Contents

Chapter 12

On the Performance of Wireless Indoor Localization Using Received Signal Strength

Jie Yang, Stevens Institute of Technology, email: jyang@stevens.edu
Yingying Chen, Stevens Institute of Technology, email: yingying.chen@stevens.edu
Richard P. Martin, Rutgers University, email: rmartin@cs.rutgers.edu
Wade Trappe, Rutgers University, email: trappe@winlab.rutgers.edu
Marco Gruteser, Rutgers University, email: gruteser@winlab.rutgers.edu

This chapter focuses on Received Signal Strength (RSS) based indoor positioning. It first provides an overview of various indoor localization techniques employing RSS including lateration methods, machine learning classification, probabilistic approaches, and statistical supervised learning techniques. It then compares their performance in terms of localization accuracy through measurement studies in real office building environments under representative WiFi and ZigBee wireless networks. It further surveys techniques and methods that are developed to support robust wireless localization and improve localization accuracy, including real-time RSS calibration via anchor verification, closely spaced multiple antennas, taking advantage of robust statistical methods to provide stability to contaminated measurements, utilizing linear regression to characterize the relationship between RSS and the distance to anchors, and exploiting RSS spatial correlation. Finally, it concludes with a discussion of popular location-based applications.

12.1 Introduction

The widespread deployment of indoor wireless technologies is resulting in a variety of location-based services. For instance, in the public arena doctors want to use location information to track and monitor patients in medical facilities and first responders need to track each other and locate victims during emergency rescues. On the other hand, in the enterprise domain, location-based access control is desirable for accessing proprietary corporate materials in restricted areas or rooms. For example, during corporate meetings, certain documents can only be received by laptops reside within the involved conference rooms, which requires location-aware content delivery. In addition, asset tracking also relies on location information. To ensure the wide deployment of these pervasive applications, accurate positioning is important as the location is a critical input to many high-level tasks supporting these applications. In addition, in indoor environments, such as shopping malls, hospitals, warehouses, and factories, where Global Positioning System (GPS) devices generally do not work, indoor localization systems promise the benefits of accurate location estimates for wireless devices such as handheld devices, electronic badges and laptop computers.

Compared to various physical modalities for localization, such as Time of Arrival (ToA) [1], Time Difference of Arrival (TDoA) [2], and Angle of Arrival (AoA), using the Received Signal Strength (RSS) [3–5] is an attractive approach to perform localization since it can reuse the existing wireless infrastructures, and thus presents tremendous cost savings over deploying localization-specific hardware. Also, all current standard commodity radio technologies, such as WiFi, ZigBee, active RFID, and Bluetooth provide RSS measurements, and consequently the same algorithms can be applied across different platforms.

Performing RSS-based localization is a challenging task due to multipath effects in unpredictable indoor settings. These effects include shadowing, i.e., blocking a signal, reflection, i.e., waves bouncing off an object, diffraction, i.e., waves spreading in response to obstacles, and refraction, i.e., waves bending as they pass through different mediums. Thus, the RSS measurements will be attenuated in unpredictable ways due to these effects. To tackle these challenges, recent studies have resulted in a plethora of methods for localizing wireless devices using RSS. This chapter provides an overview of various indoor localization techniques employing RSS including lateration methods, machine learning classification, probabilistic approaches, and statistical supervised learning techniques. We then study the localization performance by presenting representative evaluation metrics. We further compare the performance of various localization algorithms in terms of localization accuracy through real office building environments using prevailing WiFi and ZigBee wireless networks.

Furthermore, we survey techniques and methods that are developed to support robust wireless localization and improve localization accuracy, including real-time RSS calibration via anchor verification, multiple closely spaced antennas, taking advantage of the robust statistical methods to provide stability to contaminated measurements, utilizing linear regression to characterize the relationship between RSS and the distance to anchors more accurately, and exploiting RSS spatial correlation. Finally, we conclude the chapter by presenting current and emerging applications that leverage the location information.

12.2 RSS-based Localization Algorithms

An indoor positioning system consists of a set of anchor nodes (e.g. WiFi access points or traffic sniffers) placed at known locations in the area of interest, and a wireless device carried by the person or attached to the object that needs to be localized. During the operation of localization, radio signals are transmitted between the wireless device and multiple anchors. Based on the received wireless signal at either the anchors or the wireless device, a localization algorithm estimates the position of the wireless device. We next describe a generalized localization model that the localization algorithms are based upon to map the observed RSS in signal space to the physical location in the physical space [5].

In a generalized localization model, let us suppose that we have a domain D in two-dimensions, such as an office building, over which we wish to localize wireless devices. Within D, a set of n anchor points are available to assist in localization. A wireless device that transmits with a fixed power in an isotropic manner will cause a vector of n signal strength readings to be measured by the n anchor points. In practice, these n signal strength readings are averaged over a sufficiently large time window to remove statistical variability. Therefore, corresponding to each location in D, there is an n-dimensional vector of signal readings $\mathbf{s} = (s_1, s_2, \cdots, s_n)$ that resides in a range R.

This relationship between positions in D and signal strength vectors defines a fingerprint function $F : D \rightarrow R$ that takes our real world position (x, y) and maps it to a signal strength reading $\mathbf{s}$. F has some important properties. First, in practice, F is not completely specified, but rather a finite set of positions (x_j, y_j) is used for measuring a corresponding set of signal strength vectors $\mathbf{s}_j$. Additionally, the function F is generally one-to-one, but is not onto. This means that the inverse of F is a function G that is not well-defined: There are holes in the n-dimensional space in which R resides for which there is no well-defined inverse.

It is precisely the inverse function G, though, that allows us to perform localization. In general, we will have a signal strength reading $\mathbf{s}$ for which there is no explicit inverse (e.g. perhaps due to noise variability). Instead of using G, which has a domain restricted to R, we consider various pseudo-inverses G_{alg} of F for which the domain of G_{alg} is the complete n-dimensional space. Here, the notation G_{alg} indicates that there are different *algorithmic* choices, i.e., various localization algorithms, for the pseudo-inverse.

We use signal strength to illustrate the generalized localization model is because signal strength is a common wireless signal modality used by a widely diverse set of localization algorithms. For instance, radio-frequency (RF) fingerprinting approaches utilize RSS [3, 6], and many lateration approaches [7] use it as well[1]. In spite of its several meter-level accuracy, using RSS is a natural choice because it can re-use the existing wireless infrastructure and this feature presents a tremendous cost savings over deploying localization-specific hardware. We next provide an overview of a representative set of algorithms using RSS to perform position estimation.

[1] Chapter 15 describes fingerprinting approaches in more detail and Chapter 6 describes lateration approaches in more detail.

12.2.1 Approach Overview

There are several ways to classify localization algorithms that use signal strength: range-based schemes, which explicitly involve the calculation of distances to landmarks; and RF fingerprinting schemes which originated from machine learning classification, whereby a radio map is constructed using prior measurements, and a device is localized by referencing this radio map through classification. We provide an overview of various indoor localization techniques employing RSS including lateration methods, machine learning classification, probabilistic approaches, and statistical supervised learning techniques. They can be further categorized as range-based methods, such as lateration methods and Bayesian Networks, and RF fingerprinting strategies, including fingerprinting matching using machine learning classification and maximum likelihood estimation using probabilistic approaches.

In lateration methods, the RSS measurements are used to perform ranging between the wireless device and anchor points based on fitting a signal propagation model, and then the location of the wireless device can be computed via trilateration.

For the machine learning classification via fingerprinting matching, it first builds *a priori* radio signal strength map of the localization region during the training phase based on the measured signal strengths (i.e. RF fingerprints) at different locations. By comparing the online measured RSS to the pre-constructed signal map using some optimization criterion, the location can be deduced. For example, the minimum Euclidean distance in the signal-strength vector space can be used as an optimization criterion as described in RADAR [3]. In these approaches location estimation is only considered through the value of the RSS measurements.

On the other hand, probabilistic approaches have been proposed, where the RSS is treated as a random variable that may be modeled by a log-normal distribution [8] at a specific location. In these approaches, a signal strength probability distribution based on the RSS measurements is modeled. During the localization phase, the position of the targeted wireless device is estimated using probabilistic methods, such as maximum likelihood estimation.

In addition, Bayesian Networks and Kernel methods (which is discussed in Chapter XX) are methods that utilize the statistical supervised learning approach [9]. In statistical supervised learning, it infers a function from *supervised* training data. The training data consist of a set of training examples including input vectors, such as the RSS from a set of anchors and a desired output value (i.e. the location of the wireless device). The localization algorithms that perform supervised learning analyze the training data and produce an inferred function, which is called a classifier that predicts the location of wireless device for any valid input RSS vectors. We will now examine each class of algorithms in more detail.

12.2.2 Lateration Methods

Lateration is the most common method for deriving the location of a wireless device [10–12]. By estimating the distance from a wireless device to multiple anchors, lateration approaches derive the wireless device's location based on least squares methods.

In particular, RSS is employed to estimate the distance between a wireless device and an anchor. We next show an example on how to use off-line training phase to estimate the propagation model parameters so as to derive ranging information from the measured RSS during runtime localization phase. We note that there are some approaches treat the propagation model parameters as nuisance parameters that are examined along with the position[2].

Example 12.1 Parameter estimation in the signal propagation model

In this example, we show that the radio propagation parameters can be estimated from the training data collected during the off-line training phase.

Solution: During the off-line training phase, RSS samples are collected at various known locations from multiple anchors and distances are calculated from the known locations to anchors. The measured RSS readings and the corresponding distances are then used to fit the signal propagation model based on the signal-to-distance relationship [8, 13]:

$$P(d)[dBm] = P(d_0)[dBm] - 10\gamma \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma, \tag{12.1}$$

where $P(d_0)$ represents the received power of a wireless device at the reference distance d_0, d is the distance between the wireless device and the anchor, γ is the path loss exponent and X_σ is the shadow fading, which follows zero mean Gaussian distribution with σ standard deviation. Given the RSS and distances, linear regression is usually used to fit the propagation model [14].

During the runtime localization phase, there are two steps: *ranging* and *lateration*. In the ranging step, according to the measured RSS from the targeting wireless device and the fitted signal-to-distance relationship (i.e. the propagation model parameters), the distances between the wireless device and multiple anchors can be calculated. In the lateration step, the location of the wireless device can be estimated according to distances between the wireless device and the anchors based on least squares methods. In the literature, there are two popular methods: *Non-Linear Least Square (NLS)* and *Linear Least Square (LLS)*, in the lateration step.

Non-Linear Least Square (NLS): Given the estimated distances d_i from the targeting device to anchors and known positions (x_i, y_i) of the anchors, the position (x, y) of the wireless device can be estimated by finding $(\hat{x}, \hat{y})$ satisfying:

$$(\hat{x}, \hat{y}) = \arg\min_{x,y} \sum_{i=1}^{N} [\sqrt{(x_i - x)^2 + (y_i - y)^2} - {d_i}^2)] \tag{12.2}$$

where N is the number of anchors that are chosen to estimate the position of the wireless device. Non-linear least square can be viewed as an optimization problem where the objective is to minimize the sum of the square error. This is a nonlinear least squares

[2]Chapter 11 describes lateration approaches in more detail.

problem, and usually involves some iterative searching technique, such as gradient descent or Newton method, to get the solution. Moreover, to avoid local minimum, it is necessary to re-run the algorithm using several initial starting points, and as a result the computation is relatively expensive.

Linear Least Square (LLS): The LLS approach linearizes the NLS problem by introducing a constraint in the formulation and obtain a closed form solution of the location estimate. We next show an example of how the NLS problem can be linearized by introducing a constraint in the formulation.

Example 12.2 Transforming the NLS problem to a LLS problem

Solution: Starting with the $N \geq 2$ equations:

$$\begin{aligned} (x_1-x)^2+(y_1-y)^2 &= d_1^2 \\ (x_2-x)^2+(y_2-y)^2 &= d_2^2 \\ &\vdots \\ (x_N-x)^2+(y_N-y)^2 &= d_N^2 \end{aligned} \tag{12.3}$$

Now, subtracting the constraint

$$\frac{1}{N}\sum_{i=1}^{N}[(x_i-x)^2+(y_i-y)^2]=\frac{1}{N}\sum_{i=1}^{N}d_i^2 \tag{12.4}$$

from both sides of each equation, the following set of linear equations can be obtained

$$\begin{gathered} (x_1-\tfrac{1}{N}\sum_{i=1}^{N}x_i)x+(y_1-\tfrac{1}{N}\sum_{i=1}^{N}y_i)y= \\ \tfrac{1}{2}[(x_1^2-\tfrac{1}{N}\sum_{i=1}^{N}x_i^2)+(y_1^2-\tfrac{1}{N}\sum_{i=1}^{N}y_i^2)-(d_1^2-\tfrac{1}{N}\sum_{i=1}^{N}d_i^2)] \\ \vdots \\ (x_N-\tfrac{1}{N}\sum_{i=1}^{N}x_i)x+(y_N-\tfrac{1}{N}\sum_{i=1}^{N}y_i)y= \\ \tfrac{1}{2}[(x_N^2-\tfrac{1}{N}\sum_{i=1}^{N}x_i^2)+(y_N^2-\tfrac{1}{N}\sum_{i=1}^{N}y_i^2)-(d_N^2-\tfrac{1}{N}\sum_{i=1}^{N}d_i^2)]. \end{gathered} \tag{12.5}$$

Therefore, the above can be easily rewritten using the form $\mathbf{A}\mathbf{x}=\mathbf{b}$ with:

$$\mathbf{A}=\begin{pmatrix} x_1-\frac{1}{N}\sum_{i=1}^{N}x_i & y_1-\frac{1}{N}\sum_{i=1}^{N}y_i \\ \vdots & \vdots \\ x_N-\frac{1}{N}\sum_{i=1}^{N}x_i & y_N-\frac{1}{N}\sum_{i=1}^{N}y_i \end{pmatrix} \tag{12.6}$$

and

$$\mathbf{b}=\frac{1}{2}\begin{pmatrix} (x_1^2-\frac{1}{N}\sum_{i=1}^{N}x_i^2)+(y_1^2-\frac{1}{N}\sum_{i=1}^{N}y_i^2) \\ -(d_1^2-\frac{1}{N}\sum_{i=1}^{N}d_i^2) \\ \vdots \\ (x_N^2-\frac{1}{N}\sum_{i=1}^{N}x_i^2)+(y_N^2-\frac{1}{N}\sum_{i=1}^{N}y_i^2) \\ -(d_N^2-\frac{1}{N}\sum_{i=1}^{N}d_i^2) \end{pmatrix}. \tag{12.7}$$

Note that $\mathbf{A}$ is described by the coordinates of anchors only, while $\mathbf{b}$ is represented by the distances to the landmarks together with the coordinates of landmarks. Thus, the estimated location of wireless device using linear least squares is: $\mathbf{x} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{b}$. Due to the subtraction, the solution obtained from the linear equations (12.7) is not exactly the same as the solution of the original nonlinear equation (12.2). The calculation of the linear equations solution requires low computational power and the obtained solution can serve as the starting point for the nonlinear least squares problem. In general, nonlinear searching from the linear estimate produces more accurate results than the linear method at the price of a higher computational complexity.

12.2.3 Classification via Machine Learning

The fingerprinting matching proposed in [3] can be viewed as a machine learning classification method. In [3], a K-nearest neighbors (K-NN) method is used for positioning wireless device based on the closest known locations (i.e., training points) in the preconstructed signal map. Thus, in the K-nearest neighbors method, the location estimation of the wireless device is assigned as the average location of its K nearest locations in the signal map. If $K = 1$, then the position estimate of the wireless device is simply assigned as the location of its nearest known location in the signal map.

In particular, this kind of localization process through matching includes two phases: off-line training phase, which is used to collect the training data for constructing a radio map, and on-line localization phase, which is used to estimate the position of the wireless device based on the pre-built radio map.

During the off-line phase, a mobile transmitter travels to known positions and broadcasts beacons periodically, and the RSS readings at each known position are measured at the set of anchors. The RSS readings together with the locations where the RSS readings are collected are called RF fingerprints, i.e. $\{(x_j, y_j), ss_1, ss_2, ..., ss_N\}$, where (x_j, y_j), with $j = 1, 2, ..., m$, is the location of the mobile transmitter traveling and $ss_i(x_j, y_j)$ with $i = 1, 2, ..., N$, is the RSS readings at the ith anchor. To mitigate the effect of noise, each ss_i is either the averaged value or the median value of multiple measurements collected over a time period. Collecting together the RSS readings from m different locations from a set of N anchors provides a radio map.

During the online localization phase, localization is performed by measuring the targeted wireless device's RSS at each anchor, and the vector of RSS values - a fingerprint, i.e. $\{ss'_1, ss'_2, ..., ss'_N\}$ is compared to the pre-built radio map. In the nearest neighbors method, the record in the radio map whose signal strength vector is the closest in the Euclidean sense to the observed RSS fingerprint is declared to correspond to the location of the transmitter:

$$(\hat{x}, \hat{y}) = \arg \min_{(x_j, y_j)} \sqrt{\sum_{i=1}^{N} (ss'_i - ss_i(x_j, y_j))^2} \tag{12.8}$$

Other versions of this approach return the average position (e.g., centroid) of the top K closest vectors; i.e., K-nearest neighbors (K-NN). For example, if $K = 2$, it takes the closest two candidates and returns the mid-point between them. Other techniques from the field of machine learning classification that have been used for indoor localization include neural networks [15, 16], decision trees [17, 18] and support vector machines [19, 20]. A disadvantage of this approach is that it requires a large number of training points to perform adequately and is labor intensive.

To reduce the training efforts, one approach is to use an Interpolated Map Grid (IMG) is used to build signal maps [4]. Since the quality of the signal map is sensitive to the number of known locations [21], the purpose of using an IMG is to improve the resolution of the signal map so as to obtain better localization accuracy. Directly measuring the RSS at a large number of known locations is expensive, the purpose of the interpolation approach is to improve the quality of the signal map based on the averaged RSS readings from a smaller number of known locations. We next show an example on how to build an IMG from the collected fingerprints.

Example 12.3 Building an IMG signal map

Solution: The area of interest is divided into a regular grid of equal sized tiles. The center of the tile is representative of its location. The tiles are a simple way to map the expected signal strength to locations. Building an IMG is thus similar to "surface fitting"; the goal is to derive an expected fingerprint for each tile from the collected fingerprints. There are several approaches in the literature for interpolating surfaces, such as splines and triangle-based linear interpolation. For instance, triangle-based linear interpolation, which divides the floor into triangular regions using a Delaunay triangulation [4], can be used in IMG. The expected signal strength at the center of each grid can be linearly interpolated.

When performing localization, given the observed RSS readings of a targeting wireless device, Gridded-RADAR uses IMG to build the signal map and returns the (x, y) of the nearest neighbor in the IMG as the one to localize.

12.2.4 Probabilistic Approaches

We next examine the probabilistic approaches, where the RSS is treated as a random variable that can be modeled as a log-normal distribution at a physical location [8], one category of probabilistic methods is maximum likelihood estimation (MLE) [22, 23]. Assuming the targeted wireless device is located at location $L_j = (x_j, y_j)$, given the online observed RSS values $\mathbf{s}$, i.e. $\mathbf{s} = \{ss'_1, ss'_2, ..., ss'_N\}$, the estimated location of the targeted wireless device based on the MLE is given by:

$$(\hat{x}, \hat{y}) = \arg\max_{L_j} \{p(L_j|\mathbf{s})\}, \quad (12.9)$$

where $p(L_j|\mathbf{s})$ denotes the probability that the wireless device is at location L_j.

Using Bayes' rule, the above equation is equivalent to finding the position L_j which maximizes

$$P(L_j|\mathbf{s}) = \frac{P(\mathbf{s}|L_j) \times P(L_j)}{P(\mathbf{s})}. \quad (12.10)$$

Without *a priori* information about the position of the wireless device, we can assume that the probabilities that the wireless device located at different places are equally likely. Therefore, the equation (12.10) can be rewritten as:

$$P(L_j|\mathbf{s}) = c \times P(\mathbf{s}|L_j), \tag{12.11}$$

where $c = P(L_j)/P(\mathbf{s})$ is a constant.

Equation (12.11) can be further simplified by assuming conditional independence of the measurement from all anchors:

$$P(\mathbf{s}|L_j) = P(ss'_1|L_j) \cdot P(ss'_2|L_j) \cdots P(ss'_N|L_j). \tag{12.12}$$

Assuming the RSS measurements at each location follow a log-normal distribution, the expected RSS vector at each location, i.e. $ss_i(x_j, y_j)$, $P(\mathbf{s}|L_j)$ for every location L_j can be computed. Finally, the MLE returns the location L_j which maximizes $P(\mathbf{s}|L_j)$.

Rather than returning a single location, the position estimate can instead be given as an area of confidence. For example, the area-based probability (ABP) method [4] tries to return an area bounded by a pre-defined probability level that the wireless device is within the returned area. In this approach, the pre-defined probability is called the confidence level, which is an adjustable parameter, and is represented by the parameter α. In the ABP method, it first divides the whole area of interest into a set of tiles. The RSS vector of each tile is represented by the expected RSS vector at the center of each. This can be done by using one of the aforementioned interpolation methods such as linear interpolation. Given that the wireless device must be located within the area of interest, i.e. $\sum_{j=1}^{m} P(L_j|\mathbf{s}) = 1$, ABP returns the top probability tiles up to its confidence, α.

12.2.5 Statistical Supervised Learning Techniques

The statistical supervised learning method infers a function from *supervised* training data. A localization algorithm that utilizes statistical supervised learning analyzes the training data and produces an inferred function that predicts the location of a wireless device based on the input RSS vector. We next introduce Bayesian Networks as an example of localization algorithms that employ statistical supervised learning.

Bayesian networks, also called *belief networks* or *probabilistic networks* are graphical models for representing the interaction between variables visually. A Bayesian network is composed of nodes and arcs between the nodes. Each node corresponds to a random variable, v, and has a value corresponding to the probability of the random variable, $P(v)$. If there is a directed arc from node v to node w, this indicates that v has a *direct influence* on w. This influence is specified by the conditional probability $P(w|v)$. The network is a *directed acyclic graph* (DAG), namely, there are no cycles. The nodes and the arcs between the nodes define the structure of the networks, and the conditional probabilities are the *parameters* given the structure.

Adopting a Bayesian Network (BN) for RSS based localization [7], BN encodes the signal-to-distance propagation model into the Bayesian Graphical Model for location estimation. In DAG, the *parents* of a vertex v, $pa(v)$, are those vertices from which

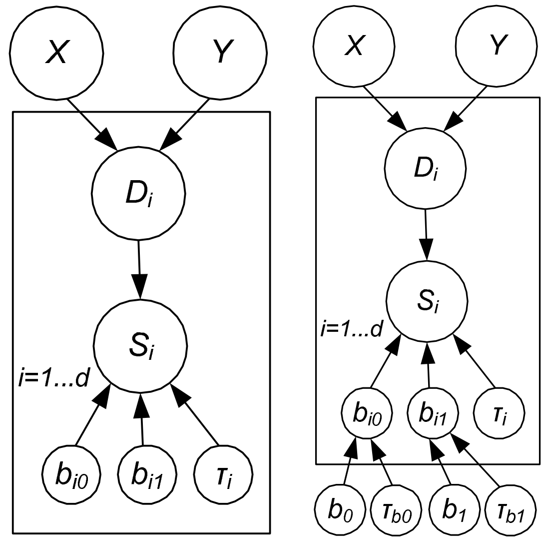


Figure 12.1: Bayesian Networks localization algorithm: Bayesian Graphical Models using plate notation

point into v. The *descendants* of a vertex v are the vertices which are reachable from v along a direct path. A vertex w is called a child of v if there is an edge from v to w. The parent(s) of v are taken to be the only nodes which have direct influence on v, so that v is independent of its non-descendant given its parents. In BN, the overall joint density of $v \in V$, where v is a random variable, only depends on the parents of v, denoted $pa(v)$:

$$p(V) = \prod_{v \in V} p(v_i|\mathrm{pa}(v_i)). \qquad (12.13)$$

Once $p(V)$ is computed, the marginal distribution of any subset of the variables of the network can be obtained as it is proportional to overall joint distribution.

Figure 12.1 presents two Bayesian Network algorithms, M1 and M2. Each rectangle is a plate, and shows a part of the network that is replicated; the nodes on each plate are repeated for each of the N anchors whose locations are known. The vertices X and Y represent location; the vertex s_i is the signal reading from the ith anchor; and the vertex D_i represents the Euclidean distance between the location specified by X and Y and the ith anchor. The value of s_i follows should a signal propagation model $s_i = b_{0i} + b_{1i} \log D_i$, where b_{0i}, b_{1i} are the parameters specific to the ith anchor. The distance $D_i = \sqrt{(X - x_i)^2 + (Y - y_i)^2}$ in turn depends on the location (X, Y) of the measured signal and the coordinates (x_i, y_i) of the ith anchor. The network models noise and outliers by modeling the s_i as a Gaussian distribution around the above propagation model, with variance τ_i:

$$s_i \sim N(b_{0i} + b_{1i} \log D_i, \tau_i). \qquad (12.14)$$

The initial parameters (b_{0i}, b_{1i}, τ_i) of the model are unknown, and the training data is used to adjust the specific parameters of the model according to the relationships encoded in the network. Through Markov Chain Monte Carlo (MCMC) simulation, the BN returns the sampled distribution of the possible location of X and Y as the localization result.

The M1 model utilizes a simple Bayesian Network model, as depicted in Figure 12.1(a), and requires location information in the training set in order to give good localization results. The M2 model extends this to a hierarchical model as shown in Figure 12.1(b), which makes the coefficients of the signal propagation model have common parents. The BN M2 algorithm can localize multiple devices simultaneously with no training set, which can significantly reduce the labor-intensive efforts of training data collection during location estimation.

12.2.6 Summary of Localization Algorithms

We summarize RSS-based localization approaches and their main characteristics in Table 12.1. Compared to the other methods, lateration based approaches are sensitive to RSS noise caused by environmental bias, e.g., multipath effects. On the other hand, the machine learning classification and probabilistic approaches are robust to environmental bias and RSS noise. However, they either need to collect a large number of training data to build RSS profiles and is thus labor intensive, or require prior knowledge of RSS distributions. Concerning localization using statistical supervised learning, it has the advantage of being able to localizing multiple devices at the same time; and can return reasonable localization accuracy even without training.

12.3 Localization Performance Study

The performance of each of the localization techniques needs to be evaluated through various aspects. In this section, we first outline the main evaluation metrics and then present a performance comparison across different localization algorithms using WiFi and ZigBee networks.

12.3.1 Performance Metrics

The set of evaluation metrics that we overview includes: *accuracy*, *precision*, *robustness*, *complexity*, and *stability*.

Accuracy. For a given localization attempt, accuracy is the Euclidean distance between the location estimate obtained from the localization system and the actual location of the targeted wireless device in the physical space. Accuracy is also referred to as localization error or distance error. Usually, the average or median distance error is adopted as the performance metric. The better the accuracy, the better the location technique. Thus, accuracy can be used to evaluate the overall performance of the localization techniques.

Precision. Precision is defined as the success probability of position estimations with respect to the predefined accuracy. Precision details the statistical characterization

Table 12.1: Summary of RSS-based localization approaches and their main characteristics.

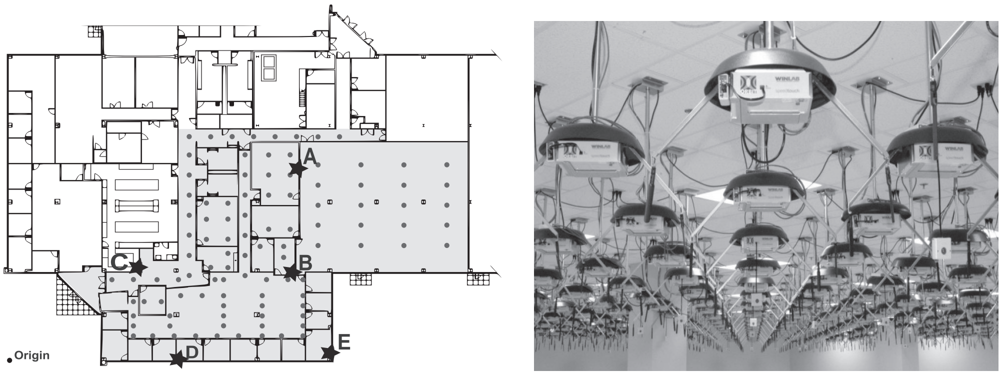
Approach	**Description**	**Algorithm Example**
Lateration	- Two steps are involved: ranging and lateration. During ranging, the distances between the wireless device and multiple anchors can be derived according to different measurement modality. RSS is one of the approaches to be used for ranging. During lateration, the location of the wireless device can be estimated according to derived distances based on least squares methods. - Pros: This method is simple to use and works well in free-space. - Cons: This method is sensitive to RSS variation, e.g., multipath effects.	Nonlinear Least Squares (NLS) and Linear Least Squares (LLS)
Machine Learning Classification	- It builds a priori radio signal map of the localization region during the training phase based on the measured signal strength (i.e. RSS fingerprints) at different locations. During localization, by comparing the online measured RSS to the pre-constructed signal map, the location can be deduced. - Pros: This approach is robust to RSS noise and environmental bias, e.g., multipath effects; - Cons: It is labor intensive to build signal map.	RADAR, K-nearest neighbors (KNN) and Support Vector Machine (SVM)
Probabilistic Approaches	- The RSS is treated as a random variable by modeling it with a log-normal distribution at each location. The position of the wireless device is returned as the most likely location with highest probability. - Pros: This approach is robust to RSS noise and environmental bias, given the priori knowledge of RSS distribution at large number of locations. - Cons: It needs to use the RSS-distribution as priori knowledge, e.g., obtain RSS-distribution through training.	Maximum Likelihood Estimation
Statistical Supervised Learning	- This approach derives the sample distribution of the estimation locations based on statistical learning. - Pros: This approach has the capability of localizing multiple devices at the same time; and returns reasonable localization accuracy even without training data. - Cons: It can be computational intensive.	Bayesian Network

of the localization error, which varies over many localization trials. It measures how consistently the localization technique works and reveals the variation of localization errors. In some works, precision is defined as the standard deviation of the localization error or the geometric dilution of precision (GDOP) [24, 25]. However, generally the Cumulative Distribution Function (CDF) of the localization error is used for measuring the precision of an indoor localization system. When the accuracies of two location algorithms are the same, the algorithm which gives better precision is preferred. In practice, precision is described by either a CDF or the percentile format. For example, one typical indoor location system has a location precision of 97% within 30 feet (the localization error is less than 30 feet with probability of 97%), and median error as 10 feet.

Robustness. A robust positioning should be able to function normally even when some signals are not available, or when some of the RSS patterns haven't been seen before. For example, the signal at an anchor point may be blocked, thus the RSS reading at that anchor point may be missing or incorrect. Sometimes, the environment changes, for example due to furniture relocation, or the presence and mobility of human beings, will change the RSS patterns. In addition, some anchors could be out of function or damaged in a harsh environment. For example, under malicious attacks, the attackers can control anchor points and manipulate the RSS measurements. The positioning techniques have to use this incomplete information to localize the wireless devices. Therefore, it is important to measure the robustness of the localization algorithms. In all these cases, if a localization technique is robust, then it should be able to provide the wireless device's position information without a reduced accuracy even if the quality of the measurement is reduced.

Complexity. Complexity of a localization system can be attributed to hardware, computing, and human intervention/efforts during deployment. The most commonly used measure of complexity is the computing complexity, which is the complexity of the localization algorithm. If the computation of the localization algorithm is performed in a distributed manner (i.e., on the wireless device side), then we would prefer localization algorithms with low complexity, which can extend the operation life of the devices which have the limited battery or power supply. Alternatively, we can use localization latency, the time it takes for a wireless device to localize. The first time and amortized latencies are different. For example, the first localization of a wireless device is long, but subsequent localization attempts in nearby areas are typically much faster.

Stability. Stability measures how much the location estimate changes in the physical space in response to small-scale movements of a wireless device. Stability is a desirable property in localization systems, since a location estimate should not move too far in the physical space if there is a small-scale movement of the wireless device. For instance, when someone works at his office desk and moves his laptop 1 feet away, the localized position of the laptop should not change too much.



(a) WINLAB (b) ORBIT room

Figure 12.2: Deployment of anchors and training locations on the experimental floors [27, 28].

12.3.2 Performance Investigation using Real Wireless Networks

We next compare the performance of different localization algorithms using WiFi and ZigBee wireless networks by focusing on accuracy and precision. We first describe the experimental scenarios in real office building environments. We then examine the experimental results across localization algorithms under different experimental scenarios.

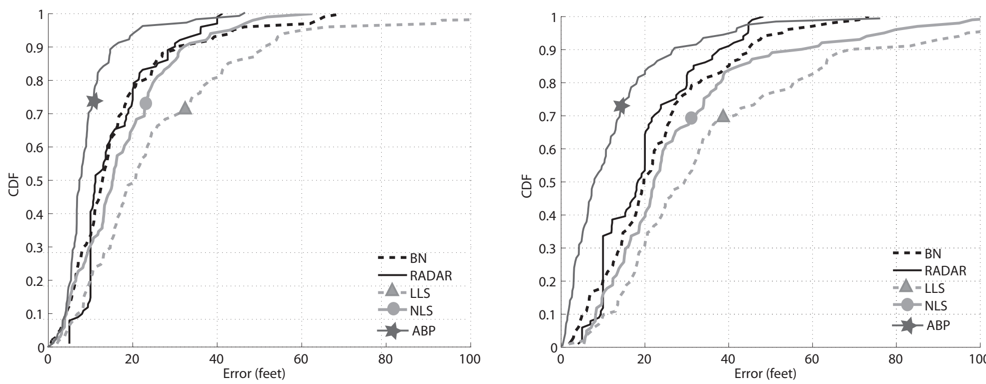
Experimental Scenarios. Figure 12.2 shows two experimental scenarios for localization performance comparison. The floor map in Figure 12.2(a) is the Wireless Network Laboratory (WINLAB) at Rutgers University, which has the floor size of 219ft by 169ft. Both a WiFi network and a ZigBee network are deployed in WINLAB. The ZigBee network is implemented using Tmote Sky motes. The second experimental setup shown in Figure 12.2(b) is the ORBIT testbed in WINLAB [26], a large scale indoor wireless testbed, which consists of 400 wireless nodes in a 20×20 regular grid with an inter-node separation of 3ft spanning a total area of 3600 sq ft. The WiFi network is used in the ORBIT testbed.

The distinctive characteristic between these two experimental scenarios are the number of anchor points and the propagation environments. In the WINLAB office setup, it is a typical indoor propagation environment with heavy multipath effects due to the walls, furniture and people movement. There are 5 anchors shown as stars in Figure 12.2(a) and denoted as A, B, C, D, and E. On the other hand, in the ORBIT testbed, there can be up to 400 anchor points and the propagation environment is free of major shadowing and with limited multipath effects. Each wireless node in Figure 12.2(b) can be configured as an anchor point.

Furthermore, the small dots in Figure 12.2(a) are the training points where the RSS readings are collected at each anchor. There are total 101 training locations in WINLAB setup. On the other hand, in the ORBIT testbed, the RSS readings are measured at each grid point. Thus, there are total 400 training locations in the ORBIT testbed. At each location, about 350 packets of RSS are collected and the averaged RSS for each location

Table 12.2: Performance Study: Localization algorithms used in WINLAB environment.

Algorithm	Abbr.	Description
Non-linear Least Squares	NLS	Lateration based algorithm, Non-linear Least Squares are used
Linear Least Squares	LLS	Lateration based algorithm, Linear Least Squares are used
Fingerprinting Matching	RADAR	Classification based method, Returns the nearest neighbor: $K = 1$
Area Based Probability	ABP	Probabilistic method using IMG, Returns the most likely area: $\alpha = 0.75$ Tile size in IMG: 10inch x 5 inch
Bayesian Network	BN	Statistical supervised learning Bayesian Graphical Model: M1



(a) WiFi network

(b) Zigbee network

Figure 12.3: Performance comparison: localization error CDF across different algorithms under the WINALB experimental scenario [27].

is used. To evaluate the different algorithms, the well-known leave-one-out approach is used to divide the data into training and testing sets. That means one location is chosen as the targeted location whereas the rest of the locations as training data. For example, in the lateration based algorithm, one location is randomly chosen to be localized and the rest of the data (i.e., training data) are used to fit the propagation model (e.g., equation 12.1) to get the propagation parameters, such as path loss parameters.

Performance Results. The algorithms under evaluation in the WINLAB office experimental scenario are: *NLS*, *LLS*, *RADAR*, *ABP*, and *BN*. These algorithms are summarized in Table 12.2.

We note that when using ABP, because it returns the most likely area in which the targeted wireless device may reside, the localization error is defined as the median localization error from the returned area.

The localization results in both WiFi network and ZigBee network are depicted as the Cumulative Distribution Function (CDF) of localization error in Figure 12.3. We observed that the overall performance of localization algorithms under the WiFi network

Table 12.3: Performance Study: Localization algorithms used in ORBIT testbed.

Algorithm	Abbr.	Description
Non-linear Least Squares	NLS	Lateration based algorithm Non-linear Least Squares are used
Fingerprinting Matching Gridded RADAR	GR	Classification based method using IMG Returns the nearest neighbor: $K = 1$ tile size in IMG: 2inch x 2inch
Highest Probability	H1	Probabilistic method using IMG Returns location with the highest probability tile size in IMG: 2inch x 2 inch
Bayesian Network	BN	Statistical supervised learning Bayesian Graphical Model: M1

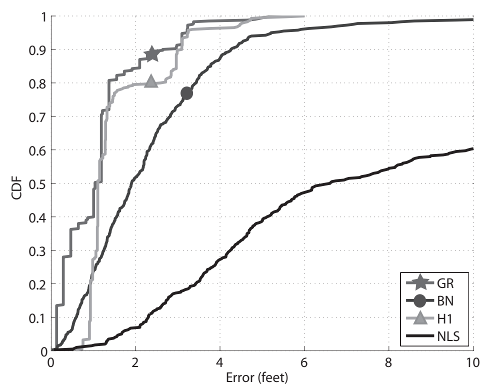


Figure 12.4: Performance comparison: localization error CDF across different algorithms under the ORBIT testbed experimental scenario [28].

is better than that of the Zigbee network due to the fact that WiFi device is more reliable than motes [27]. First, the median error from different algorithms ranges from 7 feet to 20 feet in the WiFi network, whereas it is from 8 feet to 30 feet in the Zigbee network. For both networks, ABP performs the best because it uses a fine-grained signal map via the IMG technique, while lateration based method performs the worst because of the inaccurate ranging information derived from the RSS readings affected by the unpredictable indoor setups. In addition, RADAR and BN have comparable performance in both networks. The median error is about 10 feet in the WiFi network and 20 feet in Zigbee network, respectively.

We now turn to study the algorithms under evaluation in the ORBIT testbed. The algorithms under study are summarized in Table 12.3. Figure 12.4 shows the localization error CDF across different algorithms using the ORBIT testbed with 400 anchors [28]. With the high density of anchors and benign propagation which is essentially LOS, the performance of all the algorithms improves significantly. Again, the algorithms using IMG have the best performance, while the lateration based method underperforms the other algorithms. Particularly, Gridded RADAR (GR) and Highest Probability (H1) have the best performance with median errors of 0.31 m and 0.33 m, respectively. Using

Bayesian Networks (M1) performs slightly worse, with the median error at 0.58 m. NLS performs the worst with median error of about 2 meters. This is because the lateration algorithms generally are very sensitive to data from low-quality anchors among the large number of anchors that cannot be fitted in the propagation model [28].

From the results in these two experimental scenarios, the following important conclusions can be drawn:

1. The IMG technique can improve the localization accuracy of *RADAR*, *ABP*, and *BN* localization algorithms, because it provides a fine-grained signal map that benefits these algorithms.

2. The lateration based method is sensitive to the RSS variations caused by multi-path effects, radio interference or the off-the-shelf device diversity, and have the worst localization performance. This is because the theoretical propagation model either cannot describe complicated indoor environments (e.g., multipath effect) or need to be calibrated due to device diversity (e.g., different transmitting power level), and thus the ranging information derived from RSS together with propagation model is inaccurate.

3. Except the lateration based methods, all the other localization strategies, such as fingerprinting matching, probabilistic, and statistic learning, can achieve comparable localization accuracy and precision in a typical indoor environment.

4. Finally, a benign propagation environment or increasing the density of high-quality anchors can improve the localization performance significantly.

12.4 Enhancing the Robustness of Localization

We next provide an overview of the methods that help to enhance the robustness of the localization results. These methods are either specific to one category of localization algorithms, (e.g., robust statistical methods can be applied to lateration techniques and fingerprinting matching methods, and revisiting linear regression and correlation methods are suitable for lateration methods) or generic to all the localization algorithms, for instance, employing multiple antennas to improve localization accuracy.

12.4.1 Real-time Infrastructure Calibration

In an indoor environment, radio signal propagation is affected by several factors such as multi-path effects, environmental temperature and humidity variations, door's opening and closing, and the mobility of human beings. The RSS readings at a given location will be very different over time since the indoor environment will change from time to time. To overcome the degradation of the localization performance due to environmental changes, it is necessary to calibrate the parameters of the positioning system, such as the propagation parameters and signal maps. However, collecting RSS readings to calibrate the localization system from time to time manually is labor intensive. For example, collecting RSS readings at different locations to update the signal map periodically to

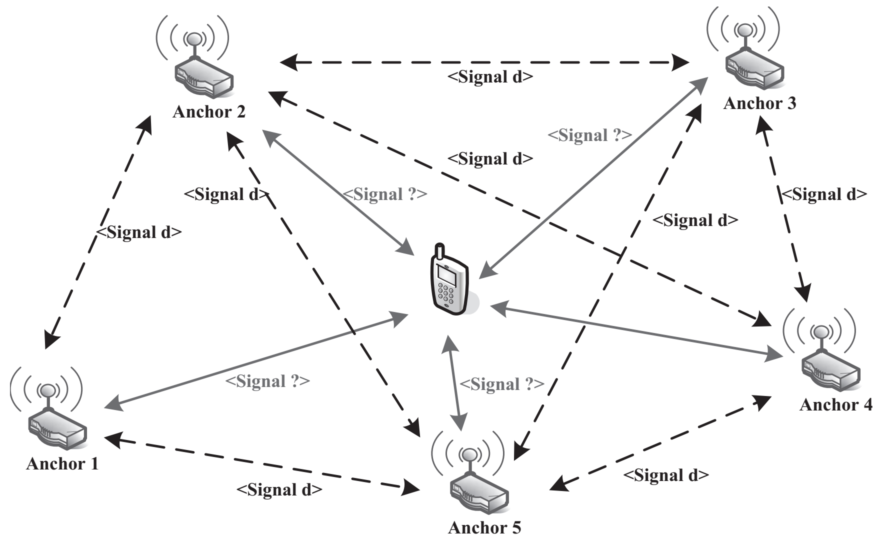


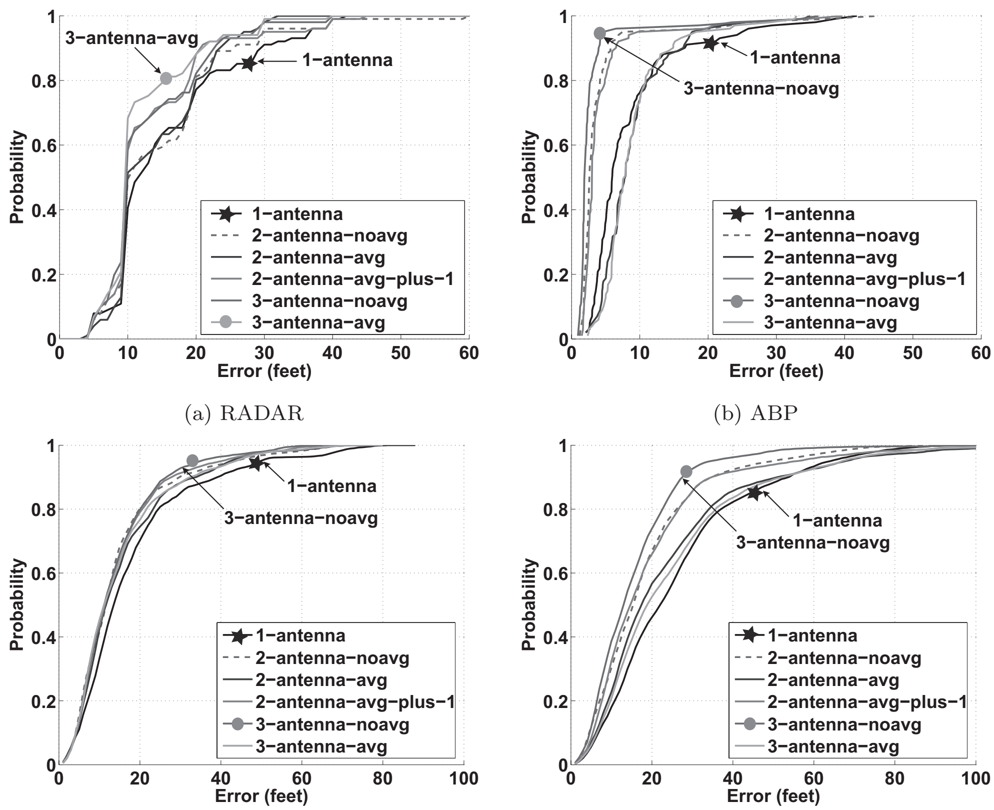
Figure 12.5: System architecture for real-time infrastructure calibration.

maintain high accuracy is extremely cumbersome. Therefore, it is desirable that a localization system is able to adapt to the environmental changes and auto-configure itself in response to environmental dynamics.

The localization system architecture presented in [29] aims to achieve real-time infrastructure calibration [30]. Figure 12.5 depicts the proposed localization system architecture from [29], in which anchors are transceivers with known locations. These anchors record the signal of beacon broadcasts from each other. These signal links between anchors are made periodically to realize fully automated and on-line calibration of the radio signal in the spatia-temporal domain. A mapping that characterizes the relationship of the signal measurements and the geographical distances to anchors is then created on-line with certain mapping rules.

To mitigate the effects of environmental dynamics, besides utilizing the measurements of signals between the wireless devices and anchors to localize the wireless devices, this localization system takes as an additional input the on-line measurements of signal strength between anchors. The on-line signal measurements between the anchors are used to capture the environmental changes, and to create a mapping between the signal measurement and the actual geographical distance. To create a signal-distance mapping, the online measured RSS readings and the corresponding distances between anchors are used to fit the radio propagation model, which is presented in Section 12.1.

In addition, techniques such as truncated the SVD pseudo-inverse method [29] can be used to improve the robustness of signal-distance mapping. Depending on the mode in which localization will be performed, either a wireless device or the localization infrastructure measures the radio signal between the wireless device and its neighboring anchors. Finally, the localization process applies the signal-distance mapping, infers the wireless device's geographical distances to anchors, and estimates its location via



(a) RADAR

(b) ABP

(c) Bayesian Network: M1

(d) Bayesian Network: M2

Figure 12.6: Localization error CDFs when using multiple antennas in the WINLAB experimental scenario under the WiFi network [27].

lateration-based methods [14, 24].

12.4.2 Effects of Employing Multiple Antennas

As mentioned previously, using the RSS to perform localization is attractive since it reuses the existing wireless infrastructure and is expected to provide tremendous cost savings over deploying localization-specific hardware. However, a significant problem with RSS is that small-scale multipath fading adds high frequency components with large amplitudes to the signal at a given location. Thus, the RSS can vary by 5-10 dB with small (a few wavelengths) changes in location. Because the small-scale fading effects occur at the level of several wavelengths (about 12 cm at 2.4 GHz), and the granularity of the localization system is typically much larger (2-3 meters), using multiple antennas spaced on the order of a few wavelengths presents the opportunity to smooth out these effects, while maintaining the same number of anchors used by the localization system [27].

To demonstrate the potential benefits, we next compare the localization accuracy across different localization algorithms when using multiple antennas [31]. Figure 12.6 presents the localization error CDFs of different localization algorithms including RADAR, ABP, and Bayesian Networks when multiple antennas are applied. The experimental scenario is the WINLAB environment under the WiFi network. The observation from Figure 12.6 is that in nearly all cases the performance of localization algorithms improved when using multiple antennas. This is because for RADAR, the collected training points are directly used to build the signal map. The distances of the training points in the experiments range from 5ft to 10ft. The improvement for the RADAR algorithm comes from the reduced RSS variation when averaging RSS readings of three antennas. However, for ABP, a fine-grained interpolated signal map is used, where the tile size is around 10inch per side. When using the fine-grained interpolated signal map, the physical distance between two RSS samples is much less (i.e. several inches) than the distance between two antennas (i.e. 1 to 2 ft in this experimental set) installed on one anchor. Thus, ABP treats each antenna as a separate anchor and achieves the best performance under the case of $3 - antenna - noavg$.

In addition, Bayesian Networks present consistent behaviors when using different Bayesian Graphical Models (i.e. M1 and M2). The performance of $3 - antenna - noavg$ is slightly better than that of others, and the $1 - antenna$ performs the worst. This is due to the fact that Bayesian Networks benefit from the smoothed out small-scale fading when averaging RSS readings of three antennas. In summary, we can conclude that the localization accuracy can be improved when employing multiple antennas in both cases of averaging and not averaging the RSS readings from multiple antennas at each anchor point.

12.4.3 Robust Statistical Methods

In indoor environment, signal strength measurements may be significantly altered by opening doorways in a hallway, or by the presence of passers by. These measurement errors caused by either unintentional human behaviors or intentional adversarial behaviors can be severe, and consequently degrade the performance of a localization method. On the other hand, these measurement errors may only be present on a small subset of anchors (i.e., specific radio links between a wireless device and the affected anchors). The idea of the robust statistical method is to utilize the redundancy in the localization infrastructure, i.e., only a portion of the RSS readings at certain anchors is affected and not all of the RSS readings are affected. The strategy is to enhance the robustness of the localization system so as to mitigate the effects of such measurement errors and adversarial behaviors by taking advantage of the redundancy in the deployment of the localization infrastructure to provide stability to contaminated measurements [32].

Statistical tools are developed to make localization techniques robust to measurement errors and adversarial data. In particular, the methods developed make use of the median as a resilient estimate of the average of aggregated data [33,34]. For example, in lateration methods, rather than minimizing the summation of the residue squares, minimizing the median of the residue squares is used to reduce the effects of measurement outliers. The lateration algorithm is vulnerable to the "outliers".

Given the distance from the wireless devices to the anchors d_i together with the anchors' location (x_i, y_i), the device location estimate $(\hat{x}_0, \hat{y}_0)$ can be found by least squares (i.e. equation (12.2)). In order to achieve accurate localization estimation when there are measurement errors present, LMS can be used instead of least squares. That is, $(\hat{x}_0, \hat{y}_0)$ can be found such that

$$(\hat{x}_0, \hat{y}_0) = \arg \min_{(x_0, y_0)} \mathrm{med}_i [\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} - d_i]^2. \tag{12.15}$$

It is known that LMS tolerates up to 50 percent measurement errors among N total measurements. The exact solution for LMS is computationally prohibitive, however, an efficient and statistically robust alternative can be found in [35].

12.4.4 Revisiting Linear Regression

Radio propagation indoors is complicated by signal reflection, refraction, shadowing and scattering due to the walls, furniture and the movement of people. According to the connectivity between the wireless device and the anchor point, the signal propagation can be classified into line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios [3, 36]. These two scenarios represent different signal propagation environments. Thus, under different scenarios, the propagation parameters are different, such as the path loss exponent and the shadow fading as described in equation (12.1). The training data collected in the area of interest usually includes both LOS and NLOS scenarios and we cannot differentiate which scenario the targeting wireless device belongs to. Thus, in the off line training phase, the fitted theoretical log-distance propagation model can not characterize both LOS and NLOS scenarios simultaneously and will result in large errors in distance estimation during the ranging step of RSS-based lateration methods, and consequently the localization accuracy is significantly affected. In Section 12.3, we observed that RSS-based lateration methods underperform other indoor localization algorithms including machine learning classification techniques, probabilistic approaches, and Bayesian Network localization.

To improve the applicability of RSS-based lateration methods in indoor environments and further provide feasible mathematical analysis for indoor localization, a regression model may be used over the theoretical log-distance propagation model to better model the relationship between the RSS and distance and improve localization accuracy for lateration methods in real-world scenarios [14].

The polynomial regression model is adapted to model the RSS to distance relationship since the polynomials dominate the interpolation theory (Weierstrass's Theorem) and they are easily evaluated [37].

Example 12.4 Using polynomial regression model to model the relationship between the RSS and distance

Solution: Given the M training points (d_i, RSS_i) collected in the area of interest, where d_i is the distance between the wireless device and an anchor point and RSS_i is the corresponding signal strength reading at the training point i. The nth-degree polynomial is fitted through the set of data and M sets of equations are obtained. The ideal nth-degree polynomial should satisfy:

$$\hat{d}_i = a_0 + a_1 * RSS_i + a_2 * RSS_i^2 + ... + a_n * RSS_i^n \tag{12.16}$$

where a_j (with $j = 0, 1, 2, ..., n$) are the coefficients of the polynomial, RSS_i is the received signal strength and $\hat{d}_i$ is the estimated distance. However, there are estimation errors e_i. Thus, we have:

$$e_i = (d_i - \hat{d}_i) = (d_i - a_0 - a_1 * RSS_i - ... - a_n * RSS_i^n), \tag{12.17}$$

where $i = 1, 2, ..., M$.
We use least squares approximation, in which the coefficients can be obtained by minimizing the sum of the error squares, which is given by

$$E(a_0, a_1, ..., a_n) = \sum_{i=1}^{i=M} (e_i)^2 \tag{12.18}$$

The equation (12.18) is a function of variables, $a_0, a_1, ..., a_n$, which minimizes the sum of error squares. We equate its partial derivatives to zero with respect to $a_0, a_1, ..., a_n$. Then, we get

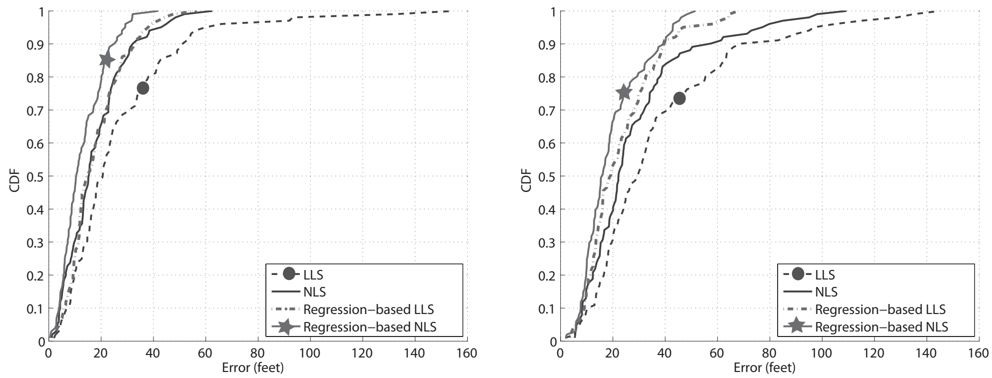
$$\frac{\partial E}{\partial a_j} = \sum_{i=1}^{i=M} -2(RSS_i^j)(d_i - a_0 - a_1 * RSS_i - ... - a_n * RSS_i^n) = 0, \tag{12.19}$$

for each j, with $j = 0, 1, 2, ..., n$. By rearranging equation (12.19), we obtain the normal equations:

$$\begin{aligned} &\sum_{i=1}^{i=M} RSS_i^j d_i = \\ &a_0 \sum_{i=1}^{i=M} RSS_i^j + a_1 \sum_{i=1}^{i=M} RSS_i^{(j+1)} + ... + a_n \sum_{i=1}^{i=M} RSS_i^{(j+n)}, \end{aligned} \tag{12.20}$$

for each j, with $j = 0, 1, 2, ..., n$. Now, the coefficients a_j, with $j = 0, 1, 2, ..., n$, can be solved by Gauss elimination of Equation (12.20).

Figure 12.7 presents the localization results of the regression-based methods using a second degree of polynomial for both the WiFi network and the Zigbee network under the WINLAB experimental scenario respectively. We observed that the regression-based lateration methods result in a better performance than the original log-distance based lateration methods. The overall improvement of median error is above 29% in both networks. In addition, the maximum localization error is significantly reduced. The overall improvement of maximum error is over 50% in two networks. These results indicate that



(a) WiFi network

(b) Zigbee network

Figure 12.7: Regression-based approach: localization error CDFs in two networks [14].

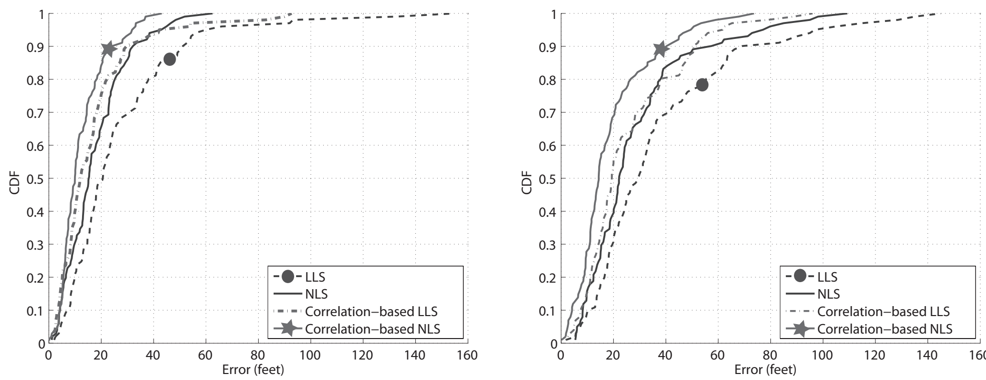
using regression-based lateration methods are effective in indoor localization.

12.4.5 Exploiting Spatial Correlation

Although radio propagation is complicated due to the placement of walls, obstacles and movement of people in indoors, the signal propagation from close-by locations in a local area to an anchor point is highly correlated as nearby locations face a similar propagation environment. For instance, the locations in a regular-sized room are facing the same radio connectivity to the access points (i.e., LOS or NLOS), and similar distance to the anchor points. Thus, the signal propagation from a local area experiences similar signal attenuation (due to distance) and penetration losses through walls. Therefore, the signal propagation model may be better fitted if we just use the data collected in the local area. Further, experimental results have provided strong evidence that shadow fading is spatially correlated in indoors [38] due to the local area facing similar obstacles. The correlation distance can range from several to many tens of meters [36, 38]. Therefore, the correlated RSS measurements collected in a local area can help to characterize the relationship between RSS and the distance, and consequently improve the localization accuracy when applying lateration methods in unpredictable indoor environments [14].

Example 12.5 Using spatial correlation to iteratively refine the localization results

The correlation-based method is developed in [14] by utilizing the correlated RSS readings that are collected from a local area to fit the theoretical log-distance propagation model. The objective is to obtain more accurate distance estimations for the wireless device, whose location belongs to that local area, based on the fitted model. The correlation-based method refines the localization result iteratively by using the correlated RSS in the local area.



(a) WiFi network

(b) Zigbee network

Figure 12.8: Correlation-based method: localization error CDFs in two networks [14].

Solution: The correlation-based method starts with a coarse-grained location estimation by using lateration with all the training data. Then, in the subsequent iterative steps, it adaptively reduces the size of the training data by just using the data that is close to the location estimate from the previous step. The iterative process stops when the estimated location falls outside of the local area where the training data comes from. The correlation method is summarized as follows.

1. **Initialize.** Using all the training points (x_j, y_j) and corresponding RSS $\{RSS_{ij}\}$ at multiple access points (x_i, y_i), with $i = \{1, 2, ..., M\}$ and $j = \{1, 2, ..., N\}$, to fit the propagation model, and obtain the initial location $(\hat{x_0}, \hat{y_0})$ according to the measured RSS $\{RSS_i\}$ of the wireless device using the lateration method.

2. **Iteration.** Refine the wireless device's location estimation $(\hat{x}_k, \hat{y}_k)$, k is the kth iteration step, by using top C_k $(C_k < C_{k-1})$ closest training points to the previous estimated location $(\hat{x}_{k-1}, \hat{y}_{k-1})$ as training data. C_k is the number of training points used in kth iteration step.

3. **Termination.** Repeat step 2 until the refined location $(\hat{x}_k, \hat{y}_k)$ falls outside of the area where the C_k training points come from. Return the wireless device's final estimated location $(\hat{x}, \hat{y}) = (\hat{x}_{k-1}, \hat{y}_{k-1})$.

Figure 12.8 shows the localization error CDFs of correlation-based lateration methods in the WiFi network and the Zigbee network under the WINLAB experimental scenario when the reducing rate of the training data size is set to 50%. The correlation-based approach significantly improves the accuracy of lateration methods in both median error as well as maximum error. The overall improvement of median error is over 33% in both networks.

12.5 Conclusion and Applications

In this chapter, we have provided an overview of RSS based localization approaches and algorithms for indoor localization. We focused in particular on the performance of these localization techniques and advanced techniques to enhance the robustness of these algorithms. Generally, the average localization accuracy of the RSS based localization techniques is around 10-20 feet. We compared the performance of these algorithms in terms of accuracy and precision, and showed by experimental evaluations that the localization accuracy can be improved via several approaches, such as the interpolated signal map and increased density of anchors. Because the radio propagation in indoors is complicated due to walls, furniture, and the movement of people, the robustness of localization algorithms is especially important in indoor environments. We thus describe techniques on enhancing the robustness of localization, in which the enhanced localization techniques are more robust to the environmental dynamics.

Location information can directly support a variety of location-based services or as input to other high-level emerging pervasive applications. For example, in the healthcare domain, doctors can directly use location information to track and monitor patients in medical facilities, or activities can be inferred using the estimated position and higher-level decision can be made accordingly. In the latter case, if a doctor and a nurse are both localized in the same room as a patient, then it will be concluded that this patient is getting treated. In the enterprise domain, location-based access control and asset tracking exploit location information directly. And the workflow management in industrial plants needs the location information as lower level input for a specific task.

Future trends make it likely that wireless indoor positioning systems will be integrated into a unified localization infrastructure, which can provide spatial positioning of wireless devices across organizational boundaries, work with diverse technologies, and support numerous applications including social networking, manufacturing, retail and security. Therefore, some important future research directions are: how to integrate the fragmented location systems belonging to different communities, such as indoor and outdoor positioning systems; how to work with diverse technologies, such as different wireless devices (i.e., WiFi, Zigbee, Bluetooth, and mote sensors) and location techniques (i.e., data fusion, hybrid location algorithms).

Bibliography

[1] P. Enge and P. Misra, *Global Positioning System: Signals, Measurements and Performance.* Ganga-Jamuna Pr, 2001.

[2] N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom)*, Aug 2000, pp. 32–43.

[3] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2000, pp. 775–784.

[4] E. Elnahrawy, X. Li, and R. P. Martin, "The limits of localization using signal strength: A comparative study," in *Proceedings of the First IEEE International Conference on Sensor and Ad hoc Communcations and Networks (SECON 2004)*, Oct. 2004, pp. 406–414.

[5] Y. Chen, K. Kleisouris, X. Li, W. Trappe, and R. P. Martin, "A security and robustness performance analysis of localization algorithms to signal strength attacks," *ACM Transactions on Sensor Networks (ACM TOSN)*, vol. 5, no. 1, February 2009.

[6] R. Battiti, M. Brunato, and A. Villani, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," University of Trento, Informatica e Telecomunicazioni, Technical Report DIT-02-086, Oct. 2002.

[7] D. Madigan, E. Elnahrawy, R. Martin, W. Ju, P. Krishnan, and A. S. Krishnakumar, "Bayesian indoor positioning systems," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2005, pp. 324–331.

[8] T. Rappaport, *Wireless Communications: Principles and Practice.* Prentice Hall, 1996.

[9] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: data mining, inference, and prediction. 2001," *NY Springer-Verlag.*

[10] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Comput. Networks*, vol. 43, no. 4, pp. 499–518, 2003.

[11] D. Niculescu and B. Nath, "Ad hoc positioning system (APS)," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2001, pp. 2926–2931.

[12] J. Yang, Y. Chen, V. Lawrence, and V. Swaminathan, “Robust wireless localization to attacks on access points,” in *IEEE Sarnoff Symposium 2009*, Princeton , NJ, USA.

[13] T. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, “A survey of various propagation models for mobile communication,” *Antennas and Propagation Magazine, IEEE*, vol. 45, no. 3, pp. 51–82, June 2003.

[14] J. Yang and Y. Chen, “Indoor localization using improved rss-based lateration methods,” in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, 2009.

[15] R. Battiti, R. Battiti, A. Villani, T. L. Nhat, T. L. Nhat, R. Villani, and R. Villani, “Location-aware computing: A neural network model for determining location in wireless lans,” Tech. Rep., 2002, tech. Rep.

[16] S.-H. Fang and T.-N. Lin, “Indoor location system based on discriminant-adaptive neural network in ieee 802.11 environments,” *Neural Networks, IEEE Transactions on*, vol. 19, no. 11, pp. 1973 –1978, nov. 2008.

[17] Y. Chen, Q. Yang, J. Yin, and X. Chai, “Power-efficient access-point selection for indoor location estimation,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 18, no. 7, pp. 877 – 888, july 2006.

[18] J. Yim, “Introducing a decision tree-based indoor positioning technique,” *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1296–1302, 2008.

[19] M. Brunato and R. Battiti, “Statistical learning theory for location fingerprinting in wireless ”lans”,” *Comput. Netw.*, vol. 47, no. 6, pp. 825–845, 2005.

[20] C.-L. Wu, L.-C. Fu, and F.-L. Lian, “Wlan location determination in e-home via support vector classification,” in *Networking, Sensing and Control, 2004 IEEE International Conference on*, vol. 2, 2004, pp. 1026 – 1031 Vol.2.

[21] J. Yang and Y. Chen, “A theoretical analysis of wireless localization using RF-based fingerprint matching,” in *Proceedings of the Fourth International Workshop on System Management Techniques, Processes, and Services (SMTPS)*, April 2008.

[22] M. Youssef, A. Agrawal, and A. U. Shankar, “WLAN location determination via clustering and probability distributions,” in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2003, pp. 143–150.

[23] M. Youssef and A. Agrawala, “Handling samples correlation in the horus system,” in *IEEE Infocom*, 2004, pp. 1023–1031.

[24] H. Lee, “A novel procedure for assessing the accuracy of hyperbolic multilateration systems,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-11, no. 1, pp. 2 –15, jan. 1975.

[25] D. Torrieri, “Statistical theory of passive location systems,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-20, no. 2, pp. 183 –198, march 1984.

[26] “ORBIT Wireless Network Testbed,” http://www.winlab.rutgers.edu/pub/docs/focus/ORBIT.html.

[27] K. Kleisouris, Y. Chen, J. Yang, and R. P. Martin, “The impact of using multiple antennas on wireless localization,” in *Proceedings of the Fifth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, June 2008.

[28] G. Chandrasekaran, M. Ergin, J. Yang, S. Liu, Y. Chen, M. Gruteser, and R. Martin, “Empirical evaluation of the limits on localization using signal strength,” in *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON ’09. 6th Annual IEEE Communications Society Conference on*, 22-26 2009, pp. 1 –9.

[29] H. Lim, L. Kung, J. Hou, and H. Luo, “Zero-configuration, robust indoor localization: Theory and experimentation,” in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, March 2006.

[30] Y. Gwon and R. Jain, “Error characteristics and calibration-free techniques for wireless lan-based location estimation,” in *MobiWac ’04: Proceedings of the second international workshop on Mobility management & wireless access protocols*. New York, NY, USA: ACM, 2004, pp. 2–9.

[31] K. Kleisouris, Y. Chen, J. Yang, and R. Martin, “Empirical evaluation of wireless localization when using multiple antennas,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. PP, no. 99, 2010.

[32] Z. Li, W. Trappe, Y. Zhang, and B. Nath, “Robust Statistical Methods for Securing Wireless Localization in Sensor Networks,” in *The Fourth International Conference on Information Processing in Sensor Networks (IPSN)*, 2005, pp. 91–98.

[33] B. Przydatek, D. Song, and A. Perrig, “”sia: Secure information aggregation in sensor networks”,” in *ACM SenSys*, Nov. 2003.

[34] D. Wagner, “Resilient aggregation in sensor networks,” in *Proceedings of the Forth ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN)*, 2004, pp. 78–87.

[35] P. Rousseeuw and A. Leroy, *Robust regression and outlier detection*. John Wiley & Sons Inc, 1987.

[36] M. Gudmundson, “Correlation model for shadow fading in mobile radio systems,” *Electronics Letters*, vol. 27, pp. 2145–2146, 1991.

[37] J. L. Buchanan and P. R. Turner, *Numerical Methods and Analysis*. New York: McGraw-Hill, 1992.

[38] N. Jalden, P. Zetterberg, B. Ottersten, A. Hong, and R. Thoma, "Correlation properties of large scale fading based on indoor measurements," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, March 2007.