Yassine *et al. EURASIP Journal on Wireless Communications and Networking* 2014, **2014**:215
http://jwcn.eurasipjournals.com/content/2014/1/215

EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH **Open Access**

Hybrid positioning data fusion in heterogeneous networks with critical hearability

Ali Yassine[1*], Youssef Nasser[1*], Senior Member IEEE, Mariette Awad[1] and Bernard Uguen[2]

Abstract

In this paper, we propose and investigate a hybrid positioning data fusion technique for heterogeneous networks in critical transmission scenarios. The focus is on two scenarios: the small indoor scenario combining Wi-Fi and cellular systems and the small-to-mid-scale scenario composed of one located Mobile Terminal (MT) and one anchor node (AN). More specifically, we investigate the effect of the availability of three metrics i.e. the time of arrival (ToA), the angle of arrival (AoA), and the received signal strength-based fingerprint (RSS) on the positioning accuracy when the number of ANs is less than three. To combine these measurements, we use a 2-level unscented Kalman Filter (UKF) in conjunction with some advanced clustering techniques based on genetic algorithms. Simulation results show that the proposed hybrid data fusion technique outperforms the techniques presented in the literature independently of the transmission conditions.

Keywords: Hybrid positioning; Data fusion; Kalman filtering; Heterogeneous networks; Clustering; Genetic algorithms

I. Introduction

Today and future communications systems aim at providing high data rates with ubiquitous service coverage. The availability of location information at the base stations, i.e. its knowledge by the operators, becomes a key factor in today's communications systems for allowing new location-based services [1,2].

In practice, localization techniques are based on time of arrival (ToA) [3], time difference of arrival (TDoA) [4], received signal strength (RSS) [5], and angle of arrival (AoA) [6]. In outdoor scenarios, the Mobile Terminal (MT) position is obtained with high accuracy due to the Global Positioning System (GPS) or the standalone cellular systems. However, the main problem of GPS or cellular systems resides in lack of hearability situations such as indoor environment where the satellite or cellular signals are weak and in scenarios with deep shadowing effect [7]. Moreover, in homogeneous networks such as cellular networks, the estimation of the positioning information (PI) of any device becomes harder as the physical communications resources are more and more valuable.

A potentially good candidate for critical scenarios resides in the class of heterogeneous approaches that combine different radio access technologies (such as cellular systems, Wi-Fi, WiMAX). Indeed, cellular and Wi-Fi network-based localization techniques have recently received increasing interests in both the localization and communication community, e.g. [8-10]. This is not only because of the request made by the Federal Communications Commission (FCC) about the accurate localization of the MTs, but also because of many other applications that are location sensitive such as billing, fleet management, mobile yellow pages, etc. [11].

Even though all positioning techniques could be exploited in indoor scenarios and homogeneous networks however, in practice, there are limitations for combing these techniques as well as on the minimal number of anchor nodes (ANs) used [12,13]. The main challenge resides in the lack of hearability between the unlocated MT (UMT) and the ANs [12]. Indeed, in many cases, only one or two ANs are communicating with the UMT. Hence, new techniques based on hybrid data fusion should be proposed and analyzed in this extreme case. In literature, many techniques have been proposed [12,13]. In [12], the authors have proposed a positioning technique based on the combination of ToA and RSS fingerprinting by using one AN and one located

---

* Correspondence: ahy04@mail.aub.edu; yn10@aub.edu.lb
[1]ECE Department, American University of Beirut, Bliss Street, Beirut, Lebanon
Full list of author information is available at the end of the article

MT (LMT). However, the accuracy of the technique proposed in [12] is not high enough. In [13], the authors present a cooperative positioning technique based on the combination of long-range measurements obtained by three ANs and short-range measurements obtained by Wi-Fi. The main problem of this technique is that it requires measurements obtained by three ANs and a Wi-Fi hotspot.

In this paper, we focus on exploiting some positioning techniques, mainly ToA, AoA, and RSS fingerprinting, in an extreme scenario where only two ANs are available. This could be seen as one AN (base station for instance) and one LMT or one AN and one Wi-Fi hotspot available for positioning. It is respectively the case of indoor short-range or small-to-mid-scale transmission scenarios where the position of the LMT could be easily obtained via GPS [7]. First of all, we propose a hybrid positioning data fusion based on ToA, AoA, and RSS fingerprinting. Contrarily to [12], we propose to use the AoA as an additional input for resolving ambiguities. Then, based on an efficient combination of the contributions presented in [12] and [13], we extend the positioning scenario to its extreme case. In contrast with [13], we assume in our work that only two ANs are available for localization. However, as proposed in [13], we investigate the combination of short-range measurements, obtained via the LMT or via the Wi-Fi hotspot, with the long-range measurements to improve the accuracy of the positioning information. Finally, we investigate different clustering approaches for the hybrid data fusion in the heterogeneous context. The main objective of clustering is to improve the PI estimation by exchanging information with the necessary MTs within a cluster while keeping a reduced overhead cost.

The contribution of this paper is summarized as follows:

- Proposition of a hybrid positioning algorithm based on the combination of ToA, AoA, and RSS fingerprinting.
- Investigation of a critical positioning scenario where only two ANs are available. Heterogeneous networks are then proposed for localization purposes.
- Based on a 2-level unscented Kalman filter (UKF), a hybrid combination between short-range and long-range measurements is proposed to improve the localization accuracy.
- Proposition and investigation of hierarchical and genetic-based clustering algorithms. The main objective of the clustering algorithms is to exchange location information with the clusters peers in such a way to improve the positioning information while keeping a reduced overhead cost.

The remainder of this paper is organized as follows. Section II includes a summary of the literature review related to positioning techniques. Section III describes the problem statement of the scenario with a critical lack of hearability. In Section IV, we investigate the decoupling and coupling algorithm in the proposed positioning technique. Section V proposes and describes two clustering techniques used for positioning improvement. Simulation results are presented in Section VI while conclusions are drawn in Section VII.

II. Literature review

In this section, we briefly describe the basic standalone positioning techniques used in the context of homogeneous networks. Then, we present a short description of the UKF adopted in the next section for tracking purposes.

A. Location estimation techniques

Training sequences sent by BSs or MTs are used for the location estimation. Basically, BSs are considered as reference nodes (anchors) for localization. There are different main techniques used for localization, among them, we distinguish the ToA-, TDoA-, AoA-, and RSS-based fingerprinting. The ToA and TDoA techniques need at least three BSs for localization. The AoA technique requires a minimum of two BSs which means that the estimation error may be large, and the ambiguities of location estimation will exist if the number of the available BSs is less than the minimum requirement.

The ToA approach includes the calculation of the time needed by the signal to travel from the UMT to the ANs. Accordingly, the UMT will be moving on a circle of center given by the AN and with a radius d estimated through the ToA. Hence, to detect the exact location of the MT, at least three ANs are required. As depicted in Figure 1, the estimated position of the UMT is simply within the region of intersection (if it exists) of the drawn circles. It could be easily obtained through any filtering technique such as least square (LS) or weighted least square (WLS).

1) Time of arrival

2) Received signal strength-based fingerprinting

The RSS approach includes two main methods: the path loss lognormal shadowing model to deduce a trilateration and the fingerprinting [12]. Basically, the RSS-based fingerprinting, as shown in Figure 2, first collects RSS fingerprints of a scene and then estimates the location of the MT by matching online measurements with the closest possible location collected by measurements in a database [8]. Therefore, for each possible location, ambiguity points could exist leading then to high estimation errors in standalone positioning scenario.

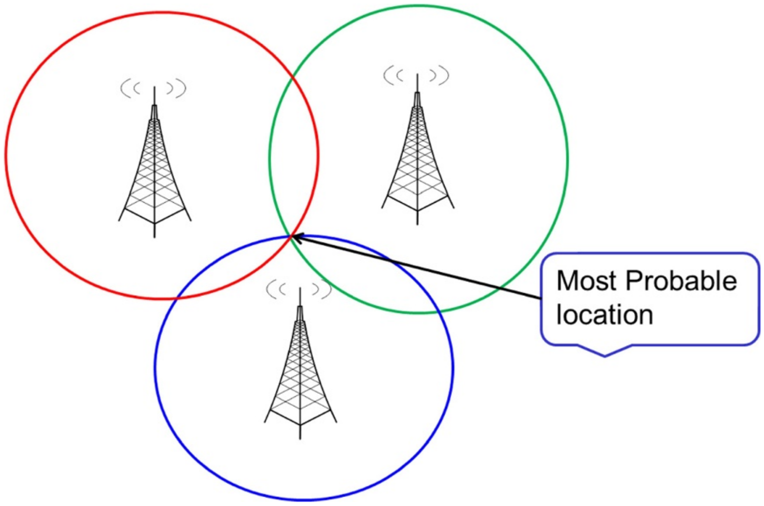

Figure 1 TOA ranging.

3) Angle of arrival

This technique includes the calculation of the angle at which the signal arrives from the UMT to the ANs [6]. Then, the region where the MT could exist can be drawn as shown in Figure 3. Basically, this region is a line having a certain angle with the ANs. Although at least two ANs are needed to estimate the location of the MT, the position estimation error could be large if a small error occurs in the AoA estimation. Therefore, an AoA-based technique is with of limited interest for positioning purposes.

B. 2-level Kalman filter (2LKF)

The paper presented by [13] introduces the 2-level Kalman filter (2LKF) as a solution for decoupling the relative localization of the users utilizing peer-to-peer *ad hoc* links from the absolute localization of the same users. Basically, the proposed algorithm in [13] analyzes the framework of positioning for cooperative schemes. This is mainly useful for scenarios with heterogeneous technologies. Thus, different timing behaviors of the channel measurement procedure can be handled for long-range and short-range technologies without any additional complexity. Hence, this algorithm is based on decoupling absolute localization that is recognized via the long-range cellular links and on relative localization that is recognized via the short-range *ad hoc* links [13].

Due to the strong non-linear behavior of the positioning measurements, UKF is highly recommended for such

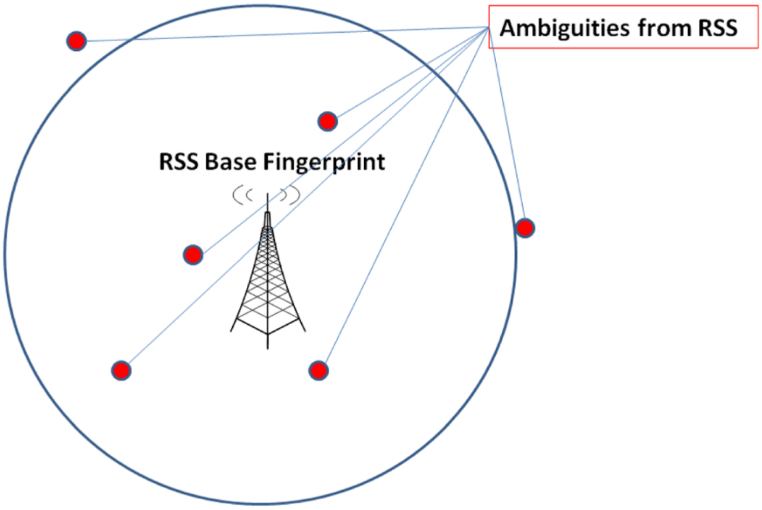

Figure 2 Ambiguities from RSS.

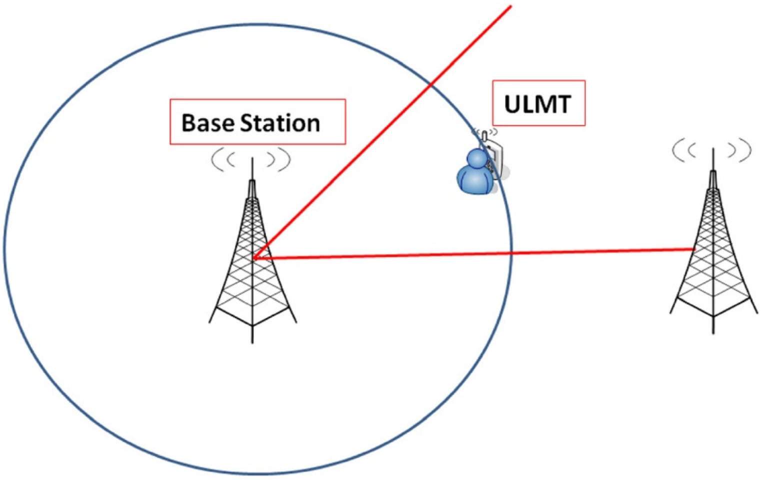

Figure 3 AOA model.

positioning and tracking scenarios. This filter is governed by a transition function and an observation function given by:

$$x_{t+1} = f(x_t) + q_t \tag{1}$$

$$y_t = h(x_t) + r_t \tag{2}$$

The transition function f is based on the state x_t at time t and the state x_{t+1} at time $t+1$. The relation between two states includes a transition noise q_t based on a predefined transition model. On the other hand, the observation function h relates between the observation y_t at time t, the state x_t, and the observation noise r_t at time t based on a predefined observation model. The observation function for short-range measurements is based on a path loss model. However, it is based on a combination between AoA and ToA models for long-range measurements. In UKF, q_t and r_t are normally distributed with zero mean and covariance Q_t and R_t, respectively.

III. Problem statement

It is well known in literature that three ANs are usually required to obtain suitable location accuracy. In GPS systems, this condition is almost satisfied since there is direct line-of-sight connection between the UMT and the different positioning satellites. However, when the UMT is in deep shadowing or indoor scenarios, the positioning information obtained through satellites will be lost and a terrestrial connection through the cellular system should be established in order to retrieve the location information.

The main problem of cellular systems however resides in the hearability restricted conditions of the UMT. With the necessary condition of three ANs, one could assume that the UMT is closer to one base station (BS - seen as AN) than others. In this case, the signal received from the other BSs will be very weak and might be interpreted as interference. Therefore, the UMT should search for some local solutions such as local Wi-Fi hotspot or cooperative communications through the interaction with a LMT to resolve the hearability conditions [14,15]. In this work, we adopt the positioning estimation scenario with restricted hearability conditions i.e. in indoor scenario or deep shadowing conditions. As depicted in Figure 4, we assume that the UMT is connected to one BS and one Wi-Fi hotspot or one LMT. We consider that all the estimated ToA and RSS fingerprinting [16] are collected at the home BS for centralized processing. We denote $a(a_x, a_y)$ as the true 2D location of the LMT or of the Wi-Fi hotspot, $u(u_x, u_y)$ as the true location of the UMT. Without loss of generality, the location of the home BS is set as O = (0, 0). Since the location of the LMT is obtained using GPS signals, this location information is imperfect. Denote $\hat{a} \triangleq (\hat{a}_x, \hat{a}_y)$ as the estimated location of the LMT using GPS. The estimated location is modeled as random with Gaussian distribution with variance σ^2_{GPS}. The Wi-Fi hotspot position is assumed to be known and without loss of generality; it could be modeled as of the LMT with a particular condition $\sigma^2_{\mathrm{GPS}} = 0$. As consequence, from now on, the LMT and the Wi-Fi Hotspot will be labeled as secondary anchor node (SAN) with coordinates $\hat{a}(\hat{a}_x, \hat{a}_y)$.

Because RSS fingerprintings are collected at the BS, ambiguities of the location estimation exist. In RSS-based fingerprinting, the ambiguities represent the possible locations given by the fingerprinting database of a MT, based on its measured RSS. The vector of ambiguities is represented as $S = [S_0, S_1, \ldots, S_M]$ while one point

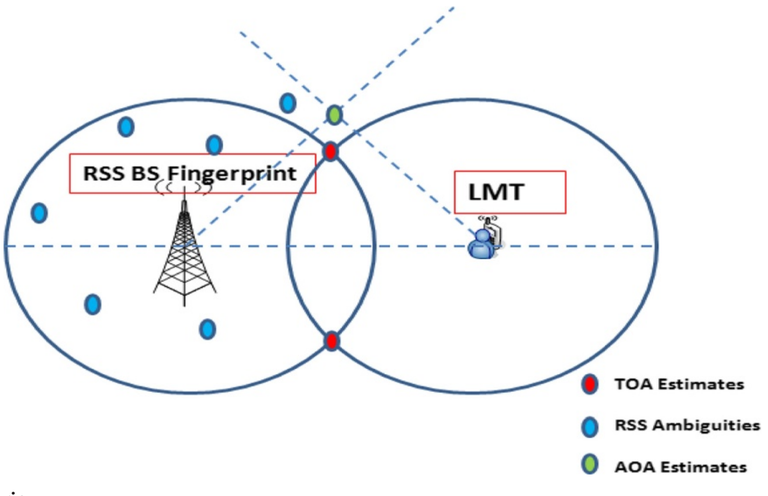

Figure 4 Positioning scenario.

is the exact position and M points represent the ambiguities. As given in [12], the (x,y) coordinates of the ambiguities are modeled as random processes with Gaussian distribution and with variance σ^2_{RSS}.

In this section, we propose to use the estimation algorithm studied in [12] as our basic coarse positioning algorithm. However, in contradiction with [12], we propose to use the AoA information to reduce the number of ambiguities analyzed through the RSS fingerprinting. The ToA-based distance estimation between the AN and the UMT could be modeled as:

$$\hat{d}_{\mathrm{BS}} = d_{\mathrm{BS}} + \omega_{\mathrm{TOA}} * \mathcal{N}(0,1) + \varepsilon_{\mathrm{NLOS}} \tag{3}$$

$$\hat{d}_{\mathrm{SAN}} = d_{\mathrm{SAN}} + \omega_{\mathrm{TOA}} * \mathcal{N}(0,1) + \varepsilon_{\mathrm{NLOS}} \tag{4}$$

where $\hat{d}_{\mathrm{BS}}$ is the estimated distance between the base station and the ULMT and d_{BS} is the true distance between the base station and the ULMT. Similarly, $\hat{d}_{\mathrm{SAN}}$ and d_{SAN} are respectively the estimated and true distances between the SAN and the ULMT. Also, ω_{TOA} is the ToA noise variance, and $\varepsilon_{\mathrm{NLOS}}$ is the error due to the non-line of sight (NLOS) following an exponential distribution with a probability distribution function (pdf) [17] of $p(b)$ that is given below [18].

$$p(b) = \begin{cases} \lambda e^{-\lambda b}, b \geq 0 \\ 0, b < 0 \end{cases} \tag{5}$$

where b denotes the NLOS error and $\mathbb{E}(b) = \frac{1}{\lambda}$.

Using ToA technique, we should be able to get the distances from different ANs to the UMT. Denote A and B as the two possible points provided by the ToA approach, one of which is the true position and the other one is an ambiguity point that refers to a wrong estimation. These solutions could be obtained by finding the intersection points of the two circles of respective centers O(0,0) and $\hat{a}(\hat{a}_x, \hat{a}_y)$ and of respective radius $\hat{d}_{\mathrm{BS}}$ and $\hat{d}_{\mathrm{SAN}}$. These intersection points could be easily written as [12]:

$$A = \begin{cases} x = \dfrac{y_{\mathrm{SAN}}\left(\gamma - \sqrt{(\gamma^2 - 4\beta\theta)}\right)}{2\beta x_{\mathrm{SAN}}} + \alpha \\ y = \dfrac{-\gamma + \sqrt{(\gamma^2 - 4\beta\theta)}}{2\beta} \end{cases}$$

$$B = \begin{cases} x' = \dfrac{y_{\mathrm{SAN}}\left(\gamma + \sqrt{(\gamma^2 - 4\beta\theta)}\right)}{2\beta x_{\mathrm{SAN}}} + \alpha \\ y' = \dfrac{-\gamma - \sqrt{(\gamma^2 - 4\beta\theta)}}{2\beta} \end{cases} \tag{6}$$

$$\alpha = \frac{d^2_{\mathrm{BS}} - d^2_{\mathrm{SAN}} + x^2_{\mathrm{SAN}} - y^2_{\mathrm{SAN}}}{2x_{\mathrm{SAN}}}$$

$$\begin{aligned} \beta &= 1 + \frac{y^2_{\mathrm{SAN}}}{x^2_{\mathrm{SAN}}} \\ \gamma &= -2\frac{y_{\mathrm{SAN}}}{x_{\mathrm{SAN}}}\alpha \\ \theta &= \alpha^2 - d^2_{\mathrm{BS}} \end{aligned} \tag{7}$$

where x_{SAN} is the abscissa of SAN, y_{SAN} is the ordinate of SAN, d_{BS} is the distance between the UMT and the base station, and d_{SAN} is the distance between the UMT and the SAN.

In order to solve for the most suitable location, the RSS fingerprinting ambiguity points could be used in combination with the ToA intersection points. The proposed solution could be obtained by taking the midpoint between the closest RSS fingerprint ambiguity point to one of the solutions A or B obtained in Equation 6, say

S. If the ToA error variances are available, then a weighted combination could be applied. In our approach, we propose to use the AoA information before exploiting the RSS fingerprinting. Indeed, as shown in [19], using the AoA estimation, we will be able to define the half-plane in which the possible A or B solution exists as shown in Figure 5.

IV. Coupling and decoupling algorithm

In [13], a 2-level Kalman filter was proposed to deal with short-range and long-range measurements and to come-up with a suitable positioning solution. The main problem of [13] is that the proposed technique was based on the long-range measurements using three BSs (in addition to the short-range measurements) from the cellular network to estimate the UMT position. Unfortunately, this scenario is quite ideal for investigation and the application of the 2-level Kalman filter would have a restricted impact. The main question would then be how much the Kalman filter and the combination of heterogeneous networks would improve the positioning estimates in critical scenarios. To deal with this scenario (Figure 4), we propose to firstly apply the coarse positioning estimation detailed in Section III and then the coupling-decoupling algorithm proposed in [13]. Contrarily to [13], we assume here that all UMT are moving with a velocity v. In addition, we assume that the non-linear RSS model is described by:

$$\hat{p}_{k|k-1^{(i,j)}} = \alpha - 10\beta \log\left(d^{(i,j)}\right) + \varepsilon_{\mathrm{NL,RSS}} \tag{8}$$

where $\hat{p}_{k|k-1^{(i,j)}}$ is the received power at the UMT j from the AN i, $d^{(i,j)}$ is the distance between the two nodes, α is a variable taking into account the shadowing effect, and β is the path loss exponent. $\varepsilon_{\mathrm{NL,RSS}}$ is a measurement error taking into account the mismatch between the path loss model and the real measurement. Contrarily to [13], we propose in this work to firstly decouple the relative localization, realized by using the short-range measurements obtained via the SAN, and the absolute localization, realized by using the long-range cellular by estimating the coordinates and orientation of the group in the cellular network. After decoupling the short- and long-range measurements, an iterative Kalman filter is applied on each measurement to improve the positioning and tracking estimates. Then, a coupling of the long- and short-range measurements is applied again to obtain the absolute positions. The proposed framework, as shown in Figure 6, runs in a cyclic way where iteration is performed when observations are available. Framework considers the following steps:

- Find the absolute coordinates of different MTs through the cellular network.
- Decouple into relative coordinates and center of mass coordinates through a transformation of coordinates.
  - Find the current transformation matrix (CTM) that corresponds to a translation followed by a rotation of the axis.
  - Assuming that the absolute and relative coordinates of MT i are defined by $x^{(i)}$ and $x^{(i)\mathrm{rel}}$ respectively and that $\mathrm{T}_{\mathrm{ctm}}$ is the CTM of the transformation of coordinates, we can write the following:

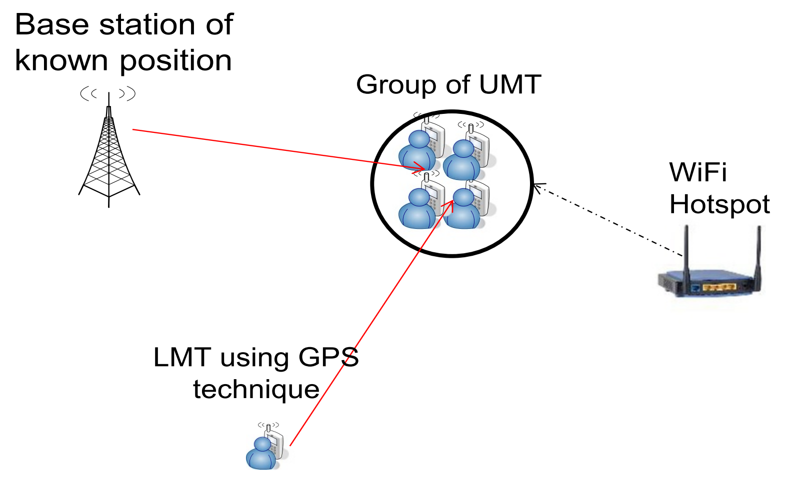

Figure 5 Combination of TOA, AOA, and RSS fingerprints.

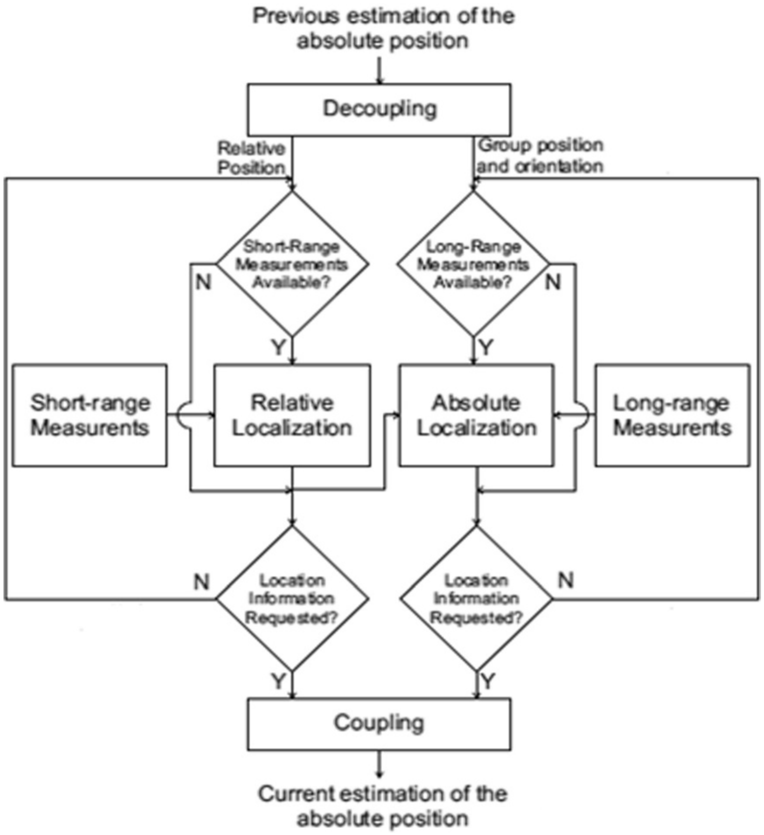

Figure 6 **Coupling and decoupling algorithm [13].**

$$\begin{bmatrix} x^{(i)\mathrm{rel}} \\ 1 \end{bmatrix} = T_{\mathrm{ctm}} \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} \tag{9}$$

Note that we added the last component of the vector on the right-hand side in order to perform transformations independent of $x^{(i)}$.

- $\mathrm{T}_{\mathrm{ctm}}$ is obtained by defining a translation equivalent to the absolute position $x^{(1)}$ of MT 1 (reference MT) followed by a rotation equivalent to the angle of the segment between MT 2 and MT 1 with respect to the absolute coordinate system ($\theta = \arctan\,(y^{(1,2)}/x^{(1,2)})$

$$T_{\mathrm{ctm}} = \begin{bmatrix} \dfrac{x^{(1,2)}}{d^{(1,2)}} & -\dfrac{y^{(1,2)}}{d^{(1,2)}} & 0 \\ y^{(1,2)}\,\substack{d^{(1,2)} \\ d\,:\,m} & \dfrac{x^{(1,2)}}{d^{(1,2)}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -x^{(1)} \\ 0 & 1 & -y^{(1)} \\ 0 & 0 & 1 \end{bmatrix} \tag{10}$$

knowing that $x^{(1,2)} = x^{(2)} - x^{(1)}$, $y^{(1,2)} = y^{(2)} - y^{(1)}$, and $d^{(1,2)}$ is the Euclidean distance between MT 1 and MT 2. Consequently, by this transformation we assumed that UMT 1 is the reference and MT 2 is on the x-axis of the relative coordinate system.

- Depending on whether available measurements are from short- or long-range technology:
  ° Single iteration of a stochastic filter for estimating relative localization.
  ° Single iteration of a stochastic filter for estimating the coordinates of the center of mass.
  ° Little iteration if it is the case of mobility where a MT is not static at all.
- Coupling estimations of relative and center of mass coordinates in order to have absolute estimators by doing the inverse of Equation 9.
- It is worth reminding that in our scenario, we assume that all UMTs are moving and then a recursive process is applied on the coupling and decoupling algorithm.

As the coupling and decoupling are applied on the center of mass of UMTs, it is then required to have a proper definition of this center for all or a set of the

MTs. Hence, we propose in this work to cluster the MTs into sets whose centers are to be efficiently defined. Therefore, the coupling and decoupling will be applied on few centers to keep the overhead cost reduced.

V. Proposed clustering algorithms

The main objective of this section is to describe and propose efficient clustering algorithms suitable for coupling and decoupling the short- and long-range measurements. Basically, the proposed clustering procedure is used to segregate MTs in such a way to improve the positioning estimation while reducing the overhead cost. In our work, we introduce three clustering techniques: the RSS-based clustering, the hierarchical clustering, and the genetic algorithm-based clustering.

A. RSS-based clustering

The RSS clustering algorithm is based on dividing the MTs into two clusters as shown in Figure 7. The first cluster center is the closest MT (strongest RSS) to the serving BS (or AN) and the second cluster center is the farthest MT (weakest RSS) from the serving BS.

B. Hierarchical binary clustering

The aim of hierarchical clustering is to build a hierarchy of clusters from a binary tree of the data that merges similar groups of points. Based on the long-range measurements, a matrix $D = (\{d_{i,j}\};\ i = 1, \dots, N;\ j = 1, \dots, N)$ of the distances between all MTs is formed. The first step in hierarchical clustering is to search for the pair of MTs that are the closest in terms of Euclidean distance. Then, a single linkage method based on the Euclidean distance between all the MTs is proposed and implemented. It is again based on the selection of MT pairs with the smallest distance. The point at which the pair of MTs is joined is called a node. Then, we repeat these steps over all MTs until we form a hierarchical binary tree (HBT) as shown in Figure 8. Basically, using this method, the distance between the merged pair and the other MTs will be the minimum distance of the pair in each case. For instance, if the distance between MT 2 and MT 1 is 5, while the distance between MT 3 and MT 1 is 6.5, we choose the minimum of the two, i.e. 5, to quantify the distance between (MT 2, MT 3) and MT 1. As a result, we will obtain the binary tree where at each step, two MTs are merged. Thus, for N MTs, we will obtain N-1 nodes. Finally, we are one step ahead from creating the clusters based on this hierarchical clustering. Mainly, we should specify a minimum and a maximum number of clusters. Then, we segregate the MTs into clusters ranging from the minimum till the maximum defined number. However, since in this work we impose a condition of at least two MTs per cluster (to perform the combination between the short-range and the long-range measurements), the condition on the maximum allowed number of clusters is removed.

Genetic clustering

In genetic-based algorithms, six main components have to be necessarily defined in order to establish the similarity with the wireless cellular networks. This includes:

- Genotype
- Population initialization
- Fitness function
- Selection operator
- Crossover operator
- Mutation operator

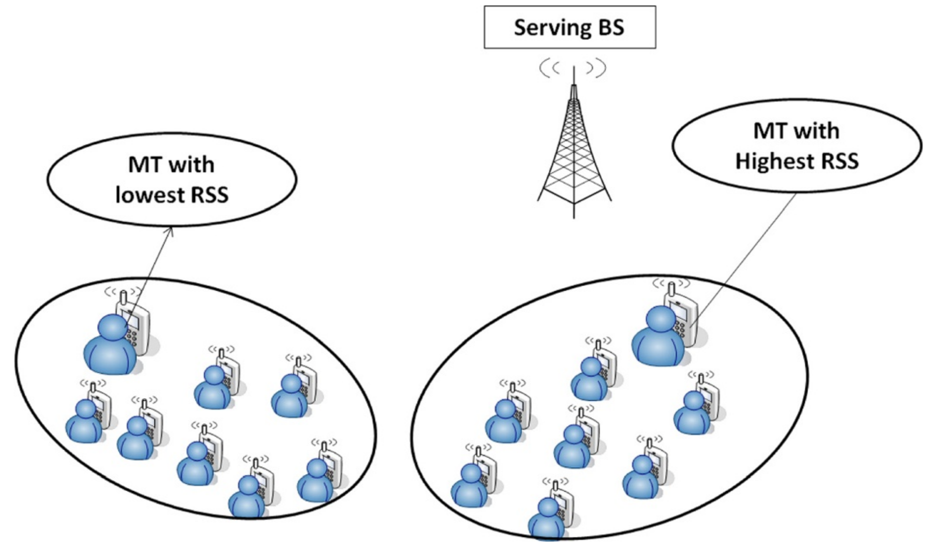

Figure 7 RSS clustering model.

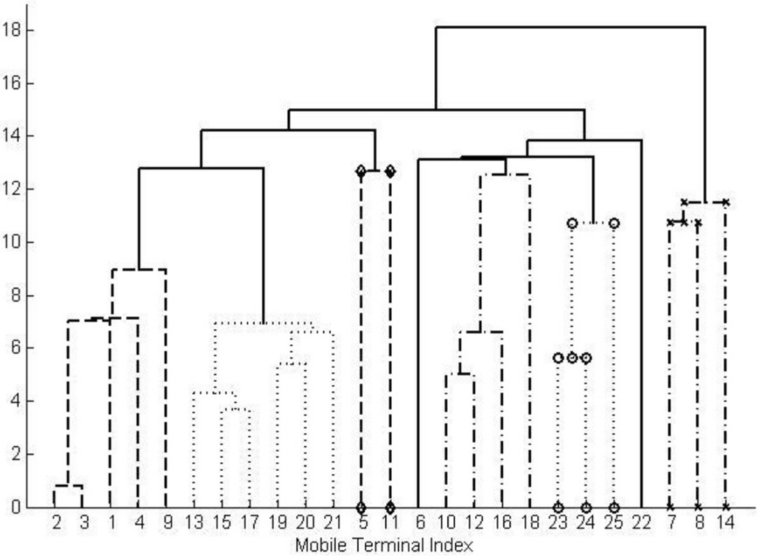

Figure 8 **Hierarchical binary tree.**

A genotype is the genetic makeup of a cell, an organism, or an individual [20]. By making an analogy from genetics to cellular networks, the chromosome of N genes represents a cell in cellular network with N MTs. Hence, a gene in a chromosome is mapped to a position of a mobile user in a cell. Secondly, the initialization of a population in genetics can also be mapped on cellular networks in such a way that the position of each UMT is selected randomly in the studied area. The fitness function is a third factor in genetic algorithms. It describes the objective function designed to satisfy some conditions such as minimizing the mean square error. In [20], Azimi et al. introduced an algorithm with two stages of fitness functions: the intra-cluster fitness and the extra-cluster fitness. Thus, the final fitness value is calculated by subtracting *Intra-Clstr-Fit* from *Extra-Clstr-Fit*. The selection operator in genetics process selects individuals from the mating pool directed by the survival of the fitness concepts of natural genetic systems. Finally, the crossover operator in genetics represents a probabilistic process that generates at least two child individuals by exchanging information between two parent individuals. Therefore, the pathway of the analogy from genetics to cellular network consists of the following main points:

- The genotype is a kind of representation for the mobile users in the cellular network.
- The population initialization for cellular network could be the initialization of the mobile users in the network by defining their positions.
- The fitness function proposed by Azimi et al. [20] could be applied on cellular networks, or another metric could be used such as RSS indicator or Euclidean distance metric.

The selection and crossover operators are the main operators required for aggregating MTs into an optimal number of clusters and finding the most probable solution. In the following, we describe the genetics-based algorithm applied in the heterogeneous context for positioning purposes. The flow chart in Figure 9 presents the proposed genetic algorithm for clustering once an initial clustering technique has been applied.

As shown in Figure 9, the entire genetic algorithm is run over many generations. This will give us the maximum possible number of clusters with at least two MTs per cluster. At the end of each generation, we select the chromosome that has the highest fitness value. Finally, at the end of all the generations, we select the generation with the highest number of clusters and the chromosome with the highest fitness. Now, we will redefine the six basic genetic components to be compatible with our wireless cellular system.

1) Genotype

A chromosome is a set of genes in genetics. In our representation, the gene is a cluster center; hence, the chromosome is a set of cluster centers. Therefore, the total number of clusters determines the size of a chromosome. An initial clustering method has to be applied on the MTs in order to proceed with the genetic algorithm. We can use one of the two clustering methods discussed

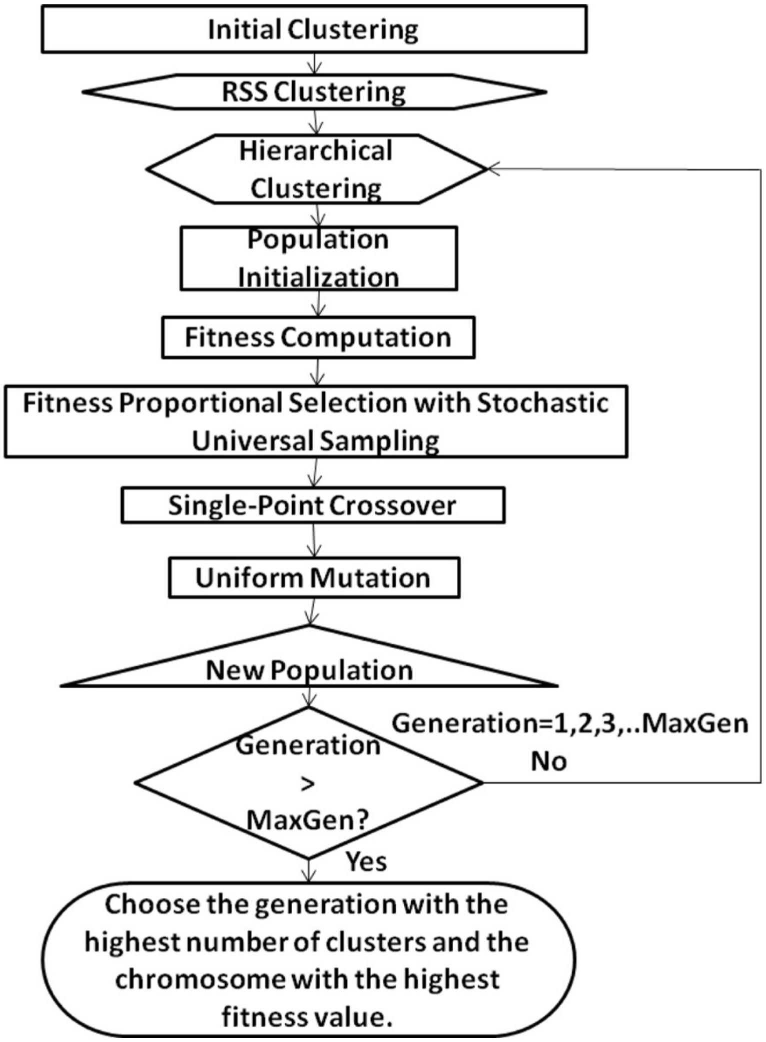

Figure 9 Genetic clustering algorithm.

before i.e. the RSS clustering and the hierarchical clustering.

2) Population initialization

The chromosome is a set of cluster centers, and the population is a set of chromosomes. Hence, the population in our case is a set of all different combinations of the cluster centers. For instance, in the case of three clusters having five, three, and four MTs respectively as cluster members, we will have $5 \times 3 \times 4 = 60$ different combinations of cluster centers. Hence, the population contains 60 chromosomes and the maximum population size will be 60.

3) Fitness function

Once an initial clustering is applied, the genetics-based algorithm requires the definition of the fitness function. In our work, we propose a clustering metric μ_i for each cluster 'i' defined by:

$$\mu i = \sum_{xj \in Ci} ||xj - zi|| \tag{11}$$

where x_j is cluster member, z_i is the cluster center, and $||\ ||$ is the Euclidian distance. Then, we define μ as the sum of the Euclidian distances of all clusters by:

$$\mu = \sum_{i=1}^{K} \mu_i \tag{12}$$

where K is the total number of clusters. The fitness function is defined by:

$$\mathrm{Fit} = \frac{1}{\mu} \tag{13}$$

In our clustering approach, we aim at maximizing the fitness function or equivalently maximizing the Euclidian distance among clusters. The fitness function plays a central role in the clustering process.

4) Selection operator

In this paper, we use the fitness proportionate selection (FPS) as the selection operator. With the FPS, the probability of an individual to be chosen increases (respectively decreases) when the fitness of the individual is greater (respectively less) than the fitness of the competitor. Knowing that a chromosome is a set of cluster centers and the fitness function is the sum of the Euclidean distance between cluster members and the cluster center among all clusters, we select the chromosome with the highest fitness function.

5) Crossover operator

We pregenerate two 'repositories' of random binary digits based on the coarse estimated positions from which the masks used in crossover will be picked up. In this work, we first use a single point crossover. Then, we copy the binary string from the beginning of chromosome to the crossover point from one parent, and the rest is copied from the second parent.

6) Mutation

The masks used in mutation are picked up from the pregenerated two 'repositories' of the random binary digits. Then, a 'xor' operation is implemented between the population and the new mask. This mutation will lead to the initiation of a new population with the same size but different binary representations, and that is the aim behind mutation. After the mutation, a new population is evolved. The population size is originally determined by the multiplication of the number of MTs in each cluster. Thus, we apply a demapping from binary to decimal. The maximum value obtained in the demapping will be the new population size. Then, a factorization will be applied as it leads to the new maximum number of the clusters. For instance, for a population size of 90, the factorization will result into $2 \times 3 \times 3 \times 5 = 90$. Hence, the new maximum number of the clusters will be 4.

VI. Simulation results

The aim of this section is to evaluate the hybrid positioning technique in parallel with the proposed clustering approaches. We use MATLAB© to develop the algorithms and run simulations. We consider $N = 50$ UMTs, uniformly distributed in a cellular area with 1 km of cell radius placing the BS (equivalently the AN) at position (0,0) and the SAN at ($\sqrt{3}$, 1) km or ($\sqrt{3}$, − 1) km. The BS with position O(0,0) will be used as a reference. Without loss of generality, we assume that all MTs are uniformly distributed around a cluster head, assumed to be the first MT, within a range radius of 50 m. The simulation was run for 3,000 realizations. The number of iterations of the UKF is equal to 2 while the number of observations is equal to 5. It is worth mentioning that the positioning estimation could be improved if the number of iterations is increased. All results are given in terms of mean square error (MSE) measured in m^2.

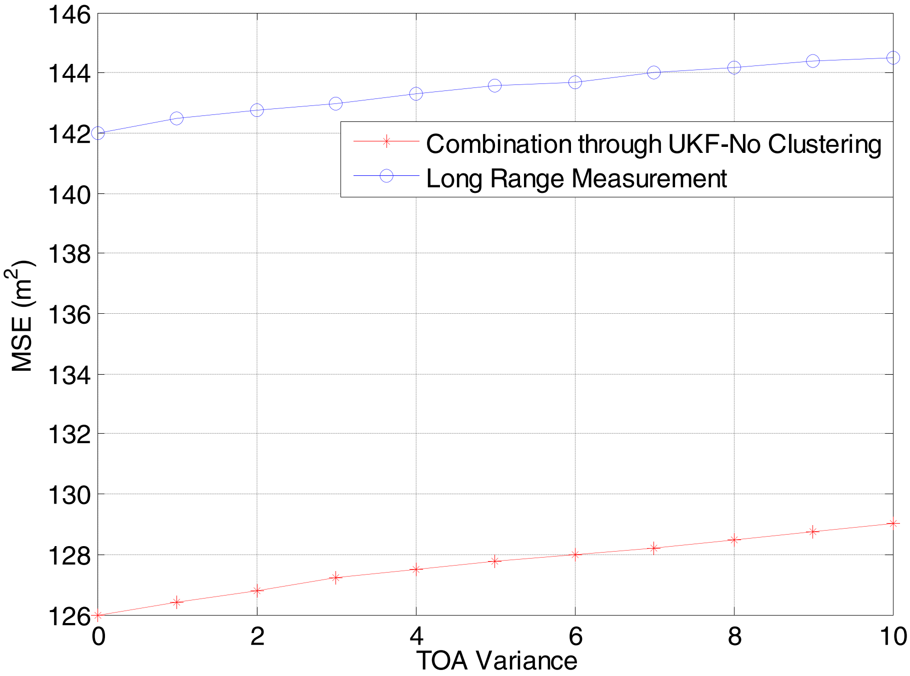

Figure 10 No-clustering: MSE vs. TOA-variance.

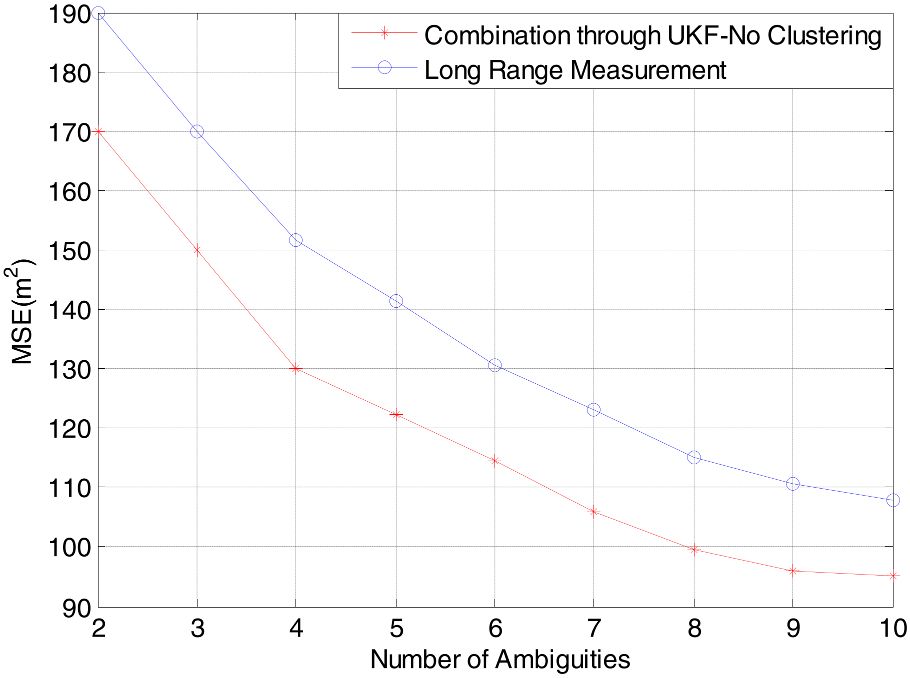

Figure 11 No-clustering: MSE vs. number of RSS ambiguities.

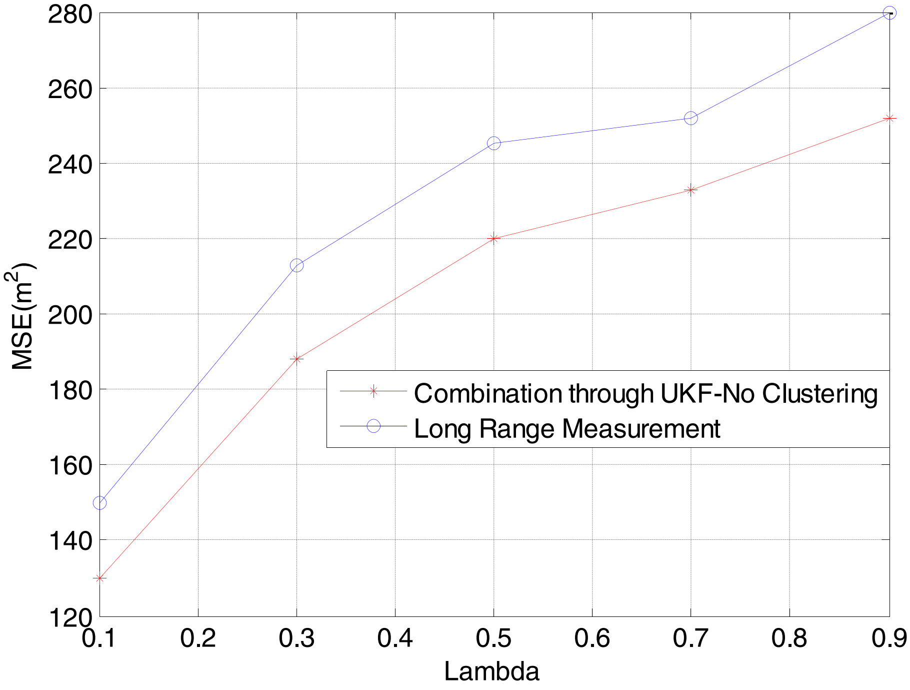

Figure 12 No-clustering: MSE vs. lambda.

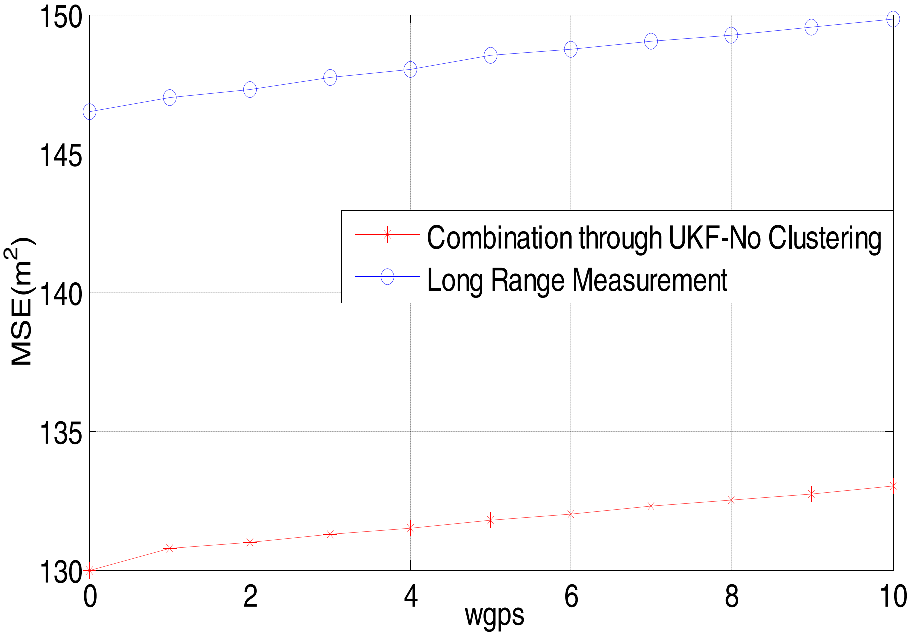

Figure 13 No-clustering: MSE vs. GPS error.

First, we calculate the distance (error) squared between the estimated position and the true position. Then, we take the average of this error over the number of observations taken at a certain realization. Finally, we calculate the average of the resulting error over different numbers of realizations.

Figures 10, 11, 12, 13, and 14 show that, independently of the parameter, the proposed hybrid data fusion using UKF coupling/decoupling significantly outperforms the stand-alone long-range measurements. In all these figures, no clustering has been applied. Figure 10 shows the improvement made by UKF in combining long-range

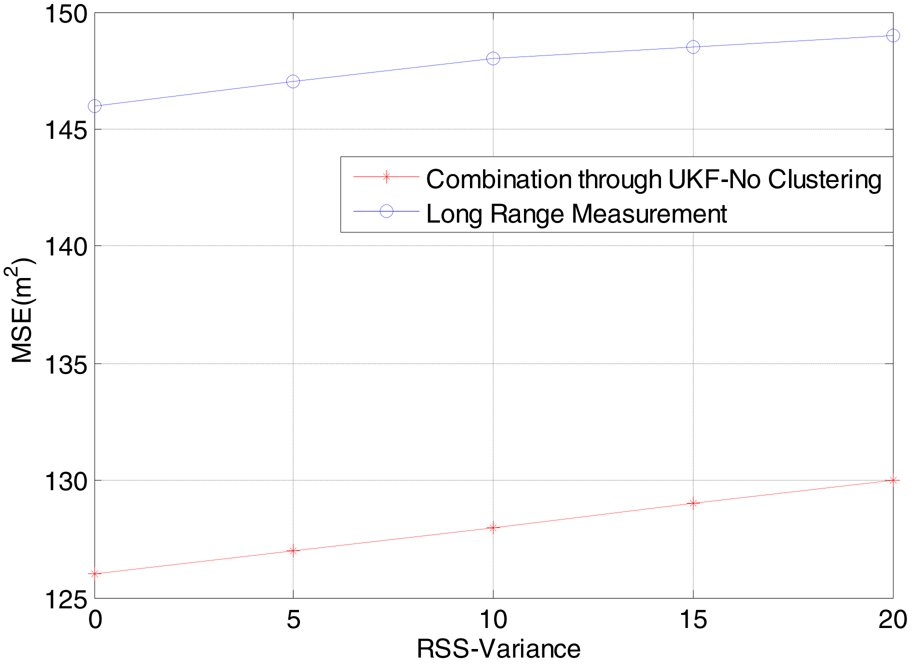

Figure 14 No-clustering: MSE vs. RSS variance.

and short-range measurements just by considering five observations per iteration. Figure 11 presents the effect of the number of ambiguities obtained from RSS fingerprints on the performance of the positioning algorithm. As expected, it is clear that higher accuracy is obtained when the number of the RSS ambiguities increases. This is true for both long-range measurements and the combination through UKF. In both figures, a noticeable positioning improvement is observed. Figure 12 explores the effect of the NLOS parameter λ on the accuracy of our positioning algorithm. λ is inversely proportional to accuracy. As shown in this figure, the MSE strictly increases from 150 till 280 for long-range measurements and from 130 till 250 for the combination, when λ increases from 0 till 0.9. Thus, the improvement made by the combination using UKF is still preserved.

Figure 13 explores the effect of the error obtained by GPS measurements. As the GPS error variance increases from 0 till 10, the MSE for the long-range measurements and for the combination slightly increases. This means that the utilization of a LMT as a SAN does not really affect the positioning estimation. Similar conclusions could be drawn from Figure 14.

Figures 15, 16, and 17 present the effect of clustering on the obtained results. It is clear that, independently of the clustering algorithm, some gain is obtained in the simulated scenario. Figure 15 shows that the RSS clustering provides a little improvement over the no-clustering case. This is compatible to our expectations due to the fact that RSS clustering leads to the formation of two clusters. That explains our motivation behind implementing a clustering technique based on advanced approach.

Figure 16 shows that the hierarchical clustering provides a significant improvement over the RSS clustering and over the no-clustering case. This improvement is due to the fact that the hierarchical clustering leads to the formation of higher number of clusters. Compared with Figure 10, the MSE is reduced from around 125 till 82 m^2 while the RSS clustering approach presents a MSE around 115 m^2.

Figure 17 shows the performance of the genetic-based clustering. It is clear that, as for the hierarchical clustering, it presents a large improvement. Nevertheless, this gain is comparable to that obtained through the hierarchical clustering. This is due to the fact that the genetic-based clustering implicitly utilizes hierarchical clustering. However, the added value provided is that genetic-based clustering generates the highest possible number of clusters whose cluster centers attained the highest fitness value. However, for larger number of MTs, we expect a significant improvement made by the genetic-based clustering as it leads to the formation of larger number of clusters than that generated by the hierarchical clustering. Moreover, cluster centers will be chosen in a better way since they are based on the maximization of the cluster centers.

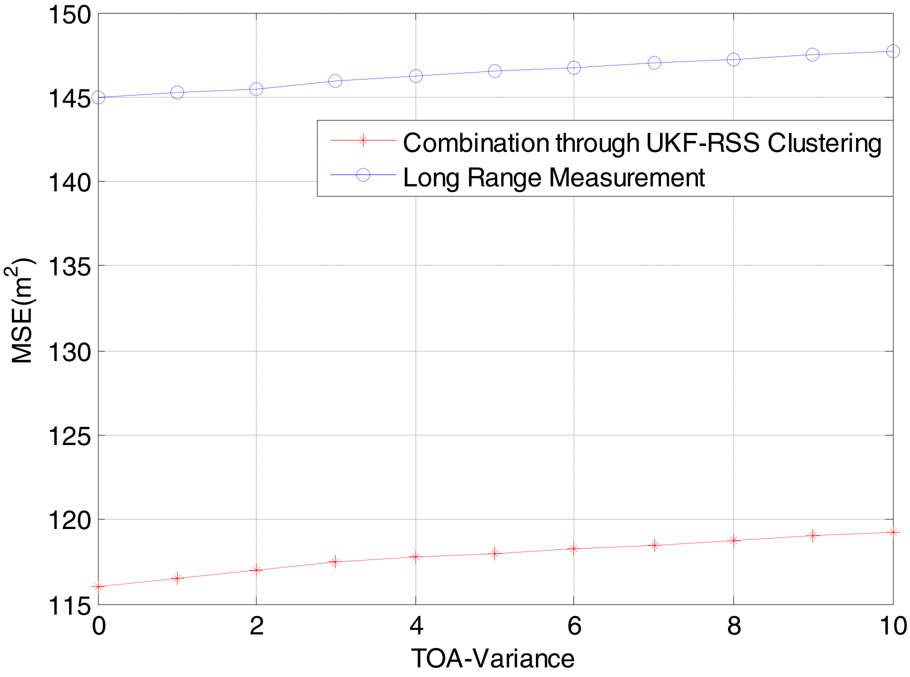

Figure 15 RSS clustering: MSE vs. TOA-variance.

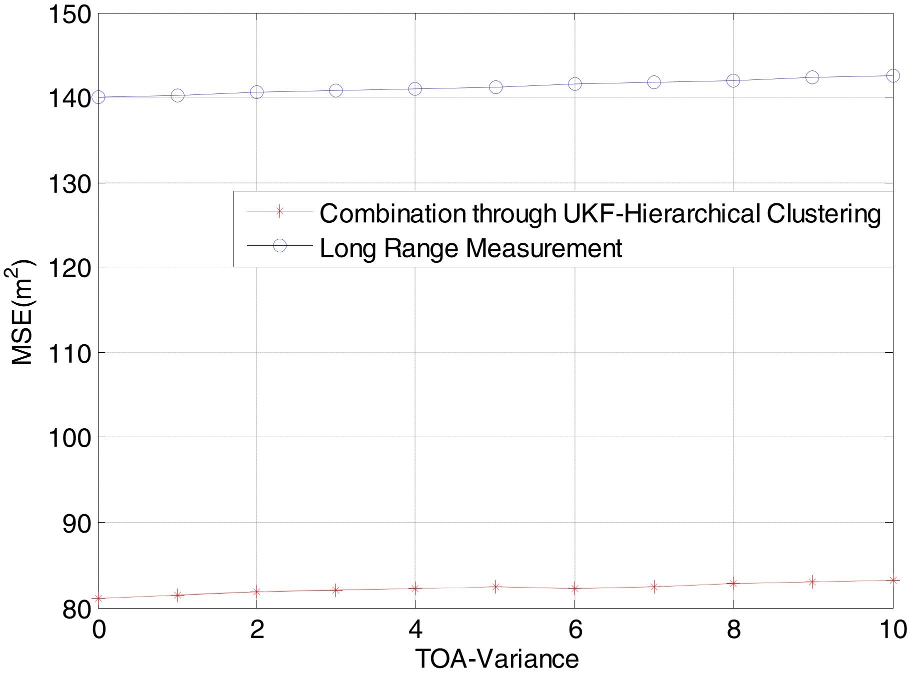

Figure 16 Hierarchical clustering: MSE vs. TOA-variance.

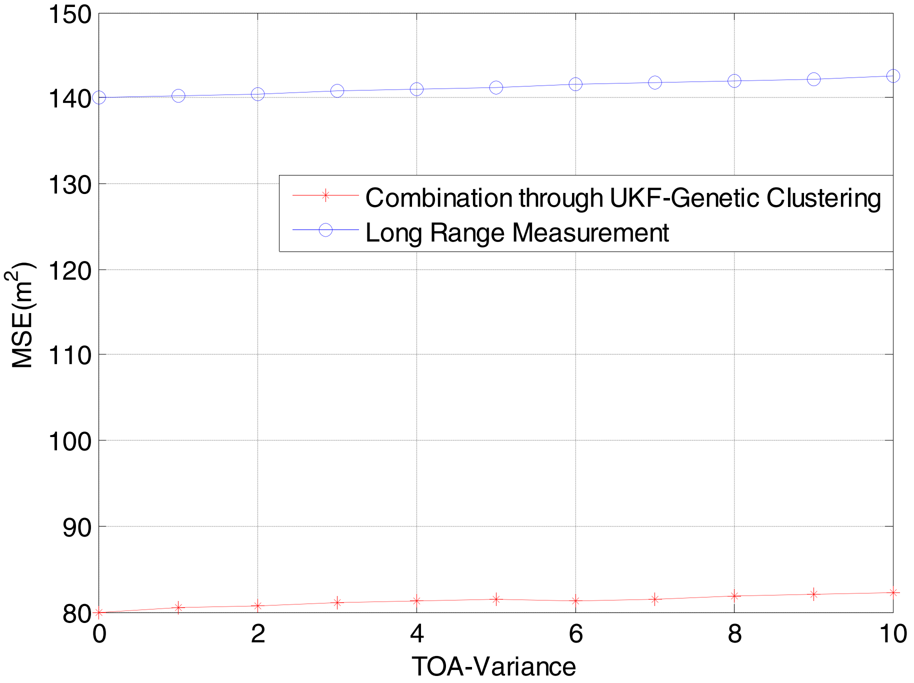

Figure 17 Genetic clustering: MSE vs. TOA-variance.

VII. Conclusions

In this paper, we have considered a particular positioning scenario with lack of hearability or with a limited number of ANs. We have presented a hybrid localization approach based on data fusion obtained from the combination of the ToA-, AoA-, and RSS-based fingerprinting techniques. Simulations have shown that the proposed hybrid approach outperforms the stand-alone ToA and RSS fingerprinting techniques in this critical transmission scenario despite the fact that they are less complex than the proposed hybrid algorithm. Based on a 2-level UKF, we have shown that significant positioning estimations could be obtained. Moreover, hierarchical and genetic-based clustering approaches have been proposed and combined with the UKF to improve the accuracy of the estimation algorithm. Simulation results have shown the outperformance of our proposed algorithm with respect to the standalone techniques proposed in the literature.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]ECE Department, American University of Beirut, Bliss Street, Beirut, Lebanon. [2]Institute of Electronics and Telecommunications of Rennes, University of Rennes 1, Rennes, France.

Received: 7 April 2014 Accepted: 13 November 2014
Published: 7 December 2014

References

1. A K¨upper, *Location-based services* (Wiley, New York, 2005)
2. AH Sayed, A Tarighat, N Khajehnouri, Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. IEEE Signal Process. Mag. **22**, 24–40 (2005)
3. J Caffery, GL Stuber, Subscriber location in CDMA cellular networks. IEEE Trans. Veh. Technol. **47**(2), 406–416 (1998)
4. Y Zhao, Standardization of mobile phone positioning for 3G systems. IEEE Commun. Mag. **40**, 108–116 (2002)
5. AJ Weiss, On the accuracy of a cellular location system based on RSS measurements. IEEE Trans. Veh. Technol. **52**(6), 1508–1518 (2003)
6. L Cong, W Zhuang, Hybrid TDOA/AOA mobile user location for wideband CDMA cellular systems. IEEE Trans. Wirel. Commun. **1**(3), 439–447 (2002)
7. FL Piccolo, A new cooperative localization method for UMTS cellular networks, in *Proc. IEEE Global Telecommun. Conf*, 2008
8. C Mensing, S Sand, A Dammann, W Utschick, Interference aware location estimation in cellular OFDM communications systems, in *Proc. IEEE Int. Conf. Commun*, 2009
9. C Mensing, S Sand, A Dammann, W Utschick, Data-aided location estimation in cellular OFDM communications systems, in *Proc. IEEE Global Telecommun. Conf*, 2009
10. 3GPP TS 36.305, *Stage 2 functional specifications of UE positioning in E-UTRAN*, 2009. http://www.3gpp.org
11. AH Sayed, A Tarighat, N Khajehnouri, Network-based wireless location. IEEE Signal Process. Mag. **22**(4), 12–23 (2005)
12. H Ziming, M Yi, R Tafazolli, A hybrid data fusion based cooperative localization approach for cellular networks, in *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, 2011, pp. 162–166. doi:10.1109/IWCMC.2011.5982409
13. J Figueiras, S Frattasi, H-P Schwefel, Decoupling estimators in mobile cooperative positioning for heterogeneous networks, in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, 2008, pp. 1–5
14. C Mayorga, F Rosa, S Wardana, G Simone, M Raynal, J Figueiras, S Frattasi, Cooperative positioning techniques for mobile localization in 4G cellular networks, in *Proceedings of the IEEE International Conference on Pervasive Services (ICPS 2007)*, 2007, pp. 39–44
15. S Frattasi, M Monti, R Prasad, Cooperative mobile user location for next-generation wireless cellular networks, in *Proceedings of IEEE International Conference on Communications, 2006. ICC '06. vol. 12*, 2006, pp. 5760–5765
16. G Sun, J Chen, W Guo, KJR Liu, Signal processing techniques in network-aided positioning. IEEE Signal Process. Mag. **22**(4), 12–23 (2005)
17. Q Cui, T Xiaofeng, Z Ping, A novel location model for 4G mobile communication networks, in *2007 IEEE 66th Vehicular Technology Conference*, 2007
18. L Cong, W Zhuang, Non-line-of-sight error mitigation in TDOA mobile location, in *Proceedings of the IEEE Global Telecommunications Conference. GLOBECOM'01, vol. 1*, 2001, pp. 680–684
19. A Yassin, M Awad, Y Nasser, On the hybrid localization in heterogeneous networks with lack of hearability, in *Telecommunications (ICT), 2013 20th International Conference* (IEEE, Morocco, 2013), p. 1. doi:10.1109/ICTEL.2013.6632158
20. R Ghaemi, N bin Sulaiman, H Ibrahim, N Mustapha, A review: accuracy optimization in clustering ensembles using genetic algorithms. Artif. Intell. Rev. **35**(4), 287–318 (2011)

doi:10.1186/1687-1499-2014-215
Cite this article as: Yassine *et al.*: **Hybrid positioning data fusion in heterogeneous networks with critical hearability.** *EURASIP Journal on Wireless Communications and Networking* 2014 **2014**:215.