

Linköping University Post Print

Robust Tracking in Cellular Networks Using HMM Filters and Cell-ID Measurements

Mussa Bshara, Umut Orguner, Fredrik Gustafsson and Leo Van Biesen

N.B.: When citing this work, cite the original article.

©2011 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Mussa Bshara, Umut Orguner, Fredrik Gustafsson and Leo Van Biesen, Robust Tracking in Cellular Networks Using HMM Filters and Cell-ID Measurements, 2011, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, (60), 3, 1016-1024.

<http://dx.doi.org/10.1109/TVT.2011.2107926>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-67307>

Robust Tracking in Cellular Networks Using HMM Filters and Cell-ID Measurements

Mussa Bshara, *Member, IEEE*, Umut Orguner, *Member, IEEE*, Fredrik Gustafsson, *Senior Member, IEEE*, and Leo Van Biesen, *Senior Member, IEEE*

Abstract—A localization algorithm based on cell identification (Cell-ID) information is proposed. Instead of building the localization decisions only on the serving base station (BS), all detected Cell-IDs (serving or non-serving) by the mobile station (MS) are utilized. The statistical modeling of user motion and the measurements is done via a Hidden Markov Model (HMM) and the localization decisions are made with Maximum A Posteriori (MAP) estimation criterion using the posterior probabilities from an HMM filter. The results are observed and compared to standard alternatives on an example whose data were collected from a WiMAX network in a challenging urban area in the Brussels capitol city.

Index Terms—Cell-ID, hidden Markov model, HMM, GPS, GSM, mobile WiMAX, navigation, LBS, positioning, positioning accuracy, WiMAX networks.

I. INTRODUCTION

LOCALIZATION in wireless networks has attracted a lot of interest as a supporting means to services that require sufficient localization accuracy (see [1] and the references therein). This includes location based services (LBSs), navigation, tracking, security applications etc.. Historically, the need of locating users with mobile devices arises to respond to security applications such as tracking users with emergency calls made by cellular phones [U.S. Enhanced 911 (E-911)]. Indeed the regulations adopted by the U.S. federal Communications Commission (FCC) require that all the emergency calls made by cellular phones have to be localized within an accuracy of 125 m in 67% of the cases [2]. Localization can be done by measuring physical quantities related to the radio signal traveling between a mobile terminal or a subscriber station (SS) and a set of base stations (BSs), such as received signal strength (RSS), time of arrival (TOA), time difference of arrival (TDOA), angle of arrival (AOA), and Cell-ID. A broad spectrum of solutions can be found in literature including those based on RSS [3–6], TOA [7], TDOA [8]

and AOA [9, 10]. More recently, a localization approach that uses a combination of the mentioned measurements has been proposed in [11]. Such hybrid techniques guarantee a higher localization accuracy. Most of the conventional localization methods have been developed to operate under line-of-sight (LOS) propagation conditions. However, due to the electromagnetic propagation properties especially in urban areas, the non-line-of-sight (NLOS) errors are very likely to corrupt the original signal. A number of methods have been used to mitigate the NLOS errors [12]. Some techniques exploit the multipath characteristics of the environment, such as the *Fingerprinting* [13–16] (See [17] for a very recent example). All the mentioned techniques above require some adjustment or modification (adding some hardware or software) to be implemented in the SS or BS. It is apparent that the accuracy and the required adjustments are tightly related. The higher the required accuracy is, the more adjustments that are required, which means additional cost and more computational burden. From this point of view, despite its low accuracy [18, 19], Cell-ID positioning is the first positioning method that has been used by GSM operators to provide LBSs. Its simplicity and low cost made it the most preferable way to position a network user when the obtained accuracy is enough for the required application. The Cell-IDs are transmitted over the control channel and easy to obtain with no extra cost. It is enough to associate the Cell-ID of the serving base station with its location (which is known to the system) to have a positioning system. This explains why this method is cheap, fast and suitable for applications requiring high capacity. The “low accuracy” has always been a characteristic of the Cell-ID positioning due to the relatively large cell size in GSM networks, especially in rural areas. Some publications proposed using enhanced Cell-ID positioning, which uses the timing advance (TA) value to reduce the cell size, and hence, improving the accuracy [20].

WiMAX networks started to be widely deployed providing the infrastructure for LBSs. The mobile WiMAX standard, which is expected to be finalized in the near future (the end of 2010), will be providing LBSs to WiMAX users. This means that the need for simple, cheap and preferably network dependent ways to locate mobile and fixed modems is becoming more urgent. In this paper, the Cell-ID localization problem is solved by an approach that is not restricted to the serving BS Cell-ID only, but it makes use of all the detected Cell-IDs by an MS at a certain time instant, with an HMM filter to locate the users. The method can be implemented either as network-centric or mobile-centric, with different requirements

Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Mussa Bshara is currently with the Department of Communication Engineering at HIAST (Higher Institute For Applied Sciences and Technology), Damascus, Syria. He has been with the the Department of Fundamental Electricity and Instrumentation, Vrije Universiteit Brussel, Brussels, Belgium up to July 2010 (e-mail: mussa.bshara@hiast.edu.sy).

Leo Van Biesen is with the Department of Fundamental Electricity and Instrumentation, Vrije Universiteit Brussel, Brussels, Belgium (e-mail: lv-biesen@vub.ac.be).

Umut Orguner and Fredrik Gustafsson are with the Department of Electrical Engineering, Linköping University, 581 83, Linköping, Sweden (e-mail: {umut,fredrik}@isy.liu.se).

Manuscript received 01,2011

on the mobile's activity. In the former case, the base station can force the mobile to make and send measurement reports as long as the phone is on, it has not to be active. In the latter case, the mobile can always decide this. We present our results along with remarks on WiMAX networks which were the main motivation and the illustrative case study for this research. However, it must be emphasized that our results equally apply to other types of networks. The importance of the contributions of this paper can be summarized as follows:

- The method proposed in this paper makes use of all the available Cell-IDs to the user in the localization, which is contrary to the current methods that utilize only that of the serving BS.
- A simple but practical way to form an HMM model from a database obtained by measurement campaigns is proposed and illustrated on a real life example.

The rest of the paper is organized as follows. Section II discusses the discretization of the area under study and the database construction both of which are quite important when one works with HMMs. Both the HMM model and the filter, which form the main contribution of this paper, are detailed in Section III. The results with localization accuracy assessment and comparison with other techniques are provided in Section IV which also presents some experimental studies investigating the effects of different algorithm parameter selections. Section V draws some conclusions which finalizes the paper.

II. AREA DISCRETIZATION AND DATABASE CONSTRUCTION

A. Area Discretization

In classical Cell-ID based positioning, the area under study is divided into cells according to the strongest transmission received in this area; i.e., the cell of a particular BS is the area where its transmission is the strongest transmission that can be detected. The cell of a BS differs from its coverage area by means of the transmission power level compared to the rest of the detected BSs. In other words, the coverage area is the area where the BS transmission is detected, but it is not necessarily the strongest transmission. The shape of an actual BS's cell is usually irregular and highly depends on the propagation environment. Sometimes it can even consist of multiple disconnected areas. In terms of localization performance depending on Cell-ID classical approach, the smaller the cell sizes are, the better accuracy one can get from Cell-ID based localization [18, 19]. Therefore, an investigation of cell sizes would give one a rough idea about the accuracy that can be obtained and therefore, the LBSs that can be provided. In order to carry out such an analysis, the cell sizes of the two main cellular networks operating in Belgium have been studied using the provided data by Clearwire for the Pre-WiMAX network, and by Proximus for the GSM network. The results are depicted in Figure 1 for the Pre-WiMAX network, where the cell size cumulative distribution function (cdf)¹ has been calculated for the regions that are covered by the Pre-WiMAX network in Belgium, and the same analysis can be

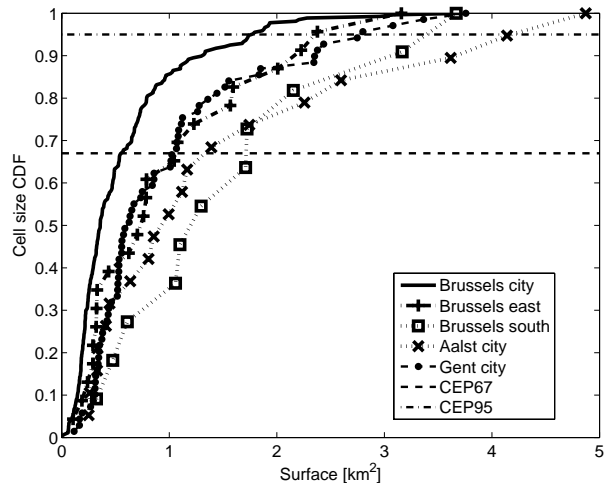


Fig. 1: Cell size analysis for the Pre-WiMAX network in Belgium.

found in [20] for the GSM network. The obtained results show smaller cells in the Pre-WiMAX network than in the GSM one. For example, the WiMAX cell size in Brussels is about 0.5 km^2 compared to 1 km^2 in GSM for CEP67² (67% of the cells), and 1.8 km^2 compared to 3 km^2 for 95% of the cells. Hence, using the Cell-ID positioning in the Pre-WiMAX network in Belgium will provide better accuracy than in the GSM network. Although this might not be the case for the other parts of the world, the trend of the wireless broadband future networks is towards having much smaller cells than what is used today (Femto cells), and WiMAX is one of the best candidates for the IMT-advanced next generation mobile networks³. Therefore, it is expected to have much smaller cells in the future WiMAX networks (where broadband is needed) than in GSM ones all over the world. The accuracy improvement (by using smaller cells, in our case the WiMAX cells), however, is expected to be limited. In order to obtain a considerable positioning accuracy improvement depending on Cell-IDs (or BS identification number BSID), a novel approach has been proposed in this paper that uses all the available BSIDs (not only the serving BSID) in the diversity set⁴ of an MS [21]. This approach could be a terminal-based or a network-based one as this information is known by the network as well⁵. The MS shall transmit a *MOB_SCN - REP* message to report the scanning results to its serving BS [21]. The scanning report contains the number of active BSs in the diversity set ($N_{\text{current_BSs}}$) which takes a maximum value of 7 (the serving BS and 6 neighboring ones), and also all the neighboring BSID (which is 48 bits long) [21]. This message can be event-triggered or periodical. When considering all the BSs in the diversity set,

²CEP67 and CEP95: Circular Error Probability for 67% and 95% of the cases respectively.

³Source: Intel Corporation and IEEE 802.16m System Requirements Document. Copyright Intel Corporation ©2008.

⁴The diversity set is a list of the active BSs to the MS.

⁵In this regard, we are not concerned about the complete message flow, but the availability of this information.

¹A cdf is the integral of a probability density function given as $\text{cdf}(x) = \int_{-\infty}^x \text{pdf}(x') dx'$ which makes $\text{cdf}(x)$ a unitless quantity.

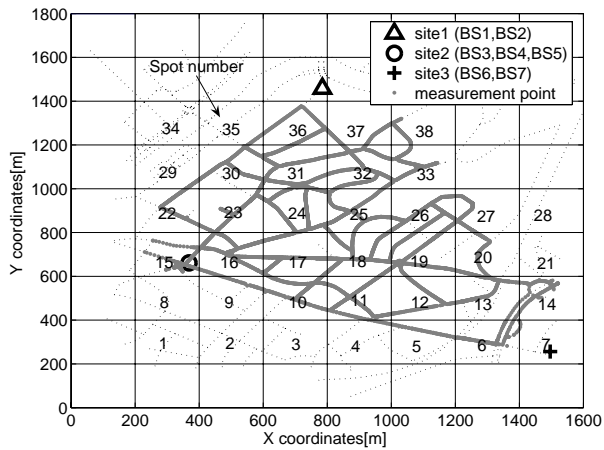


Fig. 2: The area under study is divided into equal squares (Spots). The main BS positions are also illustrated. There is also another main BS called as BS8, whose coordinates are (2364m,2839m) and is not illustrated in the figure.

the distinction between the candidate points (for the location of an MS) is much more than the classical case when only the serving BS is considered. In the case of Cell-IDs, many neighboring points will have the same diversity sets. Thus, it is rather logical to consider areas instead of single points. Therefore, the area under study has been divided into areas such that the diversity set of an MS, in each area is distinctive from the others as much as possible. The cells of the BSs can be quite irregular and actually disconnected which is quite difficult to handle in an automated localization algorithm like ours. Hence, we, in this study, will use artificial rectangular and equal areas that are distributed regularly over the area under study. From now on, in order to avoid a confusion with cells, each rectangular area will be called a “Spot”, and the message that would flow to different applications might be the coordinates of the center of the maximum probability Spot obtained from the HMM filter. Figure 2 illustrates an example about dividing the area under study to (200×200 m) Spots.

B. Spot size selection

The size of the Spots will be affected directly by the structure of the studied network which determines the diversity set of an MS. Selecting too small Spots will cause the diversity sets in close Spots to be almost the same and will basically cause a loss of computational load. In addition, the minimal Spot size is also determined by the number of the off-line measurements that are conducted in that Spot, and used to compute the BSs detection probabilities as explained in sub-section II-C. On the other hand, too large Spots will reduce the performance by causing extreme inhomogeneity in the Spots, and by reducing the effect of the motion model which is based on the transition probability matrix as explained in sub-section III-1.

C. Database construction

In order to obtain a reliable HMM model to be used in the localization, a database of Cell-IDs in the area under study

is essential. The database can be constructed using different ways. One way is to conduct real measurements using WiMAX modems. Another way is to use radio planning tools to predict the required Cell-IDs [1], or to use the principle of *wardriving* [22], where the users with positioning capabilities (for instance GPS) report their position and observations to a database [23, 24]. In this study, the first method was used and measurements were collected along the streets in the area under study using a standard WiMAX modem. Each measurement contains all the available Cell-IDs at that time instance and location. During the measurement campaigns, the true position of each measurement was also obtained using a GPS sensor. We show these known positions and the corresponding Cell-ID measurements in the database as $p_{db}^{(i)}$ and $Y_{db}^{(i)}$ respectively for $i = 1, \dots, N_{db}$ where N_{db} is the number of measurements in the database. In the case of Spot size of (200×200 m), about 50 measurements were obtained in each of the $N_s = 38$ Spots mentioned in Figure 2, and hence $N_{db} \approx 38 \times 50 = 1900$. The main Cell-IDs (BSs) are shown in Figure 2. There is also another main BS called as BS8, whose coordinates are (2364m,2839m) and is not illustrated in the figure. Those BSs illustrated have the strongest transmission in the area under study and can be received in the majority of the Spots. In order to differentiate the Spots, i.e., to raise their diversity, weak transmissions generated by far BSs (not shown on the figure) are also used. Different weak transmissions can be received in different Spots, which has a big impact on raising the diversity of the Spots. The measurement campaigns showed that there are $N_{ID} = 14$ different Cell-IDs that can be detected in the area under study. Each position $p_{db}^{(i)}$ in the database is simply composed of the x and y coordinates of the measurement when the corresponding Cell-ID measurement vector $Y_{db}^{(i)}$ is obtained. The corresponding Cell-ID measurement vector $Y_{db}^{(i)}$ is a N_{ID} -vector. The elements of $Y_{db}^{(i)}$ can be either one or zero depending on whether the corresponding Cell-ID was detected or not at $p_{db}^{(i)}$, i.e.,

$$\left[Y_{db}^{(i)} \right]_j = \begin{cases} 1, & \text{if the } j\text{th Cell-ID is detected at } p_{db}^{(i)} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where the notation $[\cdot]_j$ denotes the j th element of a vector. On average, each measurement $Y_{db}^{(i)}$ in the database (modem’s diversity list) contains about 4 non-zero elements, i.e., Cell-IDs (minimum 1 and maximum 7). Each measurement was triggered by a correct GPS reading. The average time between two consecutive measurements was about 2-3 seconds. This time depends on the GPS fix availability, and the time needed for the software to retrieve the information from the modem.

III. HMM MODELING AND FILTER

Hidden Markov models (HMMs) have become the workhorse of discrete estimation since their introduction (see the tutorial [25] and the references therein). In this section, we are going to model the Cell-ID estimation problem as an inference problem with an underlying HMM structure. For this purpose, we use mostly the terminology that was presented in [26]. We define the state vector $X \in \mathcal{X}_{N_s} \triangleq \{e_1, \dots, e_{N_s}\}$

where e_i are unit vectors in \mathbb{R}^{N_s} which has all zero elements except the i th element which is unity. The integer N_s represents the number of Spots present. Notice that, with this state vector, the event $X_k = e_i$ represents the case where the target is in the i th Spot at time k . We assume that the sequence $\{X_k\}$ is Markov and we have the equality

$$\mathbb{E}[X_{k+1}|X_k] = \Pi^T X_k \quad (2)$$

where $\Pi \triangleq [\pi_{ij}]$ is the so-called probability transition matrix of size $N_s \times N_s$ with

$$\pi_{ij} \triangleq P(X_{k+1} = e_j | X_k = e_i). \quad (3)$$

We define the measurement vector $Y \in \mathbb{R}^{N_{ID}}$ similarly to database measurement vectors $Y_{db}^{(i)}$ defined in the previous section. Notice that the classical HMM framework allows only one element of Y to be unity and the others should be all zero. However, in our work, since we can collect multiple Cell-IDs at the same time, we allow multiple nonzero elements. We assume that the elements of the measurement Y_k are independent given the state X_k , where k represents the time instant that the measurement Y_k is collected and we have the equality

$$\mathbb{E}[Y_k|X_k] = H X_k \quad (4)$$

where the matrix $H \triangleq [h_{ij}]$ of size $N_{ID} \times N_s$ has the probabilities h_{ij} defined as

$$h_{ij} = P([Y_k]_i = 1 | X_k = e_j). \quad (5)$$

At this point, we have another distinction from the classical HMM framework which is, the probabilities h_{ij} do not have to satisfy $\sum_{i=1}^{N_{ID}} h_{ij} = 1$ (more on this in the modeling subsection III-2).

The Cell-ID estimation problem related to this framework can be stated as finding the state estimate

$$\hat{X}_{k|k} = \mathbb{E}[X_k | Y_{0:k}] \quad (6)$$

where $Y_{0:k}$ denotes all the measurement obtained between time 0 and k , i.e., $Y_{0:k} \triangleq \{Y_0, Y_1, \dots, Y_k\}$. Notice here that the solution $\hat{X}_{k|k}$ might not be in the original (discrete) state space \mathcal{X}_{N_s} but it satisfies $\sum_{i=1}^{N_s} [\hat{X}_{k|k}]_i = 1$ and hence the elements $[\hat{X}_{k|k}]_i$ of the estimate can be interpreted as the posterior probabilities $P(X_k = e_i | Y_{0:k})$ i.e., the probability that the target is in the i th Spot given all the measurements.

The recursive solution of the problem (6) is given by the so called HMM-filter [25, 26]. This algorithm is summarized below.

Algorithm 1 (HMM Filter):

- 1) Initialization: Select an initial estimate $\hat{X}_{0|0} = \bar{X}$. Set $k = 1$.
- 2) Prediction update: Predict the state estimate using the model (2) as

$$\hat{X}_{k|k-1} = \Pi^T \hat{X}_{k-1|k-1} \quad (7)$$

- 3) Measurement update: Calculate the measurement updated estimate $\hat{X}_{k|k}$ from the predicted estimate $\hat{X}_{k|k-1}$ using the model (4) and the measurement Y_k as

$$\hat{X}_{k|k} = \frac{\mathcal{L}_{Y_k} \odot \hat{X}_{k|k-1}}{\sum_{i=1}^{N_s} [\mathcal{L}_{Y_k} \odot \hat{X}_{k|k-1}]_i} \quad (8)$$

where the likelihood vector $\mathcal{L}_{Y_k} \in \mathbb{R}^{N_s}$ is defined with the elements

$$[\mathcal{L}_{Y_k}]_i \triangleq P(Y_k | X_k = e_i) \quad (9)$$

and the sign \odot denotes the Hadamard product (element-wise multiplication) of the vectors.

- 4) If a location (Spot) estimate \hat{c}_k is to be found, one selects the Spot corresponding to the maximum element of $\hat{X}_{k|k}$, i.e.,

$$\hat{c}_k = \arg \max_i [\hat{X}_{k|k}]_i. \quad (10)$$

- 5) If there is a measurement Y_{k+1} , set $k = k + 1$ and go to step 2. Otherwise, stop.

Having defined the HMM filtering, in the following parts of this section, we are going to concentrate on the modeling part and examine how the model parameter matrices Π , H and the likelihood vector \mathcal{L}_{Y_k} are to be formed.

1) *Transition Probability Matrix* Π : The transition probability matrix used in the prediction step of the HMM filter is constructed using the road network properties. Once the target is in a specific Spot, it is much more probable that it is going to stay in the same Spot rather than moving into another one. This property results in a diagonally dominant transition probability matrix. The probabilities of Spot-to-Spot transitions can be arranged using the road network information or Spot proximities when one lacks the road information. It is generally a reasonable idea to reduce the transition probabilities when the corresponding Spots get farther. Very far Spots could be assigned zero transition probabilities. However, this would make the algorithm not be able to recover from wrong estimates due to highly noisy measurements. Hence it is reasonable that the lower bound of the probabilities can be selected to be slightly larger than zero so that every Spot transition is possible (although with a very low probability) and the HMM filter can make quick corrections to its estimated Spot via measurements. Overall we use the following simple algorithm for this purpose.

Algorithm 2 (Transition Probability Selection): For each Spot i ,

- 1) Determine the set of neighbor Spots $\Lambda_i = \{\lambda_1, \lambda_2, \dots, \lambda_{|\Lambda_i|}\}$ using the road network and/or Spot proximity. Here the integer $|\Lambda_i|$ denotes the cardinality of the set Λ_i i.e., the number of neighbors of Spot i .
- 2) Select two probabilities $0 < p_1, p_2 < 1$ such that

$$p_1 > |\Lambda_i| p_2 \quad \text{and} \quad p_1 + |\Lambda_i| p_2 < 1. \quad (11)$$

3) Assign the probabilities $\{\pi_{ij}\}_{j=1}^{N_s}$ as

$$\pi_{ij} = \begin{cases} p_1 & j = i \\ p_2 & j \in \Lambda_i \\ \frac{1-p_1-|\Lambda_i|p_2}{N_s-|\Lambda_i|-1} & \text{otherwise} \end{cases} \quad (12)$$

Notice that in the above algorithm, the probabilities corresponding to the neighbor Spots were selected all equal (as p_2) for the sake of simplicity but each neighbor can actually have a different probability based on the road network information if available⁶. In such a case the terms $|\Lambda_i|p_2$ in (11) and (12), which stand for the total probability mass of the neighbor Spots should be replaced with the summation $\sum_{j \in \Lambda_i} p_2^j$ where p_2^j denotes the specific probability assigned to the j th neighbor Spot. Figure 3 depicts graphically the transition matrix which serves as a discrete motion model. In the figure, the probabilities with the arrows show the selected Spot transition probabilities for the corresponding transitions from the center Spot. Note that the highest probability (0.9) at the central Spot stands for the probability of staying at the central Spot (i.e., self transition). The very small lower bound mentioned above is not illustrated in the figure for the sake of simplicity.

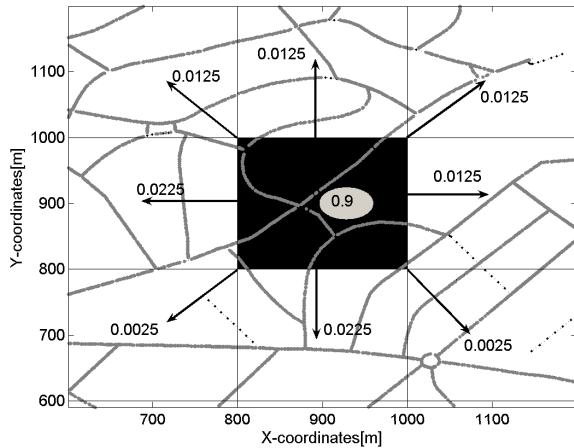


Fig. 3: The motion model presented by the transition probability matrix. The sum of all probabilities has to be equal to one.

2) *Measurement Matrix H* : The measurement matrix H which contains the probabilities defined in equation (5) is obtained from the database $\{p_{db}^{(i)}, Y_{db}^{(i)}\}_{i=1}^{N_{db}}$ described in Section II. Denoting the areas of the Spots by $\{\mathcal{S}_i\}_{i=1}^{N_s}$, the columns of the matrix H , shown as $[H]_{:,j}$, are calculated as

$$[H]_{:,j} = \frac{1}{|\{i|p_{db}^{(i)} \in \mathcal{S}_j\}|} \sum_{\{i|p_{db}^{(i)} \in \mathcal{S}_j\}} Y_{db}^{(i)}. \quad (13)$$

Notice that this calculation is the result of a frequentist interpretation for the probabilities h_{ij} . It also requires the implicit assumption that the behavior of the Cell-IDs are

⁶For example, the road directions can be used to assign higher probabilities to some Spots and lower probabilities to others.

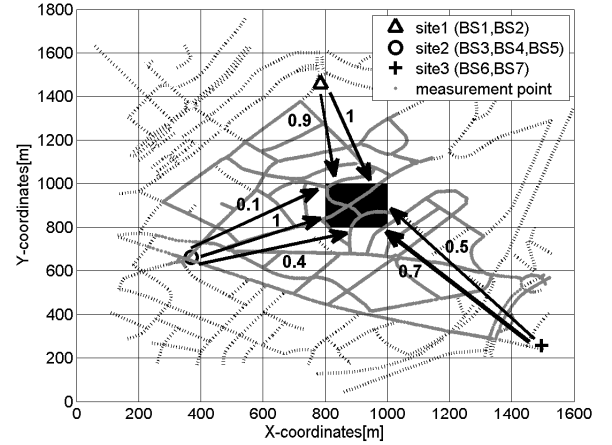


Fig. 4: The sensor model presented by the measurement matrix. Each arrow indicates the probability of detecting a certain CELL-ID (BS) in the studied (dark) Spot. The arrow starts from the site where the BS is located and ends in the studied Spot. Note that the sum of all probabilities doesn't necessarily equal to one.

homogeneous inside the Spots. For the i th Cell-ID, we simply set h_{ij} as the ratio of the number of times it has been detected at the positions $p_{db}^{(\cdot)}$ inside the j th Spot to the total number of positions $p_{db}^{(\cdot)}$ (measurements) inside the j th Spot. It is also important to emphasize that, since more than one Cell-IDs can be collected at a single position $p_{db}^{(\cdot)}$, it is not necessarily correct that the elements of $[H]_{:,j}$ sum up to unity. Figure 4 depicts graphically the elements of the H matrix which serves as the sensor model. The probabilities with the arrows in the figure represent $[H]_{i,j}$, i.e., the probability of measuring the i th Cell-ID at the j th Spot shown as the central Spot. A probability 1 means that in the central Spot, the corresponding BSs Cell-ID is expected to be measured all the time.

3) *Likelihood Vector \mathcal{L}_{Y_k}* : The likelihood vector is defined as in (9). Assuming independence over the detections from different BSs given the current Spot, we can write

$$[\mathcal{L}_{Y_k}]_j = \prod_{i=1}^{N_{ID}} \underbrace{P([Y_k]_i | X_k = e_j)}_{\triangleq \tilde{h}_{ij}(Y_k)} \quad (14)$$

By the definition of h_{ij} given by (5) we can calculate $\tilde{h}_{ij}(Y_k)$ as

$$\tilde{h}_{ij}(Y_k) = \begin{cases} h_{ij} & [Y_k]_i = 1 \\ 1 - h_{ij} & [Y_k]_i = 0 \end{cases} = |1 - h_{ij} - [Y_k]_i|. \quad (15)$$

Hence, the standard way to calculate the likelihood vector $\mathcal{L}_{Y_k} \in \mathbb{R}^{N_s}$ is given by

$$[\mathcal{L}_{Y_k}]_j = \prod_{i=1}^{N_{ID}} \tilde{h}_{ij}(Y_k). \quad (16)$$

However, what has been observed in preliminary experiments, is that this likelihood calculation mechanism is very sensitive to the non-homogeneous behavior of the probabilities h_{ij}

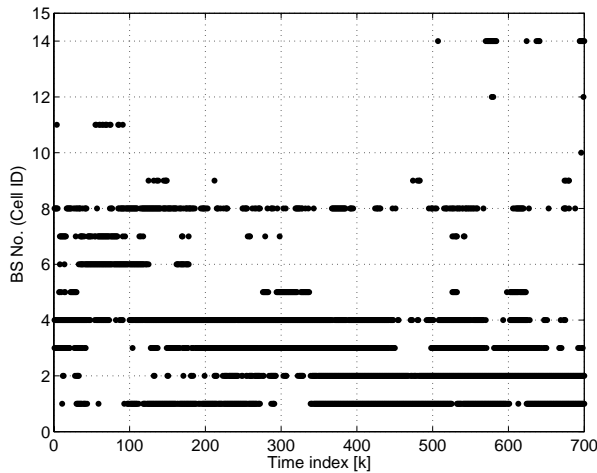


Fig. 5: The measurement vector as a function of time. Each black dot at a time value represents the existence of a measured ID from the corresponding BS at that time.

inside the Spots. Hence another likelihood calculation mechanism, which has been seen to be more robust, is suggested here as

$$[\mathcal{L}_{Y_k}]_j = \|Y_k - [H]_{:,j}\|_2^{-1} = \frac{1}{\sqrt{\sum_{i=1}^{N_{ID}} (Y_k)_i - h_{ij})^2}}. \quad (17)$$

This likelihood function has been observed to behave better than (16) in the vicinity of noisy measurements and erroneous database information. The center of the Spot with the highest posterior probability is assumed to be the estimated position.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed HMM based Cell-ID localization method was run on a test scenario whose trajectory covers almost all the roads in the area under study. The measured Cell-IDs on the test trajectory have been plotted as a function of time in Figure 5 to provide an overview of the data (Cell-IDs). For each time instant in Figure 5, the existence of a black dot at the ID number represents the measured ID of the corresponding BS. The time instance itself is not important, but the important is the simultaneous CELL-IDs. i.e., the CELL-IDs that can be detected simultaneously. Figure 6 depicts the relation between the obtained positioning accuracy and the number of available BSs (the length of the measurement vector). On average, the trajectory points (locations) with low positioning error have relatively long measurement vectors ($\text{length} \geq 4$) as shown in Figure 6. Therefore, the positioning accuracy is expected to be high in dense networks where the measurement vectors are long because long vectors can differentiate the Spots better than short vectors.

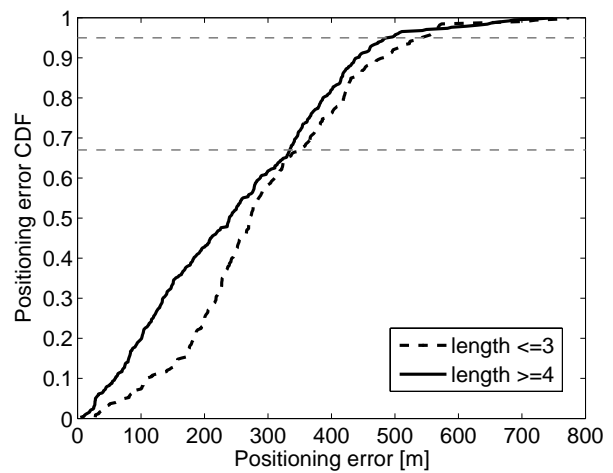


Fig. 6: The relation between the length of the measurement vector (the number of the detected BSs) and the positioning accuracy.

The estimated Spots of the HMM filter with respect to the true target Spots are illustrated in Figure 7, and the average HMM Cell-ID estimation performance is below compared with

- 1) A recent fingerprinting based particle filtering approach that uses the received signal strength indices (RSSI) [17].
- 2) The same method as above that uses SCORE values instead of RSSI values. SCORE values are related directly to RSSI values and used by WiMAX modems to evaluate the connection quality to the available BSs. From positioning point of view, SCORE values are less accurate than RSSI values, because of the effect of the *Viterbi* decoder, which plays an opposing role in this case, on them.
- 3) The mentioned classical Cell-ID positioning method (yellow pages).

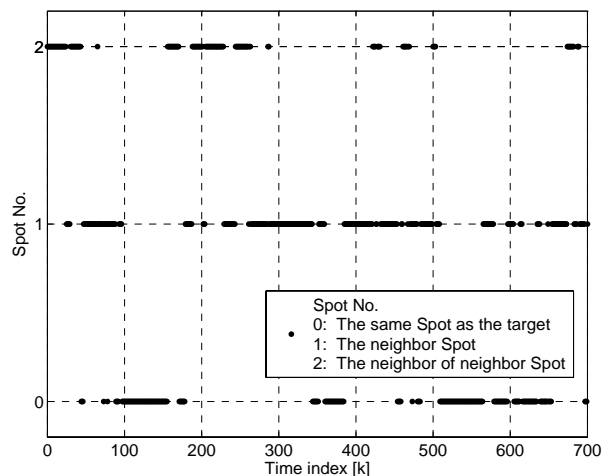


Fig. 7: The estimated Spots of the HMM filter with respect to the true target Spot.

The used fingerprinting based particle filtering approach in the comparisons is a recent approach that was proposed in

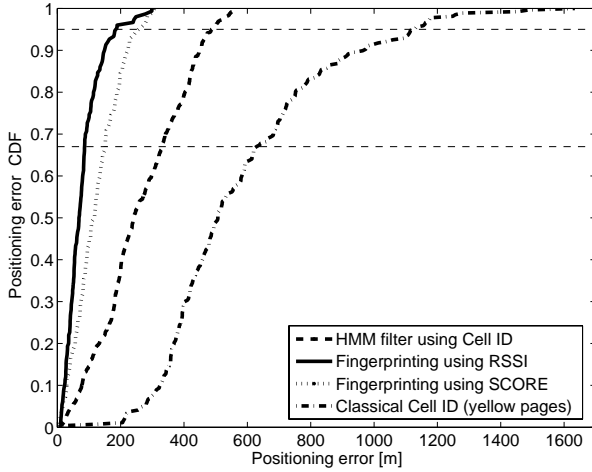


Fig. 8: Positioning accuracy assessment. A comparison between the proposed approach, Fingerprinting (using RSSI and SCORE) and the classical Cell-ID approach.

[17]. The classical fingerprinting, which involves comparing the online measurement vectors with a previously obtained database to make localization, is known for a long time [16]. The method derived in [17] gives an integration of the fingerprinting, which is well-known to be able to model the multi-path effects and fast fading sufficiently, with the particle filters (PFs) [27, 28] yielding much better results than the PFs equipped only with classical log-power model (called as Okumura-Hata model in the literature [29, 30]).

The cumulative distribution functions (cdfs) of the position estimation errors for all algorithms are depicted in Figure 8. The cdf which is obtained using the novel HMM approach (see Figure 8) shows an error of about 300 m for 67% of the cases and an error of about 480 m for 95% of the cases⁷. The fingerprinting approach, as a result of the fact that it uses much more information (RSSI or SCORE values) in addition to the Cell-IDs, provided the highest accuracy. Although having lower performance than fingerprinting approaches, HMM based approach is significantly better than the classical Cell-ID positioning. The accuracy improvement gained by using the HMM based approach is a factor of two compared to the classical one.

Considering the computational power, this approach is more efficient than fingerprinting as it requires much less storage space. It keeps one measurement vector for each Spot which acts as a sufficient statistics for the collected data in the Spot. But in fingerprinting a measurement vector has to be kept (saved in a database) for a fairly large number of points in each Spot. Hence in terms of storage, this algorithm requires much less space. Compared to the classical approach, this approach requires more computational power as the classical approach uses the information of only one Cell-ID (the serving BS).

Although the performance of our method has been shown above to be significantly better than the classical approach, our selection of the specific parameters, namely, Spot size, motion model (Spot transition probabilities) and the initial

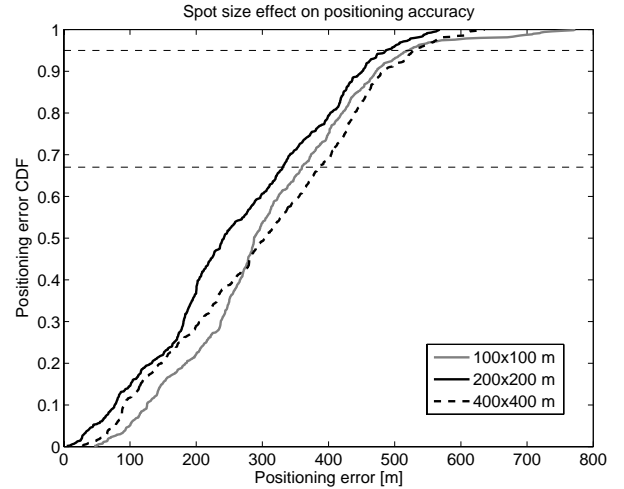


Fig. 9: The Spot size effect on positioning accuracy.

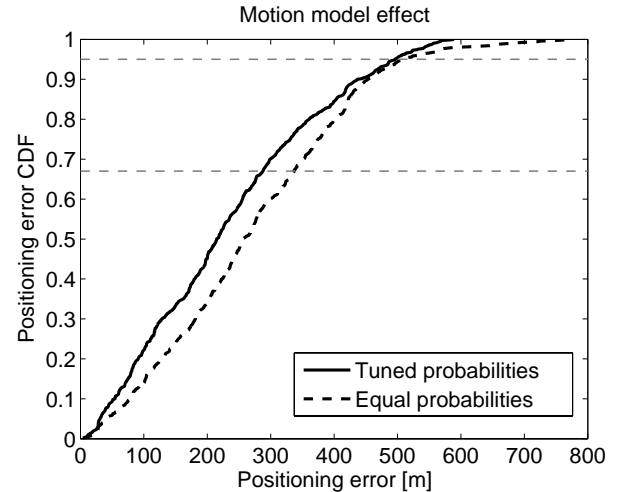


Fig. 10: The effect of the motion model (Spot transition probabilities) selection on the positioning accuracy.

estimate in the algorithm can raise robustness issues. Hence, for investigating the effect of these parameter selections on the position estimation performance, we present the results of the further experiments we have conducted below.

For illustrating the effect of the Spot size, the HMM filter was run using three different Spot sizes (100×100 m), (200×200 m) and (400×400 m). The cdfs of the corresponding position estimation errors are illustrated in Figure 9 which shows that the size of (200×200 m) provides the best accuracy. The existence of some kind of approximate local optimality at this Spot size is clearly illustrated in the figure. Although the illustrated Spot size changes caused only about 50m increase in the average positioning error, one must, in general, take into account that the increases might be overwhelming in the case of extremely small or large Spot sizes. The consideration of the trade-offs mentioned in Section II-B is hence essential in the Spot size selection.

The positioning error depends, to some extent, also on the motion model, i.e., Spot transition probabilities. In the above results we considered that the target can move to all

⁷The error here is calculated from the center of the estimated Spot.

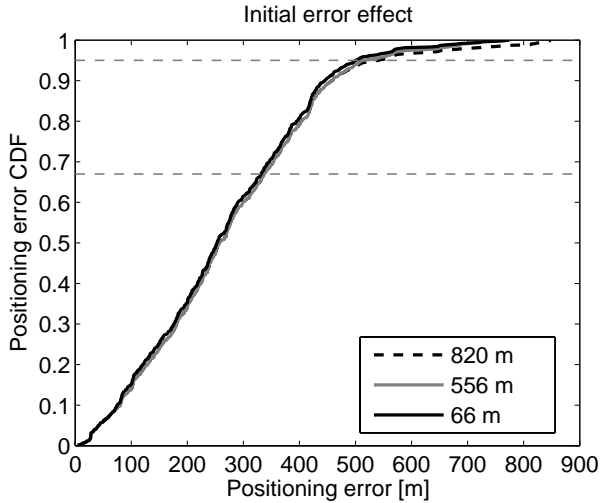


Fig. 11: The effect of the initial estimate on the positioning accuracy for different initial errors.

the neighboring Spots with equal probabilities. But one can give different probabilities to different Spots depending on geographical, demographical or road information if the target is known to be on road. In this case, some neighbor Spots can be assigned higher probabilities than the rest of the Spots and the accuracy can be improved, but this technique requires using extra information especially in the form of a road network database. Figure 10 shows the accuracy improvement achieved by assigning higher probabilities to the Spots that are more likely to be picked up by the target which were decided manually based on the road network. The results in the figure show that a manual tuning of the transition probabilities can bring on average up to 50m more accuracy into the estimation although for large databases an automatic transition probability setting still seems more feasible.

The initial estimates in the above simulations were selected by using the classical Cell-ID approach. The dependency of the results on these initial estimates has been examined in Figure 11 for different initial errors which shows that the proposed approach is quite robust against the initial estimate error. This is a manifestation of the fact that the Spot transition probabilities in our algorithm were selected to be above zero for all transitions which makes the algorithm able to forget about and in fact correct the initial errors.

Finally, we here would like to emphasize that, in the cases where the accuracy requirements are satisfied by both HMM approach and the fingerprinting approaches, HMM approach has advantages over the latter in the following ways:

- The Cell-ID off-line database can be constructed using radio planning programs with almost the same accuracy as using direct measurements but the RSSI (or SCORE) off-line database obtained by radio planning programs (which is required by the corresponding particle filtering algorithms) is much less accurate than the one obtained by direct measurements.
- Obtaining Cell-IDs is simple and fast. No modifications on the handset or BS are required.

- Using Cell-IDs requires less computational power and data storage. This fact is also evident from a simple comparison between the computational and storage requirements of HMMs in this work and those of complicated PFs with RSSI or SCORE data as in [17].

Therefore, if the accuracy provided by the novel approach satisfies the application requirements, this approach can be considered as a simple and efficient alternative even over the fingerprinting approach. The location estimates depending only on Cell-IDs derived in this work can be used in all location-dependent applications whose accuracy requirements are satisfied by our method. If not, our methodology can serve for initiating more accurate localization algorithms. In fact, all location-based services that depend on Cell-ID positioning can be enhanced and extended by using the novel approach, such as: Friend Finder, finding the closest place, “where am I?” application etc.. In addition to serving as a robust and low complexity application that supplies initial conditions for more complicated and computation costly trackers such as the one mentioned in [17]. For example, the current coarse tracker can be used as a proposal density for the particle filter explained in [17].

V. CONCLUSIONS

This paper discussed HMM-based coarse localization depending on Cell-IDs with specific emphasis on WiMAX networks. The proposed HMM is obtained from measurement campaigns easily and is different from conventional HMMs in several aspects. An assessment of the achieved accuracy has been provided with a comparison to a high accuracy approach (fingerprinting) and to the classical Cell-ID approach. The obtained results suggest the preference of the HMM based Cell-ID positioning over the classical approach. It is argued that when the accuracy requirements are satisfied with the HMM based approach, it can be preferred even over the high performance fingerprinting approaches, thanks to its simplicity and low cost.

In this study, the Spot size and model were static. The Spot size of (200×200 m) gave the best accuracy comparing to (400×400 m and 100×100 m) because it is indeed, a fair trade-off between having a sufficient number of measurements in each Spot and creating too inhomogeneous Spots. Interesting subjects for further research is to use clustering algorithms applied to the measurements to design irregular Spot sizes with similar (homogeneous) measurement vectors. Another aspect is to make the model adaptive using the principle of *wardriving*, where the users equipped with accurate localization capabilities (GPS) contribute to the database model. First, the Cell-ID measurements can be used to update the map h_{ij} . Second, the shape of the Spots can be adaptively refined when the map information is improved. Finally, the observed user mobility can be used to adapt the transition probabilities π_{ij} .

ACKNOWLEDGMENTS

The second and the third authors gratefully acknowledge funding from SSF Strategic Research Center MOVIII and the Vinnova/FMV TAIS project ARCUS.

REFERENCES

- [1] F. Gustafsson and F. Gunnarsson, "Possibilities and fundamental limitations of positioning using wireless communication networks measurements," *IEEE Signal Process. Mag.*, vol. 22, pp. 41–53, 2005.
- [2] *Revision of the commissions rules to ensure compatibility with enhanced 911 emergency calling systems*, Washington DC, RM-8143, CC Docket 94-102, U.S. FCC, 1996.
- [3] W. D. Wang and Q. X. Zhu, "RSS-based Monte Carlo localisation for mobile sensor networks," *IET Communications*, vol. 2, pp. 673–681, 2008.
- [4] A. Heinrich, M. Majdoub, J. Steuer, and K. Jobmann, "Real-time path-loss position estimation in cellular networks," in *Proceedings of International Conference on Wireless Networks (ICWN'02)*, Jun. 2002.
- [5] J. Shirahama and T. Ohtsuki, "RSS-based localization in environments with different path loss exponent for each link," *Vehicular Technology Conference*, pp. 1509–1513, 2008.
- [6] L. Mihaylova, D. Angelova, S. Honary, D. Bull, N. Canagarajah, and B. Ristic, "Mobility Tracking in Cellular Networks Using Particle Filtering," *IEEE Trans. Wireless Communications*, vol. 6, no. 10, pp. 672–681, 2007.
- [7] X. Wang, Z. Wang, and B. O'Dea, "A TOA-Based Location Algorithm Reducing the Error Due to Non-Line-of-Sight (NLOS) Propagation," *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 112–116, Jan. 2003.
- [8] J. J. Caffery and G. L. Stuber, *Wireless Location in CDMA Cellular Radio Systems*. Norwell, MA: Kluwer, 1999.
- [9] S. Sakagami, S. Aoyama, K. Kuboi, S. Shirota, and A. Akeyama, "Vehicle position estimates by multi-beam antennas in multipath environment," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 63–67, Feb. 1992.
- [10] R. Klukas and M. Fattouche, "Line-of-sight angle of arrival estimation in the outdoor multipath environment," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 342–351, Feb. 1998.
- [11] S. Venkatraman and J. J. Caffery, "Hybrid TOA/AOA techniques for mobile location in non-line-of-sight environments," *IEEE Wireless Communication Networking Conf, WCNC 2004*, vol. 1, pp. 274–278, 2004.
- [12] Y. Qi, H. Kobayashi, and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE trans. Wireless Commun.*, vol. 5, no. 3, pp. 672–681, 2003.
- [13] O. Sallent, R. Agusi, and X. Cavlo, "A mobile location service demonstrator based on power measurements," *Vehicular Technology Conference*, vol. 6, no. 1, pp. 4096–4099, Sep. 2004.
- [14] C. Takenga and K. Kyamakya, "Mobile positioning based on pattern matching and tracking approaches," *ISAST Transactions on Communications and Networking*, vol. 1, no. 1, pp. 42–51, 2007.
- [15] A. Taok, N. Kandil, S. Affes, and S. Georges, "Fingerprinting localization using ultra-wideband and neural networks," *Signals, Systems and Electronics*, vol. 54, no. 4, pp. 529–532, Aug. 2007.
- [16] M. Meurer, S. Heilman, D. Reddy, and P. Baier, "A signature based localization technique relying on covariance matrices of channel impulse responses," in *Proceedings of Workshop on Positioning, Navigation and Communication (WPNC)*, 2005.
- [17] M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen, "Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on wimax networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 283–294, Jan. 2010.
- [18] Cello consortium report. [Online]. Available: <http://www.telecom.ntua.gr/cello/documents/CELLO-WP2-VTT-D03-007-Int.pdf>
- [19] E. Trevisani and A. Vitaletti, "Cell-ID location technique, limits and benefits: an experimental study," *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*, pp. 51–60, 2004.
- [20] N. Deblauwe, "GSM-based Positioning: Techniques and Application," Ph.D. dissertation, Vrije Universiteit Brussel, 2008.
- [21] *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems*, IEEE Std. 802.16-2009, 2009.
- [22] K. Jones and L. Liu, "What where wi: An analysis of millions of wi-fi access points," *IEEE International Conference on Portable Information Devices*, May 2007.
- [23] K. Jones, L. Liu, and F. Alizadeh-Shabdiz, "Improving wireless positioning with look-ahead map-matching," *Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services*, Aug. 2007.
- [24] S. Byers and D. Kormann, "802.11b access point mapping," *Communications of the ACM*, vol. 46, no. 5, pp. 41–46, May 2003.
- [25] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.
- [26] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. NY, USA: Springer-Verlag, 1994.
- [27] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential simulation-based methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [28] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.
- [29] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Rev. Elec. Commun. Lab.*, vol. 16, pp. 9–10, 1968.
- [30] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, Aug. 1980.



Mussa Bshara received his Bachelor Degree in electrical engineering from Damascus University, Damascus, Syria, and the M.Sc. degree in signal processing and information security from Beijing University of Posts and Telecommunications (BUPT), Beijing. He obtained the Doctoral degree (Ph.D.) from the Vrije Universiteit Brussel (VUB), Brussels in 2010.

His research interests include localization, navigation and tracking in wireless networks, signal processing, wireless communications, power line communications and xDSL technologies.



Leo Van Biesen received the degree of Electro-Mechanical Engineer from the Vrije Universiteit Brussel (VUB), Brussels in 1978, and the Doctoral degree (PhD) from the same university in 1983. Currently he is a full senior professor. He teaches courses on fundamental electricity, electrical measurement techniques, signal theory, computer controlled measurement systems, telecommunication, physical communication and information theory.

His current interests are signal theory, PHY-layer in communication, time domain reflectometry, wireless communications, xDSL technologies, and expert systems for intelligent instrumentation. He has been chairman of IMEKO TC-7 from 1994-2000 and President Elect of IMEKO for the period 2000-2003 and the liaison Officer between the IEEE and IMEKO. Prof. Dr. Ir. Leo Van Biesen has been president of IMEKO from 2003 to September 2006, and he is now chairman of the Advisory Board of IMEKO as immediate Past-President. He is also member of the board of FITCE Belgium and of USRSI Belgium.



Umut Orguner received B.S., M.S. and Ph.D. degrees all in electrical engineering from Middle East Technical University, Ankara, Turkey in 1999, 2002 and 2006 respectively. Between 1999 and 2007, he was with the Department of Electrical and Electronics Engineering of the same university as a teaching and research assistant. Since January 2007 he has been working as a postdoctoral associate in Division of Automatic Control, Department of Electrical Engineering, Linköping University, Linköping, Sweden.

His research interests include estimation theory, multiple-model estimation, target tracking and information fusion.



Fredrik Gustafsson is professor in Sensor Informatics at Department of Electrical Engineering, Linköping University, since 2005. He received the M.Sc. degree in electrical engineering 1988 and the Ph.D. degree in Automatic Control, 1992, both from Linköping University. During 1992-1999 he held various positions in automatic control, and 1999-2005 he had a professorship in Communication Systems. His research interests are in stochastic signal processing, adaptive filtering and change detection, with applications to communication, vehicular, air-

borne, and audio systems. His work in the sensor fusion area involves design and implementation of nonlinear filtering algorithms for localization, navigation and tracking of all kind of platforms, including cars, aircraft, spacecraft, UAV's, surface and underwater vessels, cell phones and film cameras for augmented reality. He is a co-founder of the companies NIRA Dynamics and Softube, developing signal processing software solutions for automotive and music industry, respectively.

He was an associate editor for IEEE Transactions of Signal Processing 2000-2006 and is currently associate editor for EURASIP Journal on Applied Signal Processing and International Journal of Navigation and Observation. In 2004, he was awarded the Arnberg prize by the Royal Swedish Academy of Science (KVA) and in 2007 he was elected member of the Royal Academy of Engineering Sciences (IVA).