Database Correlation for Positioning of Mobile Terminals in Cellular Networks using Wave Propagation Models

D. Zimmermann, J. Baumann, M. Layh, F. Landstorfer
Institut für Hochfrequenztechnik
University of Stuttgart, Germany
http://www.ihf.uni-stuttgart.de
zimmermann@ihf.uni-stuttgart.de

R. Hoppe, G. Wölfle
AWE Communications GmbH
Gärtringen, Germany
http://www.awe-communications.com
reiner.hoppe@awe-communications.com

Abstract— **Many new services for mobile radio networks rely on a precise location of the mobile terminal. Especially in urban environments multipath propagation leads to very complex scenarios without line-of-sight between the mobile and the different base stations. In these situations a location technique based on simple delay evaluations and/or GPS is not accurate enough or not applicable. In this paper a database correlation method for the positioning of mobile terminals is presented which yields precise results in urban environments. The approach is based on propagation models utilized for the planning of mobile radio networks. The results of the propagation models during the planning process define a look-up-table. By evaluating the measured path losses and available propagation times between the mobile terminal and several base stations, the positions of mobile terminals in urban environments can reliably be predicted.**

I. INTRODUCTION

Location based services (LBS) require that the location of the mobile terminal is available to the network. This location can be gained from the inherent radio parameters of the network or by using GPS receivers. The latter possibility must be assisted by the network (A-GPS) to work properly in the mentioned field of application. This technique performs well in open areas with unimpeded view on the satellites but in indoor environments or in narrow street canyons of urban scenarios the method is not applicable without additional hardware.

The other mentioned possibility is to exploit the radio parameters of the wireless network. There are several radio parameters which can be useful for positioning. Both the representation of the physical values in the system and the available accuracy may vary in the different network types, but in principle there are only few location-sensitive parameters.

A. Common Location Techniques

In cellular systems there is always some information identifying the serving cell enabling a very rough location estimate. For more precise positioning it is advantageous that mobile terminals monitor the received power as well as (at least in TDMA systems and UMTS) propagation time to their serving base station. Received power suffers from shadowing and multipath phenomena and is therefore very unreliable for positioning purposes. Propagation time is more dependable but in cases without line-of-sight (LOS) between transmitter and receiver the distance is overestimated.

Additionally, time differences between several base stations could be measured (hyperbolic system) or the angle of arrival of in-bound waves could be determined (triangulation). Hyperbolic systems require synchronization in the network which is not fulfilled in e.g. GSM and UMTS. Additional hardware measuring the asynchrony can compensate for that but raises the costs. However, there are implementations like E-OTD for GSM.

The determination of angles of arrival requires antenna arrays. Owing to their size such arrays are not feasible for handsets and restricted to base station sites. The most severe problem is apparently that waves may arrive from directions other than the location of the mobile terminal which is caused by reflections or diffractions.

Hybrid methods utilizing more than one parameter can achieve better results but the major problem of all mentioned methods is that they commit a systematic error when LOS conditions do not apply.

B. Novel Approach

A completely different approach is the so-called database correlation. It overcomes the above mentioned problems by using a database built from measurements or predictions. The position of the mobile terminal will then be determined by evaluation of the radio measurements, which are performed by the mobile terminal and/or the base stations. These measurement data are compared to the entries of the database. The corresponding correlation calculations find the best matching database entry and thus lead to a location estimate. Making appropriate measurements over large areas for constructing the database is very expensive and therefore not considered applicable. Prediction data on the other hand can be obtained from wave propagation simulation tools, which are used for network planning anyway.

This paper deals with database correlation relying on prediction data. Various correlation algorithms and results from measurements in different environments are presented.

II. Database Correlation

A. Database Construction

Several methods and parameters for database correlation have been proposed or realized for positioning purposes in mobile communication networks. Some are exploiting previously gained measurements e.g. U.S. Wireless Corporation's RadioCamera and DCM [1]. The advantages of these techniques include avoiding accuracy degradation in non-line-of-sight propagation conditions. Their major drawbacks are the requirement of new hardware at base station sites (RadioCamera), and the need for the construction, maintenance, and storage of a large database of measurements (DCM). Any changes in buildings or city infrastructure would imply making new measurements to be included in the database. This turns the method into a costly solution.

The database in the method presented here is constructed from predicted values. During the planning process of wireless networks, propagation predictions are computed considering each base station in the whole region in order to analyze the coverage and interference scenario. The predictions are based on the semi-deterministic model developed in COST 231 for urban areas and on the empirical model according to Hata-Okumura for rural areas [2]. Information about the prediction area as building databases for urban areas is considered as well as digital height models from topographic maps. Starting from these results the database can be constructed as a look-up-table (LUT) for each prediction pixel. This LUT can contain any location sensitive parameter available in the network. Either one single parameter per cell and location or several different parameters can be used. In the latter case the cost-function, which evaluates the match of measurement and LUT entry, becomes multi-dimensional. As an example, in Table I path loss and propagation time values (marked with L and T respectively) are stored in the LUT.

In the following we will only deal with one-dimensional cost-functions based on predicted received power.

B. Matching Algorithms

Let N be a set of n cells for which the received power is stored in the LUT and N^* be a set of n^* measured cells.

To determine a location the measured values of the stored parameter of up to n^* measured cells are compared to the entries in the LUT. The value of n^* depends on system aspects of the mobile communications standard. Measurements in this paper were carried out in a GSM network, so $n^* = 7$.

TABLE I. STRUCTURE OF THE LOOK-UP-TABLE

Location		**Cell**									
		1		***2***		***3***		***i***		***n***	
x_1	y_1	$L_{1,1}$	$T_{1,1}$	$L_{2,1}$	$T_{2,1}$	$L_{3,1}$	$T_{3,1}$	…	…	$L_{n,1}$	$T_{n,1}$
x_2	y_2	$L_{1,2}$	$T_{1,2}$	$L_{2,2}$	$T_{2,2}$	$L_{3,2}$	$T_{3,2}$	…	…	$L_{n,2}$	$T_{n,2}$
x_3	y_3	$L_{1,3}$	$T_{1,3}$	$L_{2,3}$	$T_{2,3}$	$L_{3,3}$	$T_{3,3}$	…	…	$L_{n,3}$	$T_{n,3}$
…	…	…	…	…	…	…	…	…	…	…	…

For each predicted location a matching score can be computed from the deviation of the predicted entries p_i and instantaneously measured values m_i for cell i.

In [2] several methods are given to quantify the deviation of predictions and measurements at each location. In this paper we will refine the evaluation not only regarding the best matching entry but also show a way to benefit from a suitable evaluation of the residual entries.

1) LMS: Using the least mean square (LMS) approach the score S_{LMS} of a LUT entry is given by

$$S_{LMS} = \sum_{i \in N^*} (p_i - m_i)^2 = \sum_{i \in N^*} \Delta_i \tag{1}$$

where predictions and measurements are denoted by p_i and m_i, respectively. Thus the location with the minimum value of S_{LMS} matches best.

2) EXP: Another method to compute the score is motivated by the Gaussian probability distribution. Assuming that the single contributions of deviations in cells i are uncorrelated, the score S_{EXP} can be seen as the joint probability of the occurring measurement to match an entry. As can be seen from

$$S_{EXP} = \prod_{i \in N^*} e^{-\left(\frac{p_i - m_i}{\sigma}\right)^2} = e^{-\frac{\sum_{i \in N^*} \Delta_i}{\sigma^2}} \tag{2}$$

the score S_{EXP} is closely related to S_{LMS}. The parameter σ characterizes the deviation between predictions and measurements and has to be chosen appropriately. In our measurements it showed that a value of 16 dB yields good results without much change between 10 and 20 dB.

The location minimizing S_{LMS} and maximizing S_{EXP} is the same but the exponential approach seems to be more suitable for improvements. We therefore concentrate on this exponential approach which significantly devaluates bad matches.

The best matching location is found at maximum score. Typically, the measurements will deviate from the predictions so that S_{EXP} diminishes with the number of factors considered. If the LUT is incomplete, i.e. not all cells are predicted at any location, this penalizes locations with a higher number of comparable entries to those with less comparable entries. However, this counteracts the idea that a location which does not provide the respective entries is very unlikely to be the wanted location. Therefore, the probability for the measurement to match the LUT entry is a function not only of S_{EXP} but also considers the number of available cells n^*. P_{EXP} is regarded as the probability for the measurement to match the LUT entry.

$$P_{EXP} = \sqrt[n^*]{S_{EXP}} \tag{3}$$

In order to benefit from the set of residual cells N' in the LUT which were not received by the mobile terminal a penalty coefficient is introduced. This penalty coefficient P_{Pen} consists of penalty contributions $P_{Pen,i}$ which are computed for all n' residual cells predicted to be stronger than the weakest measured cell m_{min}.

$$P_{Pen} = \sqrt[n']{\prod_{i \in N'} P_{Pen,i}} = \sqrt[n']{\prod_{i \in N'} e^{-\left(\frac{p_i - m_{min}}{\sigma}\right)^2}} \qquad (5)$$

The matching probability P of each location is consequently a function of P_{EXP} and P_{Pen}.

$$P = \sqrt{P_{EXP} \cdot P_{Pen}} \qquad (6)$$

The location with highest matching probability is finally chosen as location estimate.

III. Propagation Models

A. Suburban and Rural Predictions

For the suburban/rural scenario the model according to Hata-Okumura was used to compute required prediction data. This simple empirical model processes only the transmitter and receiver height. Additionally, terrain obstacles between the two terminal antennas are taken into account using the Epstein-Petersen Knife Edge Model. Diffraction losses are determined under consideration of the terrain profile and added to free space loss. For the LUT these suburban/rural predictions are created with a resolution of 20 m.

B. Urban Predictions

The prediction data for the urban scenario is based on the Extended Walfisch-Ikegami Model, developed in COST 231 [5]. The dominant propagation mechanism of this model is the propagation over the rooftops with following diffraction into street canyons which is shown in Fig. 1. This semi-deterministic propagation model yields good accuracy for transmitters on roof tops, but especially wave guiding effects in street canyons are not considered. The prediction data used for the LUT was generated with a resolution of 10 m under consideration of a building database.

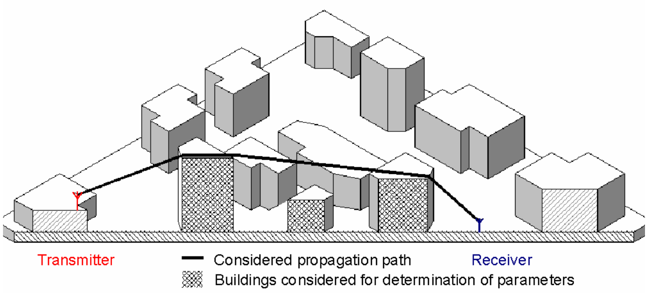

Figure 1. COST 231-Walfisch/Ikegami Model

IV. GSM-Measurements

A. GSM Network Monitor

An easy way to access measurement data of GSM terminals is the network monitor function. It provides, among other parameters, the CellID and BCCH power level of the serving cell as well as the BCCH power levels of up to six neighbor cells. The related CellIDs can be derived from the base station color codes (BCC). In dedicated mode the timing advance (TA) of the serving cell is measured and displayed additionally.

B. Measurement Set-up

In order to obtain data in normal user situations a measurement campaign is analyzed. The measurements were carried out by a standard mobile phone controlled by a notebook PC using AT-commands [4]. Additionally the PC was used for storing the data. A GPS receiver to provide a location stamp for each measurement and to calculate location errors is connected, too. For this experiment GPS errors are neglected even though expected to be in the range of 10 m in the considered scenarios. Still, location accuracy of GPS is expected to be substantially higher than the accuracy of the location methods discussed here.

Measurements were performed by a pedestrian moving at walking speed inside the considered areas along two routes. Both in the urban and in the suburban/rural scenario the routes were 2 km long. With measurements performed every five seconds, the two routes consist of 365 and 350 measurement points, respectively. The mobile phone was carried inside a backpack which also contained the notebook PC. The GPS-mouse was mounted on top of the backpack to enable good location fixes.

C. Exploited Data

GSM emphasizes on received power level measurements (RxLev) and supports only very rough timing information (TA). Each TA step corresponds to an equivalent distance of approximately 550 m. Not only is this quantization of TA measurements rather coarse but also are the measurements inaccurate. The measurement campaign showed that values of TA are often at least 2 steps higher than they are supposed to be according to the real distance between base station and mobile terminal. Even additional delays owing to NLOS do not account for this deviation. Therefore, we did not exploit timing information for the methods discussed in this paper. However, in areas with base station sites located more distant from each other TA may be helpful to improve location accuracy.

V. Results

An urban area which is usually covered by a dense network will enable higher location accuracies compared to a wide meshed network in rural environments. Absolute values of achievable accuracies are therefore not appropriate to be used as a significant capability measure of a location technique. The use of basic location methods as a reference is suggested to allow an objective assessment of more sophisticated approaches in terms of additional accuracy.

A basic reference method is to employ the active site. The position of the currently serving base station is taken as location estimate. This method simply interprets the CellID and gets the appropriate site from the location server.

An urban and a suburban/rural scenario are investigated. The GSM network in the city-centre of Stuttgart (Germany) was chosen as an example for an urban area with a real operating network. The considered area is approximately 10 km² large and consists of 109 cells with both omni-directional and sector antennas. Cell sizes are between 50 m and 1 km.

The rural/suburban scenario is located in the vicinity of Stuttgart with mostly suburban building structure and forest. There are 43 cells inside an area of 50 km² with distances between the base stations varying from 1 to 5 km. As no building database for propagation predictions was available only predictions according to the Hata-Okumura model could be performed.

Fig. 2 to 4 show location results from the urban scenario. The displayed maps are 2.5 km x 2 km covering the central part of the prediction area of the urban scenario. The route according to GPS is marked in blue. The brown and the red line in Fig. 2 and 3 connect the location estimates obtained with the database correlation method. The brown line is acquired with LMS, the red line with EXP as described in subsection II.B. In Fig. 2 the black dots indicate the antenna locations which were serving cell for some measurements on the route. The black line in between the dots does not mark a route but connects the locations of active sites in the order of appearance during the measurements.

The EXP approach combined with the penalty algorithm derived earlier leads to a route which has significantly less outliers than the one obtained from the LMS algorithm.

The cumulative distribution function (CDF) in Fig. 5 compares the location errors of the three methods. Location errors are computed as horizontal distance between location estimate and GPS.

Table II summarizes statistics not only for the urban scenario but also for the suburban/rural case. The CDF in Fig. 6 belongs to this scenario as well. As can be seen in Table II LMS and EXP performed almost equally. Therefore only EXP is displayed in the CDF in Fig. 6.

The CDF of the method using the position of the serving base station as location estimate features some steep edges. Such rapid increases of the location error emerge from the location of the measurement route inside the cell structure of the network. The distance which is covered by the measurements is small compared to the distance between base station sites. When the serving cell changes to a significantly more distant site while the mobile terminal has hardly been moved the location error increases discontinuously. As the route never gets close enough to this distant site no values in between are measured. With the mobile phone switching to dedicated mode for each measurement in our experiment such changes of the active site may occur both for traffic and for fading reasons.

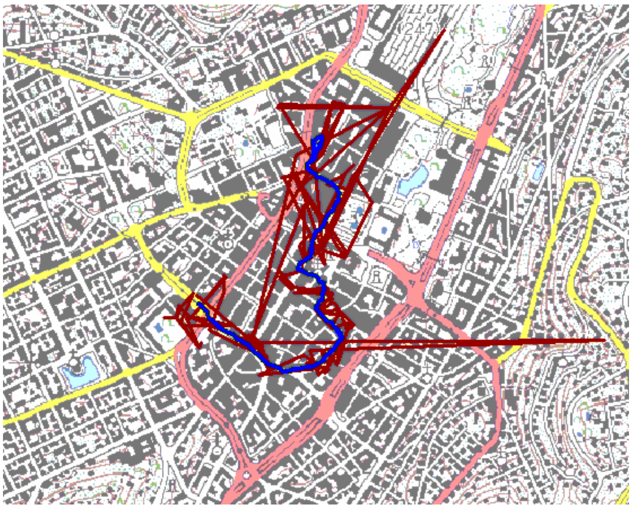

Figure 2. Real (blue) and estimated route (brown) of a walk through the city-centre of Stuttgart (Germany); depicted area: 2.5 km x 2 km

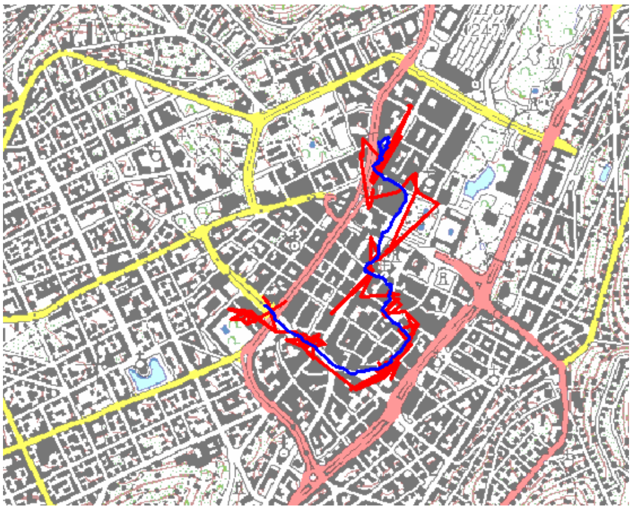

Figure 3. Real (blue) and estimated route (red) of a walk through the city-centre of Stuttgart (Germany); depicted area: 2.5 km x 2 km

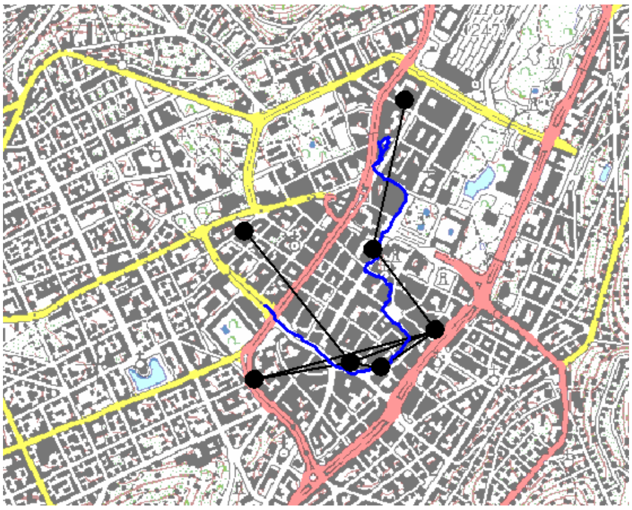

Figure 4. Real (blue) and estimated route (black) of a walk through the city-centre of Stuttgart (Germany); depicted area: 2.5 km x 2 km

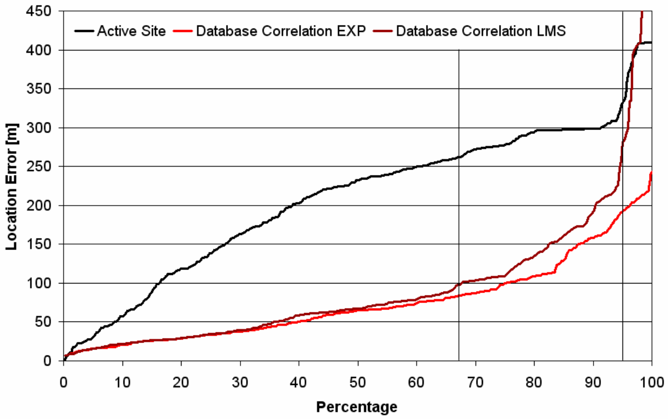

Figure 5. Comparison of the three cumulated distribution functions of location errors in the urban scenario

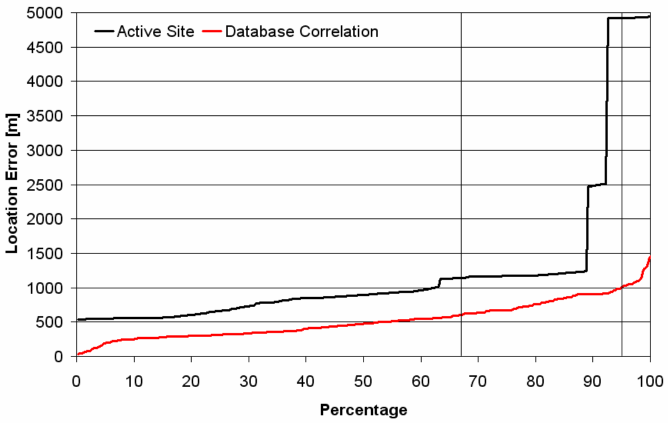

Figure 6. Comparison of the two cumulated distribution functions of location errors in the suburban/rural scenario

TABLE II. STATISTICAL EVALUATION OF LOCATION ACCURACY

Location Error		**Database Correlation**		**Active Site**
		LMS	***EXP***	
Urban scenario	***67%***	98 m	83 m	262 m
	95%	282 m	192 m	332 m
	Mean Error	99 m	75 m	209 m
Suburban/rural scenario	***67%***	602 m	607 m	1142 m
	95%	1023 m	1021 m	4927 m
	Mean Error	532 m	524 m	1238 m

VI. CONCLUSIONS

The universal database correlation method is utilized for network-based radio location. This technique does not rely on LOS-conditions but benefits from the more distinct radio patterns of multipath environments. In order to avoid costly measurements, the database is constructed with values generated by wave propagation software which is used for network planning anyway. Only network inherent parameters are used so that no hardware modifications are required and legacy handsets can be used.

The results of the measurement campaigns which were performed in an operating GSM network in Stuttgart (Germany) yield different results in the investigated scenarios. In the urban scenario location accuracy is in the range of 80 m in 67% of the measurements and still below 200 m for the 95% percentile. A comparison to the simple method employing the location of the serving base station as location estimate shows a significantly higher accuracy achievable with database correlation.

On the other hand, the additional effort which is required for the database correlation method does not bring about such a gain in accuracy in the suburban/rural scenario. In almost 90% of the measurements the location error is not even reduced to half its value. The reason for this is that predictions according to the Hata-Okumura model do not provide distinct patterns of signal strength. Deterministic or semi-deterministic models like the one used in the urban scenario are superior in this respect as they consider the buildings of the prediction area.

ACKNOWLEDGMENT

We wish to thank Vodafone Germany for supporting our work by providing the GSM network data.

REFERENCES

[1] H. Laitinen, J. Lähteenmäki, T. Nordström, "Database Correlation Method for GSM Location", Proceedings of the IEEE VTC conference, Spring 2001, Rhodes, Greece, pp. 2504-2508 vol. 4

[2] F.M. Landstorfer, "Wave Propagation Models for the Planning of Mobile Communication Networks", Invited Paper at 29th European Microwave Conference (EuMC) 1999, Munich, pp. 1-6.

[3] R. Collmann, "Evaluation of Methods for Determining the Mobile Traffic Distribution in Cellular Radio Networks", IEEE Transactions on Vehicular Technology, Vol. 50, No. 6, pp. 1629-1635, November 2001

[4] Digital cellular telecommunications system (Phase 2+); AT Command set for GSM Mobile Equipment. European Telecommunications Standards Institute, 03 2003 (ETSI TS 100 916 V7.8.0 (3GPP TS 07.07 version 7.8.0))

[5] E. Damosso, COST 231 (Digital mobile radio towards future generation systems), Final report, European Comission, Brussels, 1999