



TECHNICAL UNIVERSITY OF CRETE
SCHOOL OF ELECTRONIC AND COMPUTER ENGINEERING,
TELECOMMUNICATION DIVISION

Cell Tower Discovery via Particle Filtering

PROGRESS REPORT

Emmanouil Alimpertis

July 2013

1 Motivation

The motivation for cell tower localization is derived from the fact that the locations of cell towers, i.e. base transceiver stations (BTSs), are unknown and unlisted apart from few exceptions. It would be a great challenge to discover the locations of the cell tower through the collected measurements from the community, thus revealing the social and informational aspect of the MySignals platform. The positions of cell towers are of an extreme interest. The necessity of a dense cell tower network is emphasized. For example, we can show the high transmitted power (next to the user's head) of a user who is far away from cell tower and respectively demonstrate the lower transmitted power of a user who is closest to the same BTS. Additionally, the positions of BTSs can be exploited for developing GPS-free localization techniques for mobile users, by using the mobile's observed received signal strength indicator (RSSI). The RSSI indirectly contains information regarding the distance between the BTS and the mobile phone, and thus has become immensely popular for localization techniques. However, this association comes with extremely high noise due to scattering, fading, GSM power control etc.

2 Background and Prior Work

Modelling of the propagation loss (alternative, path loss), from the transmitter to the receiver is very important and a prerequisite in order to predict the RSS in a specific region and then develop emitter localization algorithms. Propagation loss estimation and modelling has generated immense interest over the last decades and work done briefly summarizes on the following: The propagation loss is affected by several physical mechanisms such as 1) reflection, 2) diffraction, 3) scattering and 4) multipath interference which causes fading (distinguished in small and large scale fading) as well as shadowing (i.e. shadow fading). An extended presentation and characterization of the fading in wireless channels can be found in [1].

Two basic approaches for modelling the propagation loss are adopted. In the statistical approach, the propagation loss is typically modelled by two factors: deterministic function of distance, which represent average path-loss on the given distance (logarithmic attenuation is adopted), and one random variable which models the variance around the mean due to fading and shadowing effects [1], [6]. This random variable follows the normal distribution in dBm, i.e. log-normal distribution expressed in mW. In [7] it is shown that real world collected RSS measurements fit this model.

For a transmitter i and receiver j , the following model is widely adopted in RSS-based localization [18]:

$$P_t^{[i,j]} = P_0(t) - P_{\text{LOSS}}^{[i,j]}(t) + S, [dBm] \quad (1)$$

where:

- $P_0(t)$ is the received power at reference distance d_0 (also called breaking point). Assuming isotropic radiator, it can be calculated using the free-space path loss formula.
- $P_{\text{LOSS}}^{[i,j]}(t)$ is the propagation loss, at moment t , for the radio link.
- $S \sim \mathcal{N}(0, \sigma_{dB}^2)$.

Thus,

$$P_t^{[i,j]} \sim \mathcal{N} \left(\overline{P_t^{[i,j]}}, \sigma_{dB}^2 \right), \quad (2)$$

and

$$\overline{P_t^{[i,j]}} = P_0(t) - P_{\text{LOSS}}^{[i,j]}(t) = P_0(t) - 10n_p^{[j]} \log_{10} \left(\frac{d_{i,j}}{d_0} \right), \quad (3)$$

where:

- $n_p^{[j]}$ is the path loss exponent (PLE) of the link between emitter i and the receiver j .
- $d_{i,j}$ is the distance between the transmitter i and receiver j .

The second approach aims to model the environment between the emitter and the receiver as well as the cellular network parameters which may affect the propagation. Thus, many empirical-parametric models have been developed the last decades such as the Okumura-Hata model [5] and the Walfisch-Ikegami COST-231 [2]. The Okumura-Hata model equations take into account cellular system characteristics, such as antenna height, operating frequency, mobile user height and it incorporates the effects of diffraction, reflection and scattering caused by city structures. The model varies for suburban and open areas. Similarly, Walfisch-Ikegami COST-231 extends the Okumura-Hata model and develops urban propagation model. In [4] a detailed survey and presentation on parametric-environment dependent models can be found. It is emphasized that both COST-231 and Okumura-Hata models are appropriate for outdoor communications. There are no well established analogous models for indoor communications [6]. For indoor mobile communications the log-normal model 2 seems promising for realistic propagation loss modelling, since the PLE indicates a general signal attenuation [6].

The problem of estimating the location of an emitter, i.e. source discovery, is an open problem requiring the estimation or elimination of many unknown and time variant parameters such as the following:

1. the transmitted power level (P_{TX}) is usually unknown and almost always time variant. For several reasons, network operators sometimes do not provide technical information about their BTS, including P_{TX} . In general, in a real world environment, an arbitrary wireless emitter may have many levels of transmitted power for various reasons, therefore this parameter has to be treated as unknown. For example, the BTS may decrease the P_{TX} in order to reduce energy consumption. Such need arises when the network is idle or has low traffic, as well as for avoiding interferences with neighbouring BTSs and other users. Of course the BTS may increase the P_{TX} in order to guarantee a good communication quality with a user (e.g. GSM power control).
2. unknown PLE (n_p), indicates the attenuation of the radio link between receiver and transmitter and depends on the environment as well as user's location. The majority of bibliography and prior art assumes that the n_p is constant across users in different locations. However, each user has a completely different environment which rapidly affects the attenuation. Therefore, assuming the same $n_p^{[j]}$ per user j is an oversimplification.

The majority of the prior art for emitter discovery in wireless environments (wireless cellular networks or wireless sensor networks -WSN-) mainly focuses on general wireless emitters. For example no antenna directionality is taken into account which is an oversimplification for the GSM environment where the majority of the antennas are directional. In addition, prior art mainly focuses on simulated data rather than real world collected measurements. In [14] the unknown source location estimation is considered as a semidefinite programming problem (SDP) however assuming that P_{TX} is unknown but constant. In [15] a solution using weighted least squares for the emitter localization problem was proposed, considering the following scenarios: unknown P_{TX} , then under unknown PLE and finally both P_{TX} and PLE unknown (but always constant and no time variant). Finally, in [17] emitter location discovery is also considered under the same scenarios. The paper simulation and solution is based on that measurements are taken while the sensor is moving around the BTS.

On the other hand, only a minority of the prior art works directly on the GSM cell tower localization problem [10] or estimates the cell tower positions in order to provide localization to mobile phones based on RSSI and cell-ID [11], [12], [13]. The localization techniques applied utilize wardriving data, i.e., RSS measurements with their respective location by GPS are recorded while driving into town. The most common approaches are presented below:

1. Strongest RSS: The position of the BTS is the location where the strongest measurement was recorded. If strongest RSS is observed at many locations, the mean of these location is taken into account.
2. Weighted Centroid RSS: Choose the top- K measurements and average over of their corresponding locations in order to estimate the BTS position. Usually each location has a weight proportionally to the value of the recorded RSSI.

3 Our Approach and Methodology

State of the problem definition

We rely on Particle Filtering (PFs) to estimate the multi-state belief which describes our domain and manipulates all the parameters of a realistic GSM environment,:

$$\boldsymbol{\theta} \triangleq \left[x_{BS} \quad y_{BS} \quad \phi \quad P_{TX}^{[\text{status}]} \quad n_p^{[1]} \quad \dots \quad n_p^{[N]} \right]^T, \quad (4)$$

- x_{BS}, y_{BS} : are the 2D-Cartesian coordinates of the Cell Tower Position.
- ϕ : GSM antenna orientation with respect to the x-axis.
- $P_{TX}^{[\text{status}]}$: is the defined status from the GSM power control for the transmitted power of the BTS. The actual P_{TX} is determined by the combination of the $P_{TX}^{[\text{status}]}$, ϕ .
- $n_p^{[j]}$: The path loss exponent (PLE) of a specific user (i.e. an anchor), with respect to the serving BTS.
- N : the number of the mobile phones in the system.

The outline of our work summarizes on the following:

1. Coarse estimation of $n_p^{[j]}$ for each user j using the algorithm proposed in [19]. The latter relies on a grid voting procedure.
2. Estimation of the BTS location via PFs. PFs offer the probability distribution of θ conditioned on the recorded RSS measurements. Through the aforementioned posterior one can calculate various estimates of θ such as the minimum mean squared error (MMSE) estimate. It is noted that PFs perform an approximation of sequential state estimation and thus incorporates the time variant nature of the P_{TX} as well as the GSM antenna directionality.
3. Outlier detection on K independent estimated locations from step 2¹ The outlier detection algorithm [22] is utilized which assigns each estimate position a score proportional to the probability of an estimated point being an outlier. Averaging the L points with the lowest score yields the final estimation of the cell tower's position.

It is noted that in sharp contrast to the prior art, our method does not rely on war-driving but it utilizes measurements from fixed points and is completely blind.

3.1 System Model & Measurements

In order to investigate cell tower localization from collected RSS data, a measurement campaign was conducted for collecting as many as possible RSS measurements along with their corresponding cell-ID and geo-coordinates provided by the iPhone's GPS. Measurements were recorded during many hours of a day revealing the time variant nature of P_{TX} as well as quantifying fading and scattering effects. Obviously, the GPS coordinates of the users houses were known but also the iPhone GPS was validating the operation of the iPhones on users' houses and not somewhere else.

Figure 1 displays the fixed users locations (users home) and the cell tower's position (to be estimated). Table 2 contains statistics for the collected measurements as well as details for the anchors. It is emphasized that the users were served sometimes from other cell-IDs which their locations are unknown to us. Fortunately, the overwhelming majority of the measurements were recorded from cell-IDs 6056x, Cosmote GSM network, which are served from antennas on BTS at Tzanakaki street (COSMOTE headquarters at Chania), so they can be validated.

Zero inaccuracy is assumed for the users location since a change of few meters inside their home is negligible for RSS-based localization. Thus, user locations are defined as:

$$\mathbf{l}_j \triangleq [x_j \quad y_j]^T, j = 1, \dots, N,$$

and the collection of all users location is denoted by:

$$\mathbf{l} \triangleq [\mathbf{l}_1 \quad \dots \quad \mathbf{l}_N]^T$$

Measurements separation according cell-ID:

Every cell-ID, offered by a BTS of a cellular network provider, operates at a different frequency (or in other words, ARFCN: absolute radio frequency channel number). Additionally, one more

¹In this work $K = 2000$.

critical detail for performing cell tower localization is that a cell-ID corresponding to a specific BTS usually indicates the different orientation of the antennas on the BTS establishment. GSM antennas are directional, hence several antennas (usually 3 since the main beam of the antenna is approximately 120 degrees) are established to serve a different orientation, i.e. sector, in order to offer cellular telephony in a specific region. Thus, the collected measurements are separated according to their cell-ID. The measurement's timestamp, which is expressed in *sec* from a reference point in the past, is very important and must be considered because of the time variant network traffic and the continuous environment changes during the day.

For a fixed anchor j , the measurements vector from a specific cell-ID, with their corresponding timestamps, is defined as:

$$\mathbf{z}^{[j,\text{cID}]} \triangleq \left[P_{t_1}^{[j,\text{cID}]} \quad \dots \quad P_{t_i}^{[j,\text{cID}]} \quad \dots \quad P_{t_K}^{[j,\text{cID}]} \right]^T$$

$$\mathbf{t}^{[j,\text{cID}]} \triangleq \left[t_1^{[j,\text{cID}]} \quad \dots \quad t_i^{[j,\text{cID}]} \quad \dots \quad t_K^{[j,\text{cID}]} \right]^T,$$

where:

- $P_{t_i}^{[j,\text{cID}]}$ is the recorded RSSI at time $t_i^{[j,\text{cID}]}$, from the anchor j .
- K the number of the valid RSSI readings of the anchor j observed from the cell-id "*cID*".

The collection of all users measurements is denoted by:

$$\mathbf{z}^{[\text{cID}]} \triangleq \left[\mathbf{z}^{[1,\text{cID}]} \quad \dots \quad \mathbf{z}^{[j,\text{cID}]} \quad \dots \quad \mathbf{z}^{[N,\text{cID}]} \right]^T.$$

Dataset Filtering: Remove Instantaneous Noise (Small Scale Fading)

RSSI reading is retrieved from the iPhone's baseband and the fetching procedure usually lasts around 1 second. Unfortunately, the procedure of the calculation is unknown, with no public available documentation. Details such as the exact component of the hardware/software responsible for the RSSI reading, as well as the refresh rate or if the baseband averages the reading within this 1 second are completely unknown.

We speculate that the hardware initially averages the RSS and then returns the RSSI. In that way, a filtering of the small scale fading is performed by the mobile phone itself. Small scale fading due to small changes in the position or small changes in the environment usually lasts msec or seconds. In general the RSS, will suffer from large fluctuations in a small time window due to small scale fading (also is named fast fading). A **non stationary** environment (e.g. user movement inside home) leads to small scale fading, therefore our measurements are highly affected from **instantaneous noise**, i.e. an intuitively description of the small scale (or fast) fading. On the other hand, large scale fading, i.e. shadowing fading, due to motion over large areas leads to changes which last for larger time windows. For example, a user every night closes their home's windows leading to a significant drop in the RSS for the whole night. Large scale fading is modelled satisfactorily from the log-normal model (1, 2, 3).

The effects of the small scale fading must be minimized in order to proceed with to cell tower localization. In any case, the measurements must be considered when the effects from the environment seems to be minimized because both the BTS and mobile phones are placed at a fix location. When the environment seems to be **non stationary**, the "bad" measurements must

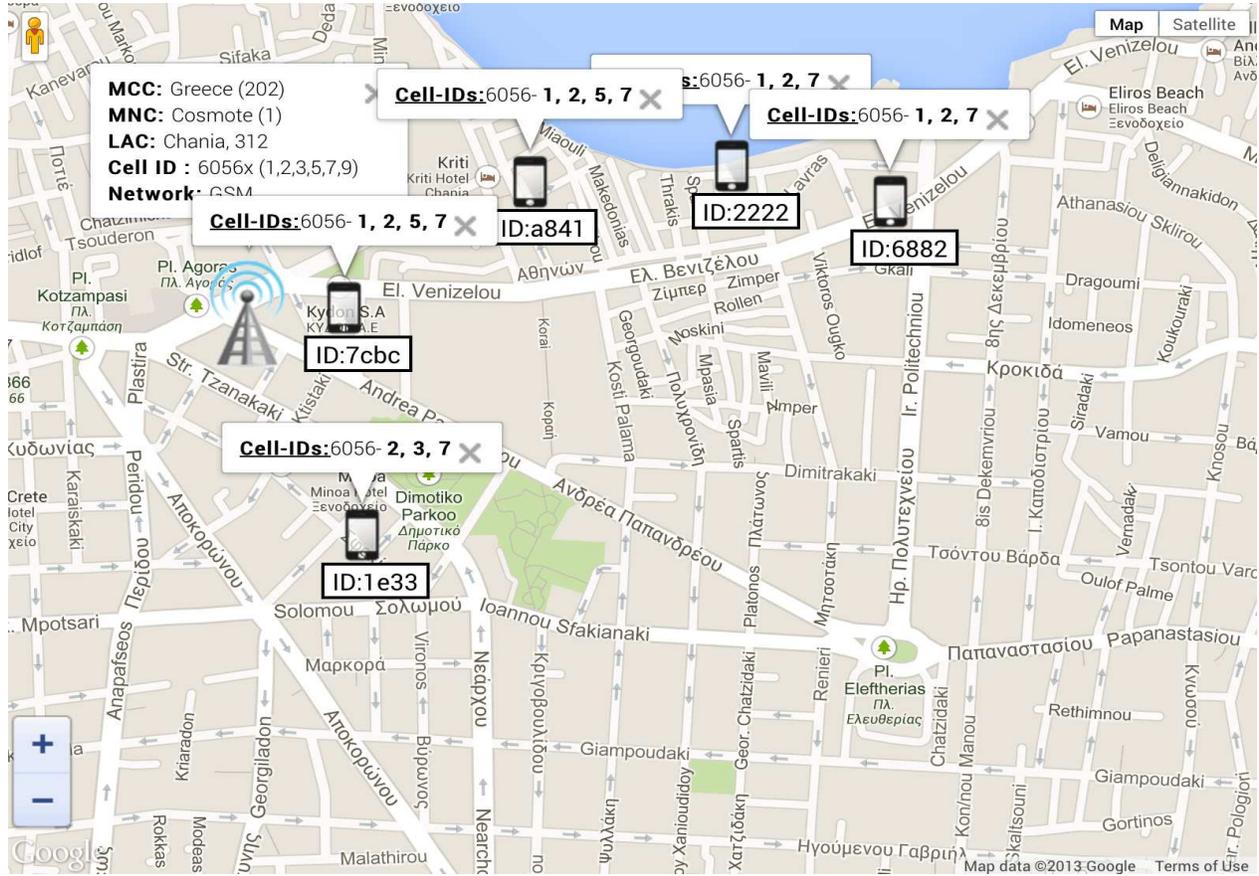


Figure 1: Map and topology of the system: Anchors and the cell tower location to be estimated.

be ignored. A first approach to limiting instantaneous fading is to apply a 4dB threshold on the standard deviation (σ_{dB}) of the measurements within a short time window (e.g. tested blocks of 10 continuous measurements), which is presented by Algorithm 1.

Algorithm 1: Dataset Filtering: Remove Instantaneous (Fast) Fading

- (1): **for** $j = 1$ to N { *anchor index* }
- (2): **for** \forall cID \in cell-IDs
- (3): **for** $k = 1$ to K do { *with step=10, measurement index* }
- (4): { **Secure** that during $[P_{t_k}^{[j,cID]} \dots P_{t_{(k+10)}}^{[j,cID]}]$ sampling does not have been stopped }
- (5): **if** $t_{k+10}^{[j,cID]} - t_k^{[j,cID]} < 30sec$
- (6): **Compute** σ_{dB} of $[P_{t_k}^{[j,cID]} \dots P_{t_{(k+10)}}^{[j,cID]}]$
- (7): **if** $\sigma_{dB} > 4dB$
- (8): **Remove** $[P_{t_k}^{[j,cID]} \dots P_{t_{(k+10)}}^{[j,cID]}]$ from $\mathbf{z}^{[j,cID]}$
- (9): **and** $[t_k^{[j,cID]} \dots t_{k+10}^{[j,cID]}]$ from $\mathbf{t}^{[j,cID]}$
- (10): **end for**
- (11): **end for**

UserID	iPhone Model	iOS	Timespan	Measurements
1e33	iPhone 3G	4.2.1	2013-04-02 <i>To</i> 2013-04-14	cellID: 60.562, #meas: 85.499 ARFCN: 787 Freq: 1878.6MHz
				cellID: 60563, #meas: 193.658 ARFCN: 83 Freq: 1869.4MHz
				cellID: 60567, #meas: 54.826 ARFCN: 794, 879, 833 762, etc
2222	iPhone 3GS	4.2.1	2012-09-13 <i>To</i> 2013-04-23	cellID: 60561, #meas: 422.441 ARFCN: 794 Freq: 1861.6MHz
				cellID: 60562, #meas: 227.319 ARFCN: 787 Freq: 1878.6MHz
				cellID: 60567, #meas: 45.866 ARFCN: 794, 879, 833 762, etc
6882	iPhone 3G	4.2.1	2013-04-01 <i>To</i> 2013-04-19	cellID: 60561, #meas: 365858 ARFCN: 794 Freq: 1861.6MHz
				cellID: 60562, #meas: 306.344 ARFCN: 787 Freq: 1878.6MHz
				cellID: 60567, #meas: 426 ARFCN: 794, 879, 833 762, etc
7cbc	iPhone 4	4.3.3	2012-09-13 <i>To</i> 2013-04-15	cellID: 60561, #meas: 413.400 ARFCN: 794 Freq: 1861.6MHz
				cellID: 60562, #meas: 47.123 ARFCN: 787 Freq: 1878.6MHz
				cellID: 60565, #meas: 8.182 ARFCN: 869 Freq: 1876.6MHz
				cellID: 60567, #meas: 2.028 ARFCN: 794, 879, 833 762, etc
a841	iPhone 3GS	4.2.1	2013-03-08 <i>To</i> 2013-04-23	cellID: 60561, #meas: 320.162 ARFCN: 794 Freq: 1861.6MHz
				cellID: 60562, #meas: 95.607 ARFCN: 787 Freq: 1878.6MHz
				cellID: 60567, #meas: 6.609 ARFCN: 794, 879, 833 762, etc

Table 1: Statistics and information for the collected measurements and anchors details.

(12): **end for**

In general, the data pre-editing, filtering and small scale fading removal needs more consideration [20], [21] and study since fading effects and the mobile radio channels are too complicated.

3.2 Estimation of $n_p^{[j]}$

In [19], a RSSI-based emitter localization algorithm considers a scenario with a different $n_p^{[j]}$ per each user j . The proposed algorithm is based on a voting procedure for each grid point (x, y) . The algorithm averages the position of the K grid cells with the maximum vote, in order

3.3 Cell Tower Position Discovery via Particle Filtering

The problem of the cell tower localization can be approached as a sequential Bayesian state estimation problem [8], i.e., recursively estimate the posterior distribution of the problem state θ 4, conditioned on the recorded measurements, i.e. $p(\theta|\mathbf{z}^{[cID]})$. Then, $p(\mathbf{X}_{BS}|\mathbf{z}^{[cID]})$ can be calculated by marginalization. It is noted that the posterior requires analytical calculation which is infeasible for large scale applications. Thus, we rely on PFs which perform a Monte Carlo approximation of the optimal Bayesian filtering. The algorithm, implementations details, as well as an extended discussion can be found in [9]. Particle filtering consists of two phases; the **prediction operation**, where particles are drawn from the state transition model (i.e., proposal distribution) in order to consider the time variant of the state and the **correction operation** which incorporates the likelihood of the observed measurements (i.e. **update phase**). Resampling of the particles is performed after the prediction and the update phase, but may also be performed after a set of correction phases. For an extended discussion on resampling of the PFs the interested readers are referred to [9].

Only the BTS's P_{TX} 4 is considered a time variant, since the BTS coordinates $[X_{BS} \ Y_{BS}]^T$ are obviously constant and the PLE within half or one hour (the time windows used in our experiments) is assumed constant. PFs explore all the possible states and converged to the PLE values and $[X_{BS} \ Y_{BS}]^T$ which maximize the likelihood of the used RSS subset.

Proposal Distribution-Time transition model for the P_{TX}

Initially, the localization algorithm was designed to deal with unknown but constant P_{TX} . During the development of this work, was found that the maximum radiated P_{TX} by the Cosmote's BTS is equal to $22W$, i.e. $43.5dBm$ [23] which used for the development of transition models. At this point, some important details must be considered in order to understand the derivation of the proposed time transition model for the $P_{TX}(t)$.

In general, the GSM Power Control during a phone call is an extremely complicated system in which both the BTS and the mobile phone participated in achieve two basic goals: 1) Minimize the energy consumption and 2) avoid interfering other mobiles in the cellular system. During a phone call, the P_{TX} due to power control may have variations up to 30dB [25]. However, 99% of the time the mobile phone is in idle mode. In the idle mode, the mobile phone is almost always connected to the unique broadcast control channel (BCCH) of a cell-ID and the RSSI is also measured by the mobile phone on the BCCH [29]. Thus, the overwhelming majority of the RSSI readings in our system were gathered at this channel so the time transition modelling must be concentrated on BCCH.

Fortunately, the GSM power control in BCCH is less complicated than the power control during a phone call. The GSM technical documentation [24] specifies that *"the BCCH carrier shall be continuously transmitted on all timeslots and without variation of RF level. However, the RF power level may be rambed down. . . "* as well as that *"the transmission is done at full power"*. The above facts are reasonable since the BCCH of a cell typically defines the coverage region. However, the GSM specification simply defines some general guidelines and leaves the exactly implementation details to the network equipment manufacturer. Moreover, at a real world BTS equipment GSM power control in applied on BCCH which is mainly focused on reducing the energy consumption when the network traffic is low. In [25] common approaches for energy saving in BCCH are presented: Switching off some transceiver modules (TRX) on BTS and a 2 dB reduction in the BCCH carrier when low network traffic is detected. By a short investigation,

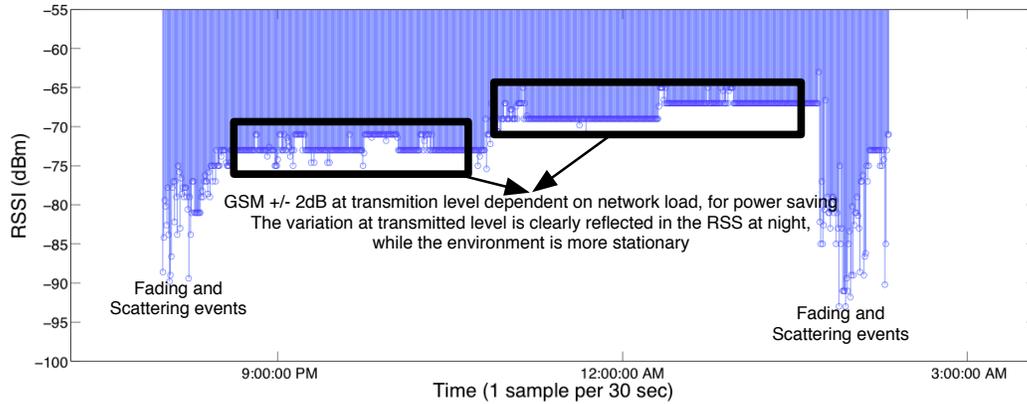


Figure 2: Demonstration of the GSM power control on the BCCH (mobile at fixed location).

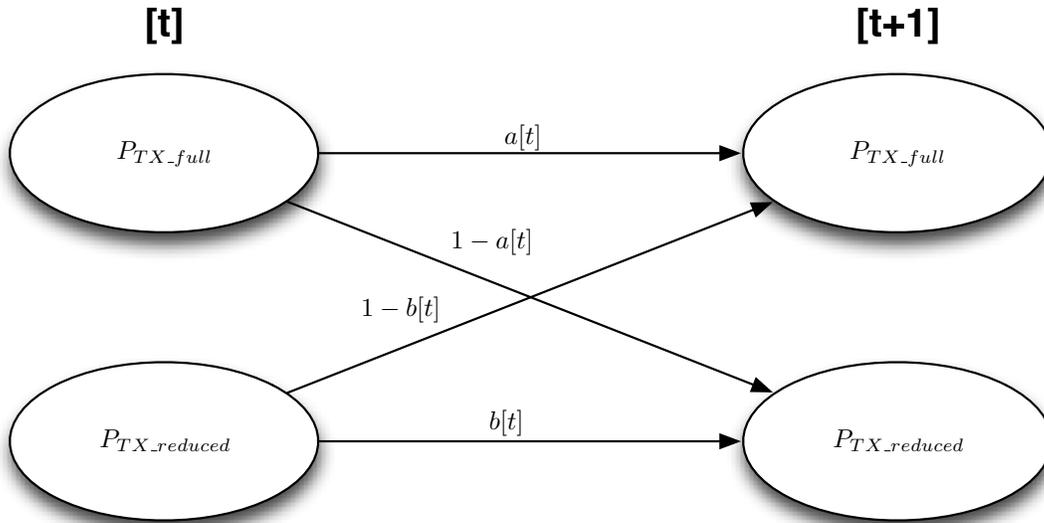


Figure 3: The two state Markov model for the transitions of the P_{TX} at BCCH.

these approaches are widely adopted at commercial BTSs [27], [28]. The aforementioned 2dB drop (or sometimes 3dB), i.e. fluctuation, which BTS performs is clearly demonstrated by the collected experimental data (more specifically at night hours where the environment is almost stationary) in Figure 2. The entire dataset follows this pattern.

In Figure 3, the two state Markov model proposed for modelling the transition model is presented. The transition probabilities, i.e. $a[t], b[t]$, are calculated as following:

1. According to [26], GSM traffic is modelled satisfactorily from Poisson arrivals with a time variant arrival rate $\lambda(t)$. As an initial approach, we assume $t = [1, \dots, 24]$ i.e. the network traffic load (as well as $a[t], b[t]$) changes every hour.
2. The $\lambda(t)$ is assigned real world values according to [25] and [26].
3. It is assumed that the BTS transmits on full power (i.e. $43.5dBm$) at the BCCH carrier if

its available slots are occupied more than 55%; otherwise the transmitter power is reduced 2dB (i.e. 41.5dBm). This approach has been adopted by [25]. Typical values for the offered capacity, i.e. slots, of a BTS per hour of the day are given also by [25].

4. Simulate GSM traffic for $K = 10000$ days, i.e. K Poisson processes realizations with $\lambda(t)$. Because the BTS offered timeslots are known, its realization offers the percentage of the occupied slots per second.
5. **If** $\text{occupiedSlots}(t) > 60\%$ **then** $P_{TX}(t) = 43.5dBm$, **otherwise** $P_{TX}(t) = 41.5dBm$. $\text{occupiedSlots}(t) = \text{newArrivals}(t) - \text{departuresFromSystem}(t) + \text{occupiedSlots}(t - 1)$. Mean duration of usage is 100 sec and obviously departures follow exponential distribution.
6. From the 10000 realizations of the stochastic process of $P_{TX}(t)$, transition probabilities $a[t], b[t]$ are estimated, depending on time t (i.e. the specific hour t), by sampling each minute and counting the transitions. Thus, $a[t], b[t]$ are calculated through a Monte Carlo (does not have a closed form solution).

In Figure 4 a realization of the GSM traffic during a day, i.e. Poisson stochastic process with arrival rate $\lambda(t)$, is demonstrated. In Figure 5 a realization of the stochastic process of $P_{TX}(t)$ constructed by the proposed two state transition Markov model 3 is compared with a realization of the $P_{TX}(t)$ in step 5. In Figure 6 a realization of a stochastic process of $P_{TX}(t)$ by 3 is compared with P_{TX} estimated by real data: A mobile phone was installed in a stationary environment on a rooftop with line of sight (LOS) with the BTS and path loss was calculated from 3. To the best of our knowledge, the proposed transition model for the BTS's P_{TX} is the first approach in bibliography which considers the time variant nature of the P_{TX} at a source discovery problem. It is straightforward that P_{TX} is a WSS stochastic process.

GSM Antenna directionality

All the aforementioned considerations for the 2dB reduction during the period of low network traffic are mentioned for the example of full transmitted power, i.e. inside the GSM antenna main lobe. Of course the 2dB reduction is reflected at any point in the radiation diagram of the GSM Antenna.

GSM antenna directionality must be considered and a first approach to model the basic assumptions follows. Directional antennas in contrast to the isotropic antennas, do not radiate uniformly the available power, but they amplify the signal in a specific direction (i.e. the main lobe) while inside the other lobes the transmitter power is lower. A typical GSM antenna has usually a 60° main lobe where antenna transmits at full power (i.e., in our case $P_{TX_full} = 43.5dBm$) and other two side lobes (approximately each lobe is 30°) where a 3dB drop in the signal level is observed. The rest of the 240° contains either secondary-minor lobes where the transmitted signal has 15 – 30dB smaller power than the main lobe or null points.

From the aforementioned facts about the directional nature of the GSM antennas, the transmitted levels provided by the Cosmote and the observations for the GSM power control, a generic radiated power according the GSM Antenna orientation and the time moment during the day can be used. The radiated pattern could be modelled with analytical expression for the gain of the antennas according to the antenna orientation with the respect to mobiles but we derive a simpler model which reflects the basic ideas/motif of all the above points, i.e. the backside lobe of 240° is assumed that attenuates 23dB the signal in average of all the secondary lobes and null points.

Assume ϕ_0 is the direction of the GSM antenna orientation with respect to the x-axis. The main lobe opening is $2\phi_s$, where $\phi_s = 30^\circ$. Thus, the antenna gain according to the ϕ_i , the angle of the orientation of the GSM antenna to the receiver i with respect to the x-axis.

$$G_H(\phi_i, \phi_0) = \begin{cases} 0 \text{ dB,} & \text{if } \phi_0 - \phi_s \leq \phi_i \leq \phi_0 + \phi_s, \text{ i.e. the main lobe} \\ -3 \text{ dB,} & \text{if } \phi_0 + \phi_s < \phi_i \leq \phi_0 + 2\phi_s \text{ or} \\ & \phi_0 - \phi_s > \phi_i \geq \phi_0 - 2\phi_s, \text{ i.e. the side lobes} \\ -23 \text{ dB,} & \text{elsewhere} \end{cases} \quad (5)$$

Finally, the radiated power obviously can be expressed as:

$$P_{TX}(\phi_i, \phi_0, t) = \begin{cases} P_{TX_full} + G_H(\phi_i, \phi_0) \text{ dBm,} & \text{if } pwrControl(t) = fullPwr \\ P_{TX_reduced} + G_H(\phi_i, \phi_0) \text{ dBm,} & \text{if } pwrControl(t) = reducedPwr \end{cases} \quad (6)$$

where the $pwrControl(t)$ is approached from the two state Markov model 3.

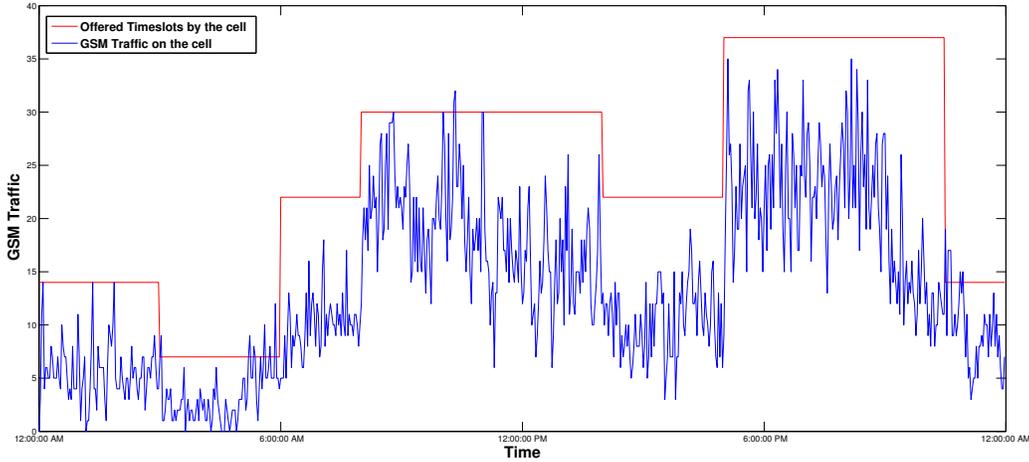


Figure 4: Realization of the (Poisson) stochastic process modelling the GSM daily traffic.

RSSI subset sampling per experiment

From the aforementioned considerations, it is clear that our algorithm must consider the timestamps of the collected RSSIs. If a specific time point into the day is fixed, the transition model for that hour is available from $a[t], b[t]$. In addition, using the entire collected dataset is not a feasible approach and thus a subset of the RSSIs must be sampled. The sampling process initially chooses random a time into the day and then samples one RSSI value per minute for $T = 30$ or 60 minutes. It is noticed that the sampling process considers **all** the collected RSSI values, **i.e. from many different days**, during this time window of T min). This approach offers a higher probability to use measurements not affected by deep fading or scattering. The algorithm 3 presents the sampling procedure which is used in this work.

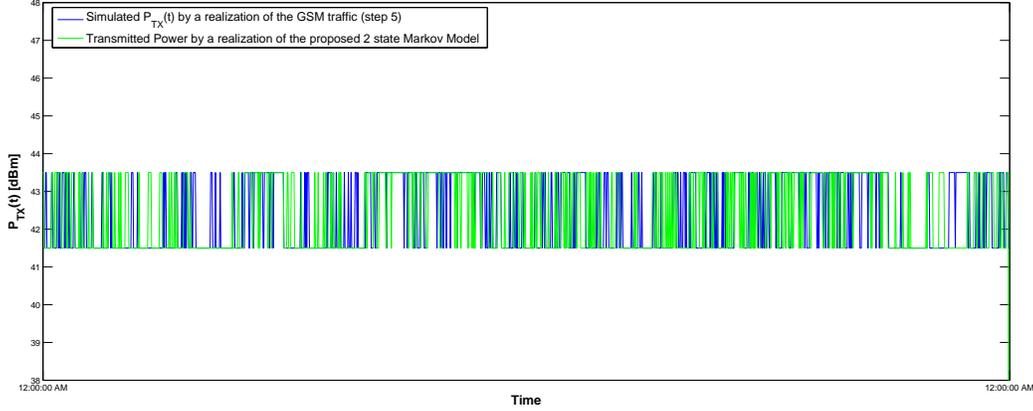


Figure 5: Realization of the stochastic process of the transmitted power of the BTS.

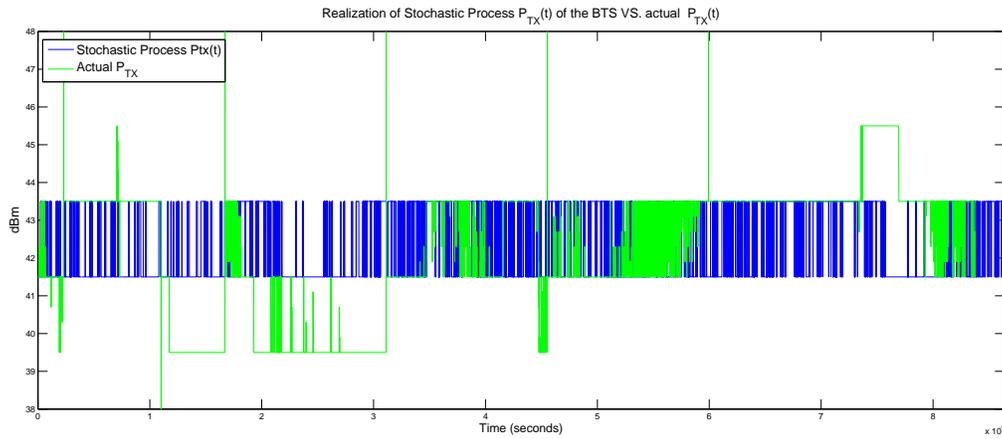


Figure 6: Realization of the Stochastic process by the proposed model VS. actual P_{TX}

Algorithm 3: Random sampling of a subset of RSS data, for a window of T minutes.

- (1): **Variables:** $T = \{30 \text{ or } 60\}$ minutes, time window duration.
 - (2): **Fixed** $cID \in \{60561 \text{ or } 60562\}$
 - (3): **Definition:** $min(t)$, returns the occurrence minute of a timestamp within a 24h day.
 - (4): $\tilde{t} = rand(0, 1440)$ i.e. select a random minute into the day ($24h * 60 = 1440min$)
 - (5): **for** $t = \tilde{t}$ to $\tilde{t} + T$ {step=1 min., index of the time window}
 - (6): **for** $j = 1$ to N {anchor index}
 - (7): **find all** C timestamps **where** $min(t_c^{[j]}) = t$
 - (8): $z_t^{[s,j]} = rand(P_{t_1}^{[j]}, \dots, P_{t_c}^{[j]}, \dots, P_{t_c}^{[j]})$, i.e. pick random from the above C RSSIs
 - (10): **end for**
 - (11): **end for**
-

Error Sensor Model (Observation Model)

The error sensor model for the RSSI observation, was modelled as a log-Normal distribution, as presented from 1, 2, 3. In the following equations, **cID** is fixed as it has been discussed earlier. Algorithm 3 gives a subset (sample) of RSS , \mathbf{z}^s , for running PFs:

$$\mathbf{z}^s \triangleq [\mathbf{z}^{[s,1]} \quad \dots \quad \mathbf{z}^{[s,j]} \quad \dots \quad \mathbf{z}^{[s,N]}]^T,$$

where, the vector of sampled RSS per user for the T minutes of sampling:

$$\mathbf{z}^{[s,j]} \triangleq [P_{t_1}^{[j]} \quad \dots \quad P_{t_b}^{[j]} \quad \dots \quad P_{t_T}^{[j]}]^T,$$

For fixed time moment t (i.e. N RSSIs at each update phase):

$$\mathbf{z}_t^s \triangleq [P_t^{[1]} \quad \dots \quad P_t^{[j]} \quad \dots \quad P_t^{[N]}]^T$$

It is straightforward that each particle is a candidate BTS position, therefore is assigned a weight during the update phase according to the likelihood of the measurements, using the error sensor model. Measurements are performed in N different and independent locations, thus are independent from each other, so for each particle m :

$$w_t^{[m]} = p(\mathbf{z}_t | \boldsymbol{\theta}_t^{[m]}) = p(P_t^{[1]} | \boldsymbol{\theta}_t^{[m]}) \dots p(P_t^{[N]} | \boldsymbol{\theta}_t^{[m]}) \quad \text{and}, \quad (7)$$

$$p(P_t^{[j]} | \boldsymbol{\theta}_t^{[m]}) = \mathcal{N}(\overline{P^{[m,j]}}(dBm), \sigma_{dB}^2) |_{P_t^{[j]}} = \frac{1}{\sigma_{dB} \sqrt{2\pi}} e^{-\frac{(P_t^{[j]} - \overline{P^{[m,j]}})^2}{2\sigma_{dB}^2}} \quad (8)$$

$\overline{P^{[m,j]}}$ is the mean value of the RSS in location \mathbf{l}_j , according to the model 2, 3, **if** BTS is on, $[x_{BS}^{[m]} \ y_{BS}^{[m]}]$. The estimated standard deviation per anchor j , $\hat{\sigma}_{dB}^{[j]}$, is calculated from measurements themselves, i.e $\hat{\sigma}_{dB}^{[j]} = \sigma_{dB}^{[j]}(\mathbf{z}^{[j]})$. In algorithm 4, the steps for get an estimation of the $\widehat{\mathbf{X}}_{BS} = [x_{BS} \ y_{BS}]$ via PFs is presented.

Algorithm 4: Particle Filtering for estimating $p(\boldsymbol{\theta}|\mathbf{z}^s)$, for a fixed **cID**

- (1): **Variables:** $M = 140000$: num. of Particles
 - (2): $X = Y = 1000m$, grid dimensions
 - (3): **Initialization of Particles:** $\boldsymbol{\theta}_{t=0}^{[m]} \forall m = 1 \text{ to } M$:
 - (4): $x_{t=0}^{[m]} = \text{rand}(1, X)$, $y_{t=0}^m = \text{rand}(1, Y)$ i.e. spread uniformly particles on the grid
 - (5): $P_{TX,t=0}^{\text{status},[m]} = \text{rand}(\text{fullPwr or reducedPwr})$
 - (6): $\phi_{t=0}^{[m]} = \text{rand}(0, 360)$ i.e. spread uniformly particles for the possible ϕ_0 of the GSM antenna
 - (7): $n_{p,t=0}^{[m][j]} \sim \mathcal{N}(\hat{n}_p^{[j]}, 0.15)$, i.e. explore values of PLE around the estimated PLE
 - (8): **Sample** $\mathbf{z}^s \leftarrow$ Algorithm 3
 - (9): **for** $t = 1 \text{ to } T$ **do** {*time-step*}
 - (10): **for** $j = m \text{ to } M$ **do** {*particle index*}
 - (11): **Sample** $P_{TX,t}^{\text{status},[m]} \sim p(P_{TX,t}^{\text{status},[m]} | P_{TX,t-1}^{\text{status},[m]})$ from 3
 - (12): Calculate $\theta_{m \leftrightarrow j}$, $\forall j$, i.e. angle between particle m and j , with respect to x-axis
 - (13): Calculate $G_H(\theta_{m \leftrightarrow j}, \phi_t^{[m]})$ from 5
 - (14): Calculate $P_{TX}(\theta_{m \leftrightarrow j}, \phi_t^{[m]}, t)$ from 5, 6
 - (15): $\overline{P}^{[m,j]} = P_{0,t}^{[m]}, -10n_{p,t}^{[m][j]} \log_{10}\left(\frac{d_{m,j}}{d_0}\right)$, $\forall j$
 - (16): $w_t^{[m]} = p(\mathbf{z}_t^s | \boldsymbol{\theta}_t^{[m]})$, from 7, 8
 - (17): **end for**
 - (18): $\boldsymbol{\theta}_t^{[1:M]} = \text{lowVarianceSampler}(\boldsymbol{\theta}_t^{[1:M]}, \mathbf{w}_t^{[1:M]})$, Particles after resampling have $\mathbf{w}_t^{[1:M]} = 1$
 - (19): **end for**
 - (20): $p(\boldsymbol{\theta}|\mathbf{z}^s) = k\text{-means_clustering}(\boldsymbol{\theta}_t^{[1:M]}, \mathbf{w}_t^{[1:M]})$ {*retrieve the continuous PDF*} [9]
-

The final step is to estimate $\widehat{\mathbf{X}}_{BS}$ from the posterior $p(\boldsymbol{\theta}|\mathbf{z}^s)$:

$$p(\mathbf{X}_{BS}|\mathbf{z}^s) = \int_{P_{TX}} \int_{n_p^{[1]}} \cdots \int_{n_p^{[N]}} p(\boldsymbol{\theta}|\mathbf{z}^s) \, dn_p \cdots dn_p \, dP_{TX}, \quad (9)$$

where: $dP_{TX} = 1$, $dn_p = 0.01$, Thus:

$$\widehat{\mathbf{X}}_{BS} = [\hat{x}_{BS}, \hat{y}_{BS}]^T = \int_x \int_y p(\mathbf{X}_{BS} = [x \ y] | \mathbf{z}^s) \cdot [x \ y]^T \, dy dx \quad (10)$$

3.4 Independent Repeats of PFs Experiments, Outlier Detection

Estimator of the position of the cell tower is provided by equation 10 for the sample \mathbf{z}^s . However, we can exploit K^3 independent experiments, i.e. independent sampled \mathbf{z}^s which intuitively will limit much more the error. Thus, except 10 the following are proposed:

³In this work $K = 2000$ is utilized.

1. $\widehat{\mathbf{X}}_{BS}^{mean} = \sum_{k=1}^K \frac{1}{K} [\widehat{x}_k \ \widehat{y}_k]$ i.e. average ALL estimated positions, $\mathbf{h}_{1:K}$, by K independent runs of Algorithm 4 (PFs).
2. $\mathbf{X}_{BS}^{filtered}$ by outlier detection (see Algorithm 5) on K independent runs of Algorithm 4 (PFs).

Algorithm 5: Outlier detection on K independent estimated positions by PFs.

- (1): **for** $k = 1$ to K $\{K = 2000\}$
 - (2): **Sample** \mathbf{z}^s using Algorithm 3
 - (3): **Calculate** $\widehat{\mathbf{X}}_{BS}^k$ by estimating $p(\boldsymbol{\theta}|\mathbf{z}^s)$ with Algorithm 4, and then solve 9, 10
 - (4): $\mathbf{h}_k = \widehat{\mathbf{X}}_{BS}^k$
 - (5): **end for**
 - (6): $scores = outlierDetection(\mathbf{h})$, i.e. assign a probability per k^{th} estimation to be a "bad" estimation [22]
 - (7): **find** the (5%) smallest scores and keep the corresponding positions in \mathbf{b}
 - (8): $\widehat{\mathbf{X}}_{BS}^{no_outliers} =$ the mean value of positions in \mathbf{b}
-

4 Results

4.1 Prior Art: Strongest RSS and Weighted Centroid

4.1.1 Strongest RSS

cell-ID	60561	60562	60563	60564	60565	60567	60569
RSE (m)	86.94	367.35	505.98	304.99	157.61	705.18	119.95
combine cells RSE (m)	648.5						

Table 2: Strongest RSS root square error (RSE) results from MySignals dataset.

RMSE=321.14m of the strongest RSS for the above cells. LOS measurements far away from antenna breaks down the strongest RSS algorithm.

4.1.2 Weighted Centroid

Results for the weighted centroid algorithm are presented in Figure 7.

4.2 Our work

The 2 cell-IDs with the most measurements were tested (60561, 60562) by considering different scenarios for the time windows T and the anchors N . The results are presented on Table 3. It is clear that larger time windows removes the bias of the estimator despite of the larger RMSE.

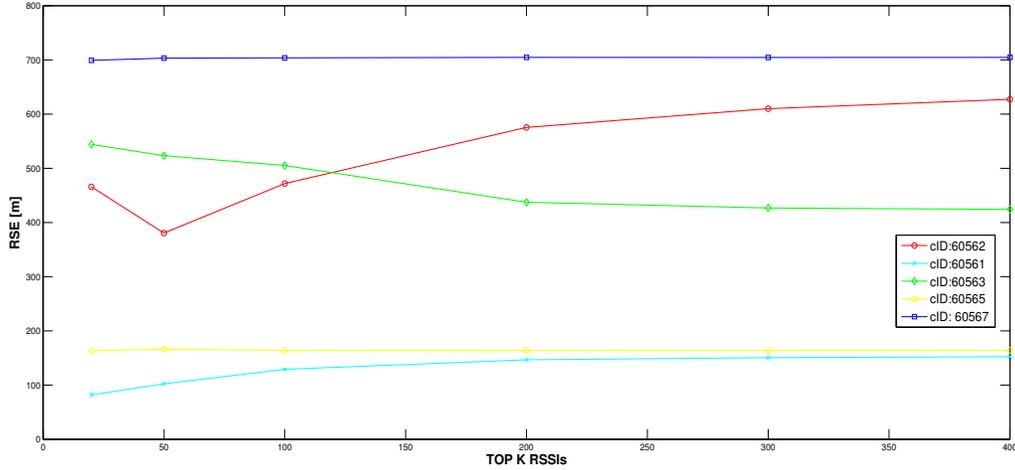


Figure 7: Weighted Centroid root square error (RSE) results from MySignals dataset.

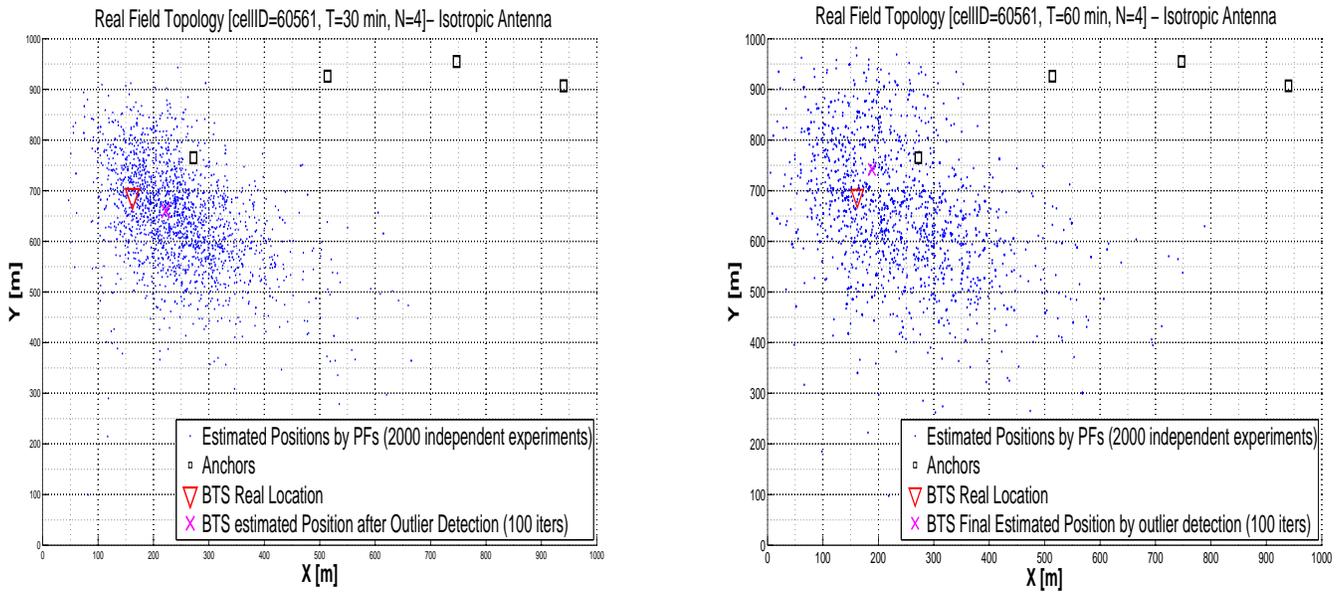


Figure 8: cID:60561, $N = 4$. Left: $T = 30min$, Right: $T = 60min$.

For the cID 60562, the algorithm was tested both for $N = 4$ and $N = 5$ (typically tested with the topology of cID 60561). In Figure 10 the RSE per experiment of the scenario 1 is presented, in order to demonstrate the high variance of the PFs estimator (a bad sampled measurements breaks down the algorithm). In Figure 8 the K estimated positions by PFs are presented for scenario 1,2 (see Table 3). Figures 9, 11 show K estimated positions are for the scenarios 4,6 and 3,5, for $T = 30min.$, $T = 60min.$ respectively: A bias in the estimation can be observed, as well as it is clear that the extra anchor helps the accuracy of the estimation. From a cloud of estimations, the estimations are limited in a circle by using the extra anchor.

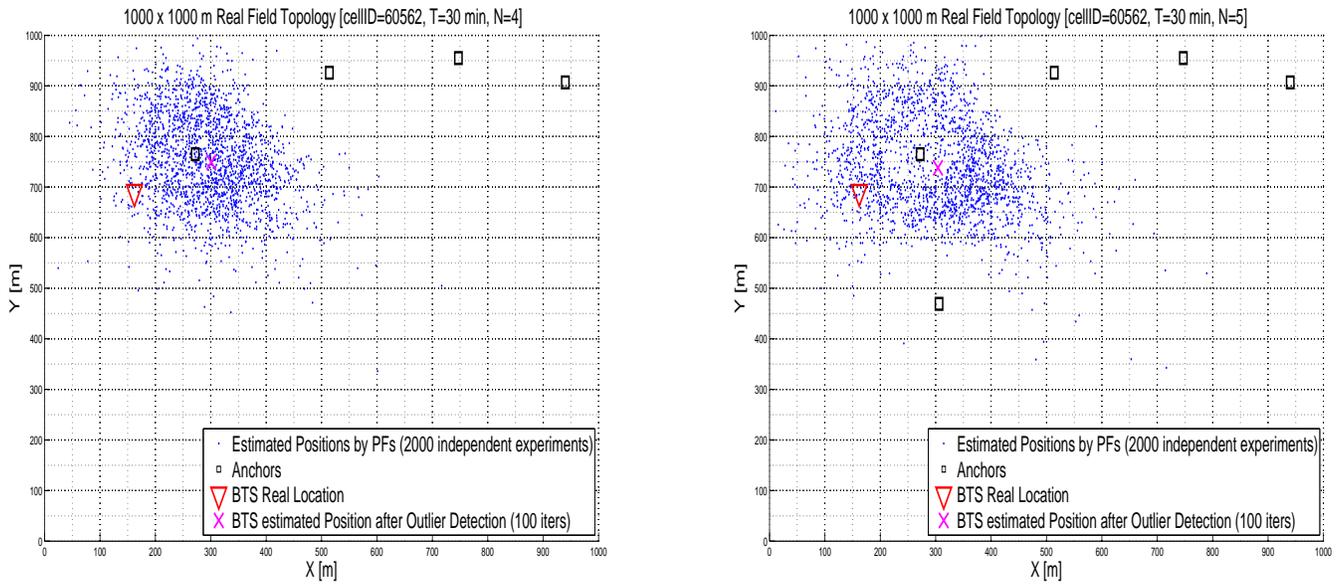


Figure 9: cID:60562. Left: $N = 4$, Right: $N = 5$. $T = 30min$.

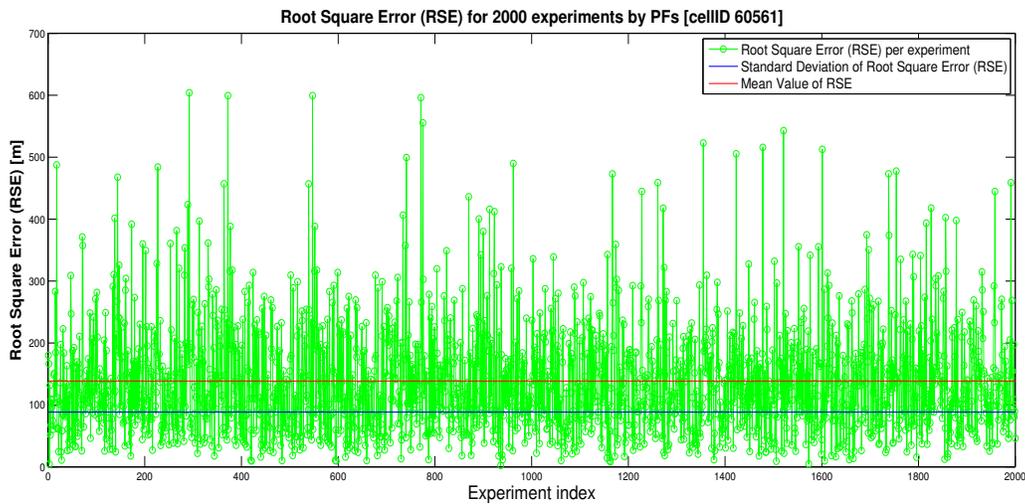


Figure 10: RSE cID 60561, 2000 independent experiments, $T = 30min$.

5 Ongoing Work

- Model and exploit directionality of the GSM cell tower antennas and incorporate it into the measurement (RSS) model used by PFs. The modelling of the antenna directionality reduces the uncertainty in the estimations and removes the bias. (see Figures 12, 13).
- Investigate and test new outlier detection algorithms to apply on the estimated positions of K independent experiments by PFs.
- Incorporate at runtime PFs a smart way of outlier and bad measurement detection following

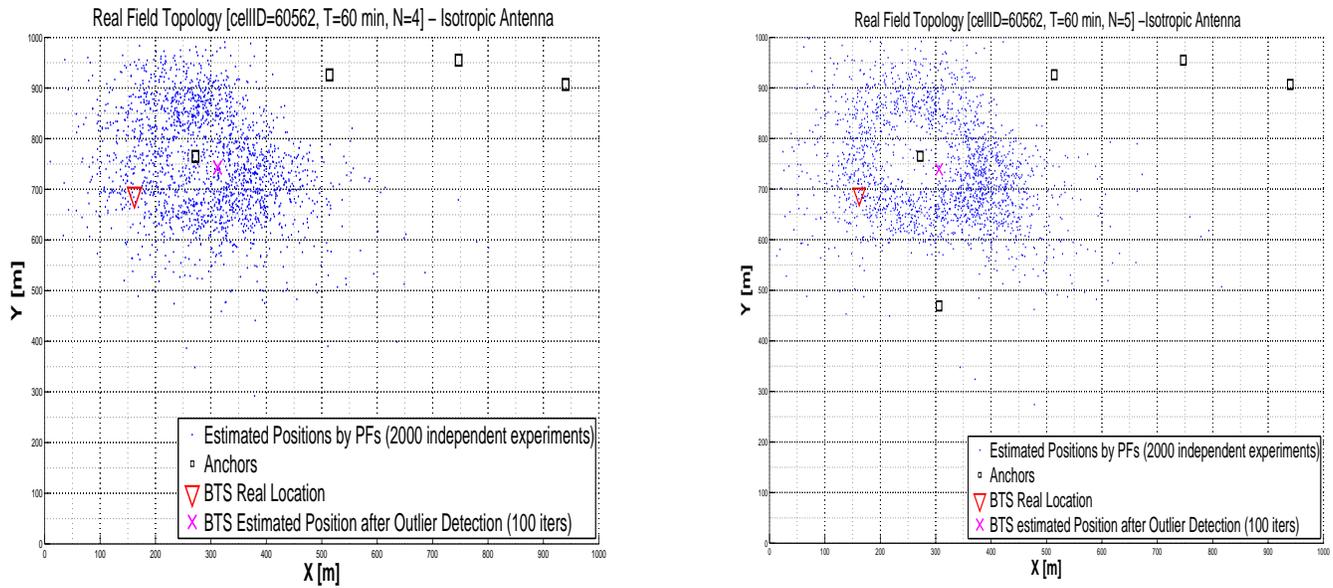


Figure 11: cID:60562. Left: $N = 4$, Right: $N = 5$. $T = 60min$. The extra anchor limits estimations to a circle.

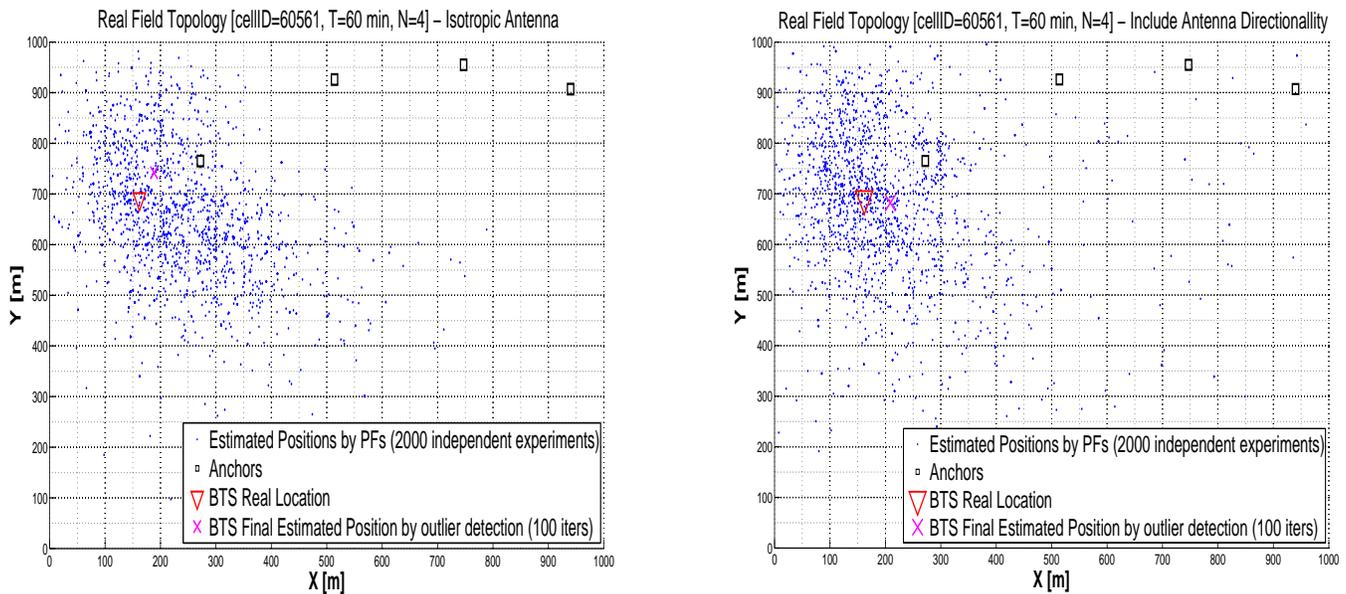


Figure 12: Incorporate GSM antenna directionality model, cID:60561. Left: Isotropic radiation, without the directionality model, Right: Incorporate to the PFs the directionality model.

the approach of the $\sigma_{dB} = 4$ (within 10 sec) threshold which is currently applied.

- Resampling in PFs is crucial. Currently, each update phase of the PFs is followed by resampling. If at the update one very "bad" measurement is taken into account, then the resampling (probably) will eliminate the correct particles which exists on the real BTS position. A better approach is to use a burst of e.g. 100 measurements for updating

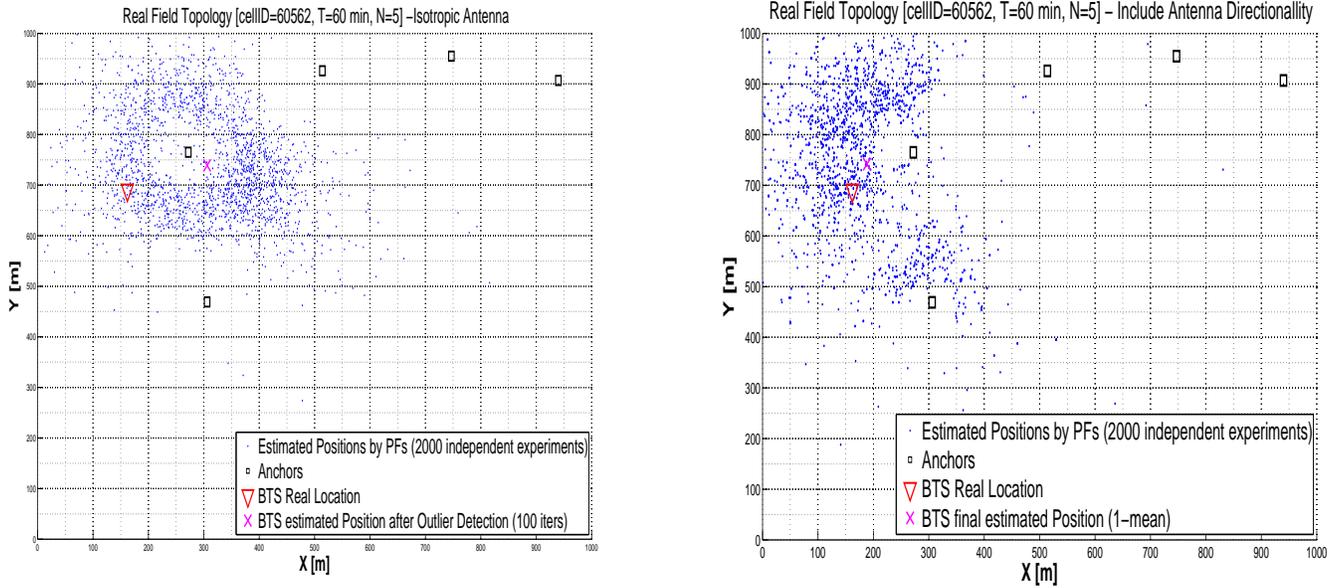


Figure 13: Incorporate GSM antenna directionality model, cID:60562. Left: Isotropic radiation, without the directionality model, Right: Incorporate to the PFs the directionality model.

Scenarios	1	2	3	4	5	6
cID	60561	60561	60562	60562	60562	60562
T (time window)	30 min	60 min	30 min	60 min	30 min	60 min
N , #anchors	4	4	4	4	5	5
A. $\widehat{\mathbf{X}}_{BS}$ $RMSE$ (m) Stand. dev.: σ_m (K independent PFs)	138.59 88.51	164.52 102.18	177.32 70.74	189.42 81.92	187.69 88.52	194.96 91.45
B. $\widehat{\mathbf{X}}_{BS}^{mean}$ RSE (m) (average K estimations)	93.60	84.18	146.51	145.67	146.11	149.59
C. $\widehat{\mathbf{X}}_{BS}^{no_outliers}$ $RMSE$ (m) {500 exper.} (outlier detection, Algo.5)	67.33	49.08	152.54	159.53	163.54	170.33
D. Strongest RSS RSE (m)	86.94	-	367.35	-	-	-
E. Combine Cells RSE (m)	648.5	Includes 6056-1,2,3,5,7,8				
F. Top K RSSIs RSE (m)	101.3	-	380.15	-	-	-

Table 3: Results of the experiments of the investigated scenarios.

the weights of the PFs and then perform the resampling. Thus, the correct particles will not be eliminated because some measurements would be outliers, but the most of the

measurements will not.

References

- [1] Bernard Sklar, Communications Engineering Services , "Rayleigh Fading Channels in Mobile Digital Communications Systems Part I: Characterization" *IEEE Communications Magazine*, pages 90-100, July 1997.
- [2] COST Action 231 , "Digital mobile radio towards future generation systems, final report" *Brussels, tech rep. European Communities*, EUR 18957, 1999.
- [3] Asharf, Mohammad Taha , "Statistical Tuning of Walfisch-Ikegami Propagation Model Using Particle Swarm Optimatization" *19th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, Nov 16, 2012.
- [4] AWE Communications, "Propagation Models & Scenarios" Urban," *available online at www.awe-communications.com/Download/AudioPresentations/Propagation_Urban.pdf*, 2012
- [5] Masaharu Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services," *IEEE Transactions on Vehicular Technology*, vol. VT.29 1, No. 3, August 1980.
- [6] Bartlomiej Blaszczyszby and Mohamed Kadhem Karray, "Linear-Regreesion Estimation of the Propagation-Loss Parameteres Using Mobiles' Measurements in Wireless Cellular Networks," *Proceedings of the 10th nternational Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012*, May 14-18, 2012, Paderborn, Germany, pp. 54-59.
- [7] D. C. Cox, R. Murray, and A. Norris, "800 MHz Attenuation Measured in and around Suburban Houses," *AT&T Bell Lab. Tech. J.*, vol. 673, no.6, July-Aug. 1984, pp. 921-954
- [8] S. Russel, P. Norvig, "Artificial Intelligence, A modern Approach", *Klidarithmos 2nd American Edition*, ch. 14, 2007
- [9] S. Thrun, W. Burgard, D. Fox, "Probabilistic Robotics," *Klidarithmos greek edition*, ch. 3, Particle Filters, 2011.
- [10] Jie Yang, Alexander Varshavsky, Hongbo Liu, Yingying Chen and Marco Gruteser, "Accuracy Characterization of Cell Tower Localization", *Proceedings of the Twelveth International Conference of Ubiquitous Computin, Ubicomp 2010*, sep. 26-29, 2010, Copenhagen, Denmark
- [11] Yu-Chung Cheng, Yatin Chawathe, Anthony LaMarca and John Krumm, "Accuracy Characterization for Metropolitan-scale Wi-Fi Localization", *ACM, MobiSys '05 Proceedings of the 3rd international conference on Mobile systems, applications, and services* , June 6-8, 2005, Seatle, USA
- [12] Alex Varshavsky, Denis Pankratov, John Krumm and Eyal de Lara, "Calibree: Calibration-free Localization using Relative Distance Estimations", *ACM, 08 Proceedings of the 6th International Conference on Pervasive Computing* , May 19-22, 2008, Sydney, Australia

- [13] Mike Y. Chen, Timothy Sohn, Dmitri Chmelev, Dirk Haehnel, Jeffrey Hightower, Jeff Hughes, Anthony LaMarca, Fred Potter, Ian Smith and Alex Varshavskym "Practical Metropolitan-Scale Positioning for GSM Phones", *Proceedings of the Eighth International Conference of Ubiquitous Computing, Ubicomp 2006*, Sep. 17-21, 2006, Orange County, California, Australia
- [14] Reza M. Vaghefi, M. R. Gholami and E.G. Strom, "RSS-Based Sensor Localization with Unknown Transmit Power", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 22-27, 2011, Prague, Czech Republic.
- [15] Gang Wang, H. Chen, Youming Li and Ming Jin, "On Received-Signal-Strength Based Localization with Unknown Transmit Power and Path Loss Exponent", *IEEE Wireless Communications letters*, vol1, No.5, pp. 536-539 ,October 2012.
- [16] Radim Zemek, Daisuke Anzai, Shinsuke Hara, Kentaro Yanagihara and Kenichi Kitayama, "RSSI-based Localization without a Prior Knowledge of Channel Model Parameters", *Springer, International Journal of Wireless Information Networks*, Volume 15, Issue 3-4, pp 128-136,,December 2008,.
- [17] A. A. Gorjl and Brian D.O. Andreson, "Emitter localization using received-strength-signal data", *Elsevier, Signal Processing Magazine*, Volume 93, Issue 5, May 2013, Pages 996–101236.
- [18] Neal Patwari, A. O. Hero, M. Perkins, N. S. Correal and R. J. O'Dea, "Relative Location Estimation in Wireless Sensor Networks", *IEEE Transactions on Signal processing*, vol 1, No. 8, August 2003.
- [19] Junichi Shirahama, Tomoaki Ohtsuki, M. Perkins, "Performance Comparison of Small Scale Fading Elimination Methods," *IEEE Transactions on Signal processing*, vol 1, No. 8, August 2003.
- [20] Tolga Kurt, Yann le Helloco, Bernard Breton, "Estimate of local average power of a mobile radio signal," *Proceedings of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, MSWiM '05*, pp. 219-222, October 10-13 2005.
- [21] Lee, W.C.Y., "Estimate of local average power of a mobile radio signal," *Vehicular Technology, IEEE Transactions on*, vol.34, no.1, pp. 22-27, February 1985.
- [22] Michael Kim, "Anomaly Detection," *available online at <http://www.mathworks.com/matlabcentral/fileexchange/39593-anomaly-detection>*, 26 Dec. 2012
- [23] Cosmote, available online at, <http://www.cosmote.gr/content/demos/gr/kinhth.html>
- [24] ETSI , "Digital cellular telecommunication system (phase 2+), Radio subsystem link Control," *ETSI, GSM Technical Specification*, BCCH Carrier, ch. 7.1 ,TS 05.08, July 1996
- [25] Mikko Saily, Guillaume Sebire, Eddie Riddington "GSM/EDGE Evolution and Performance" *WILEY publications*, ch. 13.5, Energy Savings through Software Solutions, 1st edition, 2011.

- [26] S. Bregni, R. Cioffi, M. Decina, "An empirical Study on Statistical Properties of GSM Telephone Call Arrivals," *Global Telecommunications Conference Proceedings, Globecom '06*, USA, San Francisco, Nov27-Dec. 1 2006.
- [27] Nokia Siemens Networks, "BSS20958: Energy saving mode for BCCH TRX," *available online at <http://dc199.4shared.com/doc/qnn6PMaL/preview.html>*, 2009.
- [28] Ericsson Ab 2003, "Boosting Downlink Output Power," *available online, <http://telecomfunda.com/forum/archive/index.php?t-26533.html>*, 2003
- [29] Moray Runney, "LTE and the Evolution to 4G Wireless," *Agilent Technologies, Wiley & Sons*, ch. 3.6.1.6,1st edition 2013