

# Position and Movement Detection of Wireless Sensor Network Devices Relative to a Landmark Graph\*

Keyong Li,<sup>‡</sup> Dong Guo,<sup>§</sup> Yingwei Lin,<sup>¶</sup> and Ioannis Ch. Paschalidis<sup>†</sup>

## Abstract

We present a novel probabilistic framework for reliable indoor positioning of mobile sensor network devices. Compared to existing approaches, ours adopts complex computations in exchange for high localization accuracy while needing low hardware investment and moderate set-up cost. To that end, we use full distributional information on signal measurements at a set of discrete locations, termed *landmarks*. Positioning of a mobile device is done relative to the resulting landmark graph and the device can be found near a landmark or in the area between two landmarks. Key elements of our approach include profiling the signal measurement distributions over the coverage area using a special interpolation technique; a two-tier statistical positioning scheme that improves efficiency by adding movement detection; and joint clusterhead placement optimization for both localization and movement detection. The proposed system is practical and has been implemented using standard wireless sensor network hardware. Experimentally, our system achieved an accuracy equivalent to less than 5 meters with a 95% success probability and less than 3 meters with an 87% success probability. This performance is superior to well-known contemporary systems that use similar low-cost hardware.

## Index Terms

\* Research partially supported by the NSF under grant EFRI-0735974, by the DOE under grant DE-FG52-06NA27490, and by the ODDR&E MURI10 program under grant N00014-10-1-0952.

<sup>‡</sup> Center for Information & Systems Eng., Boston University, e-mail: [likeyong@ieee.org](mailto:likeyong@ieee.org).

<sup>§</sup> Center for Information & Systems Eng., Boston University, e-mail: [dong.dongguo@gmail.com](mailto:dong.dongguo@gmail.com).

<sup>¶</sup> Center for Information & Systems Eng., Boston University, e-mail: [yingwei@bu.edu](mailto:yingwei@bu.edu).

<sup>†</sup> Corresponding author. Dept. of Electrical & Computer Eng., Division of Systems Eng., and Center for Information & Systems Eng., Boston University, 8 St. Mary's St., Boston, MA 02215, e-mail: [yannisp@bu.edu](mailto:yannisp@bu.edu), url: <http://ionia.bu.edu/>.

Wireless sensor networks, localization, probabilistic profiling, hypothesis testing, optimal deployment.

## I. INTRODUCTION

The demand for reliable indoor positioning of mobile devices is well documented and widespread. Using such a functionality, businesses can better manage their equipment and personnel; museums can provide automatically guided tours and enhance visitor experiences; hospitals can track their patients and give them more freedom; rescue workers can navigate through a disaster site more easily; and security agencies can strengthen the protections of critical assets such as nuclear and biochemical materials.

It is worth noting that technologies for outdoor positioning are relatively mature. These include mostly the GPS technology that is widely used today, but also technologies using the cellular network (see [1], [2] for the latter). Yet, translating these successes to indoor environments is far from straightforward. First, the GPS technology is hardly operational indoors due to heavy signal attenuation by the building structures. On the other hand, the cellular-based technologies for outdoor use cannot produce satisfactory positioning accuracy when confronted with the rich effects of the indoor environment on the signals.

A wide range of indoor positioning solutions have been proposed, with varying degrees of success. [3] implemented an infrared-based positioning system (*Active Badge*) for low-accuracy applications. [4] proposed an ultrasound-based system (*Cricket*) that achieved high accuracy, but the system requires the installation of a dense network of ultrasound beacons. In the wireless positioning camp, many types of signal measurements can be useful, ranging from the most basic — *Received Signal Strength Indication (RSSI)* — to more sophisticated ones including signal phase, time-of-arrival (TOA), angle-of-arrival (AOA), and multipath components (MPC) [5], [6].

Among the RSSI-based approaches, the methods in [7], [8], [9], [10] compare mean RSSI measurements to a pre-computed signal-strength map. In particular, [7] tested two methods. The first one uses a map of mean signal strength that is profiled offline, while the second assumes a signal propagation model taking into account how many walls are in the path. The former was shown to be superior, achieving an accuracy of 75% error  $< 5$  meters and 50%  $< 3$  meters. This system and others succeeded in demonstrating the feasibility of meaningful positioning services

using wireless sensor networks and injected enthusiasm into the field. Their performance, though, leaves room for improvement. Many other works followed. [11] improved upon [7] by taking the probabilistic nature of the problem into account. [8] proposed a method based on mean RSSI profiling, but used power-level diversity to achieve an accuracy of  $80\% < 3$  meters. However, further improvement of the performance within this framework (e.g., by increasing the density of “reference signatures”) seems unlikely. Another class of systems such as in [12], [13] use stochastic triangulation techniques but rely on a path loss model, thus, introducing a modeling error. In addition to the evidence provided by [7], our earlier related work based on signal strength profiling [14], [15] has also been shown to reduce the mean error distance compared to stochastic triangulation. References to many other systems can be found in [16].

Despite the rich literature, some fundamental questions remain. On one hand, there have been some efforts to understand the theoretical limit of wireless-signal-based device positioning. For example, [17] obtained a Cramer-Rao bound of wireless-based positioning, but the result was built upon the presumption that positioning is done using distance and angle of arrival measurements, and these measurements follow Gaussian distributions. A wider range of techniques were considered by [18], but similar assumptions (especially the Gaussian assumption) were put in place. Such assumptions might be overly simplifying in the indoor positioning context. On the other hand, most positioning systems proposed to date have shunned away from sophisticated decision rules. To a certain extent, our work suggests that by careful algorithm design that enables distributed processing, it is possible to implement complex decision rules to achieve robust and accurate positioning using existing hardware.

Due to the rather diverse applications of indoor positioning and the rapid evolution of the hardware capabilities found in wireless sensor network devices, the positioning problem should not be considered in just a single context. Instead, it is meaningful to explore a spectrum of techniques that reflect different constraints and trade-offs of hardware investment, computational complexity, set-up cost, and positioning accuracy. In particular, the present paper strives to achieve high accuracy with low hardware investment and moderate set-up cost, at the cost of fairly sophisticated computations. In particular, our approach does not require advanced signal measurements such as angle and time of arrival (AOA and TOA), and the amount of computation needed is kept under a threshold such that contemporary hardware suffices. One can argue that we treat computational power as a constraint and seek to minimize set-up cost and maximize

accuracy.

Our contribution goes beyond a successful positioning system with attractive experimental results. The approach differs from what has been considered in the literature and is based on a set of formal techniques that result in analytical performance guarantees. Positioning is done relative to a *landmark graph*, with node set a chosen set of landmarks, and edges defined between any two nodes corresponding to landmarks in contiguous geographical areas. The device's position is mapped either to a node of the landmark graph, if the device is in its vicinity, or to an edge if the device is in the area between two landmarks. The following topics are then covered:

A: We construct appropriate probabilistic descriptors associated with a device's position from a limited amount of RSSI measurements. This process is commonly known as *profiling*. The descriptors that we construct go beyond mean values and variances to record the shapes of the measurement distributions. Moreover, the present paper associates a parametrized family of probability density functions (pdfs) to each location, which introduces some analytical challenges, but proved to produce more robust performance. Our experimental results show that such information is of significant value to performance.

By “a limited amount” of measurements, we mainly mean that one can directly measure the RSSI distributions only for a finite number of spots during the profiling phase. Without much loss of generality, we assume that these spots coincide with the landmarks. In order to construct the probabilistic descriptors of the device's position in the contiguous space, we adopt a pdf interpolation technique originating from the completely different field of statistical physics simulation [19]. It is worth noting that we have not seen the explicit mentioning of pdf interpolation in the current localization literature. To be precise, practices that are equivalent to some form of interpolation are not foreign, but these have been carried out rather implicitly without formal design and evaluation. Thus, the present paper makes a contribution in introducing a formal technique to the field, which is also validated with experimental data.

Earlier versions of probabilistic-descriptors have been explored in [20], [14], [15]. However, the descriptors in [20] are single pdfs rather than pdf families. Although [14] used pdf-family descriptors, the way that they were constructed lacks formalism. Moreover, the technique proposed here is not only theoretically sound, but also validated experimentally. This is important, especially for a measurement-based approach such as ours. We note that there is at least one other theoretically appealing interpolation method which is based on geodesics in the space of

pdfs. But that method turns out not to perform well in practice. A related question to the profiling discussion is: “In what length-scale is pdf interpolation meaningful for RSSI signatures?” Clearly, we would like to minimize measurements but interpolating the pdfs of two very distant locations would not make sense. This question is investigated experimentally, and interestingly, the answer confirms the intuition of some other researchers.

*B:* We develop a two-tier device tracking system that relies on RSSI measurements made by a set of *clusterheads* positioned at some of the landmarks. To that end, we would like to exploit the fact that most mobile agents in indoor environments are on the move only occasionally. Fortunately, we found it possible to detect whether a device has moved from its previously known position based on observations from a single clusterhead. We call this tier *movement detection* (the lower tier). When movement of a device has been detected, the upper tier is invoked, which detects the new position of the device using multiple clusterheads. We call this tier *localization*, as it is how localization is commonly construed. The decisions of both tiers are formulated as composite hypothesis testing problems. We develop the requisite theory and characterize the probability of detection error. Movement detection has a lower run-time cost, and in many applications the device being tracked can remain at one location for a long period of time. The two-tier design thus results in significant savings. We also address the problem of optimally placing clusterheads in order to minimize the probability of making incorrect decisions. Similar results have been largely lacking in the literature except for [20], [14], [15], which considered a single tier. The present paper extends the optimization to jointly considering localization and movement detection, and establishes that this is computationally feasible.

*C:* We present a working system that demonstrates the practicality of our approach. Our system achieved an accuracy equivalent to  $95\% < 5$  meters and  $87\% < 3$  meters, which should be considered of high-quality compared with well-known contemporary systems. We also closely examine the accuracy of our formal pdf interpolation and find that interpolations with two end points 9 meters apart may replace empirically measured pdfs with very good precision. This is significant for improving the efficiency of profiling and reinforces the findings in [8].

To summarize, the differences of this contribution compared to prior work include:

- 1) The idea and experimental validation of a formal interpolation technique applied to the profiling phase is completely absent in the wireless localization literature.
- 2) The 2-tier scheme, especially the formulation of the movement detection tier and also

the joint optimization of clusterhead placement, is also novel. Although the flavor of the underlying techniques (such as composite hypothesis testing) is reminiscent of some of the co-authors' earlier work [14], [20], [15], many important details (such as the specific hypothesis test and corresponding error rate computation) are different. More importantly, the fact that movement detection can be made to work with a single clusterhead using only RSSI is a new finding.

**Notation:** We use bold lower case letters for vectors, and bold upper case letters for matrices. Our discussions will involve both probability density functions (pdfs) and probability mass functions (pmfs). With a slight abuse of terminology, we will use the term pdf throughout.

## II. TRACKING PROBLEM FORMULATION

Consider the problem of tracking a wireless sensor network device in a contiguous space  $\mathcal{X}$ , which typically corresponds to some indoor environment. As a way of discretizing this space, we consider a given set of *landmarks* and construct what we call a *landmark graph* as follows. The node set of this undirected graph is the set of landmarks  $\mathcal{V} = \{V_i \mid i = 1, \dots, M\}$ . We draw an edge between landmarks that are geographically neighboring to each other; i.e., the edge set is  $\mathcal{E} = \{E_{ij} \mid i = 1, \dots, M, j > i, V_j \in \mathcal{N}_i\}$ , where  $\mathcal{N}_i$  is the set of neighboring landmarks to  $V_i$ . There are many different ways to formally define a neighborhood and this is left to the user; one approach could be to set a radius and consider neighboring to  $i$  all landmarks within that radius. With a slight abuse of notation, we sometimes also write: (1)  $j \in \mathcal{N}_i$  if  $V_j \in \mathcal{N}_i$ , and (2)  $(i, j) \in \mathcal{E}$  when  $E_{ij} \in \mathcal{E}$ . In what follows, a *location* refers to either a node or an edge. The set of all locations will be denoted by  $\mathcal{L} = \{L_l \mid l = 1 \dots, N\}$ , where  $N = M + |\mathcal{E}|$ . The physical meaning of the nodes and edges in the landmark graph depends on the application. For example, in an office building, the nodes can correspond to offices (i.e., a landmark can be defined as a position within an office), cubicles, and positions in the corridor. For another example, in a warehouse, the nodes can correspond to sections of storage area or intersections of aisles.

The next step is profiling, i.e., to associate to various locations appropriate probabilistic descriptors of some features of the wireless signal. Here we use the RSSI, which is measured between all pairs of landmarks. (Additional RF features may also be used if available.) Let  $Y^{(k)} \in \{\eta_1, \dots, \eta_R\}$  be the RSSI received at landmark  $k$ , which takes values from an  $R$ -

dimensional discrete set. We then have a collection of empirical distributions:

$$q_i^{(k)}(y) = \text{Freq}(Y^{(k)} = y|V_i), \quad y = \eta_1, \dots, \eta_R, \quad i, k = 1, \dots, M, \quad (1)$$

where  $k$  is the index of the receiving landmark and  $i$  is the index of the transmitting landmark. In (1),  $\text{Freq}(Y^{(k)} = y|V_i)$  simply denotes the fraction of measurements for which  $Y^{(k)} = y$ . Using these empirical distributions, we build the probabilistic descriptors of all locations using methods introduced in the sections that follow. As the result of profiling, we obtain pdfs of RSSI that characterize the signals transmitted from each location and received at each landmark. In fact, for improved robustness we associate with each node and edge of the landmark graph a family of pdfs parametrized by vectors  $\theta_i$  and  $\theta_{ij}$ , respectively. These are the location descriptors or profiles:

$$\begin{aligned} p_i^{(k)}(\cdot|\theta_i), \quad i = 1, \dots, M, \quad k = 1, \dots, M; \\ p_{ij}^{(k)}(\cdot|\theta_{ij}), \quad (i, j) \in \mathcal{E}, \quad k = 1, \dots, M. \end{aligned} \quad (2)$$

Equivalently, we may list the pdf families in terms of the locations, with the notation

$$p_{Y^{(k)}|\theta_l}(\cdot), \quad l = 1, \dots, N, \quad k = 1, \dots, M. \quad (3)$$

The former notation will be used when we discuss profiling, while the latter will be used while introducing the decision rules for positioning. Clusterhead placement will place  $K \leq M$  clusterheads at some of the landmarks, which listen to the signals transmitted by the wireless device. Movement detection and localization will be done by ‘‘comparing’’ the clusterheads’ RSSI measurements with the location profiles in a statistically meaningful way.

### III. PROFILING

This section focuses on how to generate the location profiles (2) using the empirical RSSI distributions (1), each being a pdf. The key technique is the interpolation of pdfs.

#### A. Interpolation of pdfs

A naive way of interpolating pdfs is to calculate a simple weighted average. However, one may quickly find that the naive way can produce unnatural results. For example, given two Gaussian pdfs with different means, their naive interpolation always has two peaks.

A more sophisticated approach has appeared in a work on statistical physics simulation [19], which we adopt with some generalizations. Given  $K$  pdfs,  $p_1(x), p_2(x), \dots, p_K(x)$ , let

$\mu_1, \mu_2, \dots, \mu_K$  and  $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$  be their means and variances, respectively. Let  $\boldsymbol{\rho} \in \mathbb{R}^K$  have elements  $\rho_1, \rho_2, \dots, \rho_K \in [0, 1]$  satisfying  $\sum_{i=1}^K \rho_i = 1$ . We are now seeking an interpolated pdf  $p_{\boldsymbol{\rho}}(x)$ , whose mean and variance are

$$\mu_{\boldsymbol{\rho}} = \sum_{i=1}^K \rho_i \mu_i \quad \text{and} \quad \sigma_{\boldsymbol{\rho}}^2 = \sum_{i=1}^K \rho_i \sigma_i^2. \quad (4)$$

Let

$$\xi_i(x) = \frac{\sigma_i}{\sigma_{\boldsymbol{\rho}}}(x - \mu_{\boldsymbol{\rho}}) + \mu_i, \quad i = 1, \dots, K.$$

Then define

$$p_{\boldsymbol{\rho}}(x) \triangleq \sum_{i=1}^K \rho_i \frac{\sigma_i}{\sigma_{\boldsymbol{\rho}}} p_i(\xi_i(x)). \quad (5)$$

We next prove that the mean and variance of  $p_{\boldsymbol{\rho}}(x)$  are indeed given by (4). Our proof is different from that in [19], but is somewhat inspired by the discussions in that paper. We believe our proof is more complete, concise, and intuitive for the present context.

**Proposition III.1** *The pdf given in (5) satisfies (a)  $\int p_{\boldsymbol{\rho}}(x) dx = 1$ , (b)  $\int x p_{\boldsymbol{\rho}}(x) dx = \mu_{\boldsymbol{\rho}}$ , and (c)  $\int x^2 p_{\boldsymbol{\rho}}(x) dx = \sigma_{\boldsymbol{\rho}}^2 + \mu_{\boldsymbol{\rho}}^2$ .*

*Proof:* The key observation is that

$$\begin{aligned} \int x^\alpha p_{\boldsymbol{\rho}}(x) dx &= \sum_{i=1}^K \rho_i \int x^\alpha \frac{\sigma_i}{\sigma_{\boldsymbol{\rho}}} p_i(\xi_i(x)) dx \\ &= \sum_{i=1}^K \rho_i \int \left[ (\xi_i - \mu_i) \frac{\sigma_{\boldsymbol{\rho}}}{\sigma_i} + \mu_{\boldsymbol{\rho}} \right]^\alpha p_i(\xi_i) d\xi_i. \end{aligned}$$

The last inequality holds because  $d\xi_i/dx = \sigma_i/\sigma_{\boldsymbol{\rho}}$ . We will show that the integral in the above equals 1,  $\mu_{\boldsymbol{\rho}}$ , and  $\sigma_{\boldsymbol{\rho}}^2 + \mu_{\boldsymbol{\rho}}^2$ , respectively for  $\alpha = 0, 1$ , and 2.

The case of  $\alpha = 0$  is trivial. For  $\alpha = 1$ ,

$$\int \left[ (\xi_i - \mu_i) \frac{\sigma_{\boldsymbol{\rho}}}{\sigma_i} + \mu_{\boldsymbol{\rho}} \right] p_i(\xi_i) d\xi_i = (\mu_i - \mu_i) \frac{\sigma_{\boldsymbol{\rho}}}{\sigma_i} + \mu_{\boldsymbol{\rho}} = \mu_{\boldsymbol{\rho}}.$$

For  $\alpha = 2$ ,

$$\begin{aligned} &\int \left[ (\xi_i - \mu_i) \frac{\sigma_{\boldsymbol{\rho}}}{\sigma_i} + \mu_{\boldsymbol{\rho}} \right]^2 p_i(\xi_i) d\xi_i \\ &= \int \left[ (\xi_i - \mu_i)^2 \frac{\sigma_{\boldsymbol{\rho}}^2}{\sigma_i^2} + 2(\xi_i - \mu_i) \frac{\sigma_{\boldsymbol{\rho}}}{\sigma_i} \mu_{\boldsymbol{\rho}} + \mu_{\boldsymbol{\rho}}^2 \right] p_i(\xi_i) d\xi_i \\ &= \sigma_i^2 \frac{\sigma_{\boldsymbol{\rho}}^2}{\sigma_i^2} + 2(\mu_i - \mu_i) \frac{\sigma_{\boldsymbol{\rho}}}{\sigma_i} \mu_{\boldsymbol{\rho}} + \mu_{\boldsymbol{\rho}}^2 \\ &= \sigma_{\boldsymbol{\rho}}^2 + \mu_{\boldsymbol{\rho}}^2. \end{aligned}$$

In addition to the nice feature of interpolating the mean and variance, formula (5) also interpolates the shapes of the pdfs in some sense. For example, when the original pdfs are Gaussian, the interpolation is also Gaussian.

When the random variable takes discrete values, an issue is that the transformation  $\xi(x)$  may produce a value for which probability mass is not defined. An approximate formula that solves this issue is also provided by [19] for the case of interpolating two pdfs. The following is a generalized version. Assume that the probabilities are defined for values  $-\infty, \dots, -1, 0, 1, \dots, \infty$ . For integers  $j$  and  $l$ , and for  $i = 1, \dots, K$ , let

$$\gamma_{ijl} = \max \left\{ 0, \min \left\{ \xi_i(j + 0.5), l + 0.5 \right\} - \max \left\{ \xi_i(j - 0.5), l - 0.5 \right\} \right\}, \quad (6)$$

where  $\sigma_\rho$  and  $\xi_i$  are defined as before. To understand the quantity  $\gamma_{ijl}$ , let  $C_{ij,\text{tran}}$  denote the transformed cell  $[\xi_i(j - 0.5), \xi_i(j + 0.5)]$ , and  $C_{l,\text{pmf}}$  denote  $[l - 0.5, l + 0.5]$ , the cell for which the probabilities are defined. Then  $\gamma_{ijl}$  is the proportion of  $C_{l,\text{pmf}}$  that overlaps with  $C_{ij,\text{tran}}$ ; see Fig. 1. Note that the length of  $C_{l,\text{pmf}}$  is taken to be equal to 1; otherwise,  $\gamma_{ijl}$  should be normalized by the length of  $C_{l,\text{pmf}}$ . The approximate formula is then

$$p_\rho(j) = \sum_{i=1}^K \rho_i \sum_l \gamma_{ijl} \cdot p_i(l). \quad (7)$$

The formula (6) is chosen such that

$$\sum_l \gamma_{ijl} = \frac{\sigma_i}{\sigma_\rho}, \quad \forall i, j, \quad \text{and} \quad \sum_j \gamma_{ijl} = 1, \quad \forall i, l.$$

Formula (7), to which we refer to as *linear interpolation*, is what we use in the experiments. Hereafter, we denote the linear interpolation of  $K$  pdfs with the coefficient vector  $\rho \in \mathbb{R}^K$  by  $\text{Interpol}(\rho, p_1, p_2, \dots, p_K)$ .

### B. Associating pdf families to locations

It suffices to consider the RSSI profile of all locations observed by a clusterhead placed at one of the landmarks. The index of the clusterhead is thus suppressed in all formulae of this subsection.

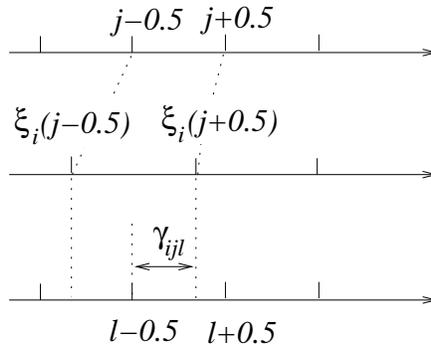


Fig. 1. Interpretation of  $\gamma_{ijl}$ .

First, we “regularize” the empirical pdfs to eliminate zero elements. This is necessary because the number of our sample measurements during profiling is finite. As a result, some RSSI value  $\eta_r$  that is possible but rare for a location  $L_i$  might not be observed during profiling, leaving the  $r$ th element of the empirical pdf equal to zero. If we use the empirical pdf directly as the probabilistic descriptor of the location, then when  $\eta_r$  appears, we would rule out  $L_i$  immediately, regardless of how many total observations are made and how the rest of the observations resemble the profile of location  $L_i$ . This is clearly undesirable. To mitigate this problem, we mix the empirical pdf with a discretized Gaussian-like pdf of the same mean and variance. Namely, let  $q$  be an empirical pdf with mean  $\mu$  and variance  $\sigma^2$ . Let  $\phi(\mu, \sigma^2)$  be a Gaussian-like pdf whose domain is discretized to the set  $\{\eta_1, \dots, \eta_R\}$ . Let  $\gamma \in (0, 1)$  be a chosen mixing factor — typically we set  $\gamma$  to a small value such as 0.1 or 0.2. Then the pdf after regularization is is

$$\tilde{q} = \text{Interpol} \left( \begin{bmatrix} 1 - \gamma \\ \gamma \end{bmatrix}, q, \phi(\mu, \sigma^2) \right).$$

Second, consider the areas represented by the landmarks. Here, we use the pdf-family framework to achieve robustness. Specifically, suppose  $V_i$  has  $I$  neighbors —  $\mathcal{N}_i = \{V_{j_1}, \dots, V_{j_I}\}$ . Let  $\boldsymbol{\theta}_i = (\theta_i^{(1)}, \dots, \theta_i^{(I)}) \in \mathbb{R}^I$ ,  $\boldsymbol{\theta}_i \geq \mathbf{0}$  elementwise, and  $\sum_{j=1}^I \theta_i^{(j)} < 1$ . Let  $\boldsymbol{\rho}_{\boldsymbol{\theta}_i} = (1 - \sum_{j=1}^I \theta_i^{(j)}, \theta_i^{(1)}, \dots, \theta_i^{(I)})$ . Then the pdf family associated with  $V_i$  can be defined as an interpolation of  $I + 1$  empirical pdfs:

$$p_i(\cdot | \boldsymbol{\theta}_i) \triangleq \text{Interpol} (\boldsymbol{\rho}_{\boldsymbol{\theta}_i}, \tilde{q}_i(\cdot), \tilde{q}_{j_1}(\cdot), \dots, \tilde{q}_{j_I}(\cdot)).$$

Last, consider the edges of the landmark graph. As will be justified by the experiments, we

associate with the edge  $(i, j)$  a pdf family defined as the interpolation of the pdf families for landmarks  $i$  and  $j$ . Let  $\vartheta_{ij} \in (0, 1)$  and  $\boldsymbol{\theta}_{ij}$  be a vector concatenating  $\boldsymbol{\theta}_i$ ,  $\boldsymbol{\theta}_j$ , and  $\vartheta_{ij}$ . The pdf family associated with edge  $(i, j)$  is

$$p_{ij}(\cdot|\boldsymbol{\theta}_{ij}) \triangleq \text{Interpol} \left( \begin{bmatrix} \vartheta_{ij} \\ 1 - \vartheta_{ij} \end{bmatrix}, p_i(\cdot|\boldsymbol{\theta}_i), p_j(\cdot|\boldsymbol{\theta}_j) \right).$$

The collection of  $p_i(\cdot|\boldsymbol{\theta}_i)$ 's and  $p_{ij}(\cdot|\boldsymbol{\theta}_{ij})$ 's corresponds to the pdfs defined in (2), which are then rearranged as in (3).

### C. An alternative interpolation technique

One may propose another theoretically attractive interpolation technique based on geodesics in the space of probability distributions. We only discuss the two-pdf case here for simplicity. A well-known geodesic that connects two pdfs  $p_0(x)$  and  $p_1(x)$  is a  $\rho$ -parameterized curve

$$p_\rho(x) = \frac{p_0^{1-\rho}(x)p_1^\rho(x)}{\int p_0^{1-\rho}(x)p_1^\rho(x)dx}, \quad (8)$$

where  $\rho \in [0, 1]$ . We will refer to this as the *geometric interpolation*. Like in the linear interpolation case, one can verify that the geometric interpolation of two Gaussian pdfs remains Gaussian. However, our experiments show that the performance of this interpolation is inferior to that of linear interpolation, which demonstrates the importance of the interpolation we have adopted.

### D. An alternative Gaussian model

In the above, we associate a family of generally shaped pdfs to each location. If a Gaussian model of the RSSI is used instead, this task can be greatly simplified. One may then ask whether using generally shaped pdfs is worth the effort. The answer to this question may depend on circumstances. However, our experiments show that significant information regarding the signals transmitted from a location is captured by our approach, but would be neglected if we assume the Gaussian model.

#### IV. TWO-TIER TRACKING SYSTEM

Independent of the method used to construct the pdf-family descriptors, we next consider how to deploy and orchestrate clusterheads to make positioning decisions.

In positioning a mobile wireless device, one typically needs to draw on measurements from multiple clusterheads. If triangulation is used, then at least three clusterheads are needed, and more clusterheads may be required due to signal noise/variability. However, a somewhat surprising observation is: if we already know the previous location of a wireless device and ask the question of whether it has moved to another location, the RSSI signature observed by a *single* clusterhead provides sufficient information. Furthermore, that clusterhead can be selected for each location based on the profiling results. We will refer to such a test that determines whether a device has moved as *movement detection*. Thus, if many mobile nodes in the system actually stay at some locations for prolonged periods of time (e.g., as is the case for office workers), then switching into the movement detection mode at times can significantly reduce the workload of the system. This motivates a two-tier positioning scheme of the following generic form with parameters  $T_L$ ,  $H$ , and  $T_M$ .

**Localization Tier:** Assuming no prior knowledge on the mobile device's position (except that it is in the coverage area), use all clusterheads to locate the device every  $T_L$  seconds. If the device's position remains the same for  $H$  rounds of localization, then switch to the movement detection tier.

**Movement Detection Tier:** Every  $T_M$  seconds, use one clusterhead (selected during the initial setup for the mobile device's last known position) to test the hypothesis that the device's position is unchanged. If the hypothesis is rejected, then return to the localization tier.

Next we address the issues of how to make the localization and movement detection decisions and how to ensure that the deployed clusterheads can perform the tasks of both tiers satisfactorily for every location. The decision rule for the localization tier has been designed in [14], [15]. Our main effort here is to design the movement detection tier. Further, we present the statistical decisions of the two tiers in a generalized framework, and optimize them jointly in the clusterhead placement phase.

### A. Statistical Decision Rules of the Two Tiers

To present the decision rules of the two tiers in a unified framework, we begin by recalling the Kullback-Leibler (KL) distance used in information theory. For two distribution functions (assuming discrete random variables)  $\mathbf{p}$  and  $\mathbf{q}$ , the KL distance of  $\mathbf{q}$  from  $\mathbf{p}$  is defined as

$$D(\mathbf{q}||\mathbf{p}) \triangleq \sum_y q(y) \log \frac{q(y)}{p(y)} \quad (\text{assuming } 0 \cdot \log \frac{0}{a} \equiv 0, \forall a \geq 0). \quad (9)$$

Intuitively, the KL distance reflects the difference between two distributions in a statistically meaningful way (see [21]). It is not a true distance metric (lacks symmetry and does not satisfy the triangle inequality) but it is non-negative and zero if and only if  $\mathbf{p} = \mathbf{q}$ . Typically,  $\mathbf{q}$  is some sample distribution and  $\mathbf{p}$  is a model distribution. For example, if a certain event is impossible according to the model distribution ( $\mathbf{p}$  has a zero element somewhere) but the event occurred in the sample (the corresponding element of  $\mathbf{q}$  is positive), then the KL distance equals  $\infty$ , and we immediately know that the sample is not drawn from that model distribution. This example also illustrates that sometimes we may want to avoid proposing a model distribution with a zero element, which is why we regularize the empirical pdfs in Section III-B.

Another important quantity from information theory is the entropy of the distribution  $\mathbf{p}$  (assume a discrete random variable taking values in  $\{\eta_1, \dots, \eta_R\}$ ) defined as:

$$H(\mathbf{p}) = - \sum_y p(y) \log p(y). \quad (10)$$

Consider an i.i.d. sequence  $\mathbf{y}^n = (y_1, \dots, y_n)$  drawn from  $\mathbf{p}$  and let  $\mathbf{q}_{\mathbf{y}^n}$  be its sample (empirical) distribution defined as in (1). It is easy to verify that the probability of  $\mathbf{y}^n$  can be written as

$$p(\mathbf{y}^n) = \prod_{i=1}^R p(\eta_i)^{nq_{\mathbf{y}^n}(\eta_i)} = e^{-n[H(\mathbf{q}_{\mathbf{y}^n}) + D(\mathbf{q}_{\mathbf{y}^n}||\mathbf{p})]}. \quad (11)$$

Recall now from Equation (3) that the pdf family associated with the RSSI from location  $l$  to clusterhead  $k$  is  $p_{Y^{(k)}|\theta_l}(\cdot)$ ,  $l = 1, \dots, N$ ,  $k = 1, \dots, M$ , as constructed in the profiling phase. For either localization or movement detection, suppose the clusterhead makes  $n$  i.i.d. observations  $\mathbf{y}^{(k),n} = (y_1^{(k)}, \dots, y_n^{(k)})$ , with a corresponding sample distribution  $\mathbf{q}_{\mathbf{y}^n}^{(k)}$ , upon which the decision rules apply. The efficiency of the decision rules to be presented will be quantified using the *error exponent*, defined as

$$d \triangleq - \lim_{n \rightarrow \infty} \sup \frac{1}{n} \log \mathbf{P}(\text{error}). \quad (12)$$

This represents the exponentially decay rate at which the probability of error converges to zero.

Next, consider the decisions that can be made by one clusterhead (hence the clusterhead index is dropped). For localization, the core task is to distinguish between two candidate locations of the mobile device. It was shown in [14], [15] that the question boils down to a binary composite hypothesis testing problem (composite because the distributions involved have unknown parameters — the  $\theta$ 's), and one can use the well-known *Generalized Likelihood Ratio Test (GLRT)*. In particular, the GLRT decides location  $i$  over  $j$  if

$$\frac{1}{n} \log \frac{\max_{\theta_j} p_{Y|\theta_j}(\mathbf{y}^n)}{\max_{\theta_i} p_{Y|\theta_i}(\mathbf{y}^n)} < \lambda_{ij},$$

for some appropriate threshold  $\lambda_{ij}$ . We have

$$\begin{aligned} \lambda_{ij} &> \frac{1}{n} \log \max_{\theta_j} p_{Y|\theta_j}(\mathbf{y}^n) - \frac{1}{n} \log \max_{\theta_i} p_{Y|\theta_i}(\mathbf{y}^n) \\ &= \max_{\theta_j} [-H(\mathbf{q}_{\mathbf{y}^n}) - D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j})] \\ &\quad - \max_{\theta_i} [-H(\mathbf{q}_{\mathbf{y}^n}) - D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_i})] \\ &= \min_{\theta_i} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_i}) - \min_{\theta_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j}). \end{aligned}$$

Thus, in terms of the KL distance, the GLRT rule can be expressed as:

$$\begin{aligned} &\text{Decide location } i \text{ relative to } j \text{ if and only if} \\ &\min_{\theta_i} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_i}) - \min_{\theta_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j}) < \lambda_{ij}. \end{aligned} \tag{13}$$

A slightly conservative estimate of the corresponding error exponent, denoted by  $d_{ij}$ , was also derived in [15] (including an explanation of why  $\lambda_{ij}$  is not always zero and how it can be selected). Moreover, [15] describes how a decision among many potential locations can be taken by using a sequence of binary tests; interestingly enough, such a decision can be taken in a distributed manner by appropriate collaboration between the clusterheads.

Let us now turn our attention to the movement detection tier, which is a new contribution of the present paper. Suppose that the last known location of a mobile device is location  $j$  and we would like to determine if the device remains at  $j$  based on the  $n$  i.i.d. observations in  $\mathbf{y}^n$  by a single clusterhead. The following *Generalized Hoeffding Test (GHT)* [22], expressed analogously to (13) using the KL distance, is applicable:

$$\begin{aligned} &\text{Report “no movement” if and only if} \\ &\min_{\theta_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j}) < \lambda_j. \end{aligned} \tag{14}$$

Let  $\alpha_j$  and  $\beta_j$  be the error probabilities of false alarm and missed detection, respectively. The following proposition shows that the GHT is optimal in a *Generalized Neyman-Pearson* sense.

**Lemma IV.1** *The test in (14) is optimal in a generalized Neyman-Pearson sense, that is,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \alpha_j < -\lambda_j, \quad \forall \boldsymbol{\theta}_j, \quad (15)$$

and  $-\limsup_{n \rightarrow \infty} \frac{1}{n} \log \beta_j$  is maximized among all tests satisfying (15) uniformly for all  $\boldsymbol{\theta}_i$  characterizing some alternative pdf.

*Proof:* Let  $\mathbf{P}_j(\cdot)$  denote a probability conditional on the device having not moved from location  $j$ . Let also  $\mathcal{Q}_n = \{\boldsymbol{\nu} \mid \boldsymbol{\nu} = \mathbf{q}_{\mathbf{y}^n} \text{ for some } \mathbf{y}^n\}$  denote the set of all empirical measures that can be obtained from an  $n$ -length observation sequence and  $T_n(\boldsymbol{\nu}) = \{\mathbf{y}^n \mid \mathbf{q}_{\mathbf{y}^n} = \boldsymbol{\nu}\}$  the set of  $n$ -length observation sequences with an empirical measure equal to  $\boldsymbol{\nu}$ .

First, letting  $\mathbf{P}_{\boldsymbol{\nu}}$  denote a probability under the measure  $\boldsymbol{\nu}$  and  $|\cdot|$  the cardinality, we have

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\nu}}[T_n(\boldsymbol{\nu})] &= \sum_{\{\mathbf{y}^n \mid \mathbf{q}_{\mathbf{y}^n} = \boldsymbol{\nu}\}} \mathbf{P}_{\boldsymbol{\nu}}[\mathbf{y}^n] \\ &= \sum_{\{\mathbf{y}^n \mid \mathbf{q}_{\mathbf{y}^n} = \boldsymbol{\nu}\}} \nu(\eta_1)^{n\nu(\eta_1)} \cdots \nu(\eta_R)^{n\nu(\eta_R)} \\ &= |T_n(\boldsymbol{\nu})| e^{-nH(\boldsymbol{\nu})}, \end{aligned}$$

which implies

$$|T_n(\boldsymbol{\nu})| \leq e^{nH(\boldsymbol{\nu})}. \quad (16)$$

Now, for all  $\boldsymbol{\theta}_j$  the false alarm probability is given by

$$\begin{aligned} \alpha_j &= \mathbf{P}_j[\{\mathbf{y}^n \mid \min_{\boldsymbol{\theta}_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\boldsymbol{\theta}_j}) \geq \lambda_j\}] \\ &= \sum_{\{\mathbf{q}_{\mathbf{y}^n} \mid \min_{\boldsymbol{\theta}_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\boldsymbol{\theta}_j}) \geq \lambda_j\}} |T_n(\mathbf{q}_{\mathbf{y}^n})| p_{Y|\boldsymbol{\theta}_j}(\mathbf{y}^n) \\ &\leq \sum_{\{\mathbf{q}_{\mathbf{y}^n} \mid \min_{\boldsymbol{\theta}_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\boldsymbol{\theta}_j}) \geq \lambda_j\}} e^{nH(\mathbf{q}_{\mathbf{y}^n})} e^{-n[H(\mathbf{q}_{\mathbf{y}^n}) + D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\boldsymbol{\theta}_j})]} \\ &= \sum_{\{\mathbf{q}_{\mathbf{y}^n} \mid \min_{\boldsymbol{\theta}_j} D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\boldsymbol{\theta}_j}) \geq \lambda_j\}} e^{-nD(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\boldsymbol{\theta}_j})} \\ &\leq (n+1)^R e^{-n\lambda_j}. \end{aligned}$$

For the first inequality above we have used (11) and (16). In the last inequality above we used the fact that the set of all possible empirical measures,  $\mathcal{Q}_n$ , has cardinality upper bounded by  $(n+1)^R$  (a symbol of length  $R$  with each element taking values from  $\{\frac{0}{n}, \dots, \frac{n}{n}\}$ ). This establishes (15).

Let now  $\mathcal{S}_n$  be some other decision rule satisfying (15). It is well known that the empirical measure is a sufficient statistic so any rule will depend only on that. Let  $\alpha_{\mathcal{S}_n}$  and  $\beta_{\mathcal{S}_n}$  denote the corresponding false alarm and missed detection probabilities. For all  $\epsilon > 0$  and all large enough  $n$  we have

$$\alpha_{\mathcal{S}_n} \leq e^{-n(\lambda_j + \epsilon)}. \quad (17)$$

Meanwhile for all  $\epsilon > 0$ , all large enough  $n$ , and any  $\mathbf{y}^n$  such that  $\mathcal{S}_n$  declares “movement” it holds

$$\begin{aligned} \alpha_{\mathcal{S}_n} &= \sum_{\{\mathbf{q}_{\mathbf{y}^n} | T_n(\mathbf{q}_{\mathbf{y}^n}) \text{ implies movement}\}} |T_n(\mathbf{q}_{\mathbf{y}^n})| p_{Y|\theta_j}(\mathbf{y}^n) \\ &\geq \sum_{\{\mathbf{q}_{\mathbf{y}^n} | T_n(\mathbf{q}_{\mathbf{y}^n}) \text{ implies movement}\}} (n+1)^{-R} e^{-nD(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j})} \\ &\geq e^{-n[D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j}) + \epsilon]}, \end{aligned}$$

where the first inequality above uses [23, Lemma 2.1.8]. Comparing the above with (17) it follows that if  $\mathbf{y}^n$  implies “movement” then for all  $\theta_j$  it should hold  $D(\mathbf{q}_{\mathbf{y}^n} \| p_{Y|\theta_j}) \geq \lambda_j$ . Therefore, the GHT in (14) should declare movement as well, which implies that  $\beta_{\mathcal{S}_n} \geq \beta_j$  for all  $\theta_i$  characterizing an alternative pdf, where  $i \neq j$ . The latter establishes that the GHT maximizes the exponent of the missed detection probability. ■

Eq. (15) provides a bound on the exponent of the false alarm probability. To bound the exponent of the missed detection probability we can use Sanov’s theorem [23, Chap. 2]. Specifically,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \beta_j \leq - \min_{\mathbf{q} \in \mathcal{H}_j} D(\mathbf{q} \| p_{Y|\theta_i}), \quad \forall \theta_i, \quad (18)$$

where  $\mathcal{H}_j = \{\mathbf{q} | \min_{\theta_j} D(\mathbf{q} \| p_{Y|\theta_j}) < \lambda_j\}$ .

Defining

$$\begin{aligned} Z_{j,\theta_i}(\lambda_j) &= \min_{\mathbf{q}} D(\mathbf{q} \| p_{Y|\theta_i}) \\ \text{s.t. } &\min_{\theta_j} D(\mathbf{q} \| p_{Y|\theta_j}) < \lambda_j, \end{aligned} \quad (19)$$

and

$$Z_j(\lambda_j) = \min_{i \neq j} \min_{\theta_i} Z_{j,\theta_i}(\lambda_j), \quad (20)$$

we can write

$$-\limsup_{n \rightarrow \infty} \frac{1}{n} \log \beta_j \geq Z_j(\lambda_j).$$

The error exponent of the whole test equals the lesser between the exponents of  $\alpha_j$  and  $\beta_j$ . On the other hand, it is straightforward to show that  $Z_j(\lambda_j)$  is monotonically decreasing in  $\lambda_j$ . If we could compute  $Z_j(\lambda_j)$ , then the solution of  $Z_j(\lambda_j) = \lambda_j$  gives the optimal value of  $\lambda_j$  as well as the best achievable error exponent. However, the computational cost of finding the exact solution is significant. This is mostly because the constraint in (19) is non-convex and also high-dimensional. To side-step this problem, we replace  $Z_{j,\theta_i}(\lambda_j)$  with an estimate based on dual relaxation:

$$\begin{aligned} \tilde{Z}_{j,\theta_i}(\lambda_j) \triangleq & \max_{\mu \geq 0} \min_{\theta_j} \min_{\mathbf{q}} D(\mathbf{q} \| p_{Y|\theta_i}) + \mu(D(\mathbf{q} \| p_{Y|\theta_j}) - \lambda_j) \\ & \text{s.t. } \sum_y q(y) = 1, \\ & q(y) \geq 0, \quad \forall y. \end{aligned} \quad (21)$$

As a relaxation, it holds that  $\tilde{Z}_{j,\theta_i}(\lambda_j) \leq Z_{j,\theta_i}(\lambda_j)$ . The computational gain comes from the fact that the minimization over  $\mathbf{q}$  is now convex. In fact, the optimal solution can be found in closed-form:

$$q^*(y) = \frac{p_{Y|\theta_i}^\rho(y) p_{Y|\theta_j}^{1-\rho}(y)}{\sum_\eta p_{Y|\theta_i}^\rho(\eta) p_{Y|\theta_j}^{1-\rho}(\eta)}, \quad \text{where } \rho = \frac{1}{1+\mu}. \quad (22)$$

Then  $\tilde{Z}_j(\lambda_j) = \min_{i \neq j} \min_{\theta_i} \tilde{Z}_{j,\theta_i}(\lambda_j)$ , which is also monotonically decreasing in  $\lambda_j$ . The solution of  $\tilde{Z}_j(\lambda_j) = \lambda_j$ , denoted by  $h_j$ , gives a near-optimal threshold and the error exponent at the same time.

### B. Using Multiple Clusterheads

From the above, for each landmark (also candidate clusterhead location)  $k$ , we have a collection of localization error exponents  $d_{ijk}$ , and movement-detection error exponents  $h_{jk}$ , all of which are optimized (note that the index of the clusterhead location is now added back). However, no matter how good the decision rule is, a single clusterhead is normally not enough for distinguishing every pair of locations in a wide coverage area, because given any single clusterhead, there may be pairs of locations whose RSSI profiles are similar. Even multiple clusterheads may fail to make decisions with small enough error probability if they are not placed appropriately. In formulating the clusterhead placement problem, our goal is to ensure:

- 1) For the localization tier, there should be at least one clusterhead  $k$  for each pair of locations  $i, j$  such that  $L_i, L_j$  can be clearly distinguished using observations of clusterhead  $k$  under the GLRT.
- 2) For the movement detection tier, there should be at least one clusterhead  $k$  for each location  $j$  such that  $L_j$  can be distinguished from alternative locations using clusterhead  $k$  under the GHT.

More precisely, for a given number  $K$  of clusterheads to deploy, we maximize the error exponent level  $\epsilon$  that is met (or exceeded) for every location in the movement detection tier, and for every pair of locations in the localization tier. Furthermore, the error exponent for each location or each pair of locations is given by the clusterhead that can best resolve the case (with greatest error exponent). Formally, let  $\mathcal{P}$  denote the set of landmarks where clusterheads are placed. The problem can be written as

$$\begin{aligned}
 \max_{\mathcal{P} \subset \mathcal{V}} \quad & \epsilon \\
 \text{s.t.} \quad & |\mathcal{P}| = K, \\
 & \min_{i,j} \max_k d_{ijk} \geq \epsilon, \\
 & \min_j \max_k h_{jk} \geq \epsilon.
 \end{aligned} \tag{23}$$

It is important to note that the optimal value of this problem provides an upper bound on the probability of error for both localization and movement detection decisions. That is, if  $\epsilon^*$  is an optimal solution, the maximum probability of error of our system satisfies

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[\text{error}] \leq -\epsilon^*. \tag{24}$$

Problem (23) can be reformulated as the *mixed integer linear programming (MILP)* problem shown in Fig. 2.

Software exists for solving generic MILPs. Furthermore, we have designed an algorithm that solves the MILP in Fig. 2 much faster than commercial general purpose MILP-solvers. Our approach is to solve this problem by an iterative *feasibility* algorithm along the lines proposed in [24]. In particular, we use a modified version of a *two-phase* algorithm proposed in [25]. Its computational advantage lies in the fact that we solve a feasibility problem in each iteration that contains only  $O(M)$  variables and  $O(N^2)$  constraints, instead of the  $O(N^2M)$  variables and  $O(N^2M)$  constraints that appear in the formulation.

---


$$\begin{aligned}
& \max \quad \epsilon \\
& \text{s.t.} \quad \sum_{k=1}^M x_k = K \\
& \quad \sum_{k=1}^M y_{ijk} = 1, \quad \forall i, j = 1, \dots, N, i < j, \\
& \quad \sum_{k=1}^M z_{jk} = 1, \quad \forall j = 1, \dots, N, \\
& \quad y_{ijk} \leq x_k, \quad \forall i, j, i < j, k = 1, \dots, M, \\
& \quad z_{jk} \leq x_k, \quad \forall j, k = 1, \dots, M, \\
& \quad \epsilon \leq \sum_{k=1}^M d_{ijk} y_{ijk}, \quad \forall i, j, i < j, \\
& \quad \epsilon \leq \sum_{k=1}^M h_{jk} z_{jk}, \quad \forall j, \\
& \quad y_{ijk} \geq 0, \quad \forall i, j, i < j, \forall k, \\
& \quad z_{jk} \geq 0, \quad \forall j, \forall k, \\
& \quad x_k \in \{0, 1\}, \quad \forall k.
\end{aligned}$$


---

Fig. 2. Clusterhead placement MILP formulation.

## V. EXPERIMENTS

Our testbed is set up on the first floor of a Boston University building (see Fig. 3), and uses MPR2400 (MICAz) motes from Crossbow Technology Inc.

### A. Testing the complete system

Our two-tier tracking system covers 10 rooms and the corridors, which are mapped to 30 landmarks, marked by either a green circle or a red square on the floor plan (Fig. 3), where the latter marks the clusterhead positions obtained from solving the optimal placement problem. The landmark graph is then constructed resulting in 39 edges, or a total of 69 locations. Hence  $N = 69$  and  $M = 30$  in this experiment. A mote is placed at each landmark location, but only some of them will serve as clusterheads. All 30 motes are connected to a base MICAz through a mesh network. The base mote is docked on a Stargate node which forwards the messages back

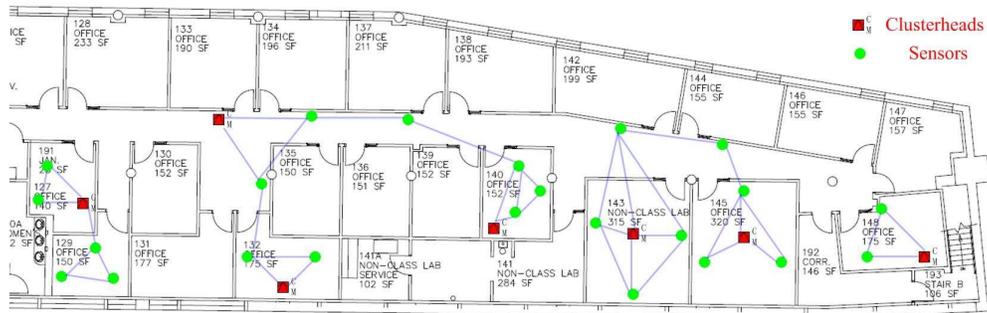


Fig. 3. Floor plan with the landmarks for the testbed.

to the server.

The experimental validation of our localization approach can be divided into the following five phases:

*Phase 1:* We obtained the empirical pdfs for the landmarks corresponding to Eq. (1). The 30 motes placed at the landmark locations took turns to broadcast packets, specifically, when one was transmitting the others were listening and recording the RSSI. A total of 200 packets were transmitted by each mote. The data collection was repeated for the combinations of two frequencies and two power levels; details will be given below.

*Phase 2:* We used the methods in Section III to construct the pdf families corresponding to Eq. (2), which are the descriptors of all 69 locations. Note that the interpolation technique allowed us to construct high quality descriptors without densely covering the area with landmarks.

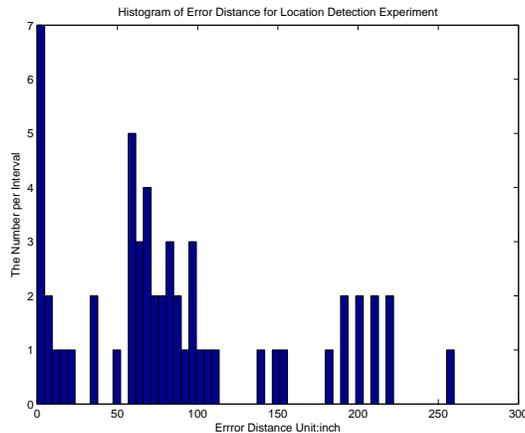
*Phase 3:* We obtained  $d_{ijk}$  and  $h_{jk}$  as described in Sec. IV.

*Phase 4:* We solved the MILP to optimize clusterhead placement and simultaneously obtained the performance guarantee of Eq. (24). In the MILP formulation, we needed to input  $K$ , the total number of clusterheads. By varying  $K$  from 1 to 30, we discovered that the performance guarantee reached a satisfactory level after  $K = 7$ , and somewhat flattens afterward. Thus, we placed clusterheads at 7 locations (again, marked by red squares in Fig 3).

*Phase 5:* We introduced mobile motes in the coverage area and let the system make localization and movement detection decisions.

We let Phase 1 (a completely automated procedure) stretch over 24 hours to acquire data under diverse conditions of the surrounding lab environment. Phase 2 takes virtually no time. Phase

3 takes another 24 hours on our computer, although further optimization of our code may reduce the computation time significantly. Phase 4 only takes about half an hour. Note that all these steps are performed only once, after which the amount of real-time computation needed for each localization decision is very small, such that the resources on the clusterheads (typically motes plugged into a wall outlet or a large enough battery) are sufficient.



$$\bar{D}_e = 87.32 \text{ inches}$$

Fig. 4. Localization result.

We know from previous experiences [15] and the literature that frequency and power diversity provide better performance. The mote to be located broadcasted 20 packets over the combination of 2 frequencies (2.410 GHz and 2.460 GHz) and 2 power levels (0 dBm and  $-10$  dBm), with 5 packets corresponding to each combination. We achieved a mean error distance of 87.32 inches, which is better than our earlier result of 96.08 inches [15] based on techniques that do not use a formal method of pdf interpolation. The percentile of errors  $< 3$  meters (118 inches) also improved from 80% to 87%. One may also count from Fig. 4 that the percentile of errors  $< 5$  meters (197 inches) is 95%.

The total coverage area (we have excluded the rooms that are in the floor plan but to which we did not have access) was 1827 feet<sup>2</sup>, that is, about 61 feet<sup>2</sup> per landmark. With a mean error distance of  $\bar{D}_e = 7.3$  feet the mean area of “confusion” was  $7.3^2 = 53$  feet<sup>2</sup>. It is evident that we were able to achieve accuracy on the same order of magnitude as the area “covered” by a landmark; this is the best possible outcome with a “discretized” system such as ours. That is,

the system was identifying the correct location or a neighboring location most of the time. We used a clusterhead density of 1 clusterhead per  $1827/7 = 261$  feet<sup>2</sup>. Note that our system is *not* based on the “proximity” to a clusterhead; the ratio of locations to clusterheads is  $69/7$ , or about 10.

We also obtained results for the movement detection tier. The mote to be monitored now transmits 20 packets at a single frequency and power level depending on its *a priori* location. The use of a single frequency/power level was selected because the diversity only benefits decisions involving multiple clusterheads. The resulting error probability was 8%. Again, we emphasize that we are exploiting only the most basic RF measurements in obtaining these results. Yet, the approach is easily generalizable to include additional information if available (in that case, instead of scalar RSSI measurements we will be dealing with an observation vector).

### *B. A closer look at pdf interpolation*

We have proposed a rather sophisticated interpolation technique for generating location profiles. One concern is: if the interpolated pdfs were merely low-quality approximations of the actual pdfs, then we might be better-off using a Gaussian approximation, which is computationally cheaper. In our experiments however, the interpolated pdfs did a very good job preserving the shape information of the empirical pdfs which were not close to a Gaussian. As will be shown, the decision accuracy using the interpolated pdfs dominates that of the Gaussian approximation by a significant margin. Another question that we attempt to answer is: At what length-scale does pdf interpolation make sense? It turns out that the interpolation is very meaningful when the two end points are about 30 feet (or 9 meters) apart, but not when they are 60 feet apart.

The first experiment is conducted in a roughly straight corridor of about 60 feet long, mapped to 5 locations roughly 15 feet apart. Labeling the locations consecutively as location 1, 2, . . . , 5, we place the clusterhead at location 1 (which is at one end of the corridor). To measure the signals transmitted from each location, one of the coauthors stood at that location holding a transmitting mote, which sends a packet every 5 seconds. We chose to have a person hold the mote because this is close to an actual application scenario. The clusterhead received the packets and recorded the RSSI values. During the experiment, a total of 150 packets were sent from each location. Due to packet loss, the number of actual samples taken by the clusterhead is less, but we still obtained more than 100 samples for each location. Then, we mix a Gaussian component

into each of the six empirical distributions as described earlier with a mixing factor of 0.2, i.e., regularized empirical distribution = 0.8 measured + 0.2 Gaussian. The empirical distributions for the six locations after regularization are denoted by  $q_1, q_2, \dots, q_5$ .

We compare three interpolation methods. First, in what is labeled “linear short interpolation,” the interpolated pdf of location  $i$  is generated using  $q_{i+1}$  and  $q_{i-1}$ :

$$p_{i,\text{short}} = \text{Interpol} \left( \left[ \begin{array}{c} 0.5 \\ 0.5 \end{array} \right], q_{i-1}, q_{i+1} \right), \quad i = 2, 3, 4.$$

Second, in what is labeled “linear long interpolation,” the interpolated pdfs are generated using  $q_1$  and  $q_5$ :

$$p_{i,\text{long}} = \text{Interpol} \left( \left[ \begin{array}{c} \frac{5-i}{4} \\ i-1 \\ \frac{4}{4} \end{array} \right], q_1, q_5 \right), \quad i = 2, 3, 4.$$

Third, we adopt the Gaussian model instead and interpolate the pdf of each location with adjacent locations:

$$\begin{aligned} p_{i,\text{gaussian}} &= \text{Interpol} \left( \left[ \begin{array}{c} 0.5 \\ 0.5 \end{array} \right], \phi(\mu_{i-1}, \sigma_{i-1}^2), \phi(\mu_{i+1}, \sigma_{i+1}^2) \right) \\ &= \phi \left( \frac{\mu_{i-1} + \mu_{i+1}}{2}, \frac{\sigma_{i-1}^2 + \sigma_{i+1}^2}{2} \right), \quad i = 2, 3, 4. \end{aligned}$$

Last, we also include the geometric interpolation (cf. Sec. III-C) for comparison, denoted by  $p_{i,\text{geo}}$ , where location  $i \in \{2, 3, 4\}$  is interpolated by location  $i - 1$  and  $i + 1$  as in the linear short case.

1) *Qualitative study*: We visually compare the various interpolation results for location 2 as an example (Fig. 5). The short interpolation seems to capture some shape information of the actual pdf that is missed by the Gaussian model. For example, the empirical pdf is skewed to the left. The interpolated pdf also exhibits the skewness, while the Gaussian pdf is always symmetrical. One may also notice that the linear long and the geometric interpolations appear very different from the actual pdf.

2) *Quantitative observation*: First, it is of interest to compare the qualities of the different interpolations using the KL distance (9) as a metric of distance between pdfs. As we have seen, the KL distance appears in both localization and movement detection error exponents.

The comparison is plotted in Fig. 6. It is very interesting to see that *the quality of linear short interpolation dominates that of the Gaussian model in the KL sense*. For example, the KL

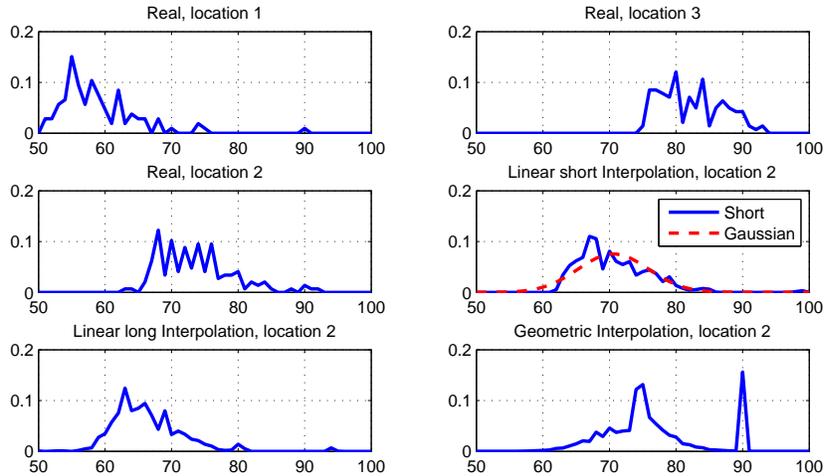


Fig. 5. Visual comparison of interpolated pdfs for location 2.

distance of short-interpolation-to-empirical for location 4 is only a little over one third of that of the Gaussian model. For locations 2 and 3, the difference is roughly a factor of 1.5, which is still significant. Similarly, the linear short interpolation appears superior to the geometric (also short) interpolation. The long interpolation on the other hand clearly departs from the actual distribution.

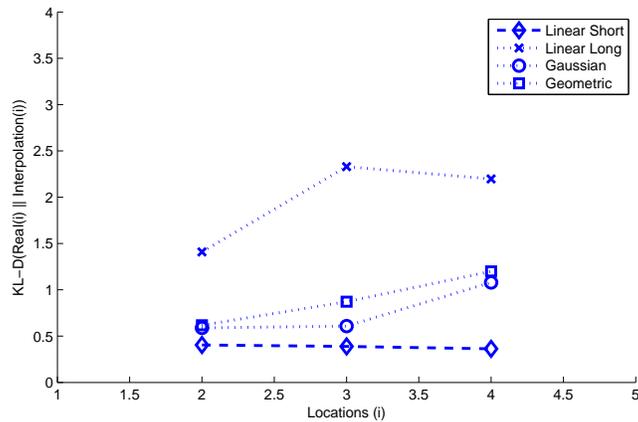


Fig. 6. Kullback-Leibler distance comparison.

Next, we attempted location distinction using these pdfs. We omit the geometric and the linear long cases as, based on the discussion above, they are not suitable for our purposes. We will hence change the label “linear short interpolation” to simply “linear interpolation”. For each location  $j = 2, 3$ , and 4, and each sample size  $n = 5, 10, 15, \dots, 30$ , we tested the hypothesis that “the wireless device is at location  $j$ ” using  $n$  RSSI measurements drawn randomly from a large pre-compiled data set. Each sample contained measurements associated with one single location  $i$ , where  $i \in \{1, \dots, 6\}$ . The GHT was used to make the decision, in which the threshold  $\lambda$  was optimized for each  $j$  and  $n$ . For each pair of  $j$  and  $n$ , we repeated the trial 4000 times and calculate the empirical error probability, which is a weighted sum of all types of errors.

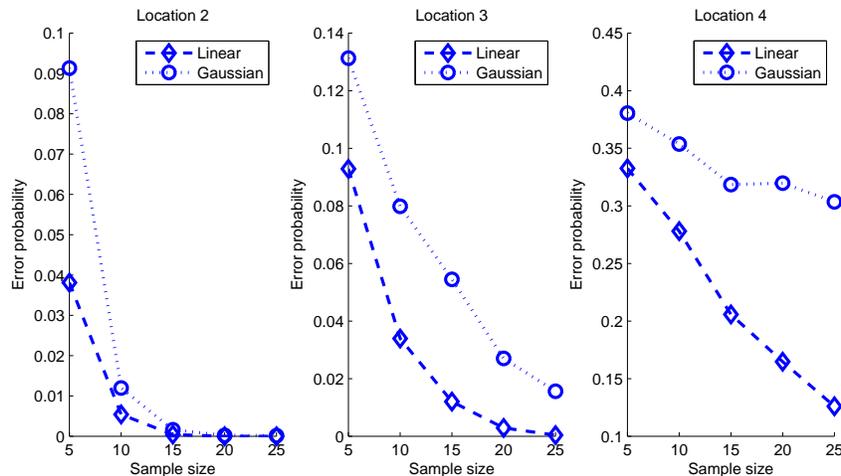


Fig. 7. Comparison of linear interpolation and the Gaussian model in terms of the error probability of GHT using the interpolated pdf.

Some observations are in order. First, very low error probability is achieved using the linearly interpolated pdfs. Second, and maybe more interestingly, the error probabilities using linearly interpolated pdfs are lower than those using the Gaussian model in all instances, and the difference is quite large in most cases. Further, as the sample size increases, the former decreases faster than the latter. These demonstrate that the approximation qualities of the interpolated pdfs are fairly notable, and the added computational effort (compared to the Gaussian model) may very much be worth our while.

Yet another interesting observation emerges in the comparison with [8], which has shown

that when the spacing of “reference signatures” goes below roughly 10 meters, the improvement in performance diminishes. The spacing of the “reference signatures” there is analogous to the distance between the two end-point locations in our pdf interpolation. The two end points in the working version (the “short”) of our pdf interpolation happen to be a little more than 9 meters (30 feet) apart. This result reinforces that of [8], as both indicate that taking empirical measurements at a spacial density of less than 9 or 10 meters apart, or roughly 1 per 25 sq. meters, carries diminishing benefit.

The interpolations above were done along a line. We also tested interpolations in a triangle, as shown in Figure 8. The clusterhead (the listener) was placed at location 1 and the profile of location 5 was interpolated using those of location 2,3, and 6. Figure 9(a) shows the KL distance from the interpolated profile of location 5 to the empirical profile of itself and other locations. The figure shows that under linear interpolation, the distance to self is clearly lower than those to others — the plot has a clear minimum at location 5 — while the distinction under Gaussian interpolation is less pronounced. Finally, Figure 9(b) shows the error probability of GHT using different profiles. Consistent with the KL distance, the linear interpolated profile out-performs the Gaussian one significantly.

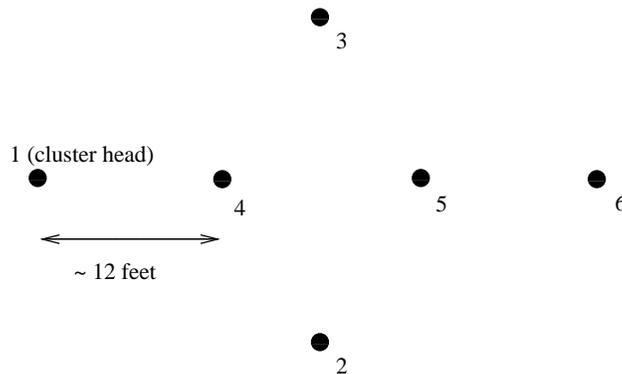
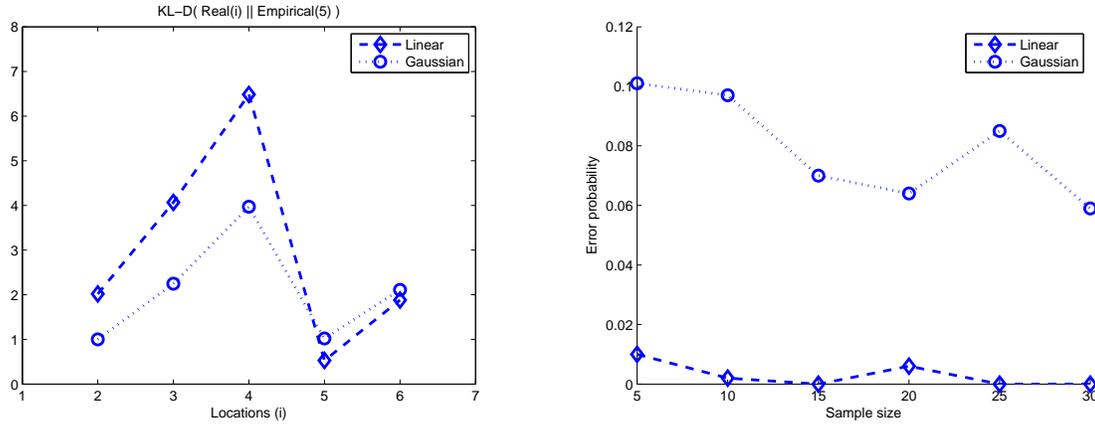


Fig. 8. Layout of the experiment. Node 1 is the clusterhead (listener).

### C. Change of signal profiles over time

Consideration of short-term changes such as the effect of people walking around has already been built into our model. With that taken care of, we can consider the signal profiles to be



(a) KL distances of the empirical pdfs to the interpolated pdf of location 5 (a clear minimum at 5 is desired). (b) Error rate of movement detection for location 5.

Fig. 9. Comparison of linear interpolation and the Gaussian model in terms of (a) the KL distance, (b) the error rate in GHT.

relatively stable. In the long run, some reprofiling would be inevitable. Our experiences indicate that the profile experiences virtually no change for the time scale of one month, but reprofiling may be needed for a period beyond 4 months.

First, to assess the effect of an outdated profile on localization accuracy, we performed a new experiment and used a current profile, an one-month-old profile, and a four-month-old profile. When we used the one-month-old profile the probability of accurately reporting the correct landmark deteriorated by 2.27% compared to the one obtained by using the current profile. On the other hand, when we used the four-month-old profile, the probability of accurately reporting the correct landmark deteriorated by 6.5% compared to the one obtained by using the current profile. Real systems are typically overprovisioned so such a performance deterioration may be tolerable. Of course, the percentage drift in the corresponding probabilities of error was substantially higher but in any case we work with small probabilities of error. Furthermore, even when we make an error the error is small since we report a neighboring landmark in virtually all of the cases.

To gain more insight, we used again the KL distance to measure the drift of signal profiles. Signal profiles of four locations that resemble a neighborhood in our landmark graph were measured first in July, then in the beginning of November (i.e., after 4 months), and then at

TABLE I  
KL DISTANCE OF SIGNAL PROFILE DRIFTING

	KL-D Early Nov vs. Late Nov	KL-D July vs. Early Nov
Location 1	0.54	0.57
Location 2	0.68	2.15
Location 3	0.58	1.12
Location 4	0.40	3.19

TABLE II  
KL DISTANCE BETWEEN LOCATIONS

KL-Distance	Location 1	Location 2	Location 3	Location 4
Location 1	-	4.42	6.44	6.49
Location 2	3.29	-	5.17	6.96
Location 3	6.86	5.48	-	3.41
Location 4	8.01	7.87	3.45	-

the end of November. Table II shows that the KL distances of signal profiles between pairs of locations are on the scale of 3 to 8. On the other hand, Table I shows that, the signal profiles changed by a KL distance of around 0.5 from the beginning to the end of November, which is negligible compared with the pairwise distances. However, the maximum drift in KL distances from July to November is around 3, which is smaller but not substantially smaller than all pairwise distances.

## VI. CONCLUSION

This paper reports a set of carefully designed rules and algorithms that played crucial roles at various junctures of a successful two-tier wireless RSSI-based positioning system. Both theoretical and experimental justifications are provided. In addition to working out the details of the design, the experimental validation of the formal pdf interpolation and that of the single-clusterhead-based movement detection are valuable pieces of new information. It is worth noting that the use of advanced decision theory does not make our approach more difficult to use; quite the contrary. Admittedly, we record full pdf information and use formal optimization to decide

clusterhead placement, which makes the initialization of the system a little more complex (but actually very doable) than other methods. But after writing these pieces of code, deploying the system is very easy exactly because the approach is “formal” at every step, and consequently there is little need for trial-and-error type of adjustments. The fact that our system produced a performance that is in the same ballpark as other state-of-the-art methods validates that our approach is not a mere theoretical exercise. The accuracy we demonstrated is rather favorable, but we recognize that the conditions of experimentation are not strictly comparable with other methods.

## REFERENCES

- [1] J. Caffery and G. Stuber, “Subscriber location in CDMA cellular networks,” *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 406–416, 1998.
- [2] A. Weiss, “On the accuracy of a cellular location system based on RSS measurements,” *IEEE Transactions on Vehicular Technology*, vol. 52, no. 6, pp. 1508–1518, 2003.
- [3] R. Want, A. Hopper, V. Falcao, and J. Gibbons, “The active badge location system,” *ACM Transactions on Information Systems*, vol. 10, pp. 91–102, January 1992.
- [4] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Mobile Computing and Networking*, 2000, pp. 32–43. [Online]. Available: “<http://citeseer.nj.nec.com/priyantha00cricket.html>”
- [5] I. Guvenc, C.-C. Chong, and F. Watanabe, “NLOS identification and mitigation for UWB localization systems,” in *Wireless Communications and Networking Conference (WCNC 2007)*, March 2007, pp. 1571–1576.
- [6] N. Patwari and S. Kasera, “Robust location distinction using temporal link signatures,” in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2007, pp. 111–122.
- [7] P. Bahl and V. Padmanabhan, “RADAR: An in-building RF-based user location and tracking system,” in *Proceedings of the IEEE INFOCOM Conference*. Tel-Aviv, Israel: IEEE, March 2000.
- [8] K. Lorincz and M. Welsh, “Motetrack: A robust, decentralized approach to RF-based location tracking,” *Springer Personal and Ubiquitous Computing, Special Issue on Location and Context-Awareness*, pp. 1617–4909, 2006.
- [9] K. Kaemarungsi and P. Krishnamurthy, “Modeling of indoor positioning systems based on location fingerprinting,” in *Proceedings of the IEEE INFOCOM Conference*, 2004.
- [10] J. Hightower, R. Want, and G. Borriello, “SpotON: An indoor 3d location sensing technology based on RF signal strength,” University of Washington, Department of Computer Science and Engineering, Seattle, WA, UW CSE 00-02-02, February 2000.
- [11] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, “A probabilistic location service for wireless network environments,” in *Proceedings of Ubicomp*. Atlanta, GA: ACM, September 2001.
- [12] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O’Dea, “Relative location estimation in wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

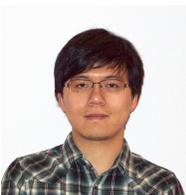
- [13] K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan, “Ecolocation: A sequence based technique for RF-only localization in wireless sensor networks,” in *The Fourth International Conference on Information Processing in Sensor Networks*, Los Angeles, CA, April 2005.
- [14] I. C. Paschalidis and D. Guo, “Robust and distributed localization in sensor networks,” in *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, Louisiana, December 2007, pp. 933–938.
- [15] —, “Robust and distributed stochastic localization in sensor networks: Theory and experimental results,” *ACM Trans. Sensor Networks*, vol. 5, no. 4, pp. 34:1–34:22, 2009.
- [16] M. A. Youssef, “Collection about location determination papers available online,” 2008, [http://www.cs.umd.edu/~moustafa/location\\_papers.htm](http://www.cs.umd.edu/~moustafa/location_papers.htm).
- [17] C. Chang and A. Sahai, “Cramer-Rao-type bounds for localization,” *Journal on Applied Signal Processing*, vol. 2006, p. 113, 2006.
- [18] S. Gezici, “A survey on wireless position estimation,” *Journal Wireless Personal Communications*, vol. 44, no. 3, pp. 263–282, 2008.
- [19] F. H. Bursal, “On interpolating between probability distributions,” *Applied Mathematics and Computation*, vol. 77, pp. 213–244, 1996.
- [20] S. Ray, W. Lai, and I. C. Paschalidis, “Statistical location detection with sensor networks,” *Joint special issue IEEE/ACM Trans. Networking and IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2670–2683, 2006.
- [21] T. Cover and J.A.Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc, 1991.
- [22] W. Hoeffding, “Asymptotically optimal tests for multinomial distributions,” *Ann. Math. Statist.*, vol. 36, pp. 369–401, 1965.
- [23] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. NY: Springer-Verlag, 1998.
- [24] M. Daskin, *Network and Discrete Location*. Wiley, New York, 1995.
- [25] F. Özsoy and M. Pinar, “An exact algorithm for the capacitated vertex p-center problem,” *Computers and Operations Research*, vol. 33, no. 5, pp. 1420–1436, 2006.



**Keyong Li** received his BS from the University of Science and Technology, China, in 1999. He received his MS and Ph.D. degrees from Boston University in September 2001 and January 2006, respectively. From August 2005 to August 2007 he was a postdoctoral associate at the department of Mechanical Engineering at Cornell University. From August 2007 to January 2010, he was a postdoctoral associate at the Center for Information and Systems Engineering at Boston University, where he applied his knowledge of optimization, control, and stochastic systems to a wide range of applications from sensor networks to protein binding computations. Since February 2010, he has been a financial software developer in Bloomberg L.P.



**Dong Guo** acquired his Ph.D. from the Division of Systems Engineering at Boston University in 2009. He received his M.E. and B.E. degrees from the Department of Automation, Tsinghua University, China, in 2004 and 2001 respectively. His research interests include localization and optimization problems in wireless sensor networks.



**Yingwei Lin** is a Ph.D. candidate in Electrical and Computer Engineering at Boston University. His research interests covers wireless sensor networks, dynamic control, embedded communication systems, software engineering, and project management. He completed his undergraduate education in Software Engineering at Tongji University, Shanghai China, in 2007.



**Ioannis Ch. Paschalidis** (M'96, SM'06) is a Professor at Boston University with appointments in the Department of Electrical and Computer Engineering and the Division of Systems Engineering. He is a Co-Director of the Center for Information and Systems Engineering (CISE) and the Academic Director of the Sensor Network Consortium. He completed his graduate education at the Massachusetts Institute of Technology (MIT) receiving an M.S. (1993) and a Ph.D. (1996), both in Electrical Engineering and Computer Science. In September 1996 he joined Boston University where he has been ever since. He has held visiting appointments with MIT, and the Columbia University Business School. His current research interests lie in the fields of systems and control, networking, applied probability, optimization, operations research and computational biology.