

Model-Free Probabilistic Localization of Wireless Sensor Network Nodes in Indoor Environments*

Ioannis C. Paschalidis^{1,2}, Keyong Li¹, and Dong Guo¹

¹ Center for Information & Systems Engineering

² Dept. of Electrical & Computer Eng., and Division of Systems Eng.
Boston University, 15 St. Mary's St., Brookline, MA 02446

yannisp@bu.edu

<http://ionia.bu.edu/>

Abstract. We present a technique that makes up a practical probabilistic approach for locating wireless sensor network devices using the commonly available signal strength measurements (RSSI). From the RSSI measurements between transmitters and receivers situated on a set of landmarks, we construct appropriate probabilistic descriptors associated with a device's position in the contiguous space using a pdf interpolation technique. We then develop a localization system that relies on these descriptors and the measurements made by a set of clusterheads positioned at some of the landmarks. The localization problem is formulated as a composite hypothesis testing problem. We develop the requisite theory, characterize the probability of error, and address the problem of optimally placing clusterheads. Experimental results show that our system achieves an accuracy equivalent to $95\% < 5$ meters and $87\% < 3$ meters.

1 Introduction

A reliable indoor positioning service gives rise to a plethora of important applications ranging from asset tracking to disaster response. The GPS technology is hardly operational for indoor use. Many other ideas have been investigated, some of which may require special hardware/infrastructure. Our primary interest is in methods that allow us to add the positioning service to an existing wireless sensor network (WSNET), using only the basic measurements of the radio frequency (RF) communications in the WSNET — specifically, the measurements of the received signal strength indication (RSSI).

For a brief (hence, incomplete) review of the RF-based positioning literature, the systems proposed by [1, 2, 3, 4] compare mean RSSI measurements to a pre-computed signal-strength map. These systems succeeded in demonstrating the feasibility of providing meaningful positioning services using WSNETs and injected enthusiasm into the field. However, their performances leave room for

* Research partially supported by the NSF under grants DMI-0330171, ECS-0426453, EFRI-0735974, and by the DOE under grant DE-FG52-06NA27490.

improvement. Many other works followed. [5] improved upon [1] by taking the probabilistic nature of the problem into account. Another class of systems such as [6, 7] use stochastic triangulation techniques relying on some path loss model, in which the modeling error can lead to inaccuracy. In contrast, our approach is model-free and based solely on actual measurements. Our earlier related work has been shown to reduce the mean error distance by a factor of 3.5 compared to stochastic triangulation; see [8]. References to many other systems can be found in [9].

Despite the rich literature, some fundamental questions remain. The present paper not only describes a successful positioning system, but also suggests a set of formal techniques that proved to work well in the real setting. Our approach is stochastic in nature. Localization is done relative to a *landmark graph*, whose nodes are a chosen set of landmarks, and whose edges exist between any two nodes if the corresponding landmarks are 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.

Choosing not to assume any model that describes signal propagation and postulates a way in which signal strength decreases with distance, our system is solely based on the measurements obtained at the landmarks. Naturally, we would want to use as few measurements as possible to achieve a desirable level of accuracy. In essence, our work suggests that the accuracy achieved is on the order of the landmark density and this provides a rule-of-thumb for designing a localization system. The interpolation technique between probability distributions we use in constructing location profiles is aimed at “generalizing” the discrete measurements at our disposal into descriptors that can cover a broader area and are not so sensitive to the exact position at which the measurements are taken.

As mentioned, the accuracy we achieve substantially outperforms stochastic triangulation approaches. It should be judged bearing in mind the density of the clusterheads used by our system. We note that the ratio of (possible discrete) locations to clusterheads is relatively high in our system compared to alternative approaches, implying that the deployment cost is low. Proximity-based systems, for instance, require many more clusterheads to achieve the level of accuracy we report.

Notation. We use bold lower case letters for vectors, bold upper case letters for matrices, and T denotes transpose. 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.

2 Problem Formulation

Consider the problem of locating a wireless sensor network device in a contiguous space \mathcal{X} , which typically corresponds to some indoor environment. First, we map this space with landmarks and areas connecting the landmarks — in other

words, a landmark graph. Denote the set of landmarks (nodes of the graph) by $\mathcal{V} = \{V_i \mid i = 1, \dots, M\}$, and the areas between landmarks (the edges) by $\mathcal{E} = \{E_{ij} \mid i = 1, \dots, M, j > i, V_j \in \mathcal{N}_i\}$, where \mathcal{N}_i is the set of neighboring landmarks of V_i . With a slight abuse of notation, we sometimes also write $j \in \mathcal{N}_i$ if $V_j \in \mathcal{N}_i$ and $(i, j) \in \mathcal{E}$ 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 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 $Y^{(k)} \in \{\eta_1, \dots, \eta_H\}$ be the RSSI received at landmark k . We then have a collection of empirical distributions:

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

where k is the index of the receiving landmark and i is the index of the transmitting landmark. 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 a pdf of RSSI that characterizes the signals transmitted from each location and received at each landmark. In fact, for improved robustness we associate with each location a family of pdfs parametrized by vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{ij}$, respectively. These are the location descriptors or profiles:

$$p_i^{(k)}(\cdot | \boldsymbol{\theta}_i), \quad i = 1, \dots, M, \quad k = 1, \dots, M; \quad (2)$$

$$p_{ij}^{(k)}(\cdot | \boldsymbol{\theta}_{ij}), \quad (i, j) \in \mathcal{E}, \quad k = 1, \dots, M. \quad (3)$$

In the above, the pdf families listed correspond to the nodes (cf. (2)) and the edges (cf. (3)) of the landmark graph, respectively. Equivalently, we may list the pdf families in terms of the locations, with the notation

$$p_{Y^{(k)} | \boldsymbol{\theta}_l}(\cdot), \quad l = 1, \dots, N, \quad k = 1, \dots, M, \quad (4)$$

where l corresponds to a location — either a node or an edge. The former notation will be used when we discuss profiling, while the latter will be used in localization. Clusterhead placement will place $K \leq M$ clusterheads (one can think of a limited clusterhead “budget”) at some of the landmarks, which will listen to the signals transmitted by the wireless device. Localization is done by “comparing” the clusterheads’ RSSI measurements with the location profiles.

3 Profiling

This section focuses on how to generate the location profiles (2), (3) using the empirical RSSI distributions (1). The key technique is the interpolation of pdfs.

3.1 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 the statistical physics literature [10], 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 $\boldsymbol{\rho} \in \mathbb{R}^K$ with elements $\rho_1, \rho_2, \dots, \rho_K \in [0, 1]$ satisfying $\sum_{i=1}^K \rho_i = 1$. We are now seeking an interpolation $p_{\boldsymbol{\rho}}(x)$, whose mean and variance are $\mu_{\boldsymbol{\rho}} = \sum_{i=1}^K \rho_i \mu_i$ and $\sigma_{\boldsymbol{\rho}}^2 = \sum_{i=1}^K \rho_i \sigma_i^2$. Let

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

When the random variable takes discrete values, an issue is that the transformation $\xi(x)$ may produce a value for which probability is not defined. An approximate formula that solves this issue is also provided as follows. 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

$$\gamma_{ijl} = \max \left\{ \begin{array}{l} 0, \\ \min \left\{ \frac{\xi_i(j+0.5)}{l+0.5} \right\} - \max \left\{ \frac{\xi_i(j-0.5)}{l-0.5} \right\} \end{array} \right\}, \quad (5)$$

The interpolation formula is then

$$p_{\boldsymbol{\rho}}(j) = \sum_{i=1}^K \rho_i \sum_l \gamma_{ijl} \cdot p_i(l). \quad (6)$$

We call this formula the *linear interpolation*. Similarly to [10], one can prove that $\mu_{\boldsymbol{\rho}}$ and $\sigma_{\boldsymbol{\rho}}^2$ are indeed the mean and variance corresponding to $p_{\boldsymbol{\rho}}(x)$. From here on, we denote the linear interpolation of the K pdfs with the coefficient vector $\boldsymbol{\rho} \in \mathbb{R}^K$ by $\text{Interpol}(\boldsymbol{\rho}, p_1, p_2, \dots, p_K)$.

3.2 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.

First, we “regularize” the empirical pdfs to get rid of zero elements. This is necessary because the size of our sample during profiling is finite. As a result, some RSSI value η_h that is possible but rare for a location L_i might not be observed during profiling, leaving the h 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 η_h 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 solve 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 μ and variance σ^2 . Let $\phi(\mu, \sigma^2)$ be a Gaussian-like pdf whose domain is discretized to the set $\{\eta_1, \dots, \eta_H\}$. Let $\gamma \in (0, 1)$ be a chosen mixing factor — typically we set γ to a small value such as 0.1 or 0.2. Then the pdf after regularization is $\tilde{q} = (1 - \gamma)q + \gamma\phi(\mu, \sigma^2)$.

Second, the landmarks are characterized by pdf-families constructed using interpolation. See [11] for more discussion on the robustness of the pdf-family framework. However, [11] did not provide a formal technique for constructing the pdf-families. Specifically, suppose V_i has I neighbors — $\mathcal{N}_i = \{V_{j_1}, \dots, V_{j_I}\}$. Let $\boldsymbol{\rho}_{\theta_i} = (1 - \sum_{j=1}^I \theta_i^{(j)}, \theta_i^{(1)}, \dots, \theta_i^{(I)})^T$, where $\boldsymbol{\theta}_i \in \mathbb{R}^I$, $\boldsymbol{\theta}_i \geq \mathbf{0}$ elementwise, and $\sum_{j=1}^I \theta_i^{(j)} < 1$. Then the pdf family associated with V_i can be defined as an interpolation of $I + 1$ empirical pdfs:

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

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 ϑ_{ij} . The pdf family associated with edge (i, j) is

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

3.3 An Alternative Gaussian Model

In the above, we focused on associating 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.

4 Localization System Design

4.1 Binary Composite Hypothesis Testing

We start our analysis by considering the simpler problem of using a single clusterhead at landmark V_k to localize a device whose location is either L_i or L_j . The pdf families associated with the two locations are $p_{\mathbf{Y}^{(k)} | \boldsymbol{\theta}_i}(\mathbf{y})$ and $p_{\mathbf{Y}^{(k)} | \boldsymbol{\theta}_j}(\mathbf{y})$, respectively. The clusterhead makes n i.i.d. observations $\mathbf{y}^{(k), n} = (\mathbf{y}_1^{(k)}, \dots, \mathbf{y}_n^{(k)})$.

The likelihood of obtaining these measurements if L_i is the true location is $p_{\mathbf{Y}^{(k)}|\theta_i}(\mathbf{y}^{(k),n}) = \prod_{l=1}^n p_{\mathbf{Y}^{(k)}|\theta_i}(\mathbf{y}_l^{(k)})$.

The problem at hand is a binary composite hypothesis testing problem for which the *Generalized Likelihood Ratio Test (GLRT)* is commonly used. The GLRT compares the normalized generalized log-likelihood ratio

$$X_{ijk}(\mathbf{y}^{(k),n}) = \frac{1}{n} \log \frac{\sup_{\theta_i \in \Omega_i} p_{\mathbf{Y}^{(k)}|\theta_i}(\mathbf{y}^{(k),n})}{\sup_{\theta_j \in \Omega_j} p_{\mathbf{Y}^{(k)}|\theta_j}(\mathbf{y}^{(k),n})}$$

to a threshold λ , and declares L_i whenever

$$\mathbf{y}^{(k),n} \in \mathcal{S}_{ijk,n}^{GLRT} \triangleq \{\mathbf{y}^n \mid X_{ijk}(\mathbf{y}^n) \geq \lambda\},$$

and L_j otherwise. There are two types of error (referred to as type I and type II, respectively) with probabilities

$$\begin{aligned} \alpha_{ijk,n}^{GLRT}(\theta_j) &= \mathbf{P}_{\theta_j}[\mathbf{y}^{(k),n} \in \mathcal{S}_{ijk,n}^{GLRT}], \\ \beta_{ijk,n}^{GLRT}(\theta_i) &= \mathbf{P}_{\theta_i}[\mathbf{y}^{(k),n} \notin \mathcal{S}_{ijk,n}^{GLRT}], \end{aligned}$$

where $\mathbf{P}_{\theta_j}[\cdot]$ (resp. $\mathbf{P}_{\theta_i}[\cdot]$) is a probability evaluated assuming that L_j (resp. L_i) is the true location. We will use the term *exponent* to refer to the quantity $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}[\cdot]$ for some probability $\mathbf{P}[\cdot]$; if the exponent is d then the probabilities approaches zero as e^{-nd} .

For any sequence of observations $\mathbf{y}^n = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, the empirical measure (or type) is given by $\mathbf{L}_{\mathbf{y}^n} = (L_{\mathbf{y}^n}(\sigma_1), \dots, L_{\mathbf{y}^n}(\sigma_{|\Sigma|}))$, where

$$L_{\mathbf{y}^n}(\sigma_i) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{\mathbf{y}_j = \sigma_i\}, \quad i = 1, \dots, |\Sigma|,$$

and $\mathbf{1}\{\cdot\}$ denotes the indicator function. We will denote the set of all possible types of sequences of length n by $\mathcal{L}_n = \{\boldsymbol{\nu} \mid \boldsymbol{\nu} = \mathbf{L}_{\mathbf{y}^n} \text{ for some } \mathbf{y}^n\}$ and the type class of a probability law $\boldsymbol{\nu}$ by $T_n(\boldsymbol{\nu}) = \{\mathbf{y}^n \in \Sigma^n \mid \mathbf{L}_{\mathbf{y}^n} = \boldsymbol{\nu}\}$. Last, recall that the Kullback-Leibler (KL) distance of $\boldsymbol{\nu}$ from another pdf $\boldsymbol{\mu}$ is

$$D(\boldsymbol{\nu} \parallel \boldsymbol{\mu}) = \sum_{i=1}^{|\Sigma|} \nu(\sigma_i) \log \frac{\nu(\sigma_i)}{\mu(\sigma_i)}. \quad (7)$$

In our previous work [8] we derived bounds on the type I and type II error probability exponents:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \alpha_{ijk,n}^{GLRT}(\theta_j) \leq -\lambda, \quad (8)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \beta_{ijk,n}^{GLRT}(\theta_i) \leq - \inf_{\mathbf{Q} \in \mathcal{D}_{ijk}} D(\mathbf{Q} \parallel \mathbf{P}_{\theta_i}), \quad (9)$$

for all $\theta_j \in \Omega_j$ and $\theta_i \in \Omega_i$, where

$$\mathcal{D}_{ijk} = \{\mathbf{Q} \mid \inf_{\theta_j} D(\mathbf{Q} \parallel \mathbf{P}_{\theta_j}) - \inf_{\theta_i} D(\mathbf{Q} \parallel \mathbf{P}_{\theta_i}) < \lambda\}.$$

4.2 Determining the Optimal Threshold

It can be seen from (8) and (9) that the exponent of the type I error probability is increasing with λ but the exponent of the type II error probability is nonincreasing with λ . We have no preference between the two types of error, thus, we wish to balance the two exponents and determine the value of λ at which they become equal.

The exponent of the type I error is simply obtained from (8). The type II error exponent from (9) is equivalent to

$$Z_{ijk}(\lambda, \boldsymbol{\theta}_i) = \min_{\mathbf{Q}} D(\mathbf{Q} \parallel \mathbf{P}_{\boldsymbol{\theta}_i}) \quad (10)$$

s.t. $\min_{\boldsymbol{\theta}_j} D(\mathbf{Q} \parallel \mathbf{P}_{\boldsymbol{\theta}_j}) - D(\mathbf{Q} \parallel \mathbf{P}_{\boldsymbol{\theta}_i}) \leq \lambda, \forall \boldsymbol{\theta}_i.$

The worst case exponent over $\boldsymbol{\theta}_i \in \Omega_i$ is given by

$$Z_{ijk}(\lambda) = \min_{\boldsymbol{\theta}_i} Z_{ijk}(\lambda, \boldsymbol{\theta}_i).$$

Problem (10) is nonconvex; we use dual relaxation to obtain a quantity that is easier to compute. Let $\bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$ be the optimal value of the dual of (10); by weak duality it follows $Z_{ijk}(\lambda, \boldsymbol{\theta}_i) \geq \bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$. It can be verified that there exists a $\lambda_{ijk}^* > 0$ such that $\bar{Z}_{ijk}(\lambda_{ijk}^*) = \lambda_{ijk}^*$. Furthermore, both error exponents in (8) and (9) are no smaller than λ_{ijk}^* .

Now suppose the clusterhead at V_k has obtained the measurements $\mathbf{y}^{(k),n}$ and seeks to decide the device location between L_i and L_j . The clusterhead has the option of using the GLRT by comparing $X_{ijk}(\mathbf{y}^{(k),n})$ to the threshold λ_{ijk}^* , or comparing $X_{jik}(\mathbf{y}^{(k),n})$ to a threshold λ_{jik}^* that can be obtained in exactly the same way as λ_{ijk}^* . We thus let

$$d_{ijk} = \max\{\lambda_{ijk}^*, \lambda_{jik}^*\}, \quad (11)$$

and set $(\bar{i}, \bar{j}) = (i, j)$ if λ_{ijk}^* is the maximizer above; otherwise set $(\bar{i}, \bar{j}) = (j, i)$. Define the maximum probability of error as

$$P_{ijk,n}^{(e)} \triangleq \max\left\{\max_{\boldsymbol{\theta}_j} \alpha_{\bar{i}\bar{j}k,n}^{GLRT}(\boldsymbol{\theta}_j), \max_{\boldsymbol{\theta}_i} \beta_{\bar{i}\bar{j}k,n}^{GLRT}(\boldsymbol{\theta}_i)\right\}.$$

The following result provides a performance guarantee.

Proposition 1. *Suppose that the clusterhead at V_k uses the GLRT and compares $X_{\bar{i}\bar{j}k}(\mathbf{y}^{(k),n})$ to d_{ijk} . Then, the maximum probability of error satisfies*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_{ijk,n}^{(e)} \leq -d_{ijk}.$$

4.3 Multiple Composite Hypothesis Testing

We assume without loss of generality that the clusterheads $1, 2, \dots, K$ are placed at positions V_1, V_2, \dots, V_K . Let d_{ijk} be the GLRT threshold obtained in Sec. 4.2 for each location pair (i, j) , $i < j$, and clusterhead k .

We make $N - 1$ binary decisions with the GLRT. Specifically, we first compare L_1 with L_2 to accept one hypothesis, then compare the accepted hypothesis with L_3 , so on and so forth. For each one of these L_i vs. L_j decisions we use a single clusterhead k as detailed in Sec. 4.1, hence the exponent of error probability is bounded by d_{ijk} . All in all we make $N - 1$ binary hypothesis tests, each involving a single (potentially different) clusterhead. These clusterheads can collaborate in a distributed fashion as we have shown in [11] to make the final decision.

4.4 Clusterhead Placement

Being able to optimize clusterhead placement is one important benefit of our hypothesis testing approach, which produced error bounds that can serve as the criterion. Consider an arbitrary placement of K clusterheads. More specifically, let \mathcal{Y} be any subset of the set of potential clusterhead positions \mathcal{B} with cardinality K . Let $\mathbf{x}(\mathcal{Y}) = (x_1(\mathcal{Y}), \dots, x_M(\mathcal{Y}))$ where $x_k(\mathcal{Y})$ is the indicator function of B_k being in \mathcal{Y} . The objective of clusterhead placement problem is to minimize the worst case probability of localization error, that is, to find ϵ^* as

$$\epsilon^* = \max_{\mathcal{Y}} \min_{\substack{i,j=1,\dots,N \\ i < j}} \max_{k: x_k(\mathcal{Y})=1} d_{ijk}. \quad (12)$$

This combinatorial optimization problem can be rewritten as a mixed integer linear programming problem (MILP). Although it is NP-hard, it can be solved efficiently by using a special purpose algorithm from [12].

We will use the decision rule outlined in Section 4.3 and for every region pair (i, j) we will rely on the clusterhead at $B_{k_{ij}^*}$ to make the corresponding decision. The following theorem establishes a performance guarantee.

Proposition 2. *Place clusterheads according $\mathcal{Y}^* \triangleq \{B_k | x_k^* = 1\}$ and for every (i, j) select one clusterhead with index k_{ij}^* so that $d_{ijk_{ij}^*} = \max_{k: x_k(\mathcal{Y})=1} d_{ijk}$. Then, the worst case probability of error for the decision rule described in Section 4.3, $P_n^{(e),opt}$, satisfies*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n^{(e),opt} \leq -\epsilon^*. \quad (13)$$

5 Experiments

Our testbed is set up on the first floor of a Boston University building (see Fig. 4), and uses MPR2400 (MICAz) motes from Crossbow Technology Inc.

5.1 Testing PDF Interpolations

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 information that resides in the shapes of the empirical pdfs. 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 75 feet apart.

Ideally, and in order to improve accuracy, one would like to place landmarks as close as possible implying that we would need to interpolate between points that are 30 feet (or less) apart. However, it turns out that interpolations over points that are less than 30 feet apart may not be worth the effort. This is consistent with results reported in [2], which have shown that when the spacing of “reference signatures” goes below roughly 10 meters, the improvement in performance diminishes. (The spacing of the “reference signatures” is analogous to the distance between the two end-point locations in our pdf interpolation.) This result reinforces that of [2], 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.

This experiment is conducted in a corridor of roughly 75 feet long, mapped to 6 locations roughly 15 feet apart. A clusterhead (the receiver) is placed at location 1. 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_6 .

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

$$p_{i,\text{short}} = \text{Interpol} \left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, q_{i-1}, q_{i+1} \right), \quad i = 2, 3, 4, 5.$$

Second, in what is labeled “long interpolation”, the interpolated pdfs are generated using q_1 and q_6 :

$$p_{i,\text{long}} = \text{Interpol} \left(\begin{bmatrix} \frac{6-i}{5} \\ \frac{i-1}{5} \end{bmatrix}, q_1, q_6 \right), \quad i = 2, 3, 4, 5.$$

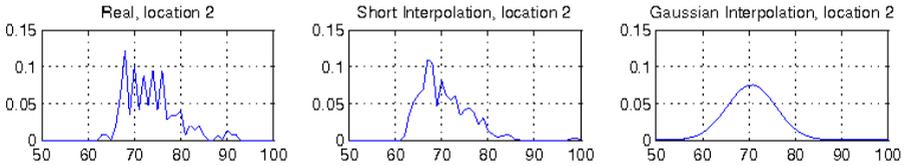


Fig. 1. Visual comparison of interpolated pdfs for location 2. From left to right: empirical pdf, short (linear) interpolation, and Gaussian approximation.

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

$$\begin{aligned}
 p_{i,\text{gaussian}} &= \text{Interpol} \left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \phi(\mu_{i-1}, \sigma_{i-1}^2), \phi(\mu_{i+1}, \sigma_{i+1}^2) \right) \\
 &= \phi \left(\frac{\mu_{i-1} + \mu_{i+1}}{2}, \frac{\sigma_{i-1}^2 + \sigma_{i+1}^2}{2} \right), \quad i = 2, 3, 4, 5.
 \end{aligned}$$

Qualitative Study. In the interest of space, we only visually compare the short interpolation and the Gaussian approximation for location 2 in Fig. 1.

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.

Quantitative Study. First, it is of interest to compare the qualities of the different interpolations using the Kullback-Leibler (K-L) distance (cf. Eq. (7)) as a metric of distance between pdfs. This information theoretic distance is closely related to statistics, including the results derived in the present paper; see Section 4.

The comparison is plotted in Fig. 2. It is very interesting to see that *the quality of short interpolation dominates that of the Gaussian model*. For example, the K-L 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. The long interpolation on the other hand clearly departs significantly from the actual distribution.

5.2 Testing the Complete System

Our localization 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. 4). The landmark graph is then constructed resulting in 39 edges, or a total of 69 locations. Hence $N = 69$, $M = 30$ and $1 \leq K \leq 30$ in this experiment. A mote is placed at the center of 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 beacon message back to server.

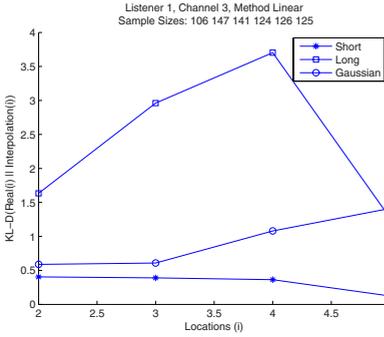
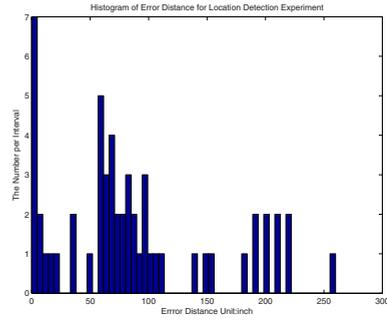


Fig. 2. Kullback-Leibler distance comparison



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

Fig. 3. Localization result

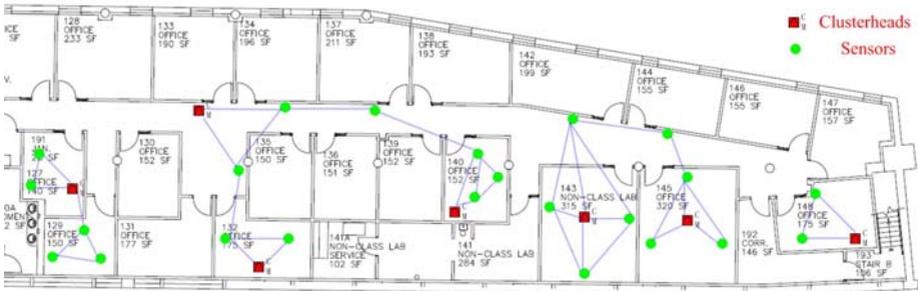


Fig. 4. Floor plan with the landmarks for the testbed

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

Phase 1. We obtained the empirical pdfs for the landmarks corresponding to Eq. (1). With 30 motes placed at each landmark location, we scheduled them such that the motes took turns to broadcast packets, and when one was transmitting the others would listen and record 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 3 to construct the pdf families corresponding to Eqs. (2), (3) which are the descriptors of all 69 locations — both the landmarks and the edges of the landmark graph. Note that the interpolation technique allowed us to construct high quality descriptors without densely covering the area while collecting empirical measurements.

Phase 3. We obtained d_{ijk} as described in Sec. 4.

Phase 4. We solved the MILP to optimize clusterhead placement and simultaneously obtained the performance guarantee (Prop. 2). 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 assigned 7 of the 30 motes as clusterheads. (Note the low clusterhead density needed by our system). The optimal placement is marked in Fig. 4 by the red squares.

Phase 5. We placed motes in the coverage area, let them broadcast messages, then possibly move some of them and let them broadcast again, and finally let the clusterheads report their 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 environment. Phase 2 takes virtually no time. Phase 3 is the most time-consuming part and 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. All these steps only need to be done once.

We know from previous experiences that frequency and power diversity provide better performance [8]. We made 56 localization tests in random positions throughout the covered area. 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 for each combination. We achieved a mean error distance of 87.32 inches, which is better than our earlier result of 96.08 inches [8] 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. 3 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 do not have access thus have not placed a mote) was 1827 feet², that is, about 61 feet² per landmark. With a mean error distance of $\bar{D}_e = 7.3$ feet the mean area of “confusion” was $7.3^2 = 53$ feet². 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 $1827/7 = 261$ feet². 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.

6 Conclusion

The paper reports a landmark-based localization system where each hypothesis is associated with a family of pdfs constructed by a pdf linear interpolation technique. Both theoretical and experimental justifications are provided.

References

1. Bahl, P., Padmanabhan, V.: RADAR: An in-building RF-based user location and tracking system. In: Proceedings of the IEEE INFOCOM Conference, Tel-Aviv, Israel (2000)
2. Lorincz, K., Welsh, M.: Motetrack: A robust, decentralized approach to RF-based location tracking. In: Springer Personal and Ubiquitous Computing, Special Issue on Location and Context-Awareness, pp. 1617–4909 (2006)
3. Kaemarungsi, K., Krishnamurthy, P.: Modeling of indoor positioning systems based on location fingerprinting. In: Proceedings of the IEEE INFOCOM Conference (2004)
4. Hightower, J., Want, R., Borriello, G.: SpotON: An indoor 3d location sensing technology based on RF signal strength. UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA (February 2000)
5. Castro, P., Chiu, P., Kremenek, T., Muntz, R.: A probabilistic location service for wireless network environments. In: Proceedings of Ubicomp, Atlanta, GA. ACM, New York (2001)
6. Patwari, N., Hero, A.O., Perkins, M., Correal, N.S., O’Dea, R.J.: Relative location estimation in wireless sensor networks. *IEEE Transactions on Signal Processing* 51(8), 2137–2148 (2003)
7. Yedavalli, K., Krishnamachari, B., Ravula, S., Srinivasan, B.: 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)
8. Paschalidis, I.C., Guo, D.: Robust and distributed stochastic localization in sensor networks: Theory and experimental results. *ACM Transactions on Sensor Network* 5(4) (November 2009)
9. Youssef, M.A.: Collection about location determination papers (2008), http://www.cs.umd.edu/~moustafa/location_papers.htm
10. Bursal, F.H.: On interpolating between probability distributions. *Applied Mathematics and Computation* 77, 213–244 (1996)
11. Paschalidis, I.C., Guo, D.: 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 (2007)
12. Ray, S., Lai, W., Paschalidis, I.C.: Statistical location detection with sensor networks. Joint special issue IEEE/ACM Trans. Networking and IEEE Trans. Information Theory 52(6), 2670–2683 (2006)