Modeling of Indoor Positioning Systems Based on Location Fingerprinting

Kamol Kaemarungsi and Prashant Krishnamurthy
Telecommunications Program
School of Information Science, University of Pittsburgh
135 N. Bellefield Avenue, Pittsburgh, Pennsylvania 15260
Email: {kakst112, prashk}@pitt.edu

***Abstract*— In recent years, positioning systems for indoor areas using the existing wireless local area network infrastructure have been suggested. Such systems make use of location fingerprinting rather than time or direction of arrival techniques for determining the location of mobile stations. While experimental results related to such positioning systems have been presented, there is a lack of analytical models that can be used as a framework for designing and deploying the positioning systems. In this paper, we present an analytical model for analyzing such positioning systems. We develop the framework for analyzing a simple positioning system that employs the Euclidean distance between a sample signal vector and the location fingerprints of an area stored in a database. We analyze the effect of the number of access points that are visible and radio propagation parameters on the performance of the positioning system and provide some preliminary guidelines on its design.**

I. Introduction

Position location systems are becoming increasingly important as add-ons to today's pervasive wireless technology. Location-aware services are based on some form of positioning techniques. Positioning systems enable context-aware computing with location awareness [1]. They are also necessary for emergency services such as E-911 for cellular systems. In the last couple of years, *location fingerprinting* techniques using existing wireless local area network (WLAN) infrastructure have been suggested for indoor areas where the global positioning system (GPS) does not work well [2]. Such schemes can provide value-added services for existing WLANs [3]. The fingerprinting technique is relatively simple to deploy compared to the other techniques such as angle-of-arrival (AOA) and time-of-arrival (TOA). Moreover, there is no specialized hardware required at the mobile station (MS). Any existing wireless LAN infrastructure can be reused for this positioning system.

Generally, the deployment of fingerprinting based positioning systems can be divided into two phases. First, in the off-line phase, the location fingerprints are collected by performing a site-survey of the received signal strength (RSS) from multiple access points (APs). The entire area is covered by a rectangular grid of points. The RSS is measured with enough statistics to create a database or a table of predetermined RSS values on the points of the grid. The vector of RSS values at a point on the grid is called the *location fingerprint* of that point. Currently, there are no guidelines to choose the grid spacing. Moreover, it is not clear how many access points need to be "seen" at a given point for particular accuracy and precision. Note that a *location accuracy* is usually reported as the error distance deviated from the actual position, while a *location precision* is reported in percentages of position information that are within the distance of accuracy [4]. Second, in the on-line phase, a MS will report a sample measured vector of RSSs from different APs to a central server (or a group of APs will collect the RSS measurements from a MS and send it to the server). The server uses an algorithm to estimate the location of the MS and reports the estimate back to the MS (or the application requesting the position information). The most common algorithm to estimate the location computes the Euclidean distance between the measured RSS vector and each fingerprint in the database. The coordinates associated with the fingerprint that provides the smallest Euclidean distance is returned as the estimate of the position. Other algorithms use neural networks [1] or Bayesian modeling [5] to relate the sample RSS vector to the fingerprint in the database. The accuracy and precision reported by most of these schemes is quite similar.

Although this type of positioning system has been investigated for quite some time, this technique still lacks theoretical basis and understanding. Some experimental results on the accuracy and precision of different algorithms for positioning have been presented in the research community. The previous studies in the literature mainly focus on the measurement and analysis of the results such as those in [1] and [3]. Recent developments have been emphasizing the algorithms used for estimating the location (that associate the fingerprints with the location coordinates) [1], [5]. However, there is a lack of clear understanding of how these systems may perform (in terms of accuracy and precision), how to design these systems (what is the impact of the architecture of a building and thus the radio propagation characteristics) and what impacts the design (what should be the spacing of the grid where location fingerprints are taken). For example, choosing a large grid spacing like 10m reduces the granularity of the position to 10m. However, choosing a smaller grid spacing may increase the granularity or accuracy, but not the precision or the probability of correctly matching the fingerprint because the fingerprints of two points on the grid that are close to one another may be very similar. It would also result in a more laborious site-survey. Certain sites may however be more tolerant to a more granular grid spacing because of the radio propagation conditions. There is also a need to evaluate the performance of location systems for indoor wireless system in a manner similar to that proposed in [6] for outdoor position location systems.

In this paper we develop a preliminary mathematical model to analyze the performance and design of indoor positioning systems based on location fingerprinting. The proposed model forms a basic framework for efficient system deployment and performance evaluation. The effects of grid spacing, the number of access points, and the effect of radio propagation (the path loss exponent and the standard deviation of the received signal strength) are considered.

This paper is organized as follows. In Section II, we develop a mathematical model of the operation of the positioning system. This model enables us to characterize the distance in signal space between the measured RSS and the location fingerprints as two types of random variables. We also use the model to determine the probability of selecting the correct location in Section III when only two location fingerprints are considered. We extend this analysis with an approximation to account for multiple location fingerprints. This analysis is used in section IV with a simplified positioning system to understand the impact of the number access points, the standard deviation of the RSS and the path loss on the probability of selecting the correct location. We discuss the limitations of our model in Section V and briefly validate our main assumption on the distribution of RSS in Section VI. Conclusions and discussions of future work are given in Section VII.

II. Modeling of the Positioning System

Consider an indoor positioning system overlaid on an indoor wireless local area network on a single floor inside a building. We assume that there are N-IEEE 802.11b access points in the area and they are all visible throughout the area under consideration. A square grid is defined over the two-dimensional floor plan and any estimate of a MS's location is limited to the points on this grid. Assuming that the grid spacing results in L points along both the x and y axes, we have $L \times L = L^2$ positions in the area. Any position can be represented by a triplet with label (x, y, z) where x and y represents the 2-D coordinates on the floor plane while z represents the height of the antenna at that particular grid position. In this study, we assume (without loss of generality) that $z = 0$ for all coordinates unless otherwise mentioned.

The results of the site-survey are collected for the predetermined points on the grid with enough statistics (we explain this below). A total of K entries are recorded in the database. If the RSS values are measured at each point on the grid, $K = L^2$. Occasionally, some points on the grid are not accessible for measurement and are left out of the database. Each entry in the database includes a mapping of the grid coordinate (x, y) to the vector of corresponding RSS values from all access points in the area. Note that the received signal strength is recommended by [3] because it exhibits a stronger correlation with location than the signal-to-noise ratio (SNR). Each element in each vector in the database is assumed to be the *true mean of the RSS* from each of the N access points in the area. Usually, this is achieved by collecting a large number of samples of the RSS for each orientation of the user and the MS [3]. Note that the finding in [3] indicates that the RSS of the same location varies depending on the user's orientation. The present analytical model thus assumes that the variations due to orientation have been averaged out when the RSS is recorded for all positions. To estimate the user location, the MS obtains a sample of the RSS from all APs at the user's current position. This sample vector is compared with all K existing entries in the database. The fingerprint entry that has the closest match to the user's sample of RSS is used by the system as the estimate of the user's current location. This simple location estimation algorithm is similar to the nearest neighbor in signal space (NNSS) technique used by the RADAR system [3].

A. Mathematical Model

Two vectors are normally used in estimating the location of the mobile station. The first vector consists of samples of the RSS measured at the mobile station from N access points in the area. We call it the *sample RSS vector* throughout the paper. This vector is denoted as: $\mathcal{R} = [\rho_1, \rho_2, \rho_3, \ldots, \rho_N]$. The indoor positioning system estimates the mobile's location using this sample RSS vector. Each component in this vector is assumed to be a random variable with the following assumptions.

- The random variables ρ_i (in dBm) for all i are mutually independent.
- The random variables ρ_i (in dBm) are normally (or Gaussian) distributed.
- The (sample) standard deviation of all the random variables ρ_i is assumed to be identical and denoted by σ (in dBm).
- The true mean of the random variable ρ_i or $E\{\rho_i\}$ is denoted as r_i (in dBm).

The second vector, that forms the fingerprint of the location, consists of the true means of all the received signal strength random variables at a particular location from the N access points and it is recorded in the location database. We call it the *location fingerprint* or the *average RSS vector* for the rest of the paper and denote it by $\tilde{\mathcal{R}} = [r_1, r_2, r_3, \ldots, r_N]$.

The rationale for assuming that the RSS is a normally distributed random variable is as follows. Measurements of the RSS in many locations over the last few decades seem to support the fact that the RSS is lognormally distributed (normal in dB) [7]. Some measurement results based on an IEEE 802.11 wireless network in [1] support this assumption as well. The measurement of the signal strength in an office room over long time periods ranging from five hours, 20 hours and one month showed that it had a standard deviation σ of 2.13 dBm. Small et. al. also reported that signal did not vary dramatically at different times of the day. Some of our own measurements over smaller durations of time indicate that the RSS may not be normally distributed. An initial data analysis on this assumption is discussed in Section VI where we found that this assumption may be valid only in certain situations. However, we use this distribution in this preliminary model for mathematical tractability. The assumption of independence is acceptable since there is no relationship between the signals transmitted by different APs.

B. Characterization of the Distance Metric

The signal distance between the sample RSS vector and the average RSS vectors is used to determine which of the points on the grid corresponds to the position of the mobile station. This simple technique selects the (x, y) coordinates corresponding to the average RSS vector with the smallest signal distance to the sample RSS vector as the estimated location. Note that the *signal distance* is not the same as the actual physical distance between the two positions in the real world. The common metric used to calculate the signal distance between the two vectors is the Euclidean distance. The Euclidean distance between $\tilde{\mathcal{R}}$ and $\mathcal{R}$ is given by:

$$Z = \left[\sum_{i=1}^{N}(\rho_i - r_i)^2\right]^{\frac{1}{2}} = \left[\sum_{i=1}^{N} q_i^2\right]^{\frac{1}{2}}. \qquad (1)$$

Interestingly, the Euclidean distance metric in (1) can be categorized into two types of random variables based on the mean value of each random component q_i (in dB). The random variable q_i has a zero mean when each of the elements in the sample RSS vector $\mathcal{R}$ has the same mean value as the corresponding element in the average RSS vector $\tilde{\mathcal{R}}$. This corresponds to the signal distance between the sample RSS vector at the location and the *true* fingerprint corresponding to this location. The random variable q_i has a non-zero mean when the the sample RSS vector is compared with a location fingerprint of another position on the grid. We will consider the characteristics of the random variable $X = Z^2$ to obtain some insights into the effects of radio propagation on the design of the positioning systems.

If the RSS is normally distributed as assumed in Section II-A, the random variable $X = Z^2$ has a *central chi-squared distribution* with N degrees of freedom [8] when the sample RSS vector has its true mean recorded in the average RSS vector. That is $E\{\rho_i\} = r_i$ or the mean of the measured RSS is exactly the true mean in the database. Thus, the distance-squared component q_i is a zero mean Gaussian random variable. The random variable X is the square of the distance between the sample RSS vector and the average RSS vector and has a probability density function (PDF):

$$p_{\chi_N^2}(x) = \frac{1}{\sigma^N 2^{N/2}\Gamma(N/2)}\, e^{-x/(2\sigma^2)} x^{(N/2)-1}, \qquad (2)$$

where $x \geq 0$.

Note that the variance of each Gaussian component in X is σ^2 and N represents the number of access points that are visible. Fig. 1 depicts the effect of σ and N on the PDF of the the random variable X. Table I summarizes the effects.

If the sample RSS vector is compared to a location fingerprint in the database that does not correspond to the correct

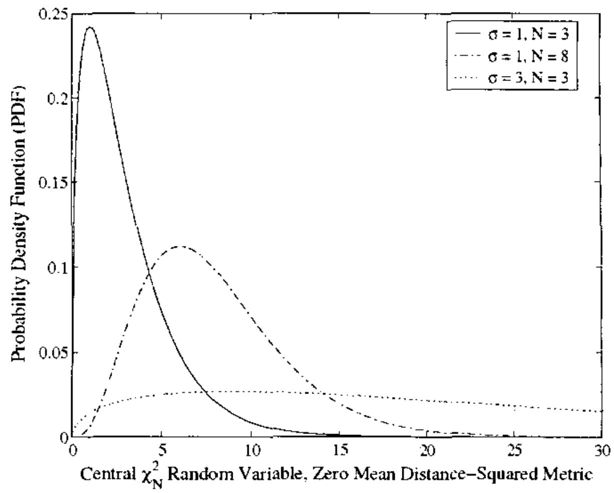

Fig. 1. PDF of central chi-squared distribution

TABLE I
PARAMETERS OF CENTRAL CHI-SQUARED DISTRIBUTION

Parameter	Effect of Larger Parameter
σ - STD. of Gaussian component	X is larger.
N - Number of access points	X is larger.

location, the random variables q_i will have a non-zero mean equal to $\mu_i = E\{\rho_i\} - r_i$. In this case, the distribution of the square of the distance between $\tilde{\mathcal{R}}$ and $\mathcal{R}$ given by $X = Z^2$ has a non-central chi-squared distribution with non-centrality parameter $\lambda = \sum_{i=1}^{N} \mu_i^2$ and N degrees of freedom. The non-centrality parameter is a measure of grid spacing because it is a function of the difference of the means of the received signal strengths at two points on the grid. The points on the grid are at different distances from different APs. The farther apart the points on the grid are, the more will be the difference in the mean received signal strengths at these locations. Thus, a larger λ means a larger physical distance between two points on the grid. The PDF of the non-central chi-squared distribution is given by:

$$p_{x;N,\lambda} = e^{-\frac{(\lambda+x)}{2\sigma^2}}\, \frac{1}{2\sigma^2}\left(\frac{x}{\lambda}\right)^{\frac{(N-2)}{4}} I_{\frac{(N-2)}{2}}\left(\frac{\sqrt{\lambda x}}{\sigma^2}\right), \qquad (3)$$

where $x \geq 0$.

Here, $I_\alpha(x)$ is the αth-order modified Bessel function of the first kind. Fig. 2 shows the effects of λ, σ, and N on the PDF of this signal distance metric. Table II summarizes the implication of the parameters λ, σ, and N to the non-zero mean signal distance metric.

Fig. 3 compares the PDFs of central chi-squared and non-central chi-squared distributions. Notice that the non-central chi-squared distribution shifts to the right for a large value of the non-centrality parameter. A larger value of the non-centrality parameter will cause the sample values of the

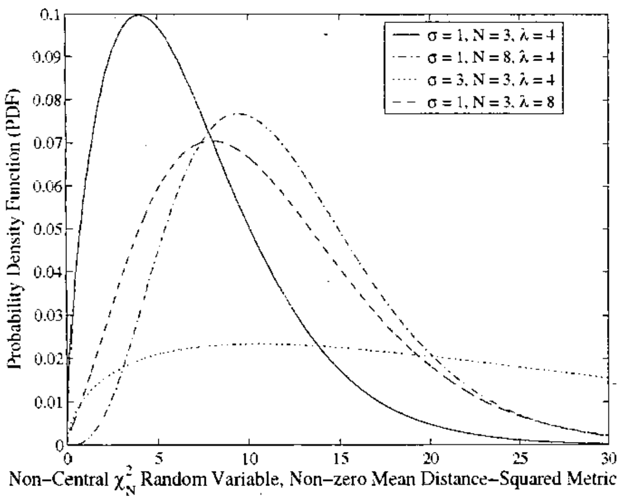

Fig. 2. PDF of non-central chi-squared distribution

TABLE II
PARAMETERS OF NON-CENTRAL CHI-SQUARED DISTRIBUTION

Parameter	Effect of Larger Parameter
σ - STD. of Gaussian component	X is larger.
N - Number of access points	X is larger.
λ - Non-centrality parameter	X is larger.

non-central chi-squared distribution to be mostly larger than the corresponding central chi-squared distribution. A larger standard deviation σ causes the two distributions to get closer to one another. For example, as the standard deviation of each of the normal RSS variables becomes as large as $\sigma = 15$ dBm, the difference between the two distributions reduces. Both distributions are almost the same and the PDFs are nearly identical as shown in Fig. 4. Note that we have kept the non-centrality parameter to a large value ($\lambda = 20$) and still see the similarity of the distributions.

We can now make some qualitative comments on the impact of some of the parameters on the design of location fingerprinting based positioning systems based on the visual results presented so far. The distance between the sample RSS vector and any location fingerprint in the database is a random variable because the received signal strengths from APs measured by the MS are all random variables. Consequently, it is possible to pick a location fingerprint in the database as being *closest* to the sample RSS vector even though it is not the location fingerprint of the correct location of the MS. This is very likely to happen if the standard deviation of the RSSs are high. Intuitively this makes sense. The location fingerprint consists of the mean values of the RSSs. If the RSSs have a large standard deviation, the probability of the sample being close to the mean is small. In fact, if the RSS has a uniform distribution, any RSS value is equally likely so that the location fingerprint that is returned as *closest* to the sample RSS vector could correspond to any point on the grid.

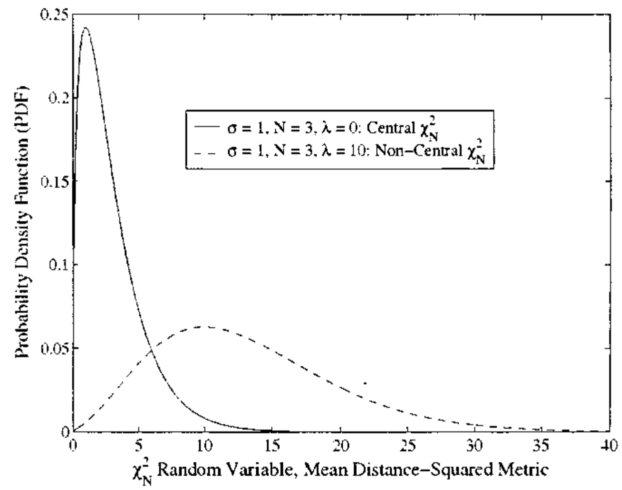

Fig. 3. Comparison of the PDFs of central and non-central chi-squared distributions

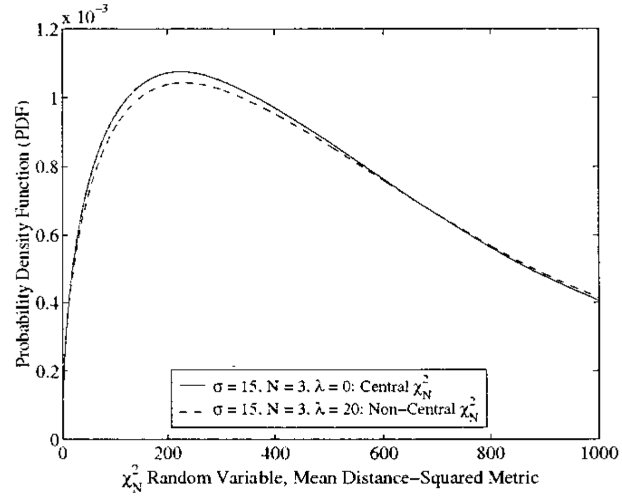

Fig. 4. Comparison of the PDFs of central and non-central chi-squared distributions with large standard deviation

Ideally then, we would like the RSSs to have a small standard deviation. Also, as the non-centrality parameter increases, the probability that an incorrect location fingerprint is returned as the closest (smaller distance) to the sample RSS vector decreases because the central chi-squared random variable has a peak in its PDF at smaller values than the corresponding non-central chi-squared random variable. In the following sections, we look at the actual probability of returning a correct location fingerprint.

III. PROBABILITY OF RETURNING THE CORRECT LOCATION FINGERPRINT

Although it is tempting to directly compare the central chi-squared random variable to the non-central chi-squared random variable, we cannot do so directly due to the dependency of the two random variables which are transformations of the same random vector $\mathcal{R}$ (the sample RSS vector). However, we can perform the following analysis to determine the probability of returning the correct fingerprint as the estimate of the location when a MS reports a sample RSS vector.

Let $\mathcal{A}$ be the square of the distance between the sample RSS vector $\mathcal{R} = (\rho_1, \rho_2, ..., \rho_N)$ and the actual location fingerprint (average RSS vector of the true location) $\tilde{\mathcal{R}} = (r_1, r_2, ..., r_N)$. Let $\mathcal{B}$ be the square of the distance between the sample RSS vector $\mathcal{R}$ and the location fingerprint $\tilde{\mathcal{S}} = (s_1, s_2, ..., s_N)$ of a neighboring point on the grid. We then denote $\{\mathcal{A} < \mathcal{B}\} = \{\mathcal{A} \leq \mathcal{B}\}$ as the event that the distance between the sample RSS vector and the correct location fingerprint is smaller than the distance between the sample RSS vector and the incorrect neighboring location fingerprint. We can determine the probability of this event. Firstly, we evaluate $\{\mathcal{A} \leq \mathcal{B}\}$ as follows:

$$
\begin{aligned}
\mathcal{A} &\leq \mathcal{B} \\
\Rightarrow \sum_{i=1}^{N}(\rho_i - r_i)^2 &\leq \sum_{i=1}^{N}(\rho_i - s_i)^2 \\
\Rightarrow \sum_{i=1}^{N}(\rho_i - r_i)^2 - \sum_{i=1}^{N}(\rho_i - s_i)^2 &\leq 0 \\
\Rightarrow \sum_{i=1}^{N}(\rho_i^2 - 2r_i\rho_i + r_i^2) - \sum_{i=1}^{N}(\rho_i^2 - 2s_i\rho_i + s_i^2) &\leq 0 \\
\Rightarrow 2\sum_{i=1}^{N}\rho_i(s_i - r_i) + \sum_{i=1}^{N}(r_i^2 - s_i^2) &\leq 0 \\
\Rightarrow 2\sum_{i=1}^{N}\rho_i\beta_i + \sum_{i=1}^{N}\Gamma_i &\leq 0, \qquad (4)
\end{aligned}
$$

where $\Gamma_i = (r_i^2 - s_i^2)$ and $\beta_i = (s_i - r_i)$.

To determine the probability of the event in (4), we first apply the properties of the sum of multiple independent Gaussian random variables [9]. The left hand side of (4), $\mathcal{C} = 2\sum_{i=1}^{N}\rho_i\beta_i + \sum_{i=1}^{N}\Gamma_i$, is a new Gaussian random variable when all ρ_i are Gaussian. The random variable $\mathcal{C}$ has following mean and variance

$$
\begin{aligned}
\mu_c &= 2\sum_{i=1}^{N} r_i\beta_i + \sum_{i=1}^{N}\Gamma_i, \\
\sigma_c^2 &= \sum_{i=1}^{N}(2\beta_i\sigma_i)^2. \qquad (5)
\end{aligned}
$$

Therefore, the probability that the system returns the correct location when it compares just two location fingerprints to the sample RSS vector is given by:

$$
\begin{aligned}
Pr\{\mathcal{C} \leq 0\} &= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(c-\mu_c)^2}{2\sigma_c^2}} dc \\
&= \frac{1}{2}\frac{2}{\sqrt{\pi}}\int_{-\infty}^{-\frac{\mu_c}{\sqrt{2}\sigma_c}} e^{-t^2} dt \\
&= \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{-\mu_c}{\sqrt{2}\sigma_c}\right). \qquad (6)
\end{aligned}
$$

From (5) and (6), we observe that the mean and variance of the new distribution are influenced by two parameters of the positioning system: the number of access points N and the standard deviation σ_i of the normally distributed RSS variables. The parameters β_i and Γ_i do not have any explicit meaning but are related to the non-centrality parameter λ discussed earlier. That is, they both depend on the difference between the mean RSS at the two locations that is determined by the path loss of the signal. In turn, the path loss of the signal depends on the site and the physical distances of the locations from the N access points and indirectly to the physical distance between the locations. We try to evaluate the effects of these parameters using the grid spacing g and the path loss exponent α as described in the next section.

In a real positioning system, the database contains several entries depending on the size of the office floor and the grid spacing. The positioning system makes comparisons between the sample RSS vector and all of these location fingerprints. The database may be arranged hierarchically. In the first phase, the APs seen by the MS are matched. Then the location fingerprints corresponding to these APs alone are compared with the sample RSS vector. In any case, each comparison depends on the same sample RSS vector. Therefore, to be able to calculate the probability of returning a correct location, we will need to know the joint probability density function (PDF) of all random variables of the form $\mathcal{C}$. Deriving an analytical model can be quite cumbersome where there may be tens or hundreds of location fingerprints being compared.

For this preliminary analysis, we consider a first cut model with only a simple approximation of this probability. Let $\mathcal{C}_k = \sum_{i=1}^{N}(\rho_i - r_i)^2 - \sum_{i=1}^{N}(\rho_i - s_{k,i})^2$ be the comparison variable. The variable $\mathcal{C}_k$ compares the distance between the sample RSS vector $\mathcal{R}$ and (a) the correct location fingerprint $\tilde{\mathcal{R}}$ and (b) the k-th incorrect location fingerprint $\tilde{\mathcal{S}}_k$. The index k runs from 1 to K excluding the correct location denoted by the index c. Once again K corresponds to the number of entries in the database. Then, we can write the probability of correct decision as:

$$
\begin{aligned}
&Prob\{\text{Correct Estimation}\} = \mathcal{P}_c \\
&= P\{\mathcal{C}_1 \leq 0, \cdots, \mathcal{C}_{c-1} \leq 0, \mathcal{C}_{c+1} \leq 0, \cdots, \mathcal{C}_K \leq 0\} \qquad (7)
\end{aligned}
$$

To compute this probability, we need the joint distribution of the random variables $\mathcal{C}_k$. To avoid the cumbersome calculation of the joint distribution, we make the assumption that they are independent. In such a case:

$$
Prob\{\text{Correct Estimation}\} = \mathcal{P}_c = \prod_{\substack{k=1 \\ k \neq c}}^{K} Pr\{\mathcal{C}_k \leq 0\}. \qquad (8)
$$

Equation (8) is an approximation to the probability of returning the correct location for an indoor location system when the correct position is c and there are K location fingerprint entries in the database. While this assumption is not correct, we compare the results of this assumption with simulation results to see how close the results are. As we shall see, the assumption of independence provides a reasonable approximation.

IV. Performance Evaluation

The major performance metrics of interest for indoor positioning systems are the accuracy and the precision in estimating a position. In this section, we investigate how the path loss and RSS characteristics influence the precision. As a measure of precision at zero-meter accuracy, we look at the probability of making a correct estimation of the location. A better measure of accuracy and precision is the distribution of the error in the location estimate that is part of our ongoing work. We use (8) as the analytical approximation of the measure of precision and also compare this with simulations. We use a simple grid in this work as described below.

A. System Model Setup

Fig. 5 illustrates an indoor positioning system with $L^2 = 25$ points on the grid. The grid at the center is labeled with a $\star$ and is assumed to be the current position of the mobile station reporting the sample RSS vector. The neighboring positions with location fingerprints recorded in the database during the site-survey are labeled with $\circ$. There are 8 neighboring positions in this system. The total number of positions within this system and in the database is $K = 9$. The outer most positions are labeled with $\square$ and reserved for placing access points only.

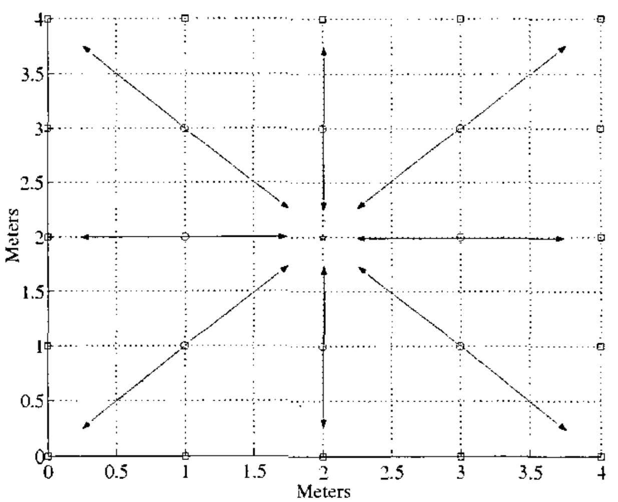

Fig. 5. A grid space for an indoor positioning system.

Initially, we place four access points at the four corners of the grid. The positions of the access points are $AP1 = (0,0)$, $AP2 = (4,4)$, $AP3 = (0,4)$, and $AP4 = (4,0)$. The position of the mobile station is $(2,2)$. Suppose the physical distance of the k-th point on the grid from the j-th AP $d_{j,k}$ meters. The true mean or expected value of ρ_j for that point on the grid can be calculated from the mean path loss given by [7]:

$$Pl(d_{j,k}) = Pl(d_0) + 10 \cdot \alpha \cdot log_{10}(d_{j,k}), \tag{9}$$

Here $Pl(d_0)$ is the free-space path loss at the reference distance of $d_0 = 1$ m (this is 41.5 dBm for line-of-sight propagation (LOS) and for 37.3 dBm non-line-of-sight propagation (NLOS) *for some reported measurements)* [10]. The variable α denotes the path loss exponent, which for indoor locations at a carrier frequency of 2.4 GHz is reported to be 2 for LOS propagation and 3.3 for NLOS propagation [10]. Under other circumstances, the indoor path loss exponent α can be between 1 and 6 [11]. Then, the mean received signal strength, $E\{\rho_j\}$, can be found by

$$E\{\rho_j\} = Pt - Pl(d_{j,k}), \tag{10}$$

where Pt is the transmit power of the access point which we will fix at 15 dBm for IEEE 802.11b based WLANs. The standard deviation of the RSS for this indoor positioning system is assumed to be σ = 2.13 dBm as reported in [1]. Other values reported for σ are 4 dB or 2.5 for different indoor radio propagation conditions [7]. A more accurate path loss prediction model could be used instead of the simple path loss model in (9). An inclusion of wall and floor attenuation factor is suggested in [3]. In summary however, the path loss equation provides us with the mean received signal strength value. We use the equation here for computational purposes. It is possible to use actual measurements as well without changing the framework here.

Initially we assume that the grid spacing is 1m (3 feet). We also look at the effect of grid spacing on the accuracy with the following grid spacing: 0.25, 0.5, 0.75, 1, 1.25, 1.5, and 1.75 meters. Note that the positions of the center of the grid and the access points are fixed for all scenarios. As an example of the database of location fingerprints, we show a sample in Table III that contains the location fingerprints of all coordinates within the system when the grid spacing is set to 1m. If only one access point is present, the fingerprints, as listed in the second column, may not be unique. this happens when two points on the grid are at the same distance from the access point. Additional access points make the fingerprints unique.

TABLE III

RSS FINGERPRINTS OF INDOOR LOCATION

Access Point	**AP1** (dBm)	**AP2** (dBm)	**AP3** (dBm)	**AP4** (dBm)
Coordinate	**(0,0)**	**(4,4)**	**(0,4)**	**(4,0)**
(2,2)	-37.6010	-37.6010	-37.6010	-37.6010
(1,1)	-27.6670	-43.4120	-39.2000	-39.2000
(1,2)	-34.2330	-41.0801	-41.0801	-34.2330
(1,3)	-39.2000	-39.2000	-43.4120	-27.6670
(2,1)	-34.2330	-41.0801	-34.2330	-41.0801
(2,3)	-41.0801	-34.2330	-41.0801	-34.2330
(3,1)	-39.2000	-39.2000	-27.6670	-43.4120
(3,2)	-41.0801	-34.2330	-34.2330	-41.0801
(3,3)	-43.4120	-27.6670	-39.2000	-39.2000

B. Results of the probability of returning the correct location for a single neighbor

Based on the model, we calculate the precision at zero-meter accuracy in terms of the probability of returning the correct position. Initially we consider only one neighboring point on the grid (we compare the location fingerprints at the positions

(2,2) and (2,1)). For this, we can simply use the analytical expression from (6).

We first look at the impact of the number of access points deployed by varying this number from one to 16 according to the system in Fig. 5. The first four access points are installed at the four corners and the rest are located at the following coordinates: $AP5$ = (2,0), $AP6$ = (4,2), $AP7$ = (2,4), $AP8$ = (0,2), $AP9$ = (1,0), $AP10$ = (4,1), $AP11$ = (3,4), $AP12$ = (0,3), $AP13$ = (3,0), $AP14$ = (4,3), $AP15$ = (1,4), and $AP16$ = (0,1). Fig. 6 shows the results of using (6) when the number of access points increases. The label "Ana" indicates calculation from the analytical equations discussed earlier. A higher number of access points improves the precision but the probability does not increase significantly for $N > 5$. We also see that a larger standard deviation of the RSS results in poorer precision especially for a smaller number of APs. For instance, if the standard deviation of the RSS changes from 1 to 4, the probability of returning the correct location drops from nearly 1 to 0.77 for three access points.

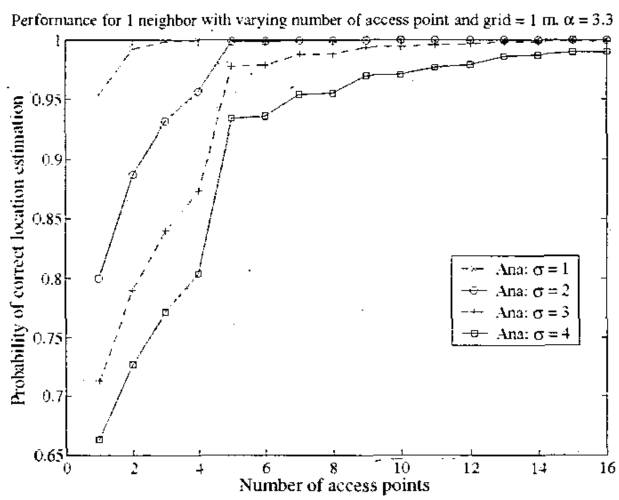

Fig. 6. Effect of number of access points on probability

We next consider the impact of the standard deviation σ of the RSS in Fig. 7. Clearly, the larger values of σ degrade the precision dramatically. Fig. 7 plots the probability of returning the correct location versus the standard deviation. The results suggest that the lower the value of standard deviation, the better the precision for any number of access points. However, this value is difficult to control because it depends on the environment. One way of improving this is to consider many RSS samples. This could contribute to the delay in obtaining the location. The results indicate that reducing the standard deviation to somewhere between 2 and 4 may be sufficient.

The second parameter that depends on the environment and cannot be controlled is the path loss exponent α. Results for the probability of returning the correct location as a function of α are shown in Fig. 8. $\mathcal{P}_c$ improves as the path loss exponent increases. This can be intuitively explained as follows. If the

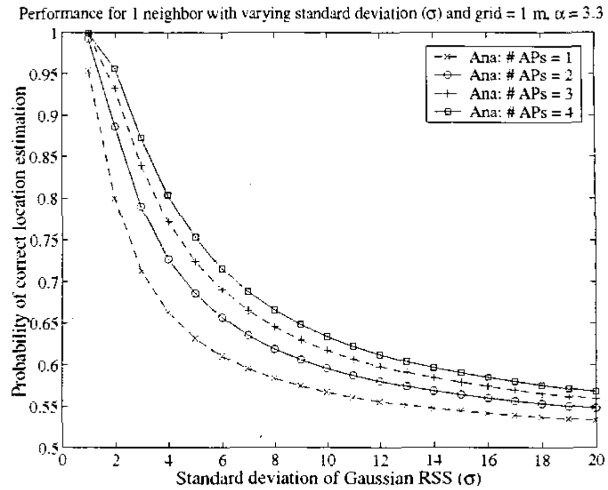

Fig. 7. Effect of RSS standard deviation on probability.

signal attenuated greatly with distance, even a small shift in the distance can result in large differences in the mean RSS. Thus, the average RSS vectors between two coordinates become easily distinguishable. This will also be the case if there are intervening obstacles like walls or floors although we have not included them in this model.

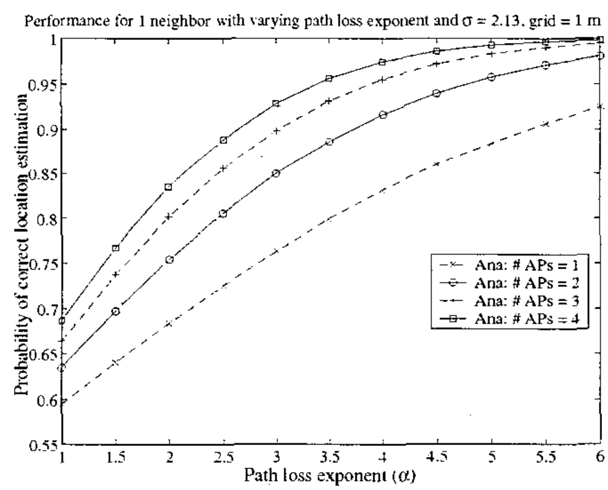

Fig. 8. Effect of path loss exponent on probability.

The last system parameter affecting the probability of returning the correct location is the grid spacing. The grid spacing can be selected during the site-survey. A large grid spacing will provide poor accuracy or granularity of the location information. On the other hand, a too small grid spacing may reduce the positioning precision. The analytical results in Fig. 9 indicate that a small grid spacing reduces the precision greatly. For a standard deviation $\sigma = 2.13$ dBm and three access points, a grid spacing of 1m results in a 90%

probability of returning the correct location.

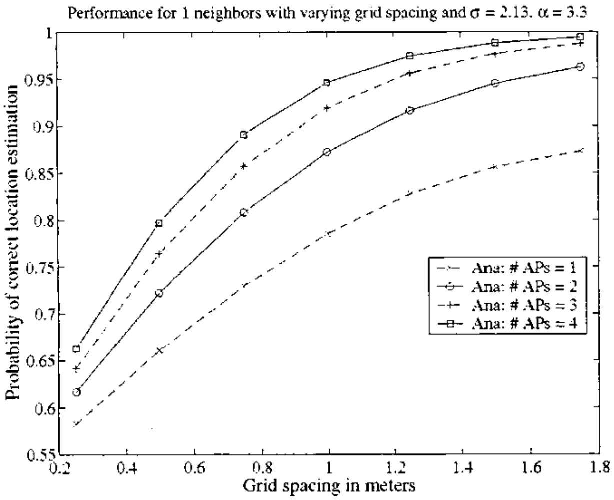

Fig. 9. Effect of grid spacing on probability

We have validated the analytical results presented so far (using (6)) with simulations. The simulations were conducted by generating a random vector of Gaussian random variables to represent the sample RSS vector with the mean values corresponding to the average RSS vector at the mobile stations correct location (the $\star$ in Fig. 5). One million sample RSS vectors were simulated for each data point. For each sample RSS vector generated, the Euclidean distances between the sample RSS vector and the two average RSS vectors (location fingerprints) in the system were calculated and then compared. The validation results confirmed that our analytical calculations exactly match the simulations.

C. *Results of the probability of returning the correct location for multiple neighbors*

In this section, we look at the probability of returning the correct location when a comparison is made not just with one neighboring fingerprint, but with the eight neighbors described previously. The MS is once again located at the center of the grid. In practice, only a few APs are visible in any WLAN configuration. This number is anywhere between 2 and 5. The results in the previous section indicate that $N = 5$ APs is sufficient for good accuracy. So we consider a maximum of four access points in this subsection. Fig. 10 compares simulation results with the approximation in (8) based on the range of standard deviation of RSS σ between 1 and 20 dBm. Note that the simulation results are labeled with "Sim". The results follow the same trend as in the case of comparing only one neighboring fingerprint. However, the lowest value of probability has been driven down to around 0.1 when compared to the previous analysis that is on the order of 0.6. The analytical approximation is found to be close to the simulation results when the number of access points is three and four. Once again, keeping the standard deviation to somewhere between 2 and 4 dBm appears to be sufficient when 4 APs are visible.

The approximation in (8) is pessimistic compared to the simulation results due to the assumption of independence between each comparison pair. This is to be expected. For example, if the sample RSS vector is close to the correct location, it also true for all comparisons.

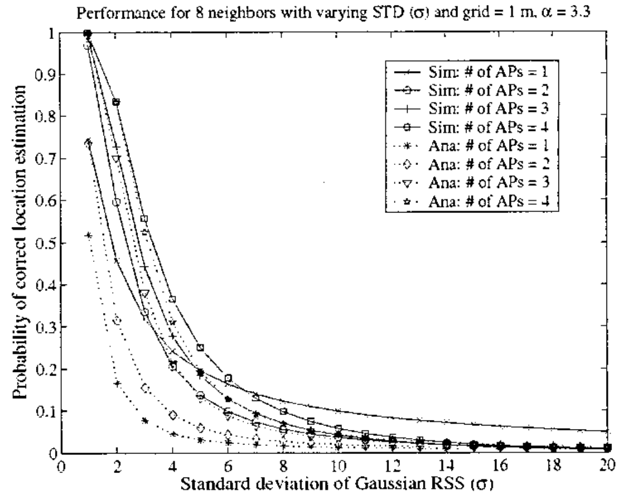

Fig. 10. Effect of RSS standard deviation on probability.

Similar trends are observed for simulation and analytical results for $\mathcal{P}_c$ as a function of the path loss exponent α and the grid spacing g as shown in Fig. 11 and Fig. 12. Larger values of the path loss exponent and grid spacing improve the precision. It is also clear from both figures that as the number of access points increases, the analytical approximation and simulation results are very close especially for larger values of α and g.

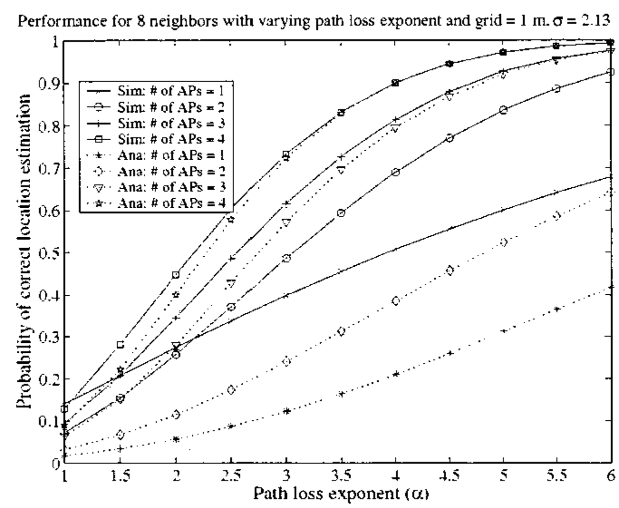

Fig. 11. Effect of path loss exponential probability

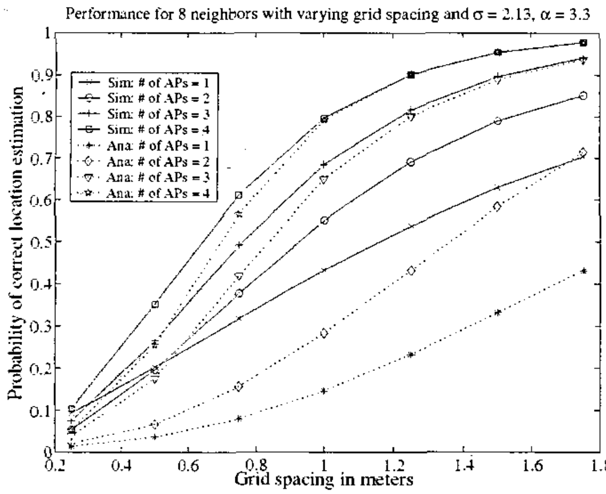

Fig. 12. Effect of grid spacing on probability

Based on the model and simplified system analysis and simulations, we can summarize the effect of system parameters on the precision and recommend a range of parameters for high precision (90%) in indoor position location systems. This is summarized in Table IV. When a WLAN is being deployed to provide communications coverage in a building, currently positioning issues are ignored. As position location becomes important, depending on the application requirements and environmental characteristics, a WLAN design can incorporate some changes to improve the accuracy and precision. For example, in a harsh environment where the RSS has large variations (large σ), the positioning system may request multiple samples of the RSS vector. Alternatively, in the design phase, APs may be placed so that locations where higher accuracy and precision are desired can "see" at least 4 APs.

TABLE IV
RECOMMENDED VALUES FOR LOCATION SYSTEM PARAMETERS

Parameters	Value Increased	Suggested Range
σ - STD. of Gaussian component	$\mathcal{P}_c$ decreases	$\sigma < 4$
N - Number of access points	$\mathcal{P}_c$ increases	$3 \leq N \leq 5$
α - Path loss exponent	$\mathcal{P}_c$ increases	$\alpha > 3.5$
g - grid spacing	$\mathcal{P}_c$ increases	$g > 1.25$ meter

The guideline provided in Table IV may not be applicable to the actual indoor positioning system because of the dense multipath effect that can cause some RSS elements to be higher at faraway locations. A more accurate path loss model with wall and floor attenuation could provide a more realistic recommendation for the deployment process using our framework proposed here. The validation of the guideline is a part of our ongoing research.

V. LIMITATIONS OF THIS WORK

In comparing this preliminary mathematical model to a real system, we can identify some differences based on the infrastructure, the survey, the real position, the database, and the location algorithm. First, in real systems, the access points are installed opportunistically and a MS may not receive signals from all access points in the building. A subset of access points may be present in the MS's view at particular time. Sample measurements indicate that some APs are never seen for significant duration of time at a location, but suddenly become visible at other times because of the dynamic changes especially in multi-floor environments. This affects the performance of positioning system but we have not incorporated this in the mathematical model. Our model also assumes a very simple grid. Our preliminary investigation indicates that the accuracy of the system depends on the placement of the AP, which is also suggested in [12]. Second, during the site-survey, the grid spacing may not be strictly uniform due to the inaccessible parts of the environment such as walls, partitions, and office furniture. Third, the actual MS positions are not limited to the grids defined in the database. Therefore the sampled RSS vector measured by the MS may not have the same mean RSS as recorded in the database. Fourth, the location estimation algorithm used here is the one that simply compares the Euclidean distance between the sampled RSS vector and the location fingerprints. Other algorithms report m location fingerprints that are closest to the sampled RSS vector. The system may compute the average of the m locations and return this as an estimate of the position. Finally, this analysis does not consider the accuracy and precision in terms of the distribution of errors in the location estimate.

VI. VALIDATION OF ASSUMPTION

We performed some preliminary measurements to validate our main assumption on the distribution of the RSS. A set of measurements using standard laptop computer equipped with 802.11b WLAN network interface card (NIC) were performed inside our office building at the University of Pittsburgh where there are access points installed opportunistically. A total of 75 RSS data sets were collected from a square grid of 25 positions on the fourth floor of the building where each position can receive signals from three access points. Each set of measurements consists of five minutes of data collected at approximately 0.25 second sampling interval. That is, approximately 1200 data points were collected for each set.

After analyzing all data sets, we found that most of the distributions are not normally distributed. The majority of distributions (67 out of 75) are left-skewed and the left-skewness usually occurs when there is a line-of-sight between an AP and a MS. Moreover, The standard deviation of the RSS varies from one AP to another and tends to vary according to the mean RSS level. The standard deviation of the RSS values from each AP is also different at different locations. In some cases, there could be more than one mode in the RSS distribution and there is at least one dominant mode. This observation of multiple modes has also been made by [13].

A sample measurement of RSS data collected from one AP is depicted in Fig. 13. We compare this short measurement with another long measurement of the signal from the same

AP. The long duration of measurement was collected over 26 days at sampling interval of five seconds and at a different location. The histogram of the long measurement is showed in Fig. 14. It is clear that both are left-skewed but the RSS values concentrate around the dominant modes. Notice that there are multiple modes in the measurement results corresponding to the 26 days in Fig. 14.

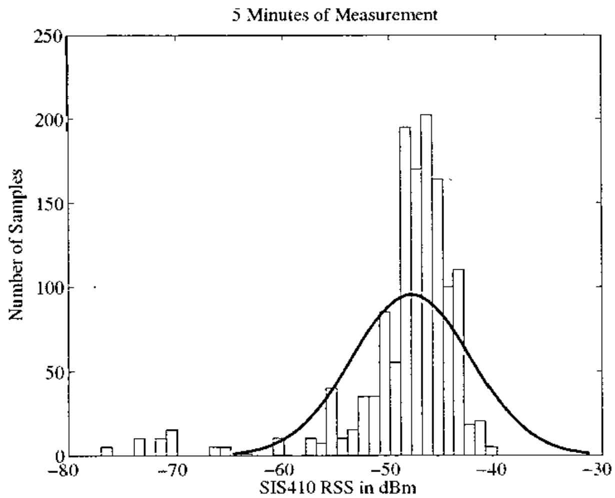

Fig. 13. Histogram of RSS measurement of SIS410 over 5 minutes

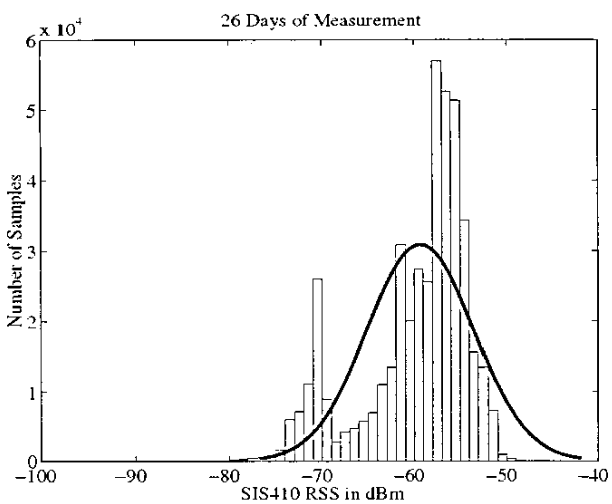

Fig. 14. Histogram of RSS measurement of SIS410 over 26 days

However, there are some data sets (21 out of 75) in our measurement that could be approximated by a Gaussian distribution where they are only slightly skewed (skewness $< |0.5|$). Fig. 15 illustrates an example of a slightly left-skewed RSS distribution measured from a different AP. We also notice that these distributions occur when the AP is far from the measurement location and do not have a direct line-of-sight. This result indicates that our model can be useful in the case where there is no line-of-sight to most of the locations which is generally true in indoor environments. This observation could also explain why the measurements in [1] reports a normal distribution where the measurement in that work is done inside an office room with no line-of-sight path. We can conclude that our assumptions regarding the distribution of the RSS could be valid to a certain degree.

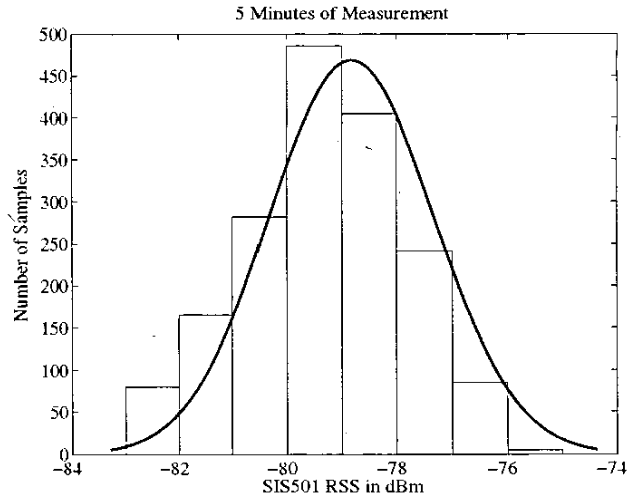

Fig. 15. Histogram of RSS measurement of SIS501 over 5 minutes

VII. Conclusions and Future Work

Given a set of system parameters and radio propagation characteristics, which are the number of access points, the grid spacing, the path loss exponent, and standard deviation of RSS, the accuracy of a positioning system can be determined in terms of the probability of returning the correct location with the model presented in this manuscript. The analysis results provide a guideline on parameters to consider in designing and deploying an indoor positioning system. The system does not require a large number of access points (no more than five) to improve the accuracy performance. The number of access points and their locations depend on the network infrastructure and the budget. The cost of IEEE 802.11b based access points is rapidly falling and it may not be a problem to add one or two more to provide additional coverage for indoor positioning. The grid spacing is closely tied to the application requirements of the indoor positioning system. The grid spacing must be selected properly and cannot be too small to provide a fine granularity or accuracy as the precision may be poor. A larger grid spacing may provide a better precision, but may render the location information very coarse. A grid spacing of 1 meter is found to be a suitable value based on our model and simulations. The standard deviation of the sampled RSS vector should be low in order to provide a better performance and a larger path loss exponent will benefit the precision. However, the standard deviation of RSS and the path loss exponent are not controllable parameters.

They vary depending on the indoor environment and building material. Larger standard deviations are reported inside large and open space buildings, while smaller standard deviations are found mostly in small and closed spaces [7]. The path loss exponent also depends on the material inside buildings. A system designer should balance all of these system parameters by adjusting the number of access points and grid spacing in order to deliver a satisfactory accuracy and precision. In our ongoing work, we are trying to address the limitations of our model discussed previously and validate our model with real measurements.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] J. Small, A. Smailagic, and D. P. Siewiorek, (2000, Dec.) Determining User Location For Context Aware Computing Through the Use of a Wireless LAN Infrastructure. [Online] Available: http://www-2.cs.cmu.edu/~aura/docdir/small00.pdf.

[2] G. M. Djuknic, and R. E. Richton, "Geolocation and Assisted GPS," *IEEE Computer*, vol. 2, pp. 123–125, Feb. 2001.

[3] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," in *Proc. IEEE INFOCOM*, 2000, pp. 775–784.

[4] J. Hightower, and G. Borriello, "Location Systems for Ubiquitous Computing," *IEEE Computer*, vol. 34, no. 8, pp. 57–66, Aug. 2001.

[5] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Room Location Service for Wireless Networked Environments," in *Proc. Ubiquitous Computing*, 2001.

[6] S. Tekinay, E. Chao, and R. Richton, "Performance Benchmarking for Wireless Location Systems," *IEEE Communications Magazine*, vol. 36, pp. 72–76, Apr. 1998.

[7] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, Upper Saddle River, New Jersey, 1996.

[8] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1-2, 2Ed, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York, 1995.

[9] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison Wesley, Reading, Massachusetts, 1994.

[10] G. J.M. Janssen, and R. Prasad, "Propagation Measurements in an Indoor Radio Environment at 2.4 GHz, 4.75 GHz and 11.5 GHz," in *Proc. IEEE VTC*, 1992, pp. 617–620.

[11] S. Y. Seidel, and T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored building," *IEEE Trans. on Antennas and Propagation*, vol. 40, pp. 207–217, Feb. 1992.

[12] Y. Chen and H. Kobayashi, "Signal Strength Based Indoor Geolocation," in *Proc. IEEE ICC*, 2002, pp. 436–439.

[13] A. M. Ladd et al, "Robotics-Based Location Sensing using Wireless Ethernet," in *Proc. MOBICOM*, 2002, pp. 227–238.