An Indoor Positioning Realisation for GSM using Fingerprinting and kNN

Ana Anastasijević, mentor: Aleksandar Nešković

Abstract — **Positioning in public land mobile networks has become increasingly popular over the past years with the development of more sophisticated mobile equipment and standards. New services which rely on location information have been developed, such as emergency call localization, navigation, location sensitive commercials/ billing, etc. This paper investigates one practical realisation of an indoor positioning model for GSM (Global System for Mobile Communications) networks, based on the received signal strength (RSS) and fingerprinting method. The achieved results are well within international localization accuracy requirements for emergency services, confirming great potential of the approach.**

Keywords — **cellular networks, Euclidean distance, fingerprint, location based services, Nearest Neighbour, positioning, received signal strength, reference points.**

I. Introduction

Public land mobile networks were first developed in order to provide reliable telephony to mobile users. However, users' demands have been growing ever since and the pressure on mobile operators to introduce new and enhanced services as well as increase data traffic rates is on-going. A relatively new type of services with growing popularity are location- based services (LBS), which were first introduced in GPRS (General Packet Radio Service), a cellular networking packet- switching service for GSM (Global System for Mobile Communications). GSM is currently the predominant standard in cellular mobile networks covering more than 90% of the world's population [1].

LBS are defined by 3GPP (The Third Generation Partnership Project) as services which use available information on the location of the user (or, more precisely the location of the mobile station) [2]. Examples of LBS are location dependent commercials and billing, navigation, tracking, improved network optimization, etc. However, the most significant among these are emergency call localization services which must meet strict requirements for the United States and the European Union (usually known as E-911 and E-112 respectively), described in detail in [3] and [4]. Crucial prerequisites for LBS deal with accuracy, reliability, latency, availability and applicability, which are expounded on in [5]. Part of the accuracy requirements for E-112 (most relevant for this paper) from [4] are presented in Table 1.

Ana Anastasijević is a student at the School of Electrical Engineering, University of Belgrade, Serbia (e-mail: aanastasijevic@gmail.com).

Table 1: Part of accuracy requirements for E-112 (indoor to suburban type of environment).

	Indoor (m)	***Urban (m)***	***Suburban (m)***
Caller can provide general information	10-50	25- 150	50- 500
Caller cannot provide any information	10-50	10-150	10- 500

Existing systems for outdoor positioning, which are not based on mobile cellular networks, such as the satellite-based positioning system GPS (Global Positioning System), which have performed well in the past, are underperforming when it comes to indoor positioning. The reasons behind this are rooted in the methods used for procuring a terminal's location, which are mainly based on lateration. Lateration requires Line of Sight (LoS) conditions with multiple (at least 4 for 3D positioning) transmitting stations (satellites for GPS, base stations for cellular- based systems, etc.) to produce reliable location estimations [6]. Although various parameters are used for lateration calculations such as Timing Advance (TA), Time Difference of Arrival (TDOA) and Angle of Arrival (AoA), LoS between transmitting stations or reference base stations and the mobile terminal is essential [7]. The term reference base stations (RBS) denotes transmitting stations, usually with known spatial coordinates, which serve as references in calculating the position of the mobile station (MS). However, LoS conditions between an MS and multiple RBSs rarely occur in indoor environments thus commanding a completely different approach for tackling the indoor positioning problem.

A great number of technologies have been proposed for indoor LBS deployment, including independent systems like cellular networks and WLAN (Wireless Local Area Network), as well as Bluetooth, computer vision and even visible light communication [8]. In spite of numerous possibilities, positioning models differ in how they measure up to evaluation criteria defined in [5]. For example, computer vision and Bluetooth- based methods require additional infrastructure and are consequently less preferable to cellular or WLAN- based positioning technologies [9]. Using existing systems for localization improves applicability, especially in the initial stages of service establishing. Moreover, apart from emergency call localization, the possibilities for commercial services based on LBS are blooming, especially with the growing trend in use of smartphones and tablets, which serve as personal assistant devices. Therefore, positioning methods that use technologies supported by these types of devices, especially cellular networks, are becoming more popular as well as more practical.

For all of the above reasons, this paper examines a practical realisation of a positioning model for GSM networks for all mobile operators available in the country. The paper is organized as follows: the second section explains the model used for positioning; the third deals with the measurement procedure, whereas the results and conclusions are presented in the last two sections. The backbone method used is fingerprinting, expounded on in the second section, which is becoming prevalent for indoor use. Tests conducted demonstrate its successfulness, thus supporting its popularity.

II. Fingerprinting and kNN

As noted previously, the most commonly used method for indoor positioning is fingerprinting. Fingerprinting consists of two phases: offline and online phase.

Offline phase involves defining specific points in the zone of interest for positioning which will serve as reference points (RPs). Reference points' coordinates (or RP positions) are known and used for calculating the coordinates for test sample points. During the offline phase, data is collected through measuring the desired parameters of the signal in RPs. The measured parameters, together with their coordinates are stored in the fingerprinting database (DB). Fingerprinting doesn't specify which parameter/s should be measured at reference and test sample points, however, Received Signal Strength (RSS) is most commonly used. RSS is preferable to other obtainable parameters (TA, TDOA, AOA, etc.) because it is not as subject to No LoS conditions and multipath fading (characteristics of indoor environment) as other abovementioned parameters [10].

The fingerprinting DB, which is the ultimate result of the offline phase, consists of fingerprint vectors for all RPs together with their coordinates. The fingerprint vector contains a list of all acquired signals at a point. Equation (1) shows the form of the fingerprint, whereas equation (2) presents an element of the fingerprinting database (x_{DBi}, y_{DBi} represent Cartesian coordinates for point i in the database).

$$\overrightarrow{RSS} = [RSS_1, RSS_2 \ldots RSS_n] \quad (1)$$

$$\overrightarrow{RSS_{DBi}} = \left[x_{DBi}, y_{DBi} \middle| RSS_{DB1,i}, RSS_{DB2,i}, \ldots, RSS_{DBn,i}\right] \quad (2)$$

The online phase entails measuring different RSS values and forming a fingerprint vector (1), which is then compared to all of the elements in the fingerprinting DB (2) in search of the closest match. The criterion according to which the closest match is chosen is defined by the algorithm employed together with fingerprinting. In this paper, kNN or k Nearest Neighbours algorithm is used.

kNN defines that for each of the fingerprints in the test sample, top k closest matches in the fingerprinting DB are found. The coordinates of these top k are then averaged to acquire the estimated position of the mobile station (3). In this paper, Euclidean metrics was used for determining "the closest match". The formula for Euclidean distance is given in (4). Different values for k were tested in order to find the optimal one (the one with the least mean error).

$$x_{MS} = \frac{\sum_{i=1}^{k} x_i}{k} \qquad y_{MS} = \frac{\sum_{i=1}^{k} y_i}{k} \quad (3)$$

$$L_{2,m} = \left(\sum_{i=1}^{n} \left|RSS_{DBi,m} - RSS_{DBi}\right|^2\right)^{\frac{1}{2}} \quad (4)$$

III. Method

Measurements were conducted on the premises of the School of Electrical Engineering, University of Belgrade. The building consists of three floors, however the measurements were limited to the ground floor (layout of the space is shown in Fig. 1.). Therefore, vertical accuracy analysis was omitted.

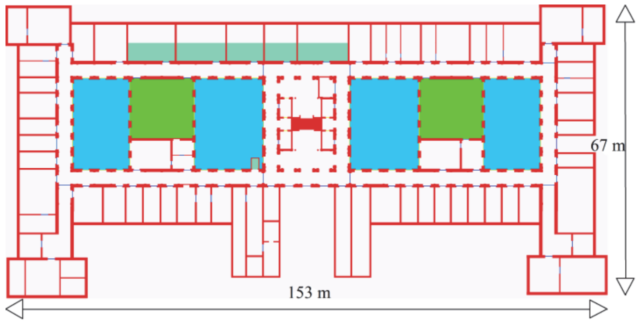

Fig. 1. Layout of the faculty premises (ground floor).

The Rohde & Schwarz TSMU Network Scanner unit was utilized for all of the measurements. The device has the ability to scan all GSM bands with the sensitivity of up to -112 dBm, and to decode system information type 1, 2, 3 and 4 from the BCCH (Broadcast Control Channel). BCCH is the main signalling channel in GSM, which contains system information needed to identify the network and gain access. These parameters include the Mobile Network Code (MNC) and the Cell Identity (CI).

The total number of measured points was 583. Out of these 583, 531 points were measured at different positions and 52 points represented repeated measurements of the initial 531 points. The repetition was done in order to test the consistency and recurrence of recorded signals. About 20% (or 106) of the different points were set aside and used as the test sample; these were chosen pseudo-randomly (random choosing was slightly modified in order to make sure that all of the test sample points were as uniformly distributed throughout the space as possible). The remaining number of points (exactly 425 of them) were used as RPs for the fingerprinting DB. The layout of all points measured is given in Fig.2, where red dots represent test sample points.

As described in II, fingerprints for each of the measured points were formed as a list of all available CIs and corresponding MNCs and averaged RSS levels. The form of the fingerprint (only the first thirteen rows out of 97 are shown for simplicity reasons) is shown in Table 2. Column 3 represents the number of samples measured, used for averaging RSS. Signals which weren't detected at a specific point have zero as a set value for the number of samples and -115dBm for RSS. This is a consequence of using the maximum number of different visible CIs in the zone of interest (97).

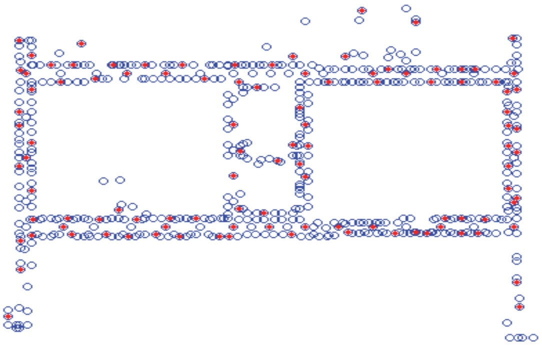

Fig. 2. Layout of the reference and test points.

TABLE 2: FINGERPRINT FORM.

CI	*RSS (dBm)*	*Number of samples*	*MNC*
18095	-28.0092	78	3
18094	-56.7005	78	3
12299	-61.1118	78	3
10144	-70.04	78	3
44771	-64.2769	78	1
30633	-66.8021	78	3
44663	-80.1877	78	1
381	-76.1128	78	5
44773	-86.2144	78	1
30144	-79.5169	78	3
15044	-71.8585	78	3
35044	-87.8523	78	3
44661	-115	0	1

Besides forming the fingerprint by using all 97 available signals, accuracy of the method was further tested by varying the number of signals used. When using a subset of the maximum 97, the signals used were those with the strongest *RSS* values.

The basic procedure for determining the accuracy of the method is described below, whereas the results thus acquired are presented in IV.

Test sample fingerprints were compared to those in the DB (Euclidean distances were calculated in accordance with (4)), whereas Cartesian coordinates were estimated in agreement with (3) for every tested k. The number of neighbours, k, used for calculations was varied between 1 and 30. Positioning error (or deviation or distance error) was determined as the distance deviated from the actual position, given in (5).

$$\Delta d = \sqrt{\Delta x^2 + \Delta y^2} \qquad (5)$$

Where:

Δd – total deviation
Δx – abscissa deviation
Δy – ordinate deviation

All of the computation was done in MATLAB.

IV. RESULTS

Three approaches were used for applying equation (4): in the first the variable *RSS* was given values in mW instead of dBm, in the second *RSS* was given dBm values and for the third approach, the values for *RSS* in every fingerprint were first normalized (the value for RSS_1 was subtracted from every other *RSS* value in the fingerprint). Results for mean, minimum and maximum error for all three approaches are given in Table 3. These results are listed for k=4 and for the maximum length of the fingerprint of 97 signals from different cells. k=4 was found to be the optimal number of neighbours for all three approaches.

TABLE 3: ERRORS FOR THREE APPROACHES FOR K=4.

Approach	**Mean error (m)**	**Minimum error (m)**	**Maximum error (m)**
1	21. 37	1. 02	40.32
2	$\boxed{6.66}$	0. 16	27. 81
3	7. 17	0. 16	30. 96

As shown in Table 3, the second approach yields the best results. Given that there is always some instability in the signal level at the receiving end, it is favourable to have changes in *RSS* mapped to smaller numbers for fingerprinting methods. This is exactly what is done by using a logarithmic scale and why, combined with the metrics defined in (4), the second approach produces superior results. Mean positioning errors for a percentage of samples (smaller than 100%) are given in Table 4.

TABLE 4: POSITIONING ERROR ACHIEVED AT A PERCENTAGE OF TEST SAMPLES.

Percentage [%]	*Mean positioning error [m]*
90	5.38
60	3.52
50	3.13

In Figs. 3-6 the distributions of mean, median, minimal and maximal positioning error for all of the tested values for k are presented (using the second approach and maximum fingerprint length). Positioning error distribution for all test sample points is displayed in Fig.7. The median error is smaller than the mean, ranging between 3.47m for k=1 to 8.67m for k=30 (Fig.4). This indicates, as shown in Fig.7, that the accuracy of the method is deteriorated by a smaller number of points with high positioning errors.

A. Consistency assessment results

In order to verify the stability of obtainable signals, consistency assessment tests were carried out in two stages: in the first, the fingerprints of the repeated 52 points were used as a test sample for the fingerprinting DB consisting of 531 points; in the second the original and repeated *RSS* were compared.

During the first stage, the original point was only identified twice as the Nearest Neighbour (NN), once as the second NN and once as the third NN. For the remaining 48 points, the original point was not in the list of top 5 NNs. The percentage of recurring signals was 84.58% on average (illustrated in Fig.8.) with the mean difference in *RSS* level per CI of -0.19 dB and mean absolute difference per CI of 4.39 dB (Fig.9.).

Consistency assessment results lead to the conclusion that by taking all of the available signals into the equation at every point, unstable signals are included. These instabilities together with extending the list of CIs in every fingerprint (from 32 on average different CIs above the sensitivity of -112dBm to 97 with values of -115dBm) are the most likely cause for poorer consistency results and, more importantly, the deterioration of the accuracy of the positioning method. Determining the threshold for incorporating a signal into the fingerprint or disregarding it, with the intention of achieving higher accuracy is the subject of further research. However, the premise that using a subset of the maximum number of available 97 signals would yield better results was tested and the results are presented in Fig. 10 and Table 5. From this analysis, it can be concluded that the optimum number of different signals to be used is around 60%, where a minimum is achieved at using 60 different signals.

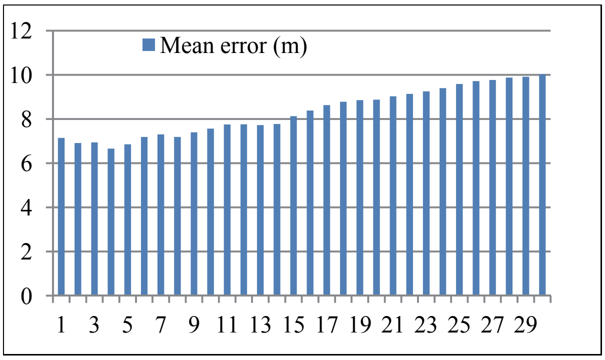

Fig. 3. Mean positioning error for k=1…30 and the second approach.

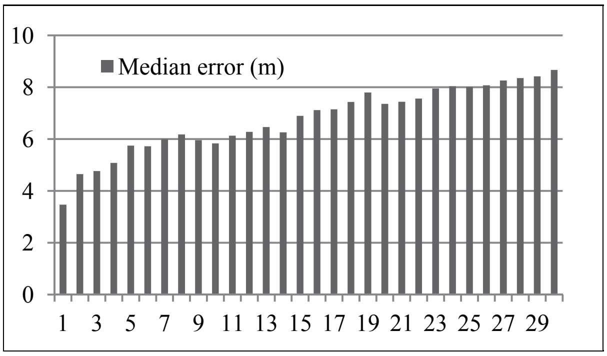

Fig. 4. Median positioning error for k=1...30 and the second approach.

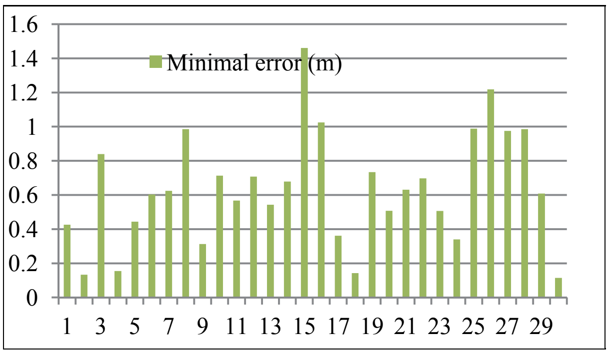

Fig. 5. Minimal positioning error for k=1...30 and the second approach.

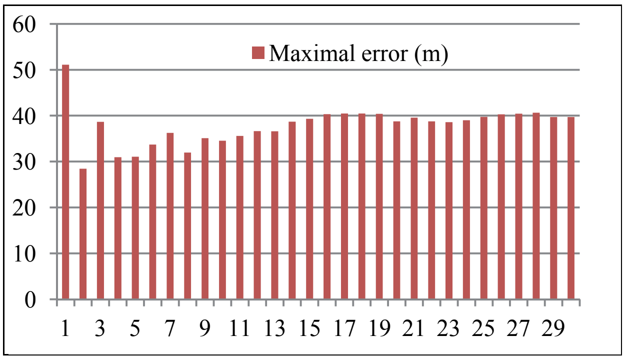

Fig. 6. Maximal positioning error for k=1...30 and the second approach.

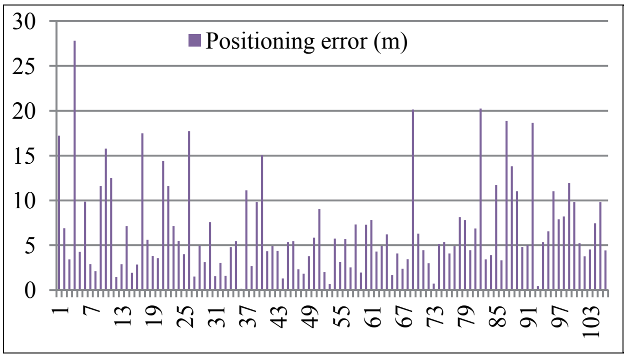

Fig. 7. Positioning error (accuracy deviation) for all test samples.

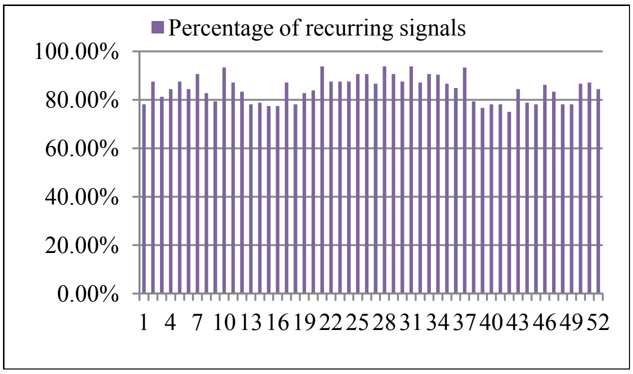

Fig. 8. Percentage of recurring signals for consistency assessment test (stage 2).

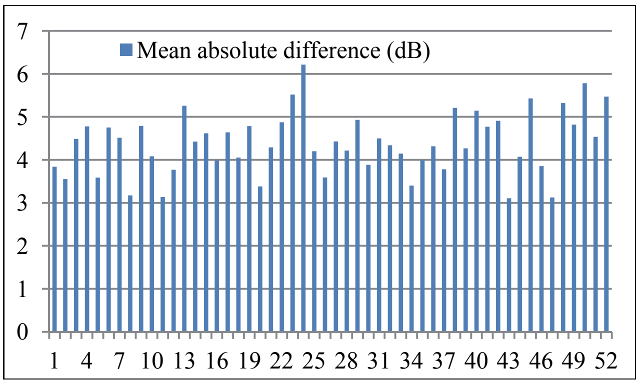

Fig. 9. Mean absolute RSS difference per CI between original and repeated point (stage 2).

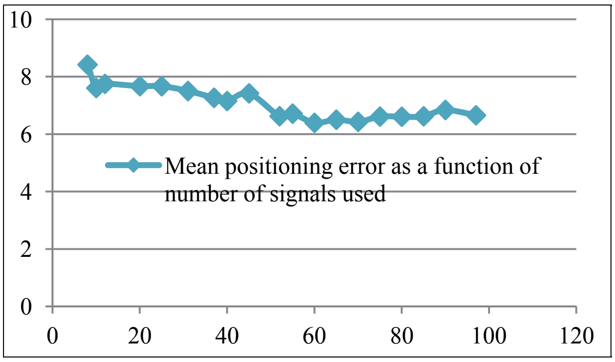

Fig. 10. Mean positioning error as a function of number of signals used.

TABLE 5: MINIMUM MEAN POSITIONING ERROR ACHIEVED (AND THE NUMBER OF NEIGHBOURS FOR WHICH IT IS ACHIEVED) WITH USING LESS THAN THE MAXIMUM AVAILABLE NUMBER OF SIGNALS.

Number of signals with different CIs used	***Mean positioning error [m]***	***Number of neighbours-k***
8	8,42	4
10	7,61	6
12	7,75	5
20	7,67	1
25	7,67	5
31	7,51	4
37	7,27	4
40	7,15	5
45	7,42	5
52	6,63	3
55	6,72	4
60	$\boxed{6,40}$	2
65	6,51	4
70	6,43	2
75	6,61	2
80	6,61	2
85	6,63	2
90	6,85	2
97	6,66	4

V. CONCLUSION

Given that the results of positioning tests (for all three approaches) meet the requirements for E-112 [4], presented it Table 1, kNN fingerprinting model was proven to be extremely successful indoors. The optimal value of NNs to be taken into calculation was found to be 4, regardless of the approach. Seeing as the accuracy of the second approach is the highest, this is the approach (using Euclidean metrics) that should be investigated further. The investigation of the dependence of mean positioning error on the number of signals used for positioning estimation confirms the observed instability of some of the signals included in the original maximum length fingerprint. However, determining the exact signals to be included should be subject to further research.

Since this method was developed for GSM, the potential for achieving great availability is substantial. Main setbacks for the method include the prerequisite of offline phase and its periodical updating, which can potentially limit implementation. However, as no changes in network infrastructure are required, the abovementioned shortcoming is outweighed in most occasions.

Bearing all this in mind, it can be concluded that, with modifications and improvements suggested below, the positioning models based on fingerprinting represent the future for indoor positioning, especially in public mobile networks.

By analysing the spatial distribution of test sample points, it can be concluded that the highest positioning error occurs for points that do not have many neighbour reference points (usually points in corners or in rooms as opposed to hallways). For the purposes of more accurate positioning, a simple modification of increasing the number of reference points (and thus achieving a more uniform distribution of reference points across the zone of interest) is recommended.

It is expected that introducing advanced data processing algorithms, such as neural networks tested in [9] would result in more accurate information on the location of the mobile device. The next step in this research is replicating the model for UMTS (Universal Mobile Telecommunications System) for comparative analysis.

ACKNOWLEDGMENT

The author would like to thank Professor Dr. Aleksandar Nešković for his mentorship throughout the project. Support and advice from the School of Electrical Engineering and their Radio Communications department, especially Majda Petrić and Mladen Koprivica is most gratefully acknowledged.

REFERENCES

[1] http://www.gsmworld.com/aboutus/gsm-technology/gsm/
[2] 3GPP TS 22.071, release 8 (v8.0.0), Location Services (LCS) Service description; Stage 1, 2007.
[3] The FCC Fact Sheet—FCC Wireless 911 Requirements. FCC (January 2001).
[4] Coordination Group on Access to Location Information for Emergency Services (CGALIES),"Report on implementation issues related to access to location information by emergency services (E112) in the European Union", (2002).
[5] I.K. Adusei, K. Kyamakya, and K. Jobmann, "Mobile Positioning Technologies in Cellular Networks: An Evaluation of Their Performance Metrics," *Proc. IEEE Military Comm. Conf.* (MILCOM), vol. 2, pp. 1239-1244, Oct. 2002.
[6] http://www.mobile.ifi.uni-muenchen.de/studium_lehre/verg_semester/ws1011/msp/07_gps.pdf
[7] M. Samiei, M. Mehrjoo, B. Pirzade "Advances of Positioning Methods in Cellular Networks", *1st International Conference on Communications Engineering* (ICCE), pp. 174- 178, Dec. 2010. Available: http://seminars.usb.ac.ir/Files/iccome/en-us/Document/1640-Dec10-2008-positioning.pdf
[8] http://www.ecma-international.org/activities/Communications/tc47-2011-028-Rev2.pdf
[9] M. Borenović, A. Nesković Positioning in Indoor Mobile Systems, Radio Communications, Alessandro Bazzi (Ed.), ISBN: 978-953-307-091-9, InTech.
[10] M.Simić and P. Pejović Positioning in Cellular Networks, Cellular Networks - Positioning, Performance Analysis, Reliability, Agassi Melikov (Ed.), ISBN: 978-953-307-246-3, InTech,