Minimizing the Average Cost of Paging on the Air Interface – An Approach Considering Privacy

Hannes Federrath*, Anja Jerichow*, Dogan Kesdogan‡, Andreas Pfitzmann*, Dirk Trossen‡

*Dresden University of Technology, Institute of Theoretical Computer Science, D-01062 Dresden, Germany
‡Aachen University of Technology, Computer Science Department IV, D-52056 Aachen, Germany

Abstract – Location Management of mobile users in a cellular network is considered from a performance and privacy point of view. Location management covers tracking functionality and paging (searching) functionality. After a risk analysis of location management w.r.t. privacy, we focus on the paging strategy. A sequential search strategy is proposed which reduces the signaling on the air interface and also considers the user's privacy.

I. INTRODUCTION

Since bandwidth is a very scarce resource in radio networks, it is necessary to subdivide the service area into many cells in order to allow the re-use of frequencies. If a call has to be established for a mobile user in a cellular network, for efficient location management it is desirable to know the current cell of the subscriber and transmit the call set-up message (called paging operation and paging message) only in this cell. Unfortunately, the "knowledge" about the current position of the mobile user is not for free: Due to the mobility of the user, it is obvious that she or he will enter and leave cells and this information must be updated in the network to establish mobile terminated calls. The higher the granularity of the managed location information (e.g. one cell) the higher is the Location Update (LUP) cost due to the update frequency of the location and vice versa. To balance the costs of LUP and paging in GSM [1] a number of cells are grouped to form a Location Area (LA). Now, the positions of the users are stored in the central databases in terms of LAs and the paging message will be broadcasted in all cells of the LA.

One current trend to minimize the LUP and paging costs uses adaptive and dynamic tracking and paging algorithms [2, 3, 4, 5, 6, 7]. The main idea of these algorithms is to collect and compute more information about the user in order to find out an individual mobility and behavior pattern of the user.

Using this information, the network should search the whereabouts of the user with minimal signaling effort and the granularity of the location knowledge should depend on the individual call rate and mobility. For example if a user is very mobile and gets few calls the location area for this user should be as large as possible, because small location areas result in wasteful LUP signaling and paging signaling occurs seldom.

We categorize these adaptive and dynamic management proposals into two types (see Table 1): user oriented and network oriented management.

A management concept is called user oriented if the user could behave dynamically according to his own collected personal mobility and communication data. For example, [2] describes an user oriented method for individual calculation of location area size in accordance with their own mobility and communication behavior. This information (the list of cells) is transmitted to the network. The network pages the user only in these cells and does not need to collect information about the individual call arrival rate or mobility behavior of the mobile user.

Otherwise, if the network collects data about its subscribers and behaves dynamically then we call the concept network oriented. Several investigations assume that the probability distribution of user location is provided. This probability distribution may be user-dependent [6]. Having such information, the paging message can be broadcasted at first time only in a portion of the LA. If the first attempt fails the user can be searched successively according to the probability distribution.

	User Oriented Management	Network Oriented Management
net-work	basic functionality: switching, paging etc.	advanced functionality: forecast the position, digital assistant etc.
mobile station	intelligent	dumb

Table 1: In the user oriented management strategies the mobile station has the intelligence (information) whereas in the second type the network has the intelligence (information and processing power). The categorization can not be always done by simple either or not decision, it is more weighing up the two categories in a proposal.

From the viewpoint of privacy and security, the network oriented management strategies are highly alarming. We discuss the risks of storing location information in centralized databases in Chapter II. In Chapter III, we describe technical means to protect location information. The described security strategies lie along the user oriented management line. Following this direction, in Chapter IV, we propose a new dynamic paging strategy (basic functionality) that reduces also signaling on the air interface and can be used in combination with the strategies presented in Chapter III. We conclude our work with a modeling approach for performance evaluation in Chapter V.

The work of H. Federrath and D. Kesdogan was supported by the Daimler- and Karl Benz-Foundation. The work of A. Jerichow was supported by the German Science Foundation (DFG).

II. CONSIDERING SECURITY IN MAINTAINING LOCATION INFORMATION

A. Security Risks

From the privacy and security point of view, location information are private data which should be protected. Interests to compromise the users exist at least for VIP's as reported from [8]. We consider this also valid in business and private environments. Potential fields of misuse of location information by third parties are many, thus we give only two examples. In business the attacker could be interested in the strategic alliances between two companies. In private environment the private life of the user may be the matter of interest. Having an access to the location information the attacker could gain information about the circle of visiting areas (e.g. home, bank, gambling house) and the intensity of visiting these areas. Furthermore, if the user has a regularity in his mobility behavior the attacker could also forecast the users actual position [9].

B. Weak Point of GSM like Systems

The weak point of GSM like networks' handling of location information is their centralized approach, i.e. location information is stored in HLR and VLR. All information about the user is accessible at one point. In the future this problem will be increased, because several network providers and several service providers must interwork. It is an old hacker's rule, that "the more people with access the better" a system can be compromised [10]. If the network accumulates more accurate location data in the central databases by using individual behavior statistics, the interest in stealing this information will increase. Consider a burglar who finds access to the profile. He would not only know the actual location of the user but also the time the user will arrive at home. Thus, we consider the management of accurate location information and storing the mobility profile of a user in the network databases as an unacceptable invasion of privacy.

C. Security Requirements

In a communication network illegal gain of information, unrecognized change of information and disturbance of the functionality by unauthorized persons must be prevented. In this regard security requirements are confidentiality, integrity and availability.

Confidentiality means that data is only available to authorized persons. Especially the protection of the profile information of a mobile user needs to be managed. Neither the potential communication partners nor third parties (including the network operator) should be able to have access to the users' location data and to generate user mobility profiles.

Integrity stands for whether data is correct, complete and current as well as for the proof that a sender has sent a message and/or the addressee has received the message.

Availability means that the communication network enables communication between all parties who wish to communicate (and who are allowed to).

III. CONFIDENTIAL LOCATION INFORMATION MANAGEMENT – THE STATE OF THE ART

All proposed confidential management strategies [11, 12, 13, 14, 15, 16] start out from user oriented management (Table 1). The network should serve the user with basic functionality and the location of the user should be protected. Again, considering the user's privacy the performance requirements should also be considered: availability of location information, the time restrictions of reaching the mobile user and the resource limitations (e.g. air interface). Of course, there is always a trade-off between privacy and performance. In this chapter, we give a brief overview about the suggested location protection methods on the basis of their common point, i.e. implicit address. In Chapter IV a new method using implicit addresses with enhanced performance characteristic is proposed.

A. The Usage of Implicit Addresses

A call set-up message can be made anonymous by broadcasting an implicit address to the recipient. An implicit address [17, 18, 19] is a meaningless pseudo randomly generated bit string which is only understandable by the recipient. Explicit addresses give no anonymity because they directly describe a place, host or person.

Broadcasting implicit addresses results in a completely anonymous establishment of mobile calls [11, 14]. Every mobile subscriber can listen, compare the received address with his identity and then answer if he is addressed. Regarding the air interface, broadcasting is very costly and because of limited bandwidth in radio networks not applicable. Therefore, location information is needed in order to make the broadcast areas smaller. In the following we sketch the main ideas of the approaches considering the protection of the location information.

B. Different Approaches

One approach [11, 12, 14] keeps the location information in a trustworthy environment. The preferred location of the trustworthy environment should be at home of the user and therefore it is called "Home Personal Computer (HPC)". The HPC has to play the role of the databases (e.g. HLR and VLR in GSM). Thus, the location registration and the location cancellation take place at the HPC rather than the VLR and the HLR. To establish a connection to a mobile user, the network requests the current location and implicit address of the user from his HPC.

The approach described in [15] uses implicit addresses as pseudonyms to anonymize the users. The network holds and manages exact location information like in GSM with the given appropriate pseudonym in HLR and VLR. The identity of the mobile user is hidden from the network by changing these pseudonyms frequently. It is proposed that the pseudonym of the user is generated synchronously in the user's HPC and his mobile station in an autonomous way. Therefore, the mobile station has to register itself periodically with the new pseudonym. Now, if a call has to be established towards the mobile user, only the implicit address (pseudonym) must be requested from the HPC. Location management can be performed like in GSM.

Another approach [16] uses special network stations (so called MIXes). These stations realize the unlinkability between input data and output data. The network gets specially prepared data packets from the mobile station which should be used only for call establishment purposes. If a call has to be established towards the mobile user the network transmits the given packet via MIXes. The call can be established in a trustworthy manner assuming the MIXes do not conspire with the network operator. Receiving the packet back from the MIXes the network could read the packet containing the implicit address. As in the approach described above the implicit address could be interpreted as pseudonym. The call can be established like in GSM if the location information of the pseudonym is known beforehand to the network databases.

IV. VARIABLE IMPLICIT ADDRESSES

Until now the implicit address was used as a unit. In this paper we investigate the usage of the implicit address, which is not used in its full length in one broadcast. We call this new type of implicit address "variable implicit address" and analyze the performance gain of this new addressing scheme for cellular networks.

A. Basic Idea

The main idea is to use the implicit address P not always in its full length $n=|P|$. P can be divided in k segments of length $l_i=|P_i|$ ($i=1\ldots k$) such that $\sum_{i=1}^{k} l_i = n$. To address somebody, the segments P_i are broadcasted step by step. C is a set of cells. C represents a group of subscribers. The paging procedure is described in the following pseudo code:

```
10 LET C = all cells of the served area
20 LET k = number of segments
30 FOR i = 1 TO k DO
    {Downlink (network – mobile station): }
        Broadcast Pi to all cells in C.
    {Uplink (mobile station – network): }
        IF (mobile station answered in all previous steps AND
        owns the broadcasted Pi )
            THEN send YES (one bit message). {The YES
            message means "I have the broadcasted segment".}
            {ELSE send nothing}
    LET C = all cells with YES message
    IF number_of_elements(C) = 1 THEN GOTO 50
40 END FOR
50 {Cell selection finished}
```

Fig. 1: Cell separation and paging procedure for segmented implicit addresses

The example in Fig. 2 shows a cell selection for 13 cells and k=4 steps. In the first Step P_1 is broadcasted over the whole served area (13 cells). In our example the YES message comes from only 10 cells. In the next step the next segment is broadcasted to this 10 cells. By sending segments a narrowing of the group C takes place. Finally, after the cell selection the called mobile subscriber is located and the communication channels can be established to a certain cell.

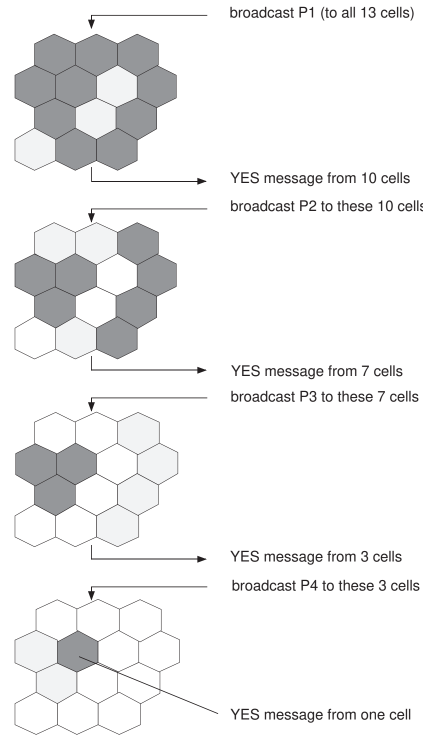

Fig. 2: Graphical example of cell separation

As smaller the segments are chosen as more steps are needed and as less bandwidth is needed for broadcasting one segment. However, the delay of the call set-up procedure increases.

The kind of addressing described is called variable because flexibility is given by sending segments and allowing answers within a certain limit. The YES message will use only a small amount of bandwidth for the message transfer from the mobile station to the network.

B. Efficient Realization of the Answering Channel

There is no need for a collision free answering channel on the uplink. It makes no difference whether one or more answers are received from one cell. One YES message already causes a broadcast in the next step. Two or more YES messages can be overlaid to one sum YES message (logical OR).

The final separation of the subscribers answering in the last step is realized by an authentication procedure, e.g. GSM Challenge Response Procedure.

In a TDMA (Time Division Multiplex Access) system the logical OR can be performed on the radio system (see. Fig. 3). We need a synchronization between the uplink and downlink TDMA channel within a cell. All YES messages for the broadcasted segment P_i at the downlink time slot t_α must be sent at the corresponding uplink time t'_α. Two or more YES messages are already overlaid on the air interface.

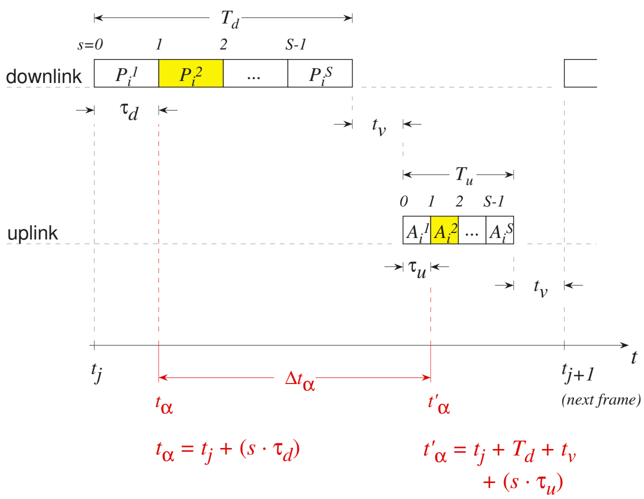

Fig. 3: Synchronization between downlink and uplink channel in the TDMA system

We denote the frame duration as $T_{d/u}$ (downlink/ uplink), the number of slots per frame as S (equal for downlink and uplink). Then, the slot duration is $\tau_{d/u} = \frac{T_{d/u}}{S}$.

P_i^s is the broadcasted segment P_i at the time slot s in the downlink frame t_j. A_i^s is the corresponding answer slot to P_i^s. The duration t_v is needed to switch the receiver/transceiver between receiving and sending, uplink and downlink channels etc.

That means, in one TDMA frame, S segments of S different implicit addresses can be broadcasted. If the mobile station knows the number of the time slot (in Fig. 3 s=1) it can compute the corresponding sending time t'_α.

C. Decreasing the Length of the Address Segments

Instead of using a constant segment length we search for the subscribers in steps with decreasing segment length in order to reduce the number of steps needed. The protocol in Fig. 4 halves the broadcasted segment length from step to step.

Compared to Fig. 1 this procedure reduces the upper bound of the number of broadcast steps and the delay to $\log_2(n)$. In every broadcast step the number of cells answering YES should on average be halved by this protocol.

10 LET C = all cells of the served area
20 LET r = n
30 WHILE (r>1 AND number_of_elements(C)>1) DO
Broadcast (the next) $\left\lceil \frac{r}{2} \right\rceil$ bits of P to C
IF (mobile station answered in all previous steps AND owns the broadcasted bits) THEN send YES
LET C = all cells with YES message
r:=r $-\left\lceil \frac{r}{2} \right\rceil$
40 END WHILE
50 Broadcast the last r bits of P
60 {Cell selection finished}

Fig. 4: Cell separation with decreasing segment length

D. Adaptive Length of the Address Segments

Considering the varying number of subscribers in a cell from time to time (or even from step to step) in every broadcast step, the length of the next segment could be dynamically adapted.

There are two cases to consider. First, a large number of cells answer YES. The next segment length should be increased in order to reduce the number of the answering parties. Second, only a few cells answer YES, then the length of the next segment should be decreased.

V. PERFORMANCE ANALYSIS

In this chapter we give a mathematical model for the basic idea of the paging algorithm. Therefore we derive the distribution and the expectation for the necessity to broadcast a segment until the unique addressee is found.

We divide the address space of l bits in k segments. The broadcast of the k segments is modelled by a game with n players at beginning. The station to which the address belongs does not play (because it will win everytime!). Each broadcast of one segment is similar to a step of the game. Each player gets a number from 1 to $2^{l/k}$ uniformly distributed in each step.

Thus a player gets a specific number with probability $p=1/2^{l/k}$. A player getting a wrong number quits the game, which excludes the player from playing the next steps. Therefore the game ends at least after k steps.

Let X_m denote the number of players who do not quit the game after m steps. This number is binomially distributed with parameters n and p^m. Thus the probabilities of the events $\{X_m=0\}$ that there are no players left after the m-th step of the game are derived by

$$P(\{X_m = 0\}) = (1 - p^m)^n \quad ; \quad 1 \le m \le k \qquad (1)$$

For m=0 define $\{X_0=0\}:=\varnothing$. The events $\{X_m=0\}$ increase according to

$$\{X_i = 0\} \subset \{X_{i+1} = 0\},\ i=0,..,k\text{-}1 \qquad (2)$$

because every event that causes the game to end after i steps causes the game to end after i+1 steps, too. Thus the events that the game ends exactly after the m-th step are given by $\{X_m=0\}\setminus\bigcup_{i=0}^{m-1}\{X_i=0\}=\{X_m=0\}\setminus\{X_{m-1}=0\}$. Define $\Omega=\{\{X_1=0\}\setminus\{X_0=0\},\ldots,\{X_k=0\}\setminus\{X_{k-1}=0\}\}$. The probability of Ω is derived by

$$P(\Omega)=\sum_{i=0}^{k-1}P(\{X_{i+1}=0\}\setminus\{X_i=0\})=P(X_k=0)$$

Let Y denote the number that the game ends exactly after the m-th step. The expectation of the number of steps necessary to end the game is derived by

$$E[Y]=\frac{1}{P(\Omega)}\cdot\sum_{i=1}^{k}i\cdot P(\{X_i=0\}\setminus\{X_{i-1}=0\}) \qquad (3)$$

This may be simplified according to (1),(2) and (3) by

$$E[Y]=\frac{1}{P(\Omega)}\cdot(k\cdot P(\{X_k=0\})-\sum_{i=1}^{k-1}P(\{X_i=0\}))$$

Fig. 5 shows the expectation of sending segments depending on the number of segments. We choose l=32 bits of address space and k=2,4,8,16,32 segments. Different numbers of stations are listed to show that by increasing this number also increase the expectation of sending segments.

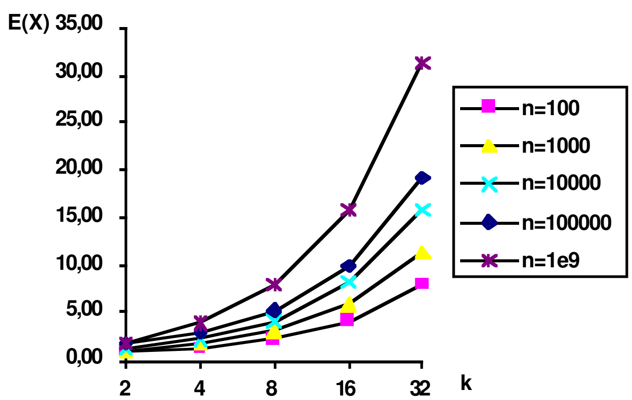

Fig. 5: Mean Expectation of the Segments sent

The not transmitted segments of the l bits represent the gain of the algorithm, i.e. $E(X)<k$. As shown in Fig. 5 the mean expectation of sending segments for smaller segments brings more gain than transmitting big segments. However, the time delay and power consumption of the mobile stations are the drawbacks of the algorithm.

In this first attempt we only give a mathematical model for the basic idea of our paging algorithm. Further investigations (simulation) are needed to evaluate the more precise protocols suggested in this paper and find out the tradeoff between the discussed parameters.

REFERENCES

[1] M. Mouly, M.-B. Pautet: The GSM System for Mobile Communications, ISBN: 2-9507190-0-7, 1992.

[2] H. Xie, S. Tabbane, D. Goodman: Dynamic Location Management and Performance Analysis, IEEE 43th VTC 1993.

[3] A. Bar-Noy, I. Kessler, M. Sidi: Mobile users: To update or not to update?, Wireless Networks I, 1995, pp.175-185.

[4] I.F. Akyildiz, J.S.M. Ho: Dynamic Mobile User Location Update for Wireless PCS Networks, ACM-Baltzer J. Wireless Networks, vol. 1, no. 2, July 1995, pp.187-196.

[6] C. Rose, R.D. Yates: Ensemble Polling Strategies for Mobile Communications Networks, IEEE 46th VTC 1996.

[7] G.P. Pollini, S. Tabbane: The Intelligent Network Signalling and Switching Costs of an Alternate Location Strategy using Memory, IEEE 43th VTC 1993.

[8] P.G. Neumann: Computer Related Risks, ACM Press, Addison Wesley, New York, NY, Jan. 1995

[9] D. Kesdogan, M. Zywiecki, K. Beulen: Mobile User Profile Generation – A Challenge between Performance and Security, Proceedings of the IFIP TC6 International Workshop on Personal Wireless Comm., D-Frankfurt, Dec. 96.

[10] K. Hafner, J. Markoff: Cyberpunk, Corgi Books, London, 1993.

[11] A. Pfitzmann: Technischer Datenschutz in öffentlichen Funknetzen. Datenschutz und Datensicherung DuD 17/8 (1993), pp. 451-463.

[12] T. Hetschold: Aufbewahrbarkeit von Erreichbarkeits- und Schlüsselinformation im Gewahrsam des Endbenutzers unter Erhaltung der GSM-Funktionalität eines Funknetzes. GMD-Studien Nr.222, Oct. 1993.

[13] M. Spreitzer, M. Theimer: Architectural Considerations for Scalable, Secure, Mobile Computing with Location Information. Proceedings of the 14th International Conf. on Distr. Sys., IEEE 1994.

[14] H. Federrath, A. Jerichow, D. Kesdogan, A. Pfitzmann: Security in Public Mobile Communication Networks. Proceedings of the IFIP TC6 International Workshop on Personal Wireless Communications, Prague (Czech Republic), April 1995, pp. 105-116.

[15] D. Kesdogan, H. Federrath, A. Jerichow, A. Pfitzmann: Location managment strategies increasing privacy in mobile communication. Informations Systems Security, IFIP SEC '96, Chapman & Hall, London, 1996, pp. 37-38.

[16] H. Federrath, A. Jerichow, A. Pfitzmann: Mixes in mobile communication systems: Location management with privacy. Inform. Hiding, LNCS 1174, Springer-Ver., Berlin 96.

[17] D.J. Farber, K.C. Larson: Network Security Via dynamic Process Renaming; Fourth Data Communications Symp., Oct. 1975, Quebec City, Canada, pp. 8-13 & 8-18.

[18] P.A. Karger: Non-Discretionary Access Control for Decentralized Computing Systems; Master Thesis, MIT, Laboratory for Computer Science, May 1977, Report MIT/LCS/TR-179.

[19] A. Pfitzmann, M. Waidner: Networks without User Observability; Computers & Security, North-Holland, 6/2 (April 1987) pp. 158-166.