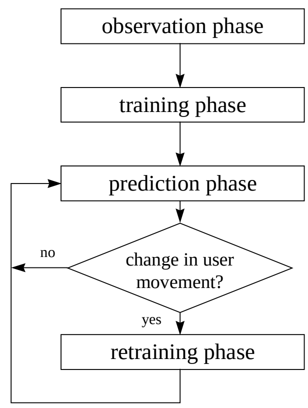

Figure 2: Flowchart of predictive methods

The disadvantage of the predictive methods are that the actual location of the user is not known. In the case that the paging attempts reach the limit without an answer, a parallel paging in the whole network is necessary. One solution to prevent this world-wide paging is to introduce *prediction areas.*

A. Prediction Areas

Prediction areas shall be defined individually for each user. Prediction areas have to be constructed by regarding the users movement pattern. The optimisation criterion can be the minimisation of the sum of prediction area updating cost and average paging cost for the subscriber. These areas constitute the maximal number of paging areas to page if the user is not found

within the limit of paging attempts. For that reason the worst case, namely the world-wide paging, is restricted to the prediction area. If a mobile terminal changes the prediction area it has to inform the network. For that reason prediction areas have similar tasks as the location areas. The main differences are that prediction areas may be geographically larger than location areas with the same paging traffic and paging is normally not carried out in the entire area. The size of the prediction areas depend on the gain obtained by the prediction procedure. This gain is dependent on the users motion process and the prediction method. The absolute value has to be determined by simulations in each specific case [3]. Paging areas have to be build as small as possible to improve the adaptive mobility management methods. The individual structure optimises the overall gain in signalling savings. The advantage of this method is that it can be implemented in existing networks.
Figure 3 shows an example how the logical structure of the different areas may look like. The paging areas are chosen as small as possible that means one cell is equal to one paging area.

B. Paging Strategy

The paging strategy depends upon the order that is formed by the prediction method. Paging messages may be used in parallel or sequentially to meet the delay constraints. For example if several locations have the same probability that the user is located actually in there, the paging messages may be sent in parallel or sequentially in groups. But if the probabilities are quite different the most probable area should be paged first and then sequentially the other ones in decreasing order of probability. There is no possibility to give a general hint to the best paging strategy, because the performance depends strongly on the actual situation. If used adaptively this method is a tool to optimise the traffic in the network. For example if there is no bottleneck in the signalling channels parallel paging can be carried out with the advantage of low delay. When the channels are congested, sequential paging may be preferred to ease the traffic, with the disadvantage of a higher paging delay.

C. Location Update Strategy

Predictive mobility management methods as proposed in this paper use dynamic registration. In the prediction phase registration messages are sent when crossing a prediction area border. This fact shows that the location areas are not necessary in the prediction phase, but they are needed in the observation phase to track the subscriber.

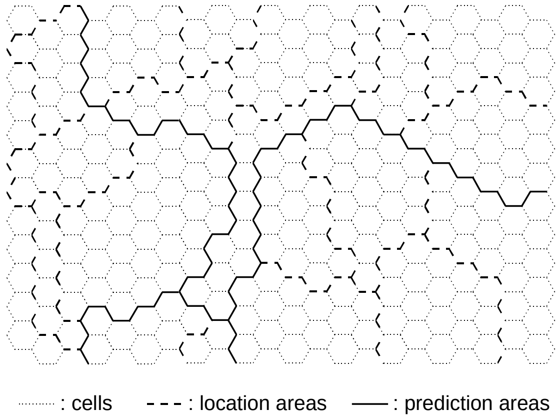

Figure 3: Example of a logical area structure

VI. PERFORMANCE ANALYSIS

To calculate the performance of the proposed algorithm in comparison to the classical method with fixed registration and paging areas the model from [4] with the extensions from [10] is taken. The cost of paging messages generated in one cell is

$$n_{pag/LA} = D \cdot n_{cells/LA} \quad (3).$$

D consists of the terminal density ρ, the cell area A, the traffic for mobile terminated calls λ and the time bandwitdh product T_pB_p necessary for paging messages:

$$D = \rho \cdot A \cdot \lambda \cdot T_p \cdot B_p \quad (4)$$

The cost of location updating messages generated in one cell on the border of a location area is:

$$n_{lup/LA} = E \cdot L_f(n_{cells/LA}) \quad (5)$$

The constant E consist of the terminal density ρ, the mean terminal speed $\bar{v}$ and the the time bandwitdh product T_lB_l necessary for location updating messages:

$$E = \frac{\rho \cdot \bar{v}}{\pi} \cdot T_l \cdot B_l \quad (6)$$

Introducing prediction methods these equation have to consider the probability p of a correct prediction. To ease the analysis only one paging attempt in one cell is permitted. It should be mentioned that this assumption reduces the performance of prediction algorithms because more than one paging attempt may be allowed.
The cost of paging extends to

$$\begin{aligned} n_{pag/PA} &= D \cdot p + D \cdot (1-p) \cdot (n_{cells/PA} - 1) \\ &= D \cdot (p + (1-p) \cdot (n_{cells/PA} - 1)) \end{aligned} \quad (7)$$

The cost of location updating messages only changes because of the higher number of cells per prediction area:

$$n_{lup/PA} = E \cdot L_f(n_{cells/PA}) \quad (8)$$

Since $L_f(n)$ is a function which descends with rising n and $n_{cells/PA} \geq n_{cells/LA}$ location updating messages are saved in every case. The saving of location updating messages at the expense of rising paging traffic. To secure that the number of paging messages is lower than in the case of location areas the following equation is used to calculate the maximum ratio a of cells per prediction area to cells per location area:
From dividing equations (3) and (7) we get the

following relation:

$$\frac{n_{pag/PA}}{n_{pag/LA}} = \frac{D\cdot(p+(1-p)\cdot(n_{cells/PA}-1))}{D\cdot n_{cells/LA}} \le 1$$
$$\Rightarrow (p+(1-p)\cdot(n_{cells/PA}-1)) \le n_{cells/LA} \qquad (9)$$

with $a = \frac{n_{cells/PA}}{n_{cells/LA}}$ and equation (9) the maximum ratio a is

$$a \le \frac{n_{cells/LA}+1-2p}{(1-p)\cdot n_{cells/LA}} \qquad (10)$$

This result shows that there is always a number of cells per prediction area which leads to a lower or the same signalling traffic for paging in comparison to the location area method. This is shown in the following limit examination:

$$p \to 0: \quad a \le \frac{n_{cells/LA}+1}{n_{cells/LA}} \approx 1 \qquad (11)$$
$$p \to 1: \quad a \to \infty$$

Figure 4 shows a comparison between the cost for *D=1, E=10, a=2, p=0,7* and square cells.

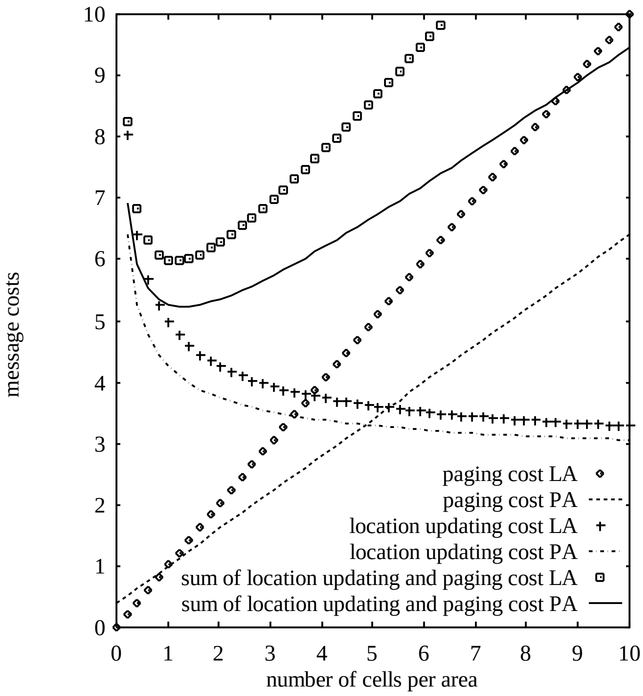

Figure 4: Example for the different cost functions

The example in Figure 4 shows that the introduction of prediction areas improves the performance of the signalling by decreasing the sum of the traffic for location updating and paging messages. As this is an example there are other combination of parameters which lead to different results. The general statement which can be stated is that by a careful choice of the ratio of the number of cells in the location and in the prediction area as given in (10) the proposed method leads to better or equal result than the classical method.

VII.CONCLUSIONS

This paper gives a short overview on mobility management procedures. It proposes the use of predictive location management methods because they increase the gain in signalling savings. To overcome the disadvantages of these methods an enhancement to limit the paging amount in the case the user is not found within the paging limit is discussed. The proposal is to create prediction areas. An analytic performance analysis shows that the signalling traffic may be reduced in introducing prediction areas.

To obtain the best method in the case of a specific realisation, simulations with real traffic pattern and user movements processes are needed.

REFERENCES

[1] J. Biesterfeld, E. Ennigrou, K. Jobmann, "Location Prediction in Mobile Networks with Neural Networks", Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications, 1997, pp. 207-214.

[2] R. Thomas, H. Gilbert, G. Mazziotto, "Influence of the Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network", in Proceedings of the Third Nordic Seminar on Digital Land Mobile Radio Communications, 1988, pp. 1-9.

[3] J. Biesterfeld, K. Jobmann, "Evaluating Different Mobility Management Methods", in Proceedings of the 3rd Workshop on Personal Wireless Communications, 1998, pp.167-174.

[4] U. Madhov, M. L. Honig, K. Steiglitz, "Optimisation of Wireless Resources for Personal Communications Mobility Tracking", IEEE Transactions on Networking, vol. 3, 1995, pp. 698-707.

[5] R. Rezaiifar, A. M. Makowski, "From Optimal Search Theory to Sequential Paging in Cellular Networks", IEEE Journal On Selected Areas in Communications, vol. 15, 1997, pp. 1253-1264.

[6] S. Tabanne, R. Nevoux, "An intelligent location tracking method for personal and terminal FPLMTS/UMTS communications", in Proceedings of the International Switching Symposium, vol. 1, 1995, A2.1.

[7] K. Lim, Y.-H. Lee, "Optimal Partitioning of Heterogeneous Traffic Sources in Mobile Communications Networks", IEEE Transactions on Computers, vol. 46, 1997, pp. 312-325.

[8] P. R. L. Gondim, "Genetic Algorithms and the Location Area Partitioning Problem in Cellular Networks", in Proceedings of the Vehicular Technology Conference, 1996, p. 1835.

[9] G. Y. Liu, G. Maguire, "Efficient Mobility Management Support for Wireless Data Services", in Proceedings of the Vehicular Technology Conference, 1995, p. 902.

[10] Chin-L. I, G. P. Pollini, R. D. Gitlin, "PCS Mobility Management Using the Reverse Virtual Call Setup Algorithm", IEEE/ACM Trans. on Networking, vol. 5, 1997, pp. 13-247