Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting

Kamol Kaemarungsi and Prashant Krishnamurthy
Telecommunications Program, School of Information Sciences, University of Pittsburgh
E-mail: kakst112,prashk@pitt.edu

Abstract

Indoor positioning systems that make use of received signal strength based location fingerprints and existing wireless local area network infrastructure have recently been the focus for supporting location-based services in indoor and campus areas. A knowledge and understanding of the properties of the location fingerprint can assist in improving design of algorithms and deployment of position location systems. However, most existing research work ignores the radio signal properties. This paper investigates the properties of the received signal strength reported by IEEE 802.11b wireless network interface cards. Analyses of the data are performed to understand the underlying features of location fingerprints. The performance of an indoor positioning system in terms of its precision is compared using measured data and a Gaussian model to see how closely a Gaussian model may fit the measured data.

Keywords: indoor, measurement, modeling, positioning system, wireless LANs

1. Introduction

Indoor positioning systems that use *location fingerprints* and existing wireless local area network (WLAN) infrastructure have been demonstrated for indoor areas [1] where the global positioning system (GPS) does not work well [2]. The fingerprinting technique is simple to deploy compared to techniques using angle of arrival (AOA) and time difference of arrival (TDOA). Instead of depending on accurate estimates of angle or distance to determine the location, location fingerprinting associates location-dependent characteristics such as the received signal strength (RSS) with a location and uses these characteristics to infer the location. In this case, there is no need for specialized hardware at the mobile station (MS) besides the wireless network interface card (NIC) and the existing WLAN infrastructure can be reused easily.

Before a positioning system can estimate the location, a *location fingerprint database* or *a radio map* [3] must be constructed. Each entry in the database is a mapping between a position and a location fingerprint. The location fingerprint can be an average value as in the RADAR system [1] or probabilistic [3]. In the average approach that we consider in this paper, the *location fingerprint* is a vector $\boldsymbol{R}$ of the *average* RSS values from multiple access points (APs) at a particular location L. A typical vector $\boldsymbol{R} = (r_1, r_2, \ldots, r_N)$ consists of N RSS values from N APs. The radio map contains all such RSS vectors for a grid of locations in the indoor area. For positioning, a MS obtains a sample RSS vector $\boldsymbol{P} = (\rho_1, \rho_2, \ldots, \rho_N)$. The *Euclidean signal distance* between the $\boldsymbol{P}$ and $\boldsymbol{R}$ for each $\boldsymbol{R}$ in the database is computed. The location is then estimated to be that L for which the Euclidean distance is the smallest. Note that the vector $\boldsymbol{P}$ is random. An error is made when the smallest Euclidean distance occurs for a location L that is not the one at which the sample $\boldsymbol{P}$ was collected. Errors occur because the measured RSS vector is a sample of a random vector while only the average RSS vector is stored in the radio map.

Understanding the statistical properties of the location fingerprint (RSS vector) is important for the design of positioning systems for several reasons. It can provide insights into how many access points are needed to uniquely identify a location [13] with a given accuracy and precision, whether preprocessing of the RSS measurements can improve the accuracy and so on. Existing literature on indoor positioning systems focuses mainly on the accuracy performance and improvement of the location estimation algorithm and ignores the study of the RSS random vector. Knowledge of the RSS properties can in fact enable the development of better algorithms to *classify* a measured RSS vector $\boldsymbol{P}$ as belonging to a particular location *L*. Although a variety of statistical radio propagation models exist, they were developed with signal coverage, communications capability and data rate in mind. Moreover, the relationship between RSS values from multiple APs is not understood very well. The distribution of RSS values, their standard deviation, their temporal variation, and the (in)dependence of RSSs from multiple access points (APs) are important for understanding and modeling the performance of fingerprint based indoor positioning systems. For instance, the distribution of the RSS is said to be normally distributed in dBm according to the study in [5]. However, our preliminary study and the study in [6] showed otherwise.

This article presents data analyses of the RSS in an indoor environment with positioning in mind. Section 2 describes the measurement setup. Section 3 explores the effect of user's body, the effect of user's orientation, the stationarity and time-dependence, and the distribution of the RSS. The independence of the RSS from multiple APs is also analyzed in Section 3. An approximate model for the RSS is applied to the study of the precision performance of positioning systems in Section 4 to evaluate the suitability of a Gaussian approximation. We conclude the paper in Section 5.

2. Measurement setup

A standard laptop computer equipped with an Orinoco WLAN card and client manager software[1] was used to collect samples of RSS from APs inside the School of Information Sciences (IS) building at the University of Pittsburgh. The WLAN card is plugged into the PCMCIA slot on the right side of the laptop. The building has 8 floors and 10 APs installed opportunistically. The dimension of each floor is approximately 76 ft × 120 ft (23 m × 37 m). All APs are from Lucent's WAVELAN and are equipped with Orinoco WLAN cards. The radio frequency channels of IEEE 802.11b are in the 2.4 GHz band which is shared by other equipment in the industrial, scientific, and medical (ISM) band such as Bluetooth. The number of non-overlapping channels for 802.11b is three [7]. We observe that the RSS value reported by the WLAN card is an average value over a sampling period and in integral steps of 1 dBm. The received signal sensitivity of the WLAN card also limits the range of the RSS to be between -93 dBm and 0 dBm [8]. Nevertheless, the highest typical value of the RSS is approximately -30 dBm at one meter from any AP.

[1] The client manager software is a site survey tool which provides the link quality and AP monitoring capabilities.

2.1. Experimental design

The measurement in each of the studies in Section 3.2 is done by sampling the RSS data every one second. The vector of RSS data at each location forms the location fingerprint with at most three RSS elements in the vector. Four locations of measurement are chosen on the fourth floor of the IS building as shown in Figure 1 denoted as *L1, L2, L3*, and *L4*. The user's orientation corresponds to the arrows at each location in this figure. The locations of APs on this floor are labeled as SIS401, SIS410, and SIS418. The radio channels used for each AP are channel 11, 6, and 1, respectively. There is one more AP on the fifth floor labeled as SIS501 with channel number 6, but there is no AP on the third floor. The MS has a direct line-of-sight to one of the APs only at *L2*. The data were collected four times with a period of approximately one hour each for every location and at different hours of the day. The total number of RSS samples would be 4 locations × 4 hours × 3 APs = 48. However, in our experiments only 46 RSS data samples were collected because at *L1* we can receive signals from only two APs at best.

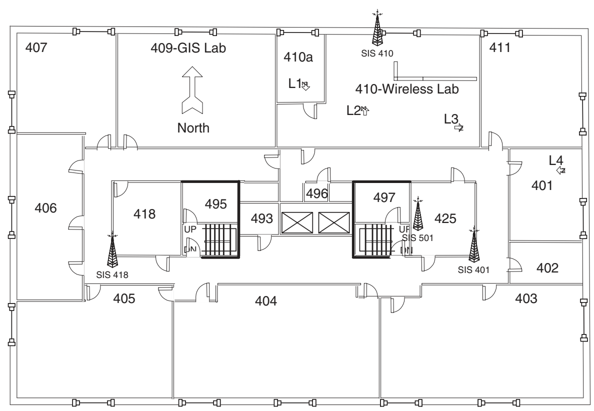

Figure 1. The fourth floor of the IS building with the locations of APs and the orientation of measurements

3. Properties of the received signal strength

Indoor radio propagation is difficult to predict because of the dense multipath environment and propagation effects such as reflection, diffraction, and scattering [9]. Multipath fading causes the received signal to fluctuate around a mean value at particular location. The received signal is usually modeled by the combined effects of large-scale fading and small-scale fading [10]. The large-scale-fading component (of interest here) describes the signal attenuation as the signal travels over a distance and is absorbed by material such as walls and floors along the way to the receiver. This component predicts the mean of the RSS and usually has a log-normal distribution [10]. Small-scale fading explains the dramatic fluctuation of the signal due to multipath fading. If there is no line-of-sight (NLOS) component, the small-scale fading is often modeled with a Rayleigh distribution. If there is a line-of-sight (LOS) component, the small-scale fading is modeled with a Rician distribution. However, these models are focused on understanding the impact of radio propagation on receiver design and signal coverage rather than from the perspective of indoor positioning systems.

The investigation of RSS data in this section is divided into four parts. All measurements were done at fixed locations. First, we consider the effect of the user's presence on a single *RSS set* (this is the set of RSS samples from one AP at a fixed location obtained over time). Second, we investigate the statistical properties of a single RSS set (the distribution, the stationarity, the time-of-day dependency, etc.). Third, we study the properties of multiple RSS sets (basically RSS values from multiple APs). We evaluate whether each RSS set is independent from the others and whether they all exhibit the same statistical properties. We also observed that in some locations the signal from certain APs is not present all the time. Finally, we compare the differences between the RSS fingerprints (these are vectors with RSS sets as their components) of two locations.

3.1. Effects of user's presence on RSS

In indoor positioning systems based on WLANs, the user typically carries the mobile station equipped with a wireless NIC. The effect of the user's presence close to the antenna plays a significant role in the mean value and the spread of the average RSS values. An observation was made in [1] that the user's orientation caused a variation in RSS level up to 5 dBm.

3.1.1. Effect of user's body. To study the effect of the user's body, we performed measurement of the signal from SIS401 at location *L1* inside the room IS 410a in Figure 1. The distance between the transmitter (AP) and the receiver (MS) is approximately 7 m and the MS does not have a clear line-of-sight to the AP. The data were recorded for two hours. During first hour, the user was present, while no user was present in the second hour. The results were analyzed by plotting histograms of the RSS for both hours. The results are shown in Figure 2.

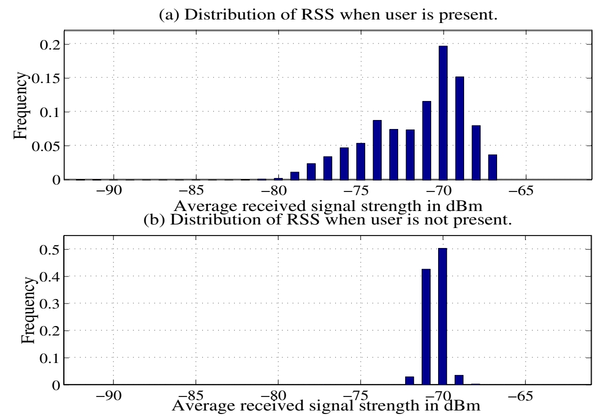

Figure 2. Comparison of histograms of RSS

Figure 2-a and 2-b depict the difference between these two distributions. The user's body influences the RSS distribution by spreading the range of RSS values by a significant amount. The standard deviation is increased from approximately 0.68 dBm to 3.00 dBm when the user is present. The mean changes from -70.4 dBm to -71.6 dBm with the user's body present. Clearly, it is essential to collect data for the radio map based on the application. When the positioning system is supposed to cater to real users, it is essential to have the user present while collecting the RSS values for the fingerprint and to take into account the effect of human's body. For applications that make use of sensors without a human presence the data should reflect that environment.

3.1.2. Effect of user's orientation. Because the resonance frequency of water is at 2.4 GHz and the human's body consists of 70% water, the RSS is absorbed when the user obstructs the signal path and causes an extra attenuation [6]. To study the effect of user's orientation, we performed another measurement at the location *L2* inside the room IS 410 in Figure 1. In this case, there is a line-of-sight between the transmitter (SIS410) and receiver and the distance between them is approximately 20 ft (6 m). Signals from SIS401 and SIS501 were also present at this location with non line-of-sight distances of 36 ft (11 m) and 22 ft (7 m), respectively. The measurement was done with four orientations (facing North, West, South, and East of the building) for a period of 15 minutes each. The results of the statistics of the RSS values from the three transmitters are shown in Tables 1, 2, and 3. The orientations that user body blocks the direct path between the AP and MS are marked with asterisks.

Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04)
0-7695-2208-4/04 $20.00 © 2004 IEEE

Table 1. LOS RSS (dBm) from SIS410 with different orientation

Statistics	North	West	South*	East
Sample Mean	-51.42	-49.73	-59.05	-53.18
Standard Deviation	4.89	4.98	3.69	3.93
Skewness	-1.51	-0.67	-0.65	-2.30

For the LOS case (Transmitter SIS410) in Table 1, when the user was facing south and the AP was behind the user, the sample mean of the RSS was lower at -59.05 dBm compared to the highest RSS of -49.73 dBm when the user faced west. The results show that the RSS can be attenuated by 9.32 dB in our case due to the obstruction from the body. This suggests that the user's orientation is crucial and should be included in computing the user location information as pointed out in [1]. The attenuation by the body of the user can even completely block the RSS from a NLOS AP as shown in Table 2 when there was *no RSS information* at all during the period that the person's back was turned towards the transmitter SIS401. This means that the location fingerprint at the same location may lack one RSS value in the vector if the user's orientation is different. The signal from SIS501 is also attenuated by 5.81 dB between the highest and the lowest RSS levels in Table 3.

Table2. NLOS RSS (dBm) from SIS401 with different orientation

Statistics	North	West*	South	East
Sample Mean	-83.12	N/A	-82.09	-83.45
Standard Deviation	1.84	N/A	2.24	1.51
Skewness	0.15	N/A	0.09	-0.01

Table3. NLOS RSS (dBm) from SIS501 with different orientation

Statistics	North	West*	South	East
Sample Mean	-79.95	-83.63	-77.82	-79.24
Standard Deviation	1.79	2.20	1.60	1.50
Skewness	-0.86	0.56	-1.19	-0.52

In what follows, we restrict the study and focus on the RSS properties when the user is present and facing one direction arbitrarily.

3.2. Statistical properties of the RSS

Traditionally, the average RSS is believed to be log-normally distributed according to popular large-scale fading models [10]. The mean value is generally predictable and believed to follow one of several standardized path loss models discussed in [9]. However, there are some conflicting conclusions regarding the RSS distribution measured at the software level by the wireless NIC for indoor radio propagation in [5] and [6]. Moreover, the standard deviation and the stationarity of the RSS are not understood very well.

3.2.1. Distribution of received signal strength. The results in [5] are based on a five second sampling period over long durations of five hours, 20 hours, and one month. Here, they conclude that the RSS is log-normally distributed (normal or Gaussian in dB) due to the similarity of the median, the mean, and the mode. However, they did not indicate whether the user was present all the time during the measurements. Thus, we suspect that the distribution of the RSS in dB that could be observed in reality may not be normally distributed as described in [5]. A recent study of a 45-second measurement period with the user's presence in [6] pointed out that the RSS distribution was non-Gaussian and asymmetric. Moreover, the histograms in [6] depicted that there could be multiple modes with one dominant mode in the distribution. The means and the modes were often different in their results.

The results of our experiment showed a similar trend as in [6]. The histogram in Figure 2-a has two modes with one dominant mode when the user is present. A visual test of Figure 2-a confirms that the RSS does not come from a normal distribution and a norm-plot test is also nonlinear. Out of the 46 RSS elements that we collected from four locations (*L1* to *L4*), most of the histograms show that the RSS does not fit the normal distribution. Only a few histograms showed a good normal approximation.

We observed that the RSS distribution tended to be left-skewed in our measurement results. The histograms that are strongly left-skewed are usually the ones with the strongest RSS out of the three APs at a particular location. The distribution of RSS with the user in Figure 2-a is also skewed to the left. The left-skew property seems to occur in most of our measurements as reported by the skewness in Tables 1, 2, and 3. This property is usually observed when the data have an upper bound which is the case for the attenuated RSS measurement. In a comparison of the 46 distributions, we saw that 39 of them had distributions that were left-skewed while five of them were almost symmetric and only two of them were right-skewed. The left-skew is the effect of the range limitation imposed by the maximum RSS at each location.

Because of the complexity of the radio propagation, the distribution of RSS is difficult to model and fit to well-known distributions. The authors in [6] conclude that they rather record the distribution of the RSS than reduce it to only the mean value. We believe that it would be a great benefit if we could find a representative or approximate distribution of the underlying RSS process for understanding location fingerprinting. This is part of our ongoing research.

3.2.2. The standard deviation of the RSS. The results in Table 1, 2, and 3 also reveal an interesting trend in the second order statistics of the RSS values. The standard deviations are quite similar for a signal from the same AP at a particular location except when the user's orientation blocks that AP. The main difference between the three APs is the distance to the measurement point *L*2. Comparing the three tables, the results indicate that the farther the AP is from the MS or the lower the received signal level is, the smaller the degree of standard deviation. Note that within the same table the RSS data which is blocked by the user has a smaller standard deviation. This occurs as a result of the signal attenuation by the user's body and the smaller range of RSS between the maximum and minimum receivable signal level. A linear regression plot (Figure 3) between the sample mean of the RSS and the standard deviation collected from 46 distributions illustrates the trend. This observation suggests that the RSS values from the same AP at two different locations may be difficult to distinguish for positioning purposes when the RSS level is high in which case it tends to have a large degree of variation. A good communications signal may not result in a good positioning signal. On the other hand, two nearby locations might be easily identified if both have low RSS levels and smaller signal variations. This good positioning case usually occurs in indoor locations with NLOS. This is rather counter-intuitive since the farther apart the WLAN receiver is from the AP, the worse the measurement accuracy or the larger the signal variation should be as suggested by [5].

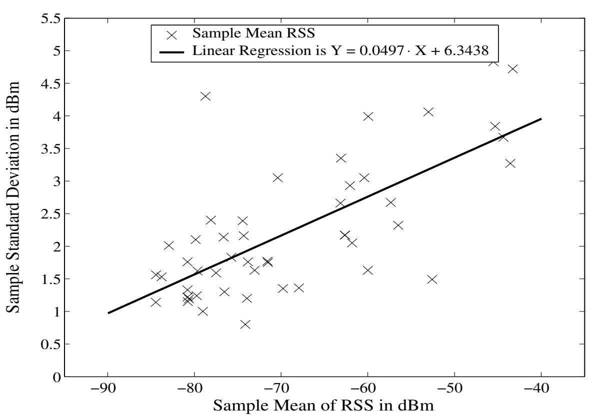

Figure 3. The relationship between average RSS and its standard deviation

3.2.3. Stationarity of the RSS. Assuming that the ergodic theorem is applied according to the Wiener definition of stationarity [11], we analyze the stationarity of the RSS element by breaking the series of RSS measurements into separate pieces over different time intervals. A random process is said to be stationary when it meets two conditions. First, its mean and variance remain the same over time. Second, its autocovariance function has the same shape for each separate time-series. We investigated this property over two time scales: pieces of 15 minutes within the same hour and pieces of one hour over five different hours.

After dividing the series of measurement data of RSS in Figure 2-a within the same hour into groups of 15 minutes, the RSS distribution within each quarter is observed to follow a similar distribution within the same group. Table 4 lists the summary statistics within each quarter. These results suggest that the RSS distribution may be stationary or time independent since the means and the sample variances of each quarter are very close together. The correlograms in Figure 4 depict the same shapes for each quarter indicating that the second condition is also met for this time scale.

Table 4. Mean and standard deviation of RSS with user (dBm)

Statistics	1st Qtr.	2nd Qtr.	3rd Qtr.	4th Qtr.
Mean	-71.71	-72.33	-71.82	-70.48
Standard Deviation	2.95	3.20	2.96	2.56
Sample Variance	8.72	10.27	8.77	6.54

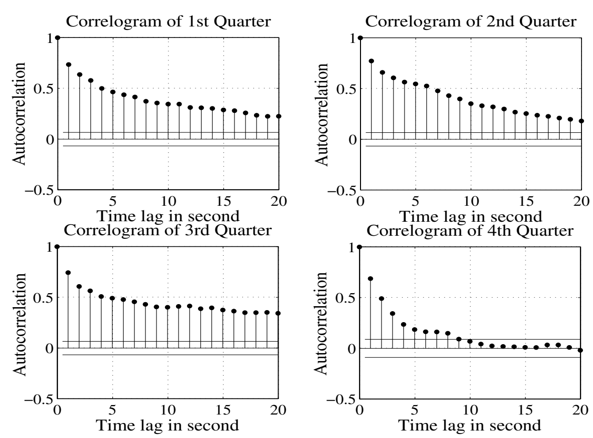

Figure 4. The correlograms of RSS within the same hour

The one-hour time scale study was made for a signal measured at *L1* over different times of day. Table 5 shows a consistent mean, but inconsistent variance values of the RSS. Therefore, the test for the first condition for stationarity fails and we conclude that the RSS random process in this case is non-stationary.

Figure 5 illustrates sample paths of average received signal strength values from the three access points measured at location *L3* over a period of one hour. Observe that the sample mean from SIS410 abruptly changes to another value (-70 to -60 dBm) which confirms our conclusion on the non-stationary property. This is common due to the changing indoor environment such as in this case when a person walked into the room and sat in the middle of the room IS 410 after approximately 30 minutes into our experiment. Notice that the rest of the received signals from other APs located outside the room are not affected by this event. Although the stationary assumption may not be valid over all time scales, there is some evidence that we could assume stationarity over small time scale for modeling purposes.

Table 5. Time dependency of RSS (dBm) from SIS410 with user's presence

Statistics	10AM	12AM	2PM	8PM	10PM
Mean	-62.68	-60.02	-61.85	-63.12	-63.18
Standard Deviation	2.17	1.63	2.05	3.35	2.66
Sample Variance	4.70	2.65	4.22	11.23	7.07

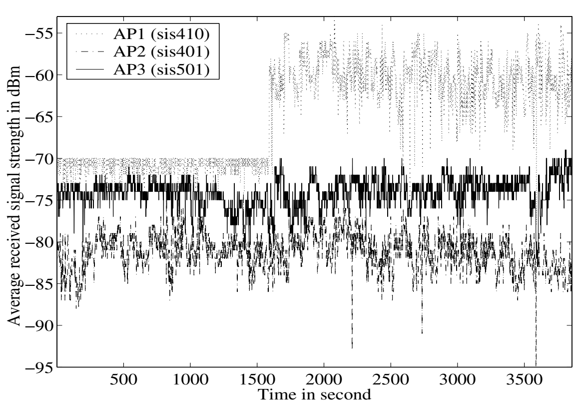

Figure 5. Samples of RSS from three APs

3.2.4. Time dependency of received signal strength. The summary of statistics in Table 5 suggests that there is some dependency of RSS on the time of day. This is due to the dynamics of the indoor environment. To understand the property better, we require more measurements and this is part of our ongoing research work.

3.3. Properties of multiple RSSs at a particular location

This subsection analyzes the dependency of multiple RSSs from multiple APs. This is to confirm an intuition that the RSS from multiple APs are actually independent. The second part of this subsection discusses the effect of interference on the RSS when there is another AP transmitting in the same frequency channel.

3.3.1. Independence of multiple RSSs. The average of the RSS from each AP is a value of a location fingerprint vector. To verify statistical independence between these values, a measurement of multiple RSS samples was collected at location 3 (*L3*) in Figure 1 where the MS can receive signals from three APs simultaneously during an afternoon hour. The distances from the three APs SIS410, SIS401, and SIS501 were approximately 8, 15, and 10 meters. We took RSS measurements for approximately one hour with the user's presence and the sample paths of RSS levels are showed in Figure 5. The standard deviations from each AP were 5.67, 2.23, and 1.81, respectively. The means were -64.52, -81.23, and -74.12 dBm. The correlation values between each pair of RSS data are $C_{(SIS410,SIS401)} = -0.02$, $C_{(SIS410,SIS501)} = 0.13$, $C_{(SIS401,SIS501)} = -0.03$. Therefore, we can conclude that the RSS from the APs are uncorrelated.

3.3.2. Interference from multiple APs. There are two APs in our experiment that use the same frequency channel number 6, which are SIS410 and SIS501. One may think that the RSSs from both APs might interfere with each other and cause difficulty in forming the location fingerprinting. However, our initial results as calculated by the correlation indicate that both RSSs are independent and do not interfere with the reception of each other. This is due to the way in which the 802.11 MAC operates where a transmission is either not heard or is deferred if a competing transmission exists.

3.4. Properties of RSS at different locations

Generally, the RSS falls linearly with the log of the distance between the transmitter and the receiver. This subsection will not focus on this common knowledge but will instead investigate how the samples of RSS fingerprints at different locations may look like. We plot the RSS patterns in two dimensions in order to analyze the clustering of the RSS data. The clustering of two RSS data at two locations is plotted in Figure 6.

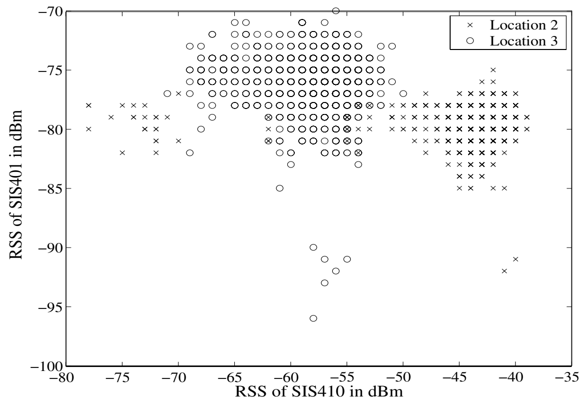

Figure 6. RSS fingerprints with two elements

The plot in Figure 6 contains all possible vectors $\boldsymbol{P} = [\rho_1\ \rho_2]$ for *L2* and *L3* with ρ_1 on the *x*-axis and ρ_2 on the *y*-axis. We see that the location fingerprints can be separated by some discriminant function or clustering technique. Note that *L*3 consists of 3,666 samples and *L*2 consists of 3,465 samples. Only certain patterns are present which implies that there are fewer unique fingerprints for each location. Figure 7 shows the density of each fingerprint. This visual study suggests that we may use the center of the cluster as a representation of the location fingerprint instead of the distribution itself as these locations can be clearly separated.

Figure 7 suggests that only two APs are sufficient to distinguish between locations for a system with small number of positions and coarse location granularity. Two locations become difficult to identify if their patterns are closer together and exhibit large variations due to the nature of standard deviation of the RSS. Increasing the number of APs is one way to further separate two location fingerprints.

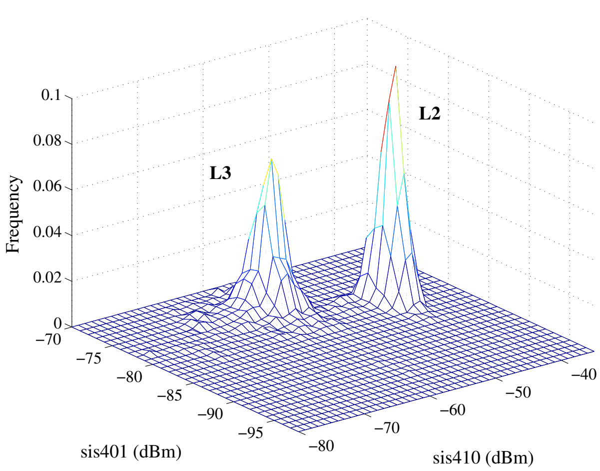

Figure 7. Frequency of occurrence of RSS patterns

3.5. Implications on positioning algorithms

The data analyses in the previous section have two major implications on selecting potential positioning algorithms. First, we observed that the RSS sample vectors exhibit clustering (they are concentrated around the center of a cluster with the greatest frequency of occurrence being roughly at the center as in Figure 7). This explains why the accuracy performance of all positioning algorithms evaluated in [12] is similar. Both the support vector machines (SVMs) approach and the *k*-nearest neighbors (or equivalently Euclidean distance) approach provide similar accuracy and precision. The reason is that location fingerprints can be simply represented by vectors of average RSSs. Second, to perform any analytical performance evaluation of pattern classifiers, we eventually need a model of location fingerprints. In the following section we suggest a simple model of location fingerprints and compare it with the empirical ones from this study based on our previous work in [13].

4. Modeling of RSS for location fingerprints and performance of the Euclidean distance

Although our preliminary finding suggests that the RSS distribution is not normal, some of them could be approximated by a normal distribution. In an attempt to model indoor positioning systems [13], we assumed a first cut Gaussian model of the RSS. A location fingerprint is denoted as a vector of *N* features of RSS, e.g. $\boldsymbol{X} = (x_1, x_2, \ldots, x_3)$. Here, we list the assumptions made in [13] that are partially supported by the work here.

- Each RSS feature in fingerprint is normally distributed and stationary over a small time scale.
- The sample standard deviations of all RSS features are assumed to be constant and unique for each RSS. (We did not model the variation in standard deviation for each RSS feature.)
- The mean of the RSS can be used as the fingerprint as samples of the RSS vector exhibit clustering.
- All RSS features in each location fingerprint are mutually independent (which is confirmed by our measurement in Section 3.3).

We compare the location estimation performance using this simplified model with the one using empirical distributions obtained from the measurement results of this paper. The simple location estimation algorithm utilizes the Euclidean distance. The number of access points *N* is three. The mathematical expression for the accuracy was developed in [13]. We use the expressions to predict the performance of two simplified positioning systems in this section.

4.1 Two-location system

As an example of a system with only two location fingerprints, we select the measurement data of locations *L2* and *L3* for analysis. Note that these are nicely distinguishable. Let $\boldsymbol{R} = (r_1, r_2, \ldots, r_3)$ and $\boldsymbol{S} = (s_1, s_2, \ldots, s_3)$ be the location fingerprints (mean values of the RSSs from the APs) of *L2* and *L3*, respectively. The actual physical distance between the two locations is 18 ft (5.5 m). Table 6 summarizes the location fingerprints and their standard deviations.

Table 6. Fingerprints of two-position system (dBm).

Location	Statistics	SIS410	SIS401	SIS501
L2	Mean	-43.60	-79.76	-79.68
	STD.	3.27	1.24	1.62
L3	Mean	-57.27	-75.77	-67.97
	STD.	2.67	1.83	1.36

Using a Monte Carlo simulation to generate the samples of location fingerprints based on the empirical distributions, we compare the error of location detection given that *L2* is the correct location with the analytical mathematical formulation in [13] and simulations with a Gaussian distribution with the same means and standard deviations. The probability of incorrectly picking location *L3* instead of *L2* is *Pe*. For a system with two positions, *Pe* can be found from the probability of returning the correct location (*Pc*) as:

$$Pe = 1 - Pc = 1 - \left(\frac{1}{2} + \frac{1}{2} erf\left(\frac{-\mu_c}{\sqrt{2}\sigma_c} \right) \right), \quad (1)$$

where $\mu_c = 2\sum_{i=1}^{N} r_i \beta_i + \sum_{i=1}^{N} \Gamma_i$, $\sigma_c^2 = \sum_{i=1}^{N} (2\beta_i \sigma_i)^2$, $\beta_i = (s_i - r_i)$, and $\Gamma_i = (r_i^2 - s_i^2)$. We performed 20 simulations of 10,000 samples each. The results are summarized in Table 7. The Gaussian approximation provides an optimistic error performance for the system. The analytical result from (1) exactly calculates *Pe*. The empirical distributions have a worse performance because there are some samples of RSS fingerprints that are closer to the location *L3*'s fingerprint than *L2*'s fingerprint. Figure 6, which represents the projection of all fingerprints into the plane of two features, clearly explains the cause of the worse performance because there are some RSS patterns on the left-side of the plot that may be wrongly interpreted as corresponding to *L3*. Table 7 also shows the results when we consider only two APs (SIS410, SIS501) and one AP (SIS410) in which the error probabilities are higher.

Table 7. Probability of returning incorrect location at 95% C.I. for two-location system.

Scenario	3 APs	2 APs	1 AP
Empirical	10.22e-3 ± 3.24e-5	10.79e-3 ± 3.02e-5	16.19e-3 ± 3.46e-5
Gaussian	0.24e-3 ± 5.37e-6	0.41e-3 ± 5.76e-6	17.56e-3 ± 4.6e-5
Analytical	0.24e-3	0.41e-3	17.5e-3

4.2 Twenty five-location system

The second system we consider consists of 25 location fingerprints where *we measured the real location fingerprints* from a grid of 25 positions. The grid spacing is approximately 1 meter. The location area of 16 m^2 covers part of the room IS410 and part of the corridor as showed in Figure 8. Each position is labeled by the square box.

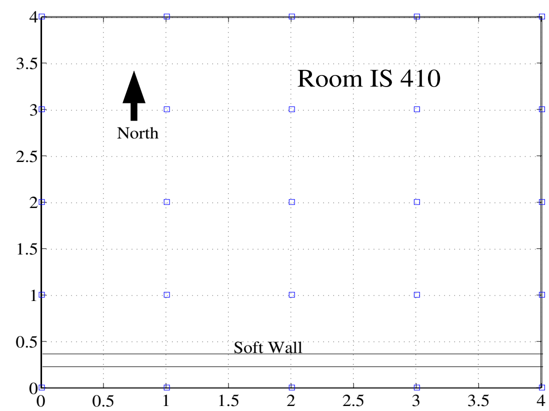

Figure 8. Grid of system with 25 locations

Due to space limitation, we will not list the measured location fingerprints here. Note that the measurement is done for only one orientation of the user (facing north). Each position can receive signals clearly from all three APs (SIS410, SIS401, SIS501) and there are approximately 1,200 samples of each RSS measured over five minutes at a rate of four samples per second.

We assume that the current MS's location is at the center of the map. A simulation was done in the same manner as the two location system. The results are summarized in Table 8. In order to find the exact analytical expression, we need to find the joint probability density of multiple location fingerprints which is rather cumbersome. We argue in [13] that this can be avoided by resorting to an approximation of the probability of returning an erroneous location. This considers only the location fingerprints of $M_N = 8$ nearest neighboring positions in two-dimensions and the correct location. The approximation works for a large number of access points and can be written as:

$$Pe \approx 1 - \prod_{j \in M_N} Pc_j \ . \qquad (2)$$

Table 8. Probability of returning incorrect location at 90% C.I. for 25 locations.

Scenario	3 APs	2 APs	1 AP
Empirical	0.67 ± 1.46e-3	0.82 ± 1.45e-3	0.84 ± 1.64e-3
Gaussian	0.85 ± 1.67e-3	0.93 ± 1.18e-3	0.95 ± 1.04e-3
Approximation	0.80	0.94	0.98

The overall performance results in this system are worse than the previous one because there are more locations to be compared with and the spacing between the locations is much closer. Depending on the standard deviation of each RSS feature, it is possible that the fingerprint patterns overlap. The results with a Gaussian model are worse than those with the empirical model and appear to be rather pessimistic in this case. This can be explained as follows: the Gaussian model tends to spread the fingerprint patterns evenly around the mean value while the real fingerprint patterns are asymmetric and more concentrated around their means in some cases. The analytical approximation is highly pessimistic and not appropriate in this case. In summary the Gaussian model provides a probability of an erroneous report on the same order as one would expect with real RSS samples, but could be either optimistic or pessimistic.

5. Conclusions

We present an initial analysis of the RSS values reported by an 802.11b NIC commonly used in indoor location systems based on location fingerprinting. We point out that the user's presence should be taken into account when collecting the location fingerprint for user related location-based services. The effect of user's orientation is significant and the orientation should be recorded in the database as demonstrated in [1]. We also analyze the statistical properties of the RSS and we find that it is stationary under certain circumstances, but in general, such a conclusion cannot be made. The distribution of the RSS is not usually Gaussian, it is often left-skewed and the standard deviation varies according to the signal level. It is clear from our measurements that signals from multiple APs are mostly independent and the interference from other APs using the same frequency does not have a significant impact on the RSS pattern. The visual presentations of the RSS patterns in Section 3.4 show that the fingerprint can be grouped together as a set of clusters. More than one cluster may represent one location because of the multimodal distribution of the RSS. In such a case, using a simple Euclidean distance as in [1] to determine the location may classify some patterns into a wrong location easily. This causes poor performance of a positioning system that uses the Euclidean distance. Finally, in the last section, we compare the error performance of a position location system with real location fingerprints and a first cut Gaussian model. The results indicate that our model of location fingerprints provide some approximations of the performance. The Gaussian model is either optimistic or pessimistic but provides values on the same order as a real system. The future work is to look for alternative models for the distribution of the RSS and to understand how they impact position location further. The results in this paper and our previous work could provide insight on the mechanism behind indoor position location systems based on location fingerprinting.

In conclusion, this study steps back and investigates the RSS pattern in a greater detail. This study raises an important aspect of designing a location fingerprinting system that is to examine the properties of the fingerprint itself before applying a pattern recognition technique to solve the positioning problem. Simple pattern recognition techniques may be suitable and more efficient than more sophisticated ones. More extensive measurement campaigns are needed to verify some of the properties listed here. However, the lessons learned from this study are used to support a theoretical model of location fingerprint which in turn can be used to create a design framework of an indoor positioning system [13]. Given a framework and theoretical explanation, any future study and design of location fingerprinting system should be more efficient and less time consuming. We could reduce the measurement time in order to fine tune the system performance such as accuracy and precision to meet any required criteria.

6. References

[1] 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.
[2] G. M. Djuknic, and R. E. Richton, "Geolocation and Assisted GPS," *IEEE Computer*, vol. 2, pp. 123–125, Feb. 2001.
[3] M. A. Youssef, A. Agrawala, and A. U. Shankar, "WLAN Location Determination via Clustering and

Probability Distributions," in *Proc. IEEE PerCom*, Mar. 2003.
[4] T. Roos et al., "A Probabilistic Approach to WLAN User Location," *Int'l Journal of Wireless Information Networks*, vol. 9, pp. 155–164, Jul. 2002.
[5] 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 at] http://www-2.cs.cmu.edu/ aura/docdir/small00.pdf.
[6] A. M. Ladd et. al., "Robotics-Based Location Sensing using Wireless Ethernet," in *Proc. MOBICOM*, 2002, pp. 227–238 .
[7] The Institute of Electrical and Electronics Engineers, Inc. IEEE Standards 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. 1999.
[8] Agere System Inc. (2003, Feb.) Agere Product Brief: WaveLAN 802.11b chip set. [Online] Available: http://www.agere.com/client/docs/PB03025.pdf.
[9] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks: A Unified Approach*, Prentice Hall PTR, Upper Saddle River, New Jersey, 2002.
[10] B. Sklar, "Rayleigh fading channels in mobile digital communication systems: I. Characterization," *IEEE Communications Magazine*, vol. 35, pp. 90-100, Jul. 1997.
[11] J. M. Gottman, *Time-series Analysis: A Comprehensive Introduction for Social Scientists*, Cambridge University Press, New York, NY, 1981.
[12] R. Battiti, M. Brunato, and A. Villani. (2002, Oct.) Statistical Learning Theory for Location Fingerprinting in Wireless LANs. Technical Report DIT-02-0086, Universit'a di Trento. [Online] Available: http://rtm.science.unitn.it/ ~battiti/archive/86.pdf
[13] K. Kaemarungsi, and P. Krishnamurthy, "Modeling of Indoor Positioning Systems Based on Location Fingerprinting," in *Proc. IEEE INFOCOM*, May 2004.