Cellular Location Determination - Reliability and Trustworthiness of GSM Location Data

Richard M. Zahoransky, Klaus Rechert, Konrad Meier,
Dennis Wehrle, Dirk von Suchodoletz

Department of Computer Science
University of Freiburg
Hermann-Herder Str. 10
79104 Freiburg

Abstract: While using mobile telephony networks, the serving network infrastructure is able to determine the mobile station's location. Until now, cellular telephony has been built on self-contained infrastructure, i.e. all network components have been certified and especially users have been unable to take over control over their mobile equipment's behavior. With the rising awareness on privacy issues, software-based mobile phone network stacks became available and thereby a new freedom degree for mobile subscribers is introduced.

While slight modification to the mobile phones behavior will not impair with the general functionality of the network, cellular location determination becomes less reliable and trustworthy. We discuss user imposed measures to detect external location determination attempts and to obfuscate generated location information. With a dedicated testbed setup, the effects of location obfuscation were evaluated.

1 Introduction

Digital wireless telephony networks have become a core communication infrastructure within the past 15 years. GSM and its successors have significantly changed the communication landscape both in developed and, with only a slight delay, in developing market economies, by far outnumbering landline connections (e.g. in Germany [Ger10]). Mobile telephony and data are a crucial part of today's communication infrastructure; moreover, they can contribute to security and safety. The mobile telephony network and its physical characteristics help to locate mobile phone users in cases of emergency [1] and may be a valuable tool for search and rescue (SAR) [CLR10]. For instance, Bengtsson et al. analyzed post-disaster population's displacement using SIM-card movements in order to improve allocation of relief supplies [BLT^{+}11]. Due to regulatory requirements but also driven by commercial opportunities, locating mobile phones gained the attention of research and industry. Furthermore, location information gathered through mobile telephony networks is now a standard tool for crime prosecution and it is enforced by the EC

[1] US Regulation on location determination in case of a emergency call: FCC Enhanced 911 Wireless Service, `http://www.fcc.gov/pshs/services/911-services/enhanced911`, [12/15/2011].

Data Retention Directive with the aim of reducing the risk of terror and organized crime [EP06]. Additionally, commercial services are based on the availability of live mobility patterns of large groups [2] or location-aware advertising [Kru10].

In general, law enforcement and commercial agencies exploiting location information have two options for utilizing location determination in mobile telephony networks: an active and a passive method. While active positioning yields immediate and more accurate results (e.g. through Uplink Time of Arrival [rGPPG02]), there are additional costs involved (e.g. network utilization) and thus, an incentive and dedicated target is required. This method is usually used to track identified individuals in criminal investigations. For instance, the police of North Rhine-Westphalia issued 225.784 location determinations on 2644 different subjects in 778 preliminary proceedings in 2010 [Min11]. Germany's federal police forces initiated 440.783 so called silent text messages. [3] On the other hand, with passive location determination techniques, all required information is generated from normal communication with the subscriber's mobile station, thereby causing no additional costs.

In this paper we investigate the user's possibilities to detect active location attempts and we lay out a scenario in which a user takes measures to provide a false position. Furthermore, measures for passive positioning methods are proposed which are capable to reduce location determination accuracy and potentially obfuscate position information in case of passive location monitoring. Finally, we evaluate and verify the location obfuscation method. For this purpose a test environment reflecting all components of a mobile telephony network was developed and deployed. The resulting mobile network infrastructure is based on real-life hardware and open-source software in order to create a realistic and defined environment which includes all aspects of the air interface in mobile telephony networks. The network is fully functional and thus provides a defined and fully controlled environment for analyzing all aspects of subscriber-provider interaction.

2 Localization Determination in Cellular Communication Networks

As an example we discuss the GSM infrastructure, because it is widely deployed and recently software and analysis tools have become available. Its successors UMTS (3G) and LTE (4G) still share most of its principal characteristics.

There is a variety of possibilities for determining a mobile station's location from the view point of the infrastructure, e.g., by Cell Origin with timing advance (TA) and Uplink Time Difference of Arrival (U-TDOA) for GSM [rGPPG09]. [4] While the latter method requires sophisticated network infrastructure, Cell Origin and TA are available in any network setup. However, both methods work without special requirements for the mo-

[2] Commercial traffic monitoring service, `http://www.vodafone.com/content/index/press/local_press_releases/germany/2008/tomtom_and_vodafone.html`, [12/15/2011].

[3] Letter of the Federal Ministry of the Interior by request of a parliamentarian, `http://www.andrej-hunko.de/start/downloads/doc_download/185-stille-sms-bei-bundesbehoerden`, [12/15/2011].

[4] For location determination options for UTRAN cf. [rGPPG10]

bile station and achieve a positioning accuracy of up to 50 m for U-TDOA in urban areas [SCGL05].

Another (non-standard) method to determine a mobile stations's (MS) location makes use of measurement results. Usually based on databases built from signal propagation models used during the planning phase of the infrastructure, this data can be used to create a look-up table for signal measurements to determine the MS's location. Based on the cell, TA and received signal strength of the serving cell as well as the six neighboring cells, Zimmermann et al. achieved positioning accuracy of below 80 m in 67% and 200 m in 95% in an urban scenario [ZBL$^+$04]. With a similar method but more generic setup, Peschke et al. report a positioning accuracy of 124 m in 67% [HUP07].

While the mobile phone is in idle mode, network-assisted positioning is not possible. The network either has to wait for the next active period of the MS (e.g. phone call, location update) or has to initiate the MS's activity. This can be done by transmitting a so called *silent message* to the MS in order to force an active communication without raising the user's awareness. The procedure is used for instance by law enforcement authorities or by location-based services based on cellular positioning [Min11].

2.1 Interference by User Controlled Mobile Network Stack

With the development of *OsmocomBB*[5] GSM baseband implementation, the basic work has been done for a fully user controlled mobile phone. For instance, such a mobile station could be modified to log and expose the location data to its user that has been gathered by the mobile communication infrastructure [RMB$^+$11].

2.1.1 Active Location Determination

A fully user controlled mobile device requires software interfaces with a network stack which controls and exposes signaling attempts (e.g. by detecting silent text messages). However, such a signaling attempt does not provide information on the purpose of paging the mobile station. Hence, it is difficult for a subscriber to decide whether the paging attempt is legitimate (i.e. incoming call or text message) or a (hidden) location determination attempt was triggered. Only after the device has reacted to the signaling the originator and the purpose of the paging becomes visible. However, by answering to the signaling, the mobile phone is getting active (i.e. sending network packages) and therefore a location measurement unit is able to determine the MS's position (e.g. through TDOA).

While active positioning requires a dedicated target and some costs, concealing the mobile station's location is also possible with some effort. Due to the usage of a full software network stack, lower network layers could be decoupled from the mobile phone. By leveraging a second communication channel, the user and his mobile station can be at a different place than the device running the physical layer and antenna, communicating directly with

[5] Open Source GSM Baseband implementation, `http://bb.osmocom.org`, [12/15/2011].

the mobile network infrastructure. This way the location of an individual SIM-card can be forged.

2.1.2 Passive Location Determination

In order to take over control on passive location monitoring, access to measurement results and the occurrence of location updates is required. Especially the density of periodic location samples makes a significant difference in the provider's possible knowledge on the user's movement pattern and thus on the user's present and future privacy risks. Such a monitor feature enables the users to select a mobile telephony provider that requests location updates less frequently or the user demands compensation for his or her loss in location privacy.

A second step to improve a user's location privacy is to reduce the observer's observation accuracy (obfuscation). A possible way to blur the exact location is to send empty or significantly altered measurement reports. Normally, the measurement reports include signal strength measurements of the surrounding BTS to support handover decision-making during active connections. Since a periodic location update requires only a very brief communication with the network, a handover between different cells is very unlikely. Thus, sending measurements of neighboring stations is technically not always required. By reducing the number of transmitted measurements the accuracy of the network's position estimation is significantly decreased. In the best case scenario (if no or false measurements were transmitted) the accuracy is decreased to the cell of origin combined with the timing advance parameter. To further decrease the accuracy of the estimated position, the MS may send with a slight timing offset. Such offsets have a direct impact on the timing advance calculation of the BTS. Consequently, this leads to an incorrect distance estimation between MS and BTS. It is also possible to report a wrong MS transmission power to the network. This influences any estimation the network draws based on the received signal strength of the MS.

The combination of manipulating measurement results, timing advance and reported transmission power makes it possible to conceal the actual position of the MS. Nevertheless, the rough location of the MS is still available through the coverage area of the serving BTS.

3 Evaluation Setup

To evaluate the proposed measures and their effects, a full mobile telephony network testbed is required. Different scenarios can be tested without interfering with the public network infrastructure.

The testbed consists of three basic components: the *Mobile Network*, the *Testbed Serving Mobile Location Center* (TB-SMLC) and the *Mobile Stations*. Figure 1 provides a schematic overview on the structure of the testbed. In combination, these components allow us to analyze all aspects of the communication between network and mobile station in

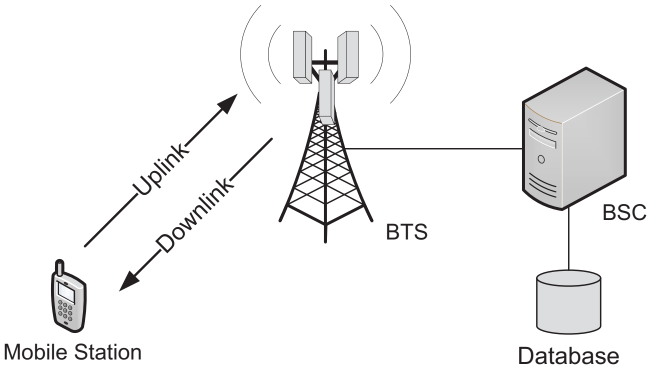


Figure 1: Overview of testbed network and transmission of measurement results.

a realistic scenario. Since the testbed implements a standard GSM network, it can simply be extended with standard GSM network components, for instance by the addition of any arbitrary mobile station. In contrast to a software simulation, such a setup allows for direct interaction with the network as a subscriber in order to get immediate feedback on status and events within the network. Since complete control over all components in the network is achieved, the subscriber's behavior as well as the impact on the infrastructure can be evaluated within the testbed. Special hardware and software is needed for the practical implementation of the testbed. A detailed description of hardware components used and the software implementation is given in earlier work [MWRv11].

3.1 Training Phase

A training phase is needed before localization can take place. For the training phase, a person equipped with a GPS receiver and a mobile phone was continually walking within the testbed's covered area. While walking, the MS was working in dedicated mode, continuously generating measurement reports that have been stored by the BSC, respectively the associated logging component. Each measurement report was assigned to its GPS coordinates. In a second step, the resolution of the measured coordinates is set. Measurements have been aggregated into tiles, with the size of the tile is the resolution of the map. However, the tile size can be set arbitrarily to a certain degree [ELM04]. An average is commuted among measurements with coordinates within the the same tile. Finally, outliners have been removed using Grubb's test [Gru69]. For our experiments the tile size of 8.52 m x 6 m have been chosen, which results in 6200 tiles for the covered testbed area. In total 171654 measurements were recorded and analyzed.

Measured received signal levels are considered as gaussian distributed. During an experiment with a stationary mobile phone 1500 MRs were created and analyzed. Results in Figure 2 show the histogram of the experiment and a fitted normal distribution. A standard

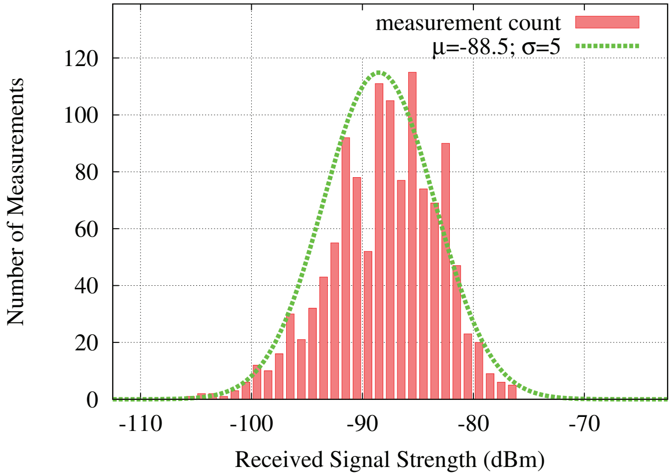


Figure 2: Histogram of observed received signal strength (RSS) of a stationary MS measuring a single BTS versus scaled Gaussian distribution with μ =-88.5 and σ =5.

deviation of $\sigma = 5$ dBm has been assumed throughout the localization calculation.

3.2 Interpolation

Since it is not feasible to take measurements for any place within the testbed coverage area, interpolation is used to approximate the RSS for places with no measurements. For this step, a Voronoi (or natural neighbor) interpolation was chosen because it shows the lowest error margin and results in a smooth interpolation (except at data points) [Suk01, LPA10]. The coordinate for which interpolation is required (denoted as point N) gets inserted into the Voronoi diagram. The resulting Voronoi region surrounding N "steals" some area from neighboring points. The stolen area size is expressed as fraction of the size of the Voronoi region of point N and treated as weighting factor. For every measurement point from which an area is stolen, the corresponding measurement is multiplied with the weighting factor. Hence, the interpolated value for point N is the sum of individual weighted measurements. Let a, b, c, d be the area size of the stolen areas by the Voronoi region of point N, with n denoting the size of N's Voronoi region, the interpolated signal strength for point N yields to $N_{RSS} = \frac{a}{n} \cdot A_{RSS} + \frac{b}{n} \cdot B_{RSS} + \frac{c}{n} \cdot C_{RSS} + \frac{d}{n} \cdot D_{RSS}$. An interpolated map is depicted in Figure 3.

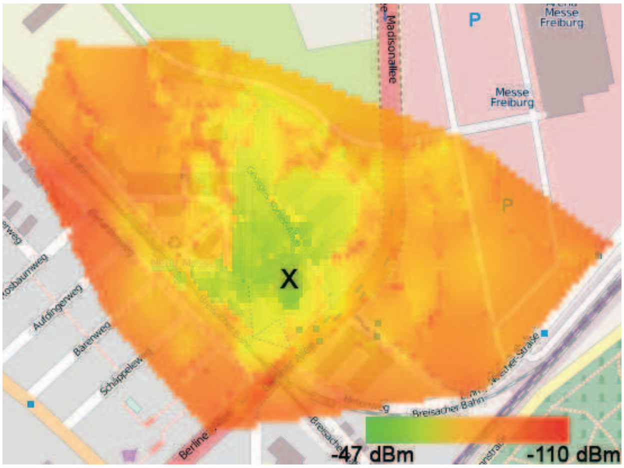


Figure 3: Interpolated GSM-map. The color denotes the receivable signal strength at that coordinate.

3.3 Location Determination

Localization calculation used is based on the Bayes' theorem. For every tile $l_{x,y}$ of the GSM map and for every receivable BTS BTS_i with $i \in 1...n$ and n denoting the number of receivable base stations, the available information is a vector of the received level of n base stations as $RSS := (BTS_1, \ldots, BTS_n)$ for the location (tile) of the MS. The probability distribution for the received signal strength RSS_i is estimated using every measurement observed within a tile: $P(RSS_i|l_{x,y}) = \mathcal{N}_i^{x,y}(\mu_i^{x,y}, \sigma_i^{x,y})$. We assume a Gaussian distribution of the signal strength measured in dBm. Figure 2 shows a histogram of observed RSS values of a stationary MS. Signal strengths from different BTSs are considered to be independent. The mean $\mu_i^{x,y}$ was already computed for every tile during training phase by averaging and outliners' removal. The variance $\sigma_i^{x,y}$ was chosen as 5 dBm, based on experiments (as shown in Figure 2).Usually, the network provides a list of the neighboring cells to the mobile phone to be monitored during an active connection in order to support a communication handover between two BTSs. In our testbed setup only one additional BTS is located on the campus, leading to limited localization possibilities. To cope with this shortcoming, we extended the neighbor list by adding additional public GSM cells receivable on the campus. By this, the mobile phones measure signal strengths of those other cells and sends the additional readings to the testbed network.

In order to locate a phone, the corresponding measurement entries are used from a prerecorded GSM signal map. An area based probability algorithm ($ABP - \alpha$) is used for location lookup [ELM04, YAUS03]. This group of algorithms return a set of the most likely map tiles, matching the actual and predetermined RSS fingerprints controlled by a confidence value α. The summed probability of the resulting set of tiles matches the required confidence value. Hence, the α-value controls the trade-off between positioning accuracy and methodical precision.

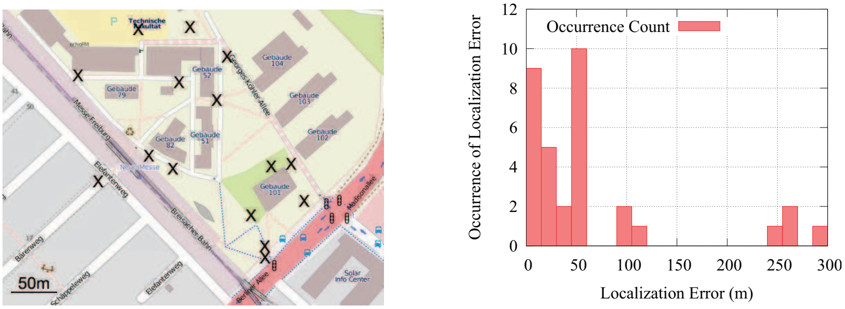
Given a received signal fingerprint vector (RSS), first we compute the probability at being at each tile's location $l_{x,y}$ using Bayes' equation

$$P(l_{x,y}|RSS) = \frac{P(RSS|l_{x,y}) \cdot P(l_{x,y})}{P(RSS)}, \tag{1}$$

with $P(RSS|l_{x,y})$ computed as multiplication over the probability distribution of BTS_i as

$$P(RSS|l_{x,y}) = \prod_{i \in 1 \dots n} \int \mathcal{N}_i^{RSS} \cap \mathcal{N}_i^{x,y} dRSS\,,$$

and $\mathcal{N}_i^{RSS}$ as the derived Gaussian distribution of the MS's received signal strength of BTS_i.



(a)

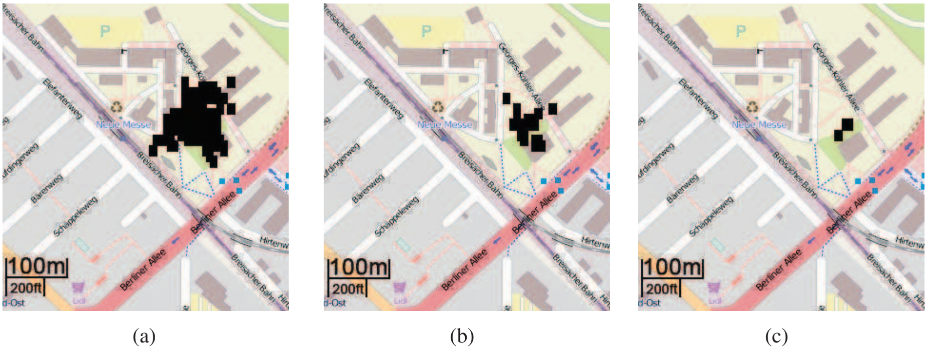
(b)

Figure 4: Results of localization experiments with different mobile phones. (a) "X" mark the locations on which localization were carried out. At some places three phones were used for testing while on other places only a subset of the available phones were tested. (b) shows the resulting localization error versus its occurrence.

A priori $P(l_{x,y})$ is considered to be equally distributed. Its value is the reciprocal of the number of tiles within the map. The probability of the fingerprint vector RSS being measured within the GSM-map is calculated as

$$P(RSS) = \sum_{x \in X, y \in Y} P(RSS|l_{x,y}) \cdot P(l_{x,y})\,.$$

Equation 1 yields the probability of being at tile $l_{x,y}$ given the fingerprint vector RSS. Since we want to return an area with a given confidence-value α, the algorithm outputs the top probability locations $l_{x,y}$ until they sum up to α. For our purposes a dedicated LMU is not necessary since the required measurement reports are generated during normal operation.



(a) (b) (c)

Figure 5: Comparison of a localization results with different number of reported base stations. The marked areas depict possible whereabouts of the mobile station. Fig. (a) shows the localization with a set of two measurement, (b) shows a reduced set of four measurements and (c) a shows the result of a full set of six measurements.

3.4 Evaluation

Localization accuracy of was measured based on thirteen different locations and a set different mobile phones. In total, 33 experiments were done. A short phone call was made with every phone and location. The actual coordinates were derived from a GPS receiver. Localization error is considered as the distance between the GPS coordinate and the most probable calculated location by the ABP-α algorithm. As shown in Figure 4, the average position error is 67 m and a median error of 47 m. With this accuracy and the possibility to locate any phone call originated in the past it is feasible to extract movement patterns of the network's users.

3.4.1 Effectiveness of Location Obfuscation

From the user's perspective it is not possible to recognize if a phone call is or will be localized. With common normal phones, a user cannot influence the creation and data hold in a MR and is therefore incapable of regaining his location privacy.

With mobile phones running a user controlled GSM network stack it is possible to fabricate and send false MRs. A strategy to regain location privacy would be trying to decrease the obtainable localization accuracy. This goal can be achieved by sending only a subset of the measurements of surrounding BTSs. Uncertainty in the position calculation rises with With less information available for the ABP-α to process, the uncertainty in localization rises. The results are shown in Figure 5.

4 Conclusion

Until now, cellular telephony was built on self-contained infrastructure, i.e. all network components were certified and especially users were unable to take over control over their mobile equipments behavior. With the rising awareness of privacy issues, software based mobile phone network stacks became available and thereby a new freedom degree for mobile subscribers is introduced.

While slight modification on the mobile phone's behavior will not impair with the general functionality of the network. However, the network based location determination becomes less reliable and trustworthy. First, attempts of law enforcement agencies are observable by using an open and user controlled mobile station. Second, by modifying the mobile phone's behavior, reliability of location information is reduced to cell size or worse, since explicitly false positions may have been generated.

References

[BLT+11] Linus Bengtsson, Xin Lu, Anna Thorson, Richard Garfield, and Johan von Schreeb. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Med*, 8(8):e1001083, 08 2011.

[CLR10] Ling Chen, M. Loschonsky, and L.M. Reindl. Characterization of delay spread for mobile radio communications under collapsed buildings. In *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 329–334, sept 2010.

[ELM04] E. Elnahrawy, Xiaoyan Li, and R.P. Martin. Using area-based presentations and metrics for localization systems in wireless LANs. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, pages 650 – 657, 2004.

[EP06] Council European Parliament. Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC. *Official Journal of the European Union*, L 105:54 – 63, 2006.

[Ger10] German Federal Network Agency (Bundesnetzagentur). Jahresbericht 2010. http://www.bundesnetzagentur.de/cae/servlet/contentblob/195950/publicationFile/10486/Jahresbericht2010pdf.pdf, 2010.

[Gru69] Frank E. Grubbs. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1–21, 1969. Available, http://www.jstor.org/stable/1266761 [last Access 10/30/2011].

[HUP07] R. Haeb-Umbach and S. Peschke. A Novel Similarity Measure for Positioning Cellular Phones by a Comparison With a Database of Signal Power Levels. In *Vehicular Technology, IEEE Transactions on*, volume 56, pages 368 –372, Jan. 2007.

[Kru10] John Krumm. Ubiquitous Advertising: The Killer Application for the 21st Century. *IEEE Pervasive Computing*, 99(PrePrints), 2010.

[LPA10] J. P. Lewis, Frédéric Pighin, and Ken Anjyo. Scattered data interpolation and approximation for computer graphics. In *ACM SIGGRAPH ASIA 2010 Courses*, SA '10, pages 2:1–2:73, New York, NY, USA, 2010. ACM.

[Min11] Ministerium für Inneres und Kommunales NRW. Drucksache 15/3300 Funkzellenauswertung (FZA) und Versenden "Stiller SMS" zur Kriminalitätsbekämpfung. `http://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument?Id=MMD15/3300(11/23/2011)`, 2011.

[MWRv11] Konrad Meier, Dennis Wehrle, Klaus Rechert, and Dirk von Suchodoletz. Testbed for mobile telephony networks. In *Resilience and IT-Risk in Social Infrastructures (RISI 2011)*, pages 661–666, 2011.

[rGPPG02] 3rd Generation Partnership Project (3GPP). TS 45.811 Technical Specification Group GSM/EDGE Radio Access Network; Feasibility Study on Uplink TDOA in GSM and GPRS (Release 6). `http://www.3gpp.org/FTP/Specs/html-info/45811.htm`, 06 2002.

[rGPPG09] 3rd Generation Partnership Project (3GPP). TS 43.059 Technical Specification Group GSM/EDGE Radio Access Network; Functional stage 2 description of Location Services (LCS) in GERAN (Release 9). `http://www.3gpp.org/ftp/Specs/html-info/43059.htm`, 11 2009.

[rGPPG10] 3rd Generation Partnership Project (3GPP). TS 25.305 Technical Specification Group Radio Access Network; Stage 2 functional specification of User Equipment (UE) positioning in UTRAN (Release 10). `http://www.3gpp.org/ftp/Specs/html-info/25305.htm`, 9 2010.

[RMB$^+$11] Klaus Rechert, Konrad Meier, Greschbach Benjamin, Dennis Wehrle, and Dirk von Suchodoletz. Assessing Location Privacy in Mobile Communication Networks. In J. Zhou X. Lai and H. Li, editors, *ISC 11*, LNCS 2001, pages 309–324. Springer, Heidelberg, 2011.

[SCGL05] Guolin Sun, Jie Chen, Wei Guo, and K.J.R. Liu. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs. *Signal Processing Magazine, IEEE*, 22(4):12 – 23, July 2005.

[Suk01] B. Semenov Sukumar, N. Moran. Natural neighbour galerkin methods. *International Journal for Numerical Methods In Engineering*, Volume 50; Part 1:1–28, 2001.

[YAUS03] M.A. Youssef, A. Agrawala, and A. Udaya Shankar. WLAN location determination via clustering and probability distributions. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 143 – 150, 2003.

[ZBL$^+$04] D. Zimmermann, J. Baumann, A. Layh, F. Landstorfer, R. Hoppe, and G. Wolfle. Database correlation for positioning of mobile terminals in cellular networks using wave propagation models. In *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, volume 7, pages 4682 – 4686 Vol. 7, sept 2004.