Sixth Annual IEEE International Conference on Pervasive Computing and Communications

Location Fingerprint Analyses Toward Efficient Indoor Positioning

Nattapong Swangmuang and Prashant Krishnamurthy

Graduate Program in Telecommunications and Networking
University of Pittsburgh
135 N. Bellefield Avenue, Pittsburgh, Pennsylvania 15260
Email: nswang@mail.sis.pitt.edu, prashant@mail.sis.pitt.edu

***Abstract*—Analytical models to evaluate and predict "precision" performance of indoor positioning systems based on location fingerprinting are lacking. Such models can be used to improve the design of positioning systems, for example by eliminating some fingerprints and reducing the size of the location fingerprint database. In this paper, we develop a new analytical model that employs proximity graphs for predicting performance of indoor positioning systems based on location fingerprinting. The model allows computation of an approximate probability distribution of error distance given a location fingerprint database based on received signal strength and its associated statistics. The performance results from the simulation and the analytical model are found to be congruent. This model also allows us to perform analysis of the internal structure of location fingerprints. We employ the analysis of the internal structure to identify and eliminate unnecessary location fingerprints stored in the database, thereby saving on computation while performing location estimation.**

I. Introduction

Location of mobile computers is essential to enable location-aware applications in wireless pervasive computing [1]. By utilizing location information, location-aware computers can enable many location-based services (LBS) possible for mobile users. The global positioning system (GPS), although operating well for outdoor, does not perform very well in urban environment especially inside buildings. Consequently, location fingerprinting based positioning systems using wireless local area networks (WLANs) have been suggested as a viable alternative to provide location information for indoor areas. With no requirement of specialized hardware modules at the mobile station (MS) nor additional spectrum for positioning, location fingerprinting based positioning can be deployed using current WLAN infrastructure. In addition, the technique is relatively simple and more robust in multipath conditions compared to techniques such as time-of-arrival (TOA) and angle-of-arrival (AOA).

The location fingerprinting technique connects location-dependent characteristics such as received signal strength (RSS) from known access points to a location and uses these characteristics to infer the location. Locations within the entire area of interest are usually expressed as a set of rectangular grid points. Fingerprinting based positioning is divided into *offline* and *online* phases. In the offline phase, by site-surveying, the RSS from multiple access points (APs) at different grid points are collected and stored in a fingerprint database, often called a *radio map*. The vector of *mean* RSS values at point on the grid is called the location fingerprint of that point [2]. In the online phase, a MS will measure a *sample* fingerprint vector of RSSs from different APs at its position. The sample fingerprint is sent to a central server in the WLAN infrastructure. The server compares the measured fingerprint to fingerprints stored in the radio map for determining the location of the MS on the grid. The estimated result is then reported back to the MS. Commonly, the Euclidean distance between the measured sample and each fingerprint in the radio map is computed and used for location estimation. The grid coordinate associated with the fingerprint that yields the smallest Euclidean distance is selected as the estimate of the position. Other methods using Bayesian modeling [3] and Statistical Learning [4] have been suggested to map sample RSS vectors to the fingerprint in the radio map.

In deploying positioning systems, several performance benchmarking metrics needed to be considered [5]. The most fundamental metric is *location accuracy*, which is usually reported as an *error distance* between the estimated location and the actual MS location. Another metric is *location precision*, which is the percentage of successful location estimates with a given accuracy. With 100% accuracy (error distance = 0m), the location precision metric will correspond to the probability of returning the correct location. Conventionally, a large number of location fingerprints are collected from a regular rectangular grid of locations in the given area during the offline phase. This process can be slow and laborious. Location fingerprints corresponding to a regular *physical grid* are typically scattered and asymmetric (depending heavily on site-specific signal propagation). The Euclidean distance between some sets of fingerprints can be very small compared to the variation of the RSSs at the corresponding locations. Collecting such fingerprints with small "signal distance" in the radio map may not improve performance and also cause extra computational effort while estimating the location of a MS. Including some fingerprints in the radio map may even reduce location precision. Thus a deployment challenge is how to efficiently collect fingerprints and construct a radio map so that it contains only necessary location fingerprints without sacrificing performance.

However, models to analyze and predict accuracy and precision performance of indoor positioning systems based on

0-7695-3113-X/08 $25.00 © 2008 IEEE
DOI 10.1109/PERCOM.2008.33

location fingerprinting are lacking. Previous work on location fingerprinting [2] considers the probability of returning the correct location, but does not provide means for computing the distribution of the error distance which is necessary to evaluate precision and accuracy. In this paper, we develop a new analytical model that employs proximity graphs for predicting performance of indoor positioning systems based on location fingerprinting. The model allows computation of an approximate probability distribution of error distance given a location fingerprint database based on received signal strength and its associated statistics. The performance results from the simulation and the analytical model are found to be congruent. This model also allows us to perform analysis of the internal structure of location fingerprints. We employ the analysis of the internal structure to identify and eliminate unnecessary location fingerprints stored in the database, thereby saving on computation while performing location estimation.

The paper is organized as follows: Related work on indoor positioning systems is described in Section II. Section III discusses the previous analytical model [2] that characterizes the Euclidean distance in signal space between the measured sample vector and fingerprints in the radio map as random vectors. This model is then used to determine the probability of selecting the correct fingerprint (error distance = 0m). In Section IV, we present a new analytical model to derive the approximate probability distribution of the error distance based on proximity graphs. Comparisons between the new model and simulations are shown in Section V for (a) a simplified radio map and (b) a real radio map based on measurement data. Finally, discussion and conclusion of the paper are presented in Sections VI and VII respectively.

II. Related Work

In the past years, many indoor positioning systems have been been studied and developed by the research community. Systems can be grouped by those utilizing existing infrastructure (e.g., 802.11-compliant) and those relying on additional hardware or technology. RADAR [6] is the first WLAN-based positioning system that computes a MS's location based on the RSS from many APs. Placelab [7] is a system that depends on information about the AP's coordinates in a database in order to predict location. Both systems use the mean RSS vector at locations on a grid to represent the fingerprint. There are variations of techniques based on the mean RSS vector that aim to improve system performance [8][9]. Some research works study the signal strength distributions from APs and use probabilistic algorithms to estimate the MS's location [10][11]. However, the accuracy of location estimation using the mean RSS vector (known as *deterministic approach*) or the RSS distribution (known as *probabilistic approach*) has been reported to be similar [9]. Where models exist, they are used to determine performance bounds (e.g., the Cramer-Rao bound is suggested for predicting localization performance in wireless sensor networks [12]). They are not geared towards predicting or testing actual system performance. None of the previous works has developed a model for predicting performance of an indoor positioning system with the exception of [2].

There are many positioning systems that require specialized hardware or additional technologies. The Active Badge system [13] uses ceiling-mounted infrared sensor detectors to detect signals from a MS's active badge (so that a central unit can process the data and determine the MS's location). The drawback of the system is that infrared signals have a limited range and are susceptible to interference from sunlight and fluorescent light indoors. The Cricket system [14] uses ultrasound and RF receivers to derive time-of-flight information and applies multilateration for location estimation. Despite an accuracy of a few centimeters, many receivers must be deployed in the system. Systems utilizing RFID-based technology (e.g., SpotON [15] and LANDMARC [16]) have also been studied. A hybrid system like [17] cooperates Zigbee radios, as additional proximity sensors, and collaborative localization in clusters to improve accuracy of the WLAN-based positioning system.

The problem of analyzing relationships among a set of scattered points has been studied using the concepts of Voronoi diagram and proximity graphs. A Voronoi diagram helps determine the "closest" region to an associated point (in comparison with other points) and it has been applied to problems in geophysics, biology, computational geometry, and engineering [18]. A proximity graph is a particular type of graph that has been used to represent neighbor relationships between such points. These concepts, although widely applied in other fields, have never been adopted in the field of wireless indoor positioning system. We use these concepts in our modeling and analysis efforts.

III. Indoor Location Fingerprint Model

In this section we describe the preliminary model described previously in [2] which is a precursor to our model. Consider an indoor positioning system overlaid on a WLAN in a single floor inside a building. We assume that there are N access points (APs) 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 location can be represented by a label (x, y) where x and y represent the 2D coordinates on the floor. We assume zero height (i.e., $z = 0$) for all coordinates.

During the offline phase a total of $K = L^2$ of the RSS vectors are collected from site-survey at predetermined grid points. All K entries are recorded in a radio map database and each entry includes a mapping of the grid coordinate (x, y) to the vector of corresponding RSS values from all APs in the area. Each element in each vector in the database is assumed to be the *mean* of the RSS from each of the N access points in the area. This is typically done by collecting a large number of samples of the RSS for different orientations of the MS, and calculating an average value. This approach reduces variations

due to orientation and time in the system. During the online phase, to determine the MS's location, a sample of the RSS from all APs at the current position is obtained. This sample vector is compared with all K existing entries in the database. The fingerprint entry that has the closest match to the users sample of RSS is used by the system as the estimate of the user's current location.

To derive mathematical models for predicting performance, two vectors are used in estimating the location of the MS, *a sample vector* and *a fingerprint vector*. The sample vector consists of samples of the RSS measured at the MS from N access points in the the area. The sample vector is denoted as: $R = [r_1, r_2, r_3, ..., r_N]$. Each component in the vector is assumed to be a random variable such that:

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

The fingerprint vector in the radio map consists of the means of all the RSS random variables at a particular location from the N access points and it is denoted as: $\tilde{R} = [\rho_1, \rho_2, \rho_3, ..., \rho_N]$.

The assumption that the RSS is a normally distributed random variable is acceptable. Our previous study in [19] observed that the RSS's distribution often exhibits left-skewness and varies according to its average value or its location. However, when the AP is far from the measurement location and the RSS contains no direct line-of-sight, the distribution can be closely approximated by a Gaussian distribution. Moreover, this assumption allows tractability of the mathematical model. Also, there is no observable relationship between the RSS variations transmitted by different APs. Hence, the assumption of independence is reasonable.

As discussed earlier, the "signal distance" between the sample RSS vector and the fingerprint is used to determine which of the points on the grid corresponds to the position of the MS. The (x, y) coordinates corresponding to the fingerprint that has the smallest distance from the sample RSS vector is returned as the estimated location. This approach is sometimes referred as *the Nearest Neighbor Point in Signal Space* (NNSS) [6]. The signal distance, being different from physical distance, is calculated by the Euclidean distance between $\tilde{R}$ and R and it is given as: $Z = [\sum_{i=1}^{N}(\rho_i - r_i)^2]^{1/2}$. A detailed analysis of the characteristics of the Euclidean distance metric Z for indoor location fingerprinting can be found in [2]. For example, Z can have either a central or non-central chi distribution.

Next we discuss the mathematical model modified from [2] for predicting the probability of selecting the correct location fingerprint when the grid system contains two locations and multiple locations.

A. Probability of Selecting the Correct Location Fingerprint from a Set of Two

Consider a grid system with two grid points, indexed as i and k, and assume a MS is at the i^{th} grid point. We define the *pairwise error probability (PEP)* as the probability that a sample vector R_i is closer to the fingerprint vector $\tilde{R}_k$ than the target fingerprint vector $\tilde{R}_i$. In fact, it is the probability that we have an incorrect estimate of the location (picking the k^{th} grid point instead of the i^{th} grid point). Given sd_{ik}, the Euclidean signal distance between $\tilde{R}_i$ and $\tilde{R}_k$, we can compute the pairwise error probability $PEP(\tilde{R}_i, \tilde{R}_k)$ between the target (correct) fingerprint vector $\tilde{R}_i$ and another (incorrect) fingerprint vector $\tilde{R}_k$ as follows:

$$\begin{aligned} PEP(\tilde{R}_i, \tilde{R}_k) &= P\{R_i \text{ is closer to } \tilde{R}_k \text{ than } \tilde{R}_i\} \\ &= P\{||\tilde{R}_k - R_i|| < ||\tilde{R}_i - R_i||\} \\ &= \int_{x=\frac{sd_{ik}}{2}}^{x=\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(\frac{-x^2}{2\sigma^2})}\, \mathrm{d}x \\ &= Q(\frac{sd_{ik}}{2\sigma}). \end{aligned} \quad (1)$$

$||.||$ denotes the magnitude (i.e., Euclidean distance) of an RSS vector. $Q(x)$ represents the right-tail probability for a standard Gaussian random variable where the random variable exceeds x. Note that $PEP(\tilde{R}_i, \tilde{R}_k) = PEP(\tilde{R}_k, \tilde{R}_i)$ (i.e., the MS is at the k^{th} grid point, but its location is estimated to be the i^{th} grid point).

The chance of the event that the distance between the sample RSS vector R_i and the correct location fingerprint $\tilde{R}_i$ is smaller than the distance between the sample RSS vector R_i and the incorrect neighboring location fingerprint $\tilde{R}_k$ is recognized as the probability of returning the correct location. When considering only two location fingerprints, the pairwise probability of returning the correct location or *pairwise correct probability (PCP)* between the correct fingerprint vector $\tilde{R}_i$ and an incorrect fingerprint vector $\tilde{R}_k$ can be computed as:

$$PCP(\tilde{R}_i, \tilde{R}_k) = 1 - PEP(\tilde{R}_i, \tilde{R}_k) = 1 - Q(\frac{sd_{ik}}{2\sigma}). \quad (2)$$

B. Probability of Selecting the Correct Location Fingerprint from a Set of Many

In a positioning system, the radio map database contains several location entries and fingerprints. To find the probability of returning a correct location, the joint probability density function (PDF) of the location fingerprints needs to be known. Let $C_k = ||\tilde{R}_i - R_i|| - ||\tilde{R}_k - R_i||$ be the comparison variable. The variable C_k compares the distance between the sample RSS vector and the correct fingerprint $\tilde{R}_i$ and the corresponding distance to the incorrect fingerprint $\tilde{R}_k$. The index k runs from 1 to K excluding the correct location index (the index i in this case). So, the probability of correct decision is described by:

$$\begin{aligned} &Prob\{\text{Correct Decision}\} = \mathcal{P}_c \\ &= P\{C_1 \le 0, \cdots, C_{i-1} \le 0, C_{i+1} \le 0, \cdots, C_K \le 0\} \end{aligned} \quad (3)$$

Unfortunately, deriving such a probability analytically proves to be difficult and may not be practical when there is a large number of location fingerprints in the database. The model in [2] applies a simple approximation that assumes independence among the many different comparison variables. So

$$Prob\{\text{Correct Decision}\} = \mathcal{P}_c = \prod_{\substack{k=1 \\ k \neq i}}^{K} Pr\{C_k \leq 0\}$$

$$Prob\{\text{Erroneous Decision}\} = \mathcal{P}_e = 1 - \mathcal{P}_c. \quad (4)$$

This simple model yields a reasonable estimation for the probability of selecting the correct location. However, the above analytical model is not sufficient to find the *probability distribution* of the *error distance*. To obtain the distribution of the error distance, we need an estimate of the probability of selecting an arbitrary location (and then associating it with the corresponding error in physical distance). Further, we want to find the chance of picking one location against other locations in order to determine the "level of importance" of a corresponding fingerprint in terms of how it impacts the probability of returning the correct location. In fact, the internal structure or distance relationships among location fingerprints has a direct impact on the performance of the positioning system. Although seemingly unique, each fingerprint will have different influence level in terms of the chance of selecting the fingerprint and thus the distance error in the estimated location. Therefore, understanding the fingerprint structure that dictates both a decision region and the probability of fingerprint selection is critical to the design of a good wireless positioning system.

In the next section, we will discuss an extended analytical model that considers fingerprint structure in order to better model the probability distribution of fingerprint selection and thus the distance error.

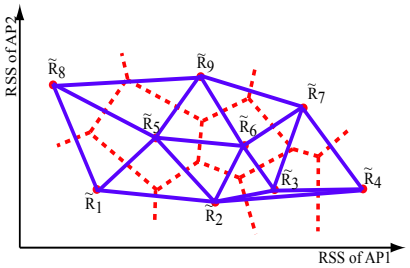

Fig. 1. Voronoi Diagram of 9 Fingerprints with DG

IV. Analytical Model for Probability Distribution of Picking Fingerprints

Toward the above mentioned goal, in this section we introduce useful tools. They are the Voronoi diagram and different proximity graphs. The Voronoi diagram is a tool introduced to find "decision regions" for each fingerprint in the system (explained below). A proximity graph is a tool that helps analyze the fingerprint structure and yield proximity

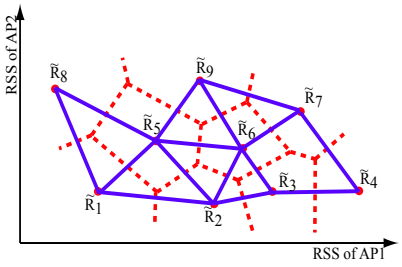

Fig. 2. Voronoi Diagram of 9 Fingerprints with GG

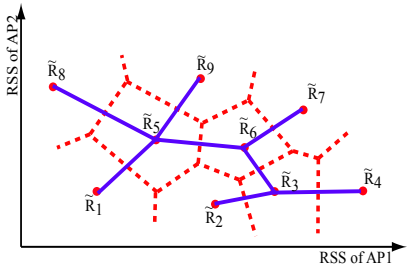

Fig. 3. Voronoi Diagram of 9 Fingerprints with RNG

information or a "neighbor set" of a given fingerprint. A neighbor is a fingerprint that is believed to be more important to the precision of selecting the target fingerprint, i.e., one that is "relatively close" to the target fingerprint in signal space. Applying these tools, we create a mathematical model for approximating the probability distribution of fingerprint selection.

A. Voronoi Diagram for the Location Fingerprint

The Voronoi diagram of a set of fingerprints $\mathcal{FP} = \{\tilde{R}_1, \tilde{R}_2, \cdots, \tilde{R}_K\}$ is defined as a division of the space according to the nearest-neighbor rules, where each fingerprint from $\mathcal{FP}$ is associated with a region of the Euclidean space closest to a given fingerprint from $\mathcal{FP}$. Such a region is called a *Voronoi region* (or a *decision region*) for a fingerprint. From the definition, the Voronoi region of a fingerprint $\tilde{R}_i$ in N-dimensional Euclidean space (N access points) can be expressed as:

$$V(\tilde{R}_i) = \{R : ||\tilde{R}_i - R|| \leq ||\tilde{R}_j - R||, \text{ for } \forall j \neq i\}.$$

Combining the Voronoi regions of all the fingerprints yields the Voronoi diagram for the radio map. Alternatively, the Voronoi diagram can be defined by a bisector $B(\tilde{R}_i, \tilde{R}_j) = \{R : ||\tilde{R}_i - R|| = ||\tilde{R}_j - R||\}$ between two fingerprints $\tilde{R}_i$ and $\tilde{R}_j$. The bisector is a line perpendicular to the line segment $\overline{\tilde{R}_i \tilde{R}_j}$ that bisects this segment in Euclidean 2D space. It is a plane (hyperplane) perpendicular to the segment $\overline{\tilde{R}_i \tilde{R}_j}$ that bisects this segment in 3D (higher-D). The Voronoi diagram is created using bisectors between any two fingerprints to derive decision regions for all fingerprints in the radio map. A bisector is also referred as a *Voronoi edge*. Fig. 1 shows an example of a Voronoi Diagram (red dashed lines) for 9

fingerprints in 2D space. If a sample RSS vector falls in the Voronoi region of a fingerprint $\tilde{R}_m$, it is closest to that fingerprint in terms of Euclidean distance. Thus, the NNSS approach will pick $\tilde{R}_m$ (or decide that location on the grid corresponding to $\tilde{R}_m$ is the correct location). The Voronoi regions can be used to determine the probability of picking a particular fingerprint given the statistics of the random RSS vector. The method of doing this is to determine the probability that the RSS vector falls in the Voronoi region. This is mathematically tractable for rectangular Voronoi regions, but not so for irregular polygonal Voronoi regions. Instead, we use the related concept of proximity graphs to approximate this probability.

B. Proximity Structure and Proximity Graphs

We use the idea of proximity graphs to extract structure information (especially proximity information such as a neighbor set) for fingerprints. Two fingerprints are "close together" and they are neighbors if there are no other fingerprints in a certain "forbidden region" defined differently by different proximity graphs. We consider three proximity graphs; the Deluanay graph (DG), the Gabriel graph (GG), and the relative neighborhood graph (RNG). Here only a brief description for each graph is provided. More details and constructing algorithms for these graphs can be found in most references on geometric graph theory [18].

DG: DG is defined based on the Voronoi diagram through the principle of duality. In the graph, there exists a Delaunay edge between two points u, v if they share the same Voronoi edge as boundary. This graph is referred to as *Delaunay triangulation*.

GG: GG is a graph that contains a Gabriel edge between two points u, v – if a diametral circle from these two points contains no other point w. Mathematically, $sd_{uv}^2 \leq sd_{uw}^2 + sd_{vw}^2$. In fact, a Gabriel edge is a Delaunay edge that actually cuts across a Voronoi edge in the Voronoi diagram.

RNG: RNG is a graph that contains an edge between two points u, v – if there is no other point w that is simultaneously closer to both points than they are to one another. Mathematically, $sd_{uv} \leq max[sd_{uw}, sd_{vw}]$.

DG, GG, and RNG are shown by solid line graphs in Figs. 1, 2, and 3 respectively. Note that RNG⊂GG⊂DG. From the particular proximity graph, we define a neighbor fingerprint as a fingerprint point that has edge connected to a given fingerprint. For example, in DG a neighbor fingerprint is one that has a Delaunay edge connected to a given fingerprint (from Fig. 1, neighbors of fingerprint $\tilde{R}_6$ are fingerprint $\tilde{R}_2$,$\tilde{R}_3$,$\tilde{R}_5$,$\tilde{R}_7$, and $\tilde{R}_9$). The same idea is applied for defining neighbors with GG and RNG.

Different proximity graphs can yield different sets of neighbors. A good graph (in our case) must give us the right set of neighbors such that they represent the top candidates for location fingerprint selection (given the location of a MS). Note that the DG yields the largest set of neighbors while an RNG yields the smallest set of the three graphs. The way we employ proximity graphs is described in the next subsection.

C. Approximate Probability Distribution using Proximity Graphs

As mentioned in III-B, to find the exact probability of selecting a fingerprint and thus a location on the grid, the joint probability density function (PDF) of fingerprints is needed. A mathematical expression for such a function is very difficult to derive. Moreover, we simply wish to find the probability of selecting one fingerprint against others in order to evaluate the influence level of a fingerprint on the probability of correctly selecting a fingerprint. Given a MS at the i^{th} grid point, the probability of selecting fingeprint $\tilde{R}_k$ is approximated using a new model as follows:

$$Prob\{\text{Selecting Fingerprint } \tilde{R}_k\} = PEP(\tilde{R}_i, \tilde{R}_k) \times \prod_{j \in \text{neighbor of } i} PCP(\tilde{R}_k, \tilde{R}_j). \quad (5)$$

The idea behind this equation is as follows. Instead of using all of the comparison variables C_k as in (4), we use only the neighbors as the most significant candidates – and still use an independence assumption. Given that the MS is at grid point i, it is the probability of selecting $\tilde{R}_k$ and not $\tilde{R}_i$, AND the probability of selecting $\tilde{R}_k$ and not any of the other neighbors of $\tilde{R}_i$. That is, the above approximation weighs the $PEP(\tilde{R}_i, \tilde{R}_k)$ with all PCPs between fingerprint $\tilde{R}_k$ and only "neighbors" (as defined by the proximity graphs) of the correct fingerprint $\tilde{R}_i$. The influence from remote fingerprints is ignored by using this approach. The set of neighbors to be employed in the approximation depends on the choice of the proximity graph. Moreover, this approach allows us to compute the probability of not only picking the correct location, but also the probability of picking any of the neighbors in the set. However, fingerprints outside the neighbor set are assumed to be never picked (although there is always a negligible probability that this may happen). To find the probability of selecting the correct fingerprint, the first line in (4) can be used. However, instead of using all K fingerprints we use only the neighbor set of the correct fingerprint in the computation, for better estimation (a claim that is validated by our results in section V).

V. Performance Evaluation

In this section, we evaluate the analytical model discussed in section IV. We study the results of the probability of fingerprint selection (for correct and incorrect fingerprints). We then look at the results of the distribution of the probability of fingerprint selection. The distribution of the error distance of the location estimate is also studied. We do this for a simple system model described below as well as for a real radio map derived from measurement.

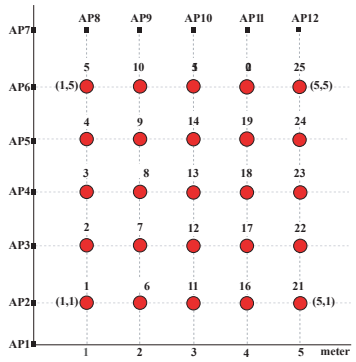

Fig. 4. 25 Grid points for an indoor positioning system

TABLE I
EXAMPLE RADIO MAP

Access Points	**AP1** (dBm)	**AP4** (dBm)	**AP8** (dBm)	**AP12** (dBm)
Coordinate	**(0,0)**	**(0,3)**	**(1,6)**	**(5,6)**
Loc7(2,2)	-57.1918	-53.1094	-63.7390	-67.0888
Loc8(2,3)	-61.4089	-51.1712	-59.1300	-64.2355
Loc9(2,4)	-65.1506	-53.1094	-53.1094	-61.4089
Loc12(3,2)	-61.4089	-59.1300	-65.1506	-65.1506
Loc13(3,3)	-64.2355	-58.2149	-61.4089	-61.4089
Loc14(3,4)	-67.0888	-59.1300	-57.1918	-57.1918
Loc17(4,2)	-65.1506	-63.7390	-67.0888	-63.7390
Loc18(4,3)	-67.0888	-63.2124	-64.2355	-59.1300
Loc19(4,4)	-69.2330	-63.7390	-61.4089	-53.1094

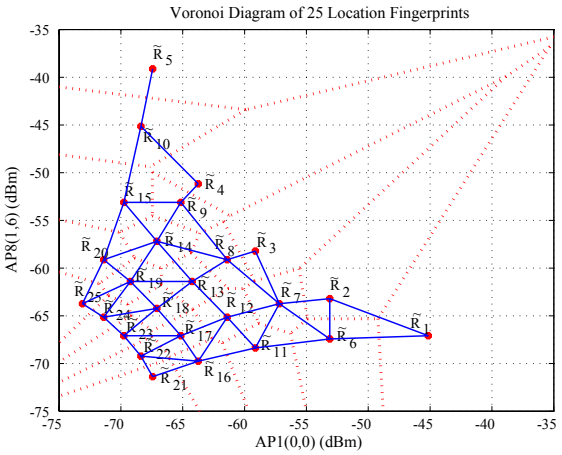

Fig. 5. Voronoi Diagram and GG of 25 Location Fingerprints

A. System Model

The system model considered for evaluation is as follows. We use a 25 grid point system (see Fig. 4) with a grid spacing of 1 meter (≈ 3 feet). We place access points along the outer most positions (small dark rectangles in Fig. 4). Initially we consider only two access points: AP1 = (0,0) and AP8 = (1,6). The position of the mobile station could be at any one of the 25 locations in Fig. 4. Suppose the physical distance of the k^{th} grid point from the j^{th} AP is $d_{j,k}$ meters. The mean or expected value of r_j for the grid point is calculated from the mean path loss given by:

$$Pl(d_{j,k}) \quad = \quad Pl(d_0) + 10.\alpha.\log_{10}(d_{j,k}). \qquad (6)$$

Here $Pl(d_0)$ is the free-space loss at the reference distance of d_0 = 1 m (i.e., 54.13 dBm for line-of-sight propagation (LOS) and 37.3 dBm for non-line-of-sight propagation (NLOS) as reported in some measurements [20]). The variable α denotes the path loss exponent, which for indoor locations could be between 1-6 [21]. The mean received signal strength $E\{r_j\}$ can be computed using:

$$E\{r_j\} \quad = \quad \rho_j = P_t - Pl(d_{j,k}) \qquad (7)$$

where P_t 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 = 4 dB as reported in [22]. Other values of σ for indoor location systems are reported in [23]. A more accurate path loss prediction model, such as those including wall and floor attenuation factors suggested in [6], could also be used. The path loss equation only provides us with the mean received signal strength value. It is possible to use values from actual measurements as well without changing our analytical model.

Table I shows the database of location fingerprints when access point AP1, AP4, AP8, and AP12 in Fig. 4 are deployed. The table contains the location fingerprints of locations 7-9, 12-14,and 17-19, which are located around the center of the system. Note that 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 location grid are at the same distance from the access point. Additional access points make the fingerprint unique. An example of a Voronoi diagram of the 25 location fingerprints when AP1 and AP8 are deployed, along with its GG, is shown in Fig. 5. Note that, although a symmetric square physical location grid is used, the resulting fingerprints are not necessary symmetric in signal space. Fingerprints ($\tilde{R}_{21} - \tilde{R}_{25}$) that are far from an AP tend to stay closer while fingerprints closer to the APs ($\tilde{R}_1 - \tilde{R}_5$) tend to be apart in signal space. The decision regions for fingerprints are also different in shape and size. Bounded (unbounded) regions are found at inner (outer) grid points.

Next, we compare the results from simulation and the analytical model. We simulated 10,000 RSS samples from a given MS location and applied the nearest neighbor computation to estimate its location. Then we computed relevant performance metrics (i.e., error probabilities, error distances) from simulations.

B. Results of Error Probability of Fingerprint Selection

We first study the precision at zero-meter accuracy in terms of the probability of error of fingerprint selection ($\mathcal{P}_e$ – the probability of not picking the correct fingerprint). In particular, the impact of the standard deviation σ of RSS on $\mathcal{P}_e$ is considered. We choose the grid point 13, in Fig. 4, as the MS's actual location (it is at the center of our grid system).

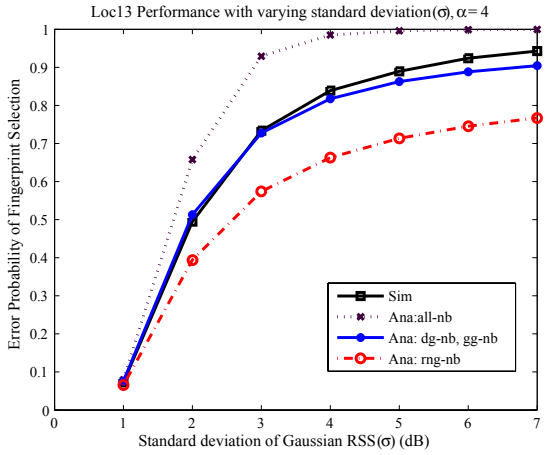

Fig. 6. Impact of the standard deviation on error probability

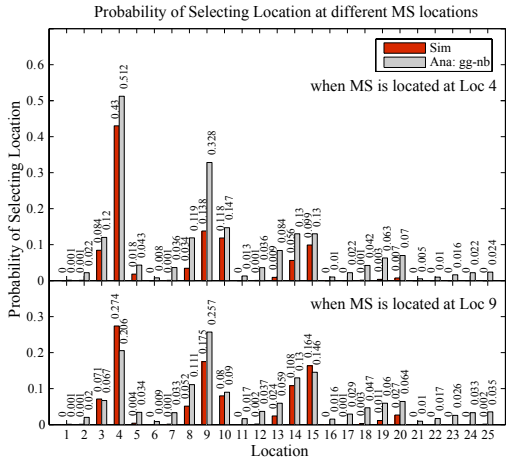

Fig. 7. Probability Distribution of Fingerprint Selection

We consider the standard deviation σ between 1-7 dB which corresponds to values seen in extensive experiment results [24]. The results from both simulation and the analytical model are given in Fig. 6.

In the case of the analytical results, we consider different approximations depending on the number of fingerprints involved in the probability estimation. First, we select all the other 24 fingerprints (*Ana:all-nb*) to compute the probability of error as given in (4). Second, we use only the neighbor set (*Ana:dg-nb, gg-nb, rng-nb*) derived from the different proximity graphs to estimate the probability value. Clearly, when the standard deviation σ increases, the error probability also increases. The error probability estimation using Delaunay neighbors shows the closest results to the simulation. The same result is obtained by using Gabriel neighbors because both graphs yield the same neighbor set for the grid point 13. In *Ana:all-nb*, remote fingerprints (those that are far away from the correct fingerprint in signal space) are included in the comparison variables $[Pr\{C_k \leq 0\}]$ in (4). Since such probabilities are multiplied assuming independence, the probability of correct decision decreases and the error probability increases, thereby giving higher results compared to simulations. Note that the random variables C_k are not really independent. On the other hand, using the relative neighborhood graph [*Ana:rng-nb*] gives a lower probability of error compared to simulation since it underestimates the number of significant neighbor fingerprints used in the approximation.

TABLE II
COMPARISON OF ERROR PROBABILITY, σ = 4

APs	Sim	Anal: all-nb	Anal: dg-nb	Anal: gg-nb	Anal: rng-nb
2	0.8334	0.9849	0.8176	0.8176	0.6631
3	0.7603	0.9409	0.8592	0.7891	0.6830
4	0.6337	0.8285	0.7770	0.6792	0.5759
5	0.5278	0.7111	0.6677	0.5815	0.4890

Table II summarizes the comparison of the error probability for estimating location 13 when the number of access points in the system increases. We study different cases when we deploy 2 APs (AP1-AP8), 3 APs(AP1-AP8-AP12), 4 APs(AP1-AP4-AP8-AP12), and 5 APs(AP1-AP4-AP8-AP10-AP12) in the system. Here, we assume the same standard deviation of RSS (σ = 4 dB) from all APs. As expected, increasing the number of APs will reduce the error probability. We can see that *Ana:all-nb* provides a poor upper bound approximation compared to simulation results. The *Ana:dg-nb* and *Ana:rng-nb* provide better upper and lower bounds as the number of APs increases. The *Ana:gg-nb* is the closest to the simulation results. The neighbor set derived from GG reflects fingerprints that have a better chance to be picked outside of the correct fingerprint. We also did an extensive study by locating the MS at different locations with different combinations of the APs. It turned out that similar results were observed. From this, we conclude that using the neighbor set derived from a Gabriel Graph makes the most sense in estimating the probability of error.

C. Results of Probability Distribution of Fingerprint Selection

Next we evaluate our analytical model to see how well it approximates the probability distribution of picking specific locations given the MS is at one grid point, as discussed in IV-C. Fig. 7 shows examples for the comparison of the distributions between the analytical model using GG and simulations, when the MS is located at location 4 and 9 respectively. Here we assume that $\sigma = 4$ as before. Note that the results from the analytical model are found to be close to the simulation results in both cases. We also found that using DG and RNG give relatively close results to the simulation with only small differences in shape and height of the histograms, but the GG is the best. We considered several different MS's locations – in each case the analytical model gives distributions that are close to those from simulation.

As it provides reasonable estimates for the probability distribution, the analytical model can now be used to find some fingerprints that, if retained in a radio map, can degrade

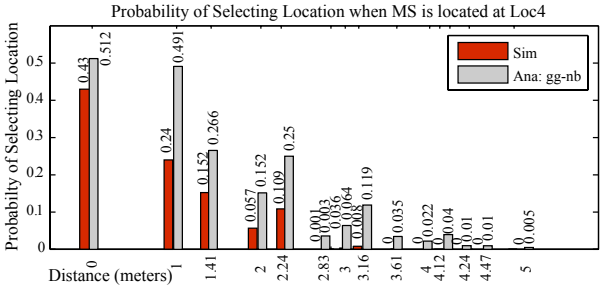

Fig. 8. Probability Distribution (by meters) of Fingerprint Selection

the overall performance of the location estimation. Those fingerprints also cause inefficient nearest neighbor computation during the online phase with a complexity of $O(n*D)$, where n is the cardinality or number of the location fingerprints and D is dimensionality (number of RSSs from the APs) in the fingerprint system. To give an example, let consider all the analytical results in Fig. 7. We can see that, when the MS is at a grid location 4, the probability of selecting the grid location 4 (0.512) is a lot higher than the next highest probability of selecting a different grid location (location 9 with probability 0.328). However, when the MS is located at the grid location 9, the two highest probability values are almost equal (0.206 for location 4 and 0.257 for location 9). This means that retaining the fingerprint at the grid location 9 in the radio map results in the system having a relatively high chance of falsely returning the grid location 4 as a correct location of the MS. In fact, the simulation results in the figure report that the system picks the grid location 4 as the correct location more often than the actual grid location 9 (0.274 over 0.175). Therefore, this suggests that we should not include location 9's fingerprint ($\tilde{R}_9$), and it should be eliminated from the database. If possible, such locations must be avoided in the laborious offline phase as well.

The fingerprint elimination procedure is described below:

1) Compute all prob. distributions of all locations L1
2) For each L1's distribution:
- *If $prob\{$picking L1$\}$ is not the highest, eliminate L1*
- *If there exists L2 where $|prob\{$picking L1$\} - prob\{$picking L2$\}| < threshold$,*
  - *Check L2's distribution. Eliminate L1 if $prob\{$picking L2$\}$ is the highest& it differs from $prob\{$picking L1$\} >$ threshold*

We can repeat this procedure for probability distributions associated with different MS grid locations and determine which fingerprints should be removed from the radio map. A threshold value of 0.2 is chosen in this work based on trial and error.

Fingerprint elimination procedures are effective and make sense when the positioning system is deployed in a large area such as entire floor of a building where hundreds of grid locations are present. Fewer numbers of fingerprints in the database mean smaller numbers of comparisons needed during the online phase. From an extensive (analytical) evaluation, in our system of 25 fingerprints, we can identify four more fingerprints ($\tilde{R}_{17}$,$\tilde{R}_{19}$,$\tilde{R}_{22}$,$\tilde{R}_{24}$) that should be eliminated. By doing so, we can save about 5/25 = 20% of computation required for nearest neighbor search during the online phase.

The probability distribution of fingerprint selection can be represented in terms of physical distance error as well (which is necessary for determining precision). Fig. 8 gives an example of the distribution plot of error distance in meters when the MS is at grid location 4. As expected, the probability decreases when error distance increases. Again, the distributions from simulation and analysis, though not perfectly precise, are close. Note that the probabilities from analysis do not add to one, since they are determined using an approximation that assumes independence between comparisons of random variables.

A question that arises is how eliminating certain fingerprints impacts performance. In Fig. 9, the cumulative probability distribution of the error distance in meters, with and without elimination of fingerprints from the radio map, are shown. The results are averaged based on simulations from all 25 locations. Again, we keep $\sigma = 4$. We can see that, after eliminating some fingerprints ($\tilde{R}_9$, $\tilde{R}_{17}$, $\tilde{R}_{19}$, $\tilde{R}_{22}$, $\tilde{R}_{24}$), the cumulative distribution is only slightly different compared to the case where all fingerprints are kept in the system. The difference in the distributions occurs only in the first few meters of error and then diminishes as the error distance increases. Note that, there is about a 70 % chance that the system still maintains the distance error within 1 m after fingerprint elimination and it is close to the case when all fingerprints are used. Therefore, by applying fingerprint elimination, we do not lose much in terms of precision (probability), and yet maintain acceptable accuracy (error distance).

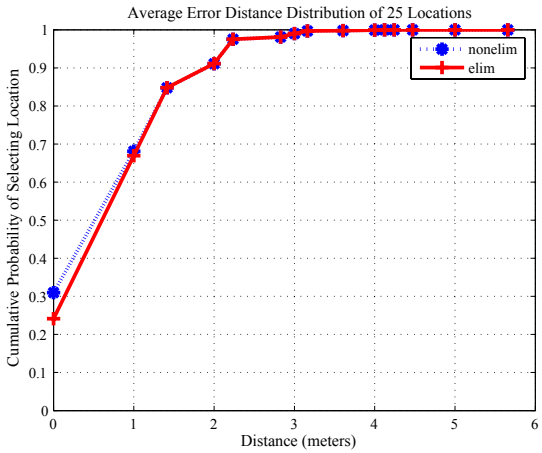

Fig. 9. Average Cumulative Distribution of Error Distance

D. Results with Fingerprint Measurement

We apply the analytical model and fingerprint elimination technique as discussed in V-C to a radio map from past measurements [24]. Unlike the simple model used for evaluation,

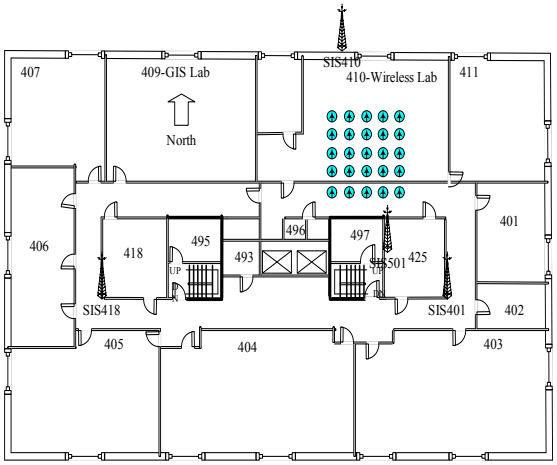

Fig. 10. Measurement Setup at the fourth floor IS building

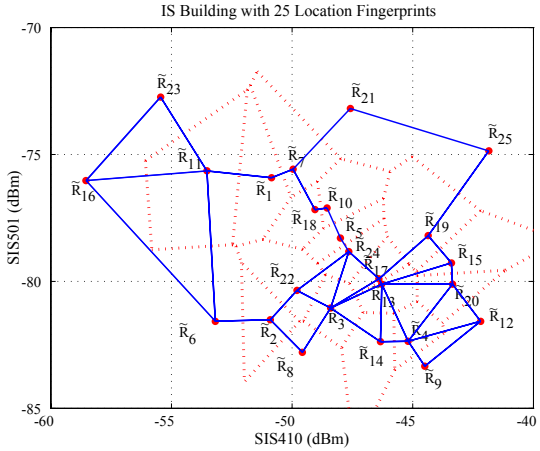

Fig. 11. Voronoi Diagram and GG of Measured Radio Map

the measured fingerprints have different σ's for different APs and locations. Therefore, instead of using (1), we approximate the new PEP by:

$$PEP(\tilde{R}_i, \tilde{R}_k) = Q(\frac{sd_{ik}^2}{2[\sum_{j=1}^{N} \beta_{ij}^2 \sigma_{ij}^2]^{1/2}}). \quad (8)$$

The above equation is a modification – in the form of a PEP – of the derivation based on a sum of multiple Gaussian variables in [2] with different σ_{ij}. $\beta_{ij} = \rho_{ij} - \rho_{kj}$ (difference of $\tilde{R}_i$ and $\tilde{R}_k$ at the j^{th} AP of fingerprint) and σ_{ij} is the standard deviation of the RSS from the j^{th} AP at the i^{th} location. The new PEP is then used in (5) to approximate the probability distribution.

The radio map measurement was conducted in an office environment on the 4th floor of the Information Sciences (IS) building as shown in Fig. 10. The measurement setup consists of a grid area of 25 locations where 20 locations are inside room 410 and 5 locations are along the corridor. The grid spacing is approximately 1 meter. The figure also shows locations of nearby APs that can be detected at grid locations. The same location labels as in the simple model are used.

Fingerprints are based on the 2 APs named SIS410 and SIS501. The fingerprints are shown in Fig. 11 with their Voronoi regions and GG. By applying the analytical model with the Gabriel Graph, the probability distributions of picking a locations given the MS's location are approximated. Pairwise comparisons are used for fingerprint elimination. It turned out that we can identify 9 fingerprints that can be eliminated. Fig. 12 shows the average CDF of the error distance from simulations. The same assumptions as in Section III were used in the simulations, except with different σ's for the RSS values. We can see a minimally different CDF when we eliminate fingerprints. This result indicates that we can sacrifice little performance but save 9/25 = 36% by reducing the search space in the online phase.

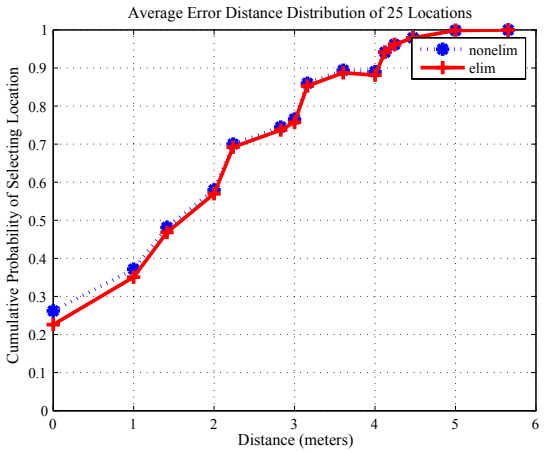

Fig. 12. Average Cumulative Distribution of Error Distance from Measurement

VI. Application to Offline Phase

Identifying and eliminating unnecessary fingerprints from a radio map database can reduce computational time spent in the online phase for fingerprinting based positioning systems. However, collecting fingerprints from all predefined grid locations in space during the offline phase proves to be very tedious and time-consuming. System designers lack hints that could help them cleverly pick locations from a site-survey during the offline phase. By studying the actual measurement data combined with the simple analytical results, we come up with a few tips that we think can be handy during pre-deployment of the system.

From the measured characteristics of the RSS, the σ of the RSS can vary from one location to another and from one AP to another. In [24], it was found that the standard deviation of the RSS is large (6-7 dB) when the MS is located near the AP with a strong RSS (-60 dBm to -40 dBm). These locations usually see a direct line of sight (LOS)

of the received signal between the AP and the MS. On the contrary, the standard deviation is small (1-2 dB) when the MS is located far from the AP with weak RSS (-95 dBm to -85 dBm). Such locations often have non line-of-sight (NLOS) of the received signal between the AP and the MS. With large deviation of the signal, the error probability of fingerprint selection is high. We believe that an "inefficient" fingerprint (one that is hardly picked as a correct location) is likely to be gathered at a location close to the AP. In other words, observing a cluster of locations with a large RSS vector's magnitude will likely contain a lot of "inefficient" fingerprints. Hence, a small grid spacing should not be used at such locations during the offline site surveying. Using a larger granularity of a grid spacing in the locations with a strong RSS vector could save on labor while potentially keeping acceptable performance. In addition, as reported in [22], large standard deviations are found inside large and open space buildings, while small standard deviations are found inside small and closed spaces. So, the designer should expect to see many inefficient fingerprints when deploying systems inside open space areas. As a consequence, they must carefully select, for example, a sparser grid spacing in open environments as compared to cluttered environments. It is still necessary to require trial-and-error of measurements before determining grid points with "good" fingerprints at this stage. We are looking at carefully evaluating some possibilities (such as the impact of different standard deviations at different locations and scalability issue of location fingerprint elimination) toward more efficient location fingerprinting. This is a part of our ongoing research study.

VII. Conclusion

In this paper a new analytical model for estimating the probability distribution of fingerprint selection in an indoor positioning systems using WLANs and location fingerprinting has been proposed. The analysis and simulation results have been compared. Both results are harmonious for a variety of system parameters. The analytical model allows a system designer to reduce the search database in the online phase and some hints for fingerprint collection in the offline phase.

Acknowledgment

The authors acknowledge the work of Dr. Kamol Kaemarungsi in providing previous measurement data as well as information relevant to this study.

References

[1] M. Satyanarayanan, "Pervasive Computing: Vision and Challenges," *IEEE Personal Comm.*, vol. 8, no. 4, pp. 10–17, 2001.

[2] K. Kaemarungsi and P. Krishnamurthy, "Modeling of Indoor Positioning Systems Based on Location Fingerprinting," in *Proc. IEEE INFOCOM*, 2004.

[3] P. Castro, P. Chiu, T. Kremenek, and R. R. Muntz, "A Probabilistic Room Location Service for Wireless Networked Environments," in *Proc. Ubiquitous Computing*, 2001, pp. 18–34.

[4] M. Brunato and R. Battiti, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," *Computer Networks*, vol. 47, no. 6, pp. 825–845, 2005.

[5] P. Krishnamurthy, "Position Location in Mobile Environments," in *Proc. NSF Workshop on Context-Aware Mobile Database Management (CAMM)*, Providence, RI, Jan. 2002.

[6] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," in *Proc. IEEE Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'00)*, Tel Aviv, Israel, Mar. 2000, pp. 775–784.

[7] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. E. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. N. Schilit, "Place Lab: Device Positioning Using Radio Beacons in the Wild." in *Proc. of International Conference on Pervasive Computing*, 2005, pp. 116–133.

[8] P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, "On Indoor Position Location with Wireless LANs," in *Proc. IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC'02)*, Lisbon, Portugal, Sept. 2002.

[9] R. Battiti, M. Brunato, and A. Villani, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," Technical Report, Oct. 2002. [Online]. Available: http://rtm.science.unitn.it/ battiti/archive/86.pdf

[10] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen, "A Probabilistic Approach to WLAN User Location Estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, July 2002.

[11] A. M. Ladd, K. E. Bekris, G. Marceau, A. Rudys, L. E. Kavraki, and D. S. Wallach, "Robotics-Based Location Sensing using Wireless Ethernet," in *Proc. ACM International Conference on Mobile Computing and Networking (MOBICOM'02)*, 2002, pp. 227–238.

[12] N. Patwari, J. Ash, S. Kyperountas, A. H. III, R. Moses, and N. Correal, "Locating the Nodes: Cooperative Localization in Wireless Sensor Networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, 2005.

[13] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The Active Badge Location System," *ACM Transactions on Information Systems*, vol. 40, no. 1, pp. 91–102, Jan. 1992.

[14] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support System," in *Proc. ACM International Conference on Mobile Computing and Networking (MOBICOM'00)*, Boston, MA, Aug. 2000, pp. 32–43.

[15] J. Hightower, R. Want, and G. Borriello, "SpotON: An Indoor 3D Location Sensing Technology Based on RF Signal Strength," University of Washington, Seattle, WA, Techical Report UW CSE 2000-02-02, Feb. 2000.

[16] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANMARC: Indoor Location Sensing using Active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, Nov. 2004.

[17] L.-W. Chan, J.-R. Chiang, Y.-C. Chen, C. nan Ke, J. Y. jen Hsu, and H.-H. Chu, "Collaborative Localization: Enhancing Wifi-Based Position Estimation with Neighborhood Links in Clusters," in *Pervasive*, 2006, pp. 50–66.

[18] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed. NYC: Wiley, 2000, 671 pages.

[19] K. Kaemarungsi and P. Krishnarmurthy, "Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting," in *Proc. IEEE First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MOBIQUITOUS'04)*, Boston, MA, Aug. 2004, pp. 14–23.

[20] 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 Vehicular Technology Conference (VTC'92)*, Denver, CO, May 1992, pp. 617–620.

[21] S. Y. Seidel and T. S. Rappaport, "914 MHz Path Loss Prediction Models for Indoor Wireless Communications in Multifloored Buildings," *IEEE Trans. Antennas Propagat.*, vol. 40, no. 2, Feb. 1992.

[22] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 1st ed. Upper Saddle River, NJ: Prentice Hall PTR, 1996.

[23] J. Small, A. Smailagic, and D. P. Siewiorek, "Determining User Location for Context Aware Computing through the Use of a Wireless LAN Infrastructure," Online, Dec. 2000. [Online]. Available: http://www-2.cs.cmu.edu/∼aura55/docdir/small00.pdf

[24] K. Kaemarungsi, "Design of Indoor Positioning Systems based on Location Fingerprinting Technique," Ph.D. dissertation, Univ. of Pittsburgh, Pittsburgh, Feb. 2005.