Pervasive and Mobile Computing 8 (2012) 292–316

Contents lists available at SciVerse ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Analysis of WLAN's received signal strength indication for indoor location fingerprinting

Kamol Kaemarungsi [a,*], Prashant Krishnamurthy [b]

[a] National Electronics and Computer Technology Center, NSTDA, Pathumthani, Thailand

[b] School of Information Sciences, University of Pittsburgh, Pittsburgh, USA

ARTICLE INFO

Article history:
Received 11 September 2009
Received in revised form 29 September 2010
Accepted 14 September 2011
Available online 24 September 2011

Keywords:
Data analysis
Indoor positioning system
Location fingerprint
Measurement
Wireless LAN

ABSTRACT

An indoor positioning system that uses a location fingerprinting technique based on the received signal strength of a wireless local area network is an enabler for indoor location-aware computing. Data analysis of the received signal strength indication is very essential for understanding the underlying location-dependent features and patterns of location fingerprints. This knowledge can assist a system designer in accurately modeling a positioning system, improving positioning performance, and efficiently designing such a system. This study investigates extensively through measurements, the features of the received signal strength indication reported by IEEE 802.11b/g wireless network interface cards. The results of the statistical data analysis help in identifying a number of phenomena that affect the precision and accuracy of indoor positioning systems.

1. Introduction

Location awareness as a part of a context-aware computing paradigm is one of the keys to the success of ubiquitous and pervasive computing [1]. With an increasing number of mobile computing devices inside buildings, new opportunities for location based applications have attracted a number of researchers to pursue technologies that could provide indoor location information. For instance, researchers in [2] had suggested a novel idea to utilize a wireless local area network's (WLAN) signal for indoor location determination.

This study focuses on a technique of indoor location determination called location fingerprinting that exploits the presumably unique relationship between the received signal strength indication (RSSI) of a WLAN and an indoor location. This technique can provide a location-aware capability for mobile devices equipped with today's prevalent WLAN interfaces [2]. The advantages of this technique are lower complexity due to the built-in RSSI measurement capability in virtually all WLAN interfaces and lower cost due to widespread deployment of WLAN infrastructure inside buildings.

The RSSIs can be viewed as sensor data that refer to indoor locations; however, the characteristics of RSSI of WLANs such as IEEE 802.11 itself have been rarely studied. Although there is extensive knowledge available regarding radio frequency (RF) phenomena and properties of the received signal in indoor environments (such as the distance-dependent property through path loss models and the fluctuation of signal because of the multipath effect), such knowledge is aimed toward communications capability and receiver design, making it limited for understanding positioning applications. A system designer needs to understand the underlying mechanism of RSSI to efficiently model, analyze and design an indoor positioning system.

Since RSSI was not originally intended to be used as a location sensor, there could be intrinsic performance limitations in terms of location determination employing RSSI. Its statistical distribution is expected to be different from other studies

* Corresponding author. Tel.: +66 0863132714.
E-mail addresses: kamol.kaemarungsi@nectec.or.th (K. Kaemarungsi), prashant@mail.sis.pitt.edu (P. Krishnamurthy).

1574-1192/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.pmcj.2011.09.003

of indoor radio propagation that use sophisticated equipment such as vector signal analyzers to collect received signal data. Moreover, it is not very well understood whether any of the hardware properties of WLAN cards will affect the performance of an indoor positioning system. The primary objective of this study is to gain an in-depth understanding of various characteristics of RSSI relevant to the location fingerprinting technique through extensive statistical data analysis.

The organization of this manuscript is as follows. First, Section 2 reviews related work investigating the properties of RSSI for an indoor positioning system. Next, Section 3 describes the details of measurement setup and experimental design for our data analysis. In Section 4, the influences of hardware factors in data collection of location fingerprints are pointed out. This is followed by Section 5, which reports on the statistical properties of RSSI such as distribution, mean, standard deviation, skewness, and stationarity based on extensive measurement results according to environmental effects. In Section 6, the causes of errors in detecting a location in such a positioning system are discussed. After that, Section 7 summarizes the findings and analyzes their implication to location determination. Finally, Section 8 concludes this manuscript.

2. Related work

The properties of RSSI which describe location fingerprinting have been pointed out by a number of researches into indoor positioning systems. In the pioneering work of [2], an observation was made that the user's orientation could cause a variation in RSSI level of up to 5 dBm. However, no analysis of the RSSI data was provided. Different orientations of user and mobile device with respect to access points could change the mean values of RSSI at a location. The authors in [2] also suggested that orientation should be included in computing the user's location. Later researches in indoor positioning systems such as COMPASS [3] and the system in [4] exploited the orientation of users through the addition of digital compasses on mobile devices to improve location accuracy. However, the increase from two to eight orientations did not provide significant improvement in the location determination performance [3].

A preliminary study of the properties of RSSI for location fingerprinting was reported in [5]. The authors performed a number of RSSI measurements which were influenced by the user's presence, orientation, and attenuation. The authors of [5] found that not only the mean values of RSSI were changed but also the RSSI values fluctuated more wildly with larger variance when a user was present during RSSI measurement. The finding emphasized the inclusion of a user's effects into the collection of RSSI for the purpose of human-related indoor positioning. Additionally, the movement of the user carrying a mobile station causes fluctuation of the received signal strength, which is called small-scale fading [6]. Larger fluctuations of RSSI values have been exploited to infer the movement of the mobile station in [7].

The modeling of RSSI based location fingerprinting is essential for location determination algorithms such as the probabilistic approach [8–10] or a preliminary analytical model in [11]. In those researches, Gaussian or log-normal distributions were used to model the randomness of RSSI. Although it is an intuitive choice of model for RSSI, there are few statistical data presented in the literature, such as [12] to validate the use of the Gaussian model. The study in [9] summarized in their large-scale measurement that the majority of RSSI histograms fitted very well with Gaussian distribution and there were a few histograms that could be fitted with bi-modal Gaussian distributions. However, most researches in the literature did not investigate further on the actual statistical properties of RSSI samples.

Observation of time-varying phenomena of RSSI was reported by Haeberlen et al. [9]. The authors pointed out that environmental effects such as interference and movement of people or mobile nodes created time-correlated fluctuations of RSSI and severely reduced the performance of the positioning system. A discussion of sample correlations was presented in [10] where the authors exploited these highly correlated RSSI samples from the same access point (AP) in their positioning system. However, there was no detailed report that investigated the stationarity of the time-varying phenomena or correlation in the time series of RSSI. Note that the RSSI from an AP was assumed stationary in [10]. Moreover, signals from APs may be absent due to the intermittent degradation of radio channels over time. This may cause a null value of RSSI in location determination and could lead to negative effects in performance, computational complexity, and scalability of the indoor positioning system [13].

In indoor localization, there are two approaches for measuring RSSI which are mobile station (MS) based measurement and access point (AP) based measurement. Research in [14] pointed out an asymmetry property in RSSI measurement between MS and AP. However, the authors in [14] concluded that the asymmetry in the link had little effect on localization as long as the measurement of RSSI was performed only in one direction. Moreover, the increase of transmitting power in a WLAN was found to decrease the performance of localization [14].

Other factors that influence location estimation algorithms were pointed out by Lemelson et al. [15]. These are the number of received signals from APs and the quality of the hardware. However, there was no elaboration on or investigation into the effects of these factors. A possible cause of difference in RSSI due to hardware could be in the design of the antenna [16]. In another related work of [17], performance limitations of localization using signal strength were reported where the authors mentioned that observed limitations were fundamental to the algorithm or inherent in the data. Most existing researches in indoor localization focused on the algorithms but did not analyze the RSSI data.

Researchers in [18,19] did recognize the problem of differences in the hardware of WLAN interfaces that could degrade the location determination performance. They proposed techniques which could reduce the impacts on the performance through transformation [18] or scaling [19] of the RSSI patterns. Finally, a use of virtual local area network (VLAN) on WLAN AP for indoor positioning was investigated in [20] wherein the authors found that, with four VLANs enabled, the precision performance was improved by up to 10% at a low error distance when compared to a WLAN without VLAN. Based on our review of related work, a comprehensive analysis of the properties of RSSI for an indoor positioning system was not performed and consolidated in any of the current literature.

3. Measurement setup

This study investigated the real measurements of the RSSI based on IEEE 802.11b/g cards and access points (APs) available to the authors. The scope of our work was limited to the measurement of wireless network interfaces on laptop computers or MS side only. No measurement of RSSI was collected using AP. Investigation of RSSI measurement for indoor positioning on personal digital assistants (PDAs) was beyond the scope of our work and could be found in [16].

The measurement experiments reported in this work were based on the following assumptions. First, we assumed that the location determination is based on a point instead of an area position. We collected RSSI samples and averaged them over the same point of position. Second, we assumed that all WLAN cards used in our work had omni-directional antennas. Third, this study performed measurements using a quasi-static laptop only. While we are aware of the effect of WLAN's mobility, it was beyond the scope of current study. Our assumption here was that typical usage of WLANs occurs under stationary conditions. Finally, the effects of a user's orientation and presence which have impact on RSSI are not emphasized in this work.

The measurements reported in subsequent sections were conducted by taking the data logging of RSSI measurements from APs "visible" to a mobile client over a certain period of time. For a particular location, vectors of RSSI were formed based on measurable signals from APs during the measurement time. Each RSSI measurement represents a calibration point instead of an area. This study assumes that the selected software tool provides the same information as a query from positioning software accessing a network interface card (NIC). The RSSI values reported by most WLAN cards are in integral steps of 1 dBm. Note that all possible RSSI values cannot be represented by a set of integer values [21].

The advancement in WLAN usage and the upgrading of major operating systems (OS) over the years have introduced a variety of methods to obtain RSSI, such as proprietary [22] or open source [23] software tools or customized code through different versions of OS's application programming interfaces (APIs) [24–27]. Versatile software tools such as Lucent's client manager [22] and Vistumber [23] are categorized as site-survey tools that provide link quality and AP monitoring capabilities. Typically, the information available to the user from a site-survey tool includes the AP's name, AP's medium access control (MAC) address, received signal (dBm), noise (dBm), signal-to-noise ratio (SNR in dB), and channel number. However, some site-survey tools limit the control abilities of users in selecting sampling rates of RSSI. Open source software tools are often based on the reverse engineering of information exchanged between the OS and device driver software which might report RSSI inaccurately due to the limited access to information of manufacturer's proprietary object identifier [23]. In this work, we mainly used the Lucent's client manager which was a proprietary software tool that should provide us with the most accurate measurement of RSSI.

To obtain RSSI for different wireless network cards, we resorted to using customized codes based on available Microsoft APIs such as Wireless Research API (WRAPI) [25], Windows Management Instrumentation/WMI Query Language (WMI/WQL) [26], and native WiFi [27] depending on the version of the operating systems. However, there is no unique standard on how to interpret the measurement of RSSI because the mapping of available information in the hardware to the RSSI is hardware specific and proprietary. The RSSI returned by Window's driver is measured in dBm and has typical values between -10 and -200 [28]. Measuring RSSI from various APs with different channels can be done in either passive scanning mode or active scanning mode or a combination of both [29]. The passive scanning is done by recording any periodical broadcasting of the 802.11 beacon or probe response frames from APs to WLAN card's cached Basic Service Set Identifier (BSSID) scan list [30]. On the other hand, the active scanning is done by broadcasting an 802.11 probe request frame on a scanning channel and recording any probe reply or any beacon frames to card's cached BSSID scan list [30]. However, some WLAN cards reported RSSI data back in positive values which were not in dBm. This may require some translation or calculation to derive a value in dBm as discussed in [23].

3.1. Experimental design

A number of factors that affected RSSI were exploited by researchers to improve the design and performance of indoor positioning systems as reviewed in the previous section. However, some of these factors and their effects were not thoroughly investigated and were included in their designs as assumptions. This work investigates in more detail the important factors that affect the statistical properties of RSSI, which we believe influence the location determination performance of an indoor positioning system. A number of experiments that vary these different factors were performed and the results were analyzed to identify the statistical properties of RSSI patterns. This work differs from other related work because we focused on the comprehensive data analysis of RSSI measurement instead of location determination algorithms.

Initially, basic factors such as user's presence/absence, user's orientation, building types and materials, and distance from transmitter were identified and known to change the properties of WLAN RSSI in the literature. Other possible hardware factors such as orientation, directionality, and type of antenna are also known to have effects on the level and quality of RSSI. These factors are by no means exhaustive. However, we will limit the scope of our work to the factors listed in Table 1, which should provide common insights into the statistical properties of RSSI. The factors in Table 1 are classified into two groups which are a hardware factor that affects data collection and a group of environmental factors that affect the statistics of RSSI. Corresponding options used in the subsequent studies are listed next to its factor.

Table 1
Factors that affect statistics of RSSI fingerprint.

Effect of	Factors	Options
Hardware	1. Make of WLAN card	Lucent Gold, Lucent Silver, SMC, Cisco, D-Link, Proxim, 3COM, Hawking Technology and Intel
Environment	2. Time of measurement	Times of day and days of week
	3. Period of measurement	Second, minute, and hour
	4. Interference	Co-channel or adjacent radio channel
	5. Building environment	Corridor, small office or large hall

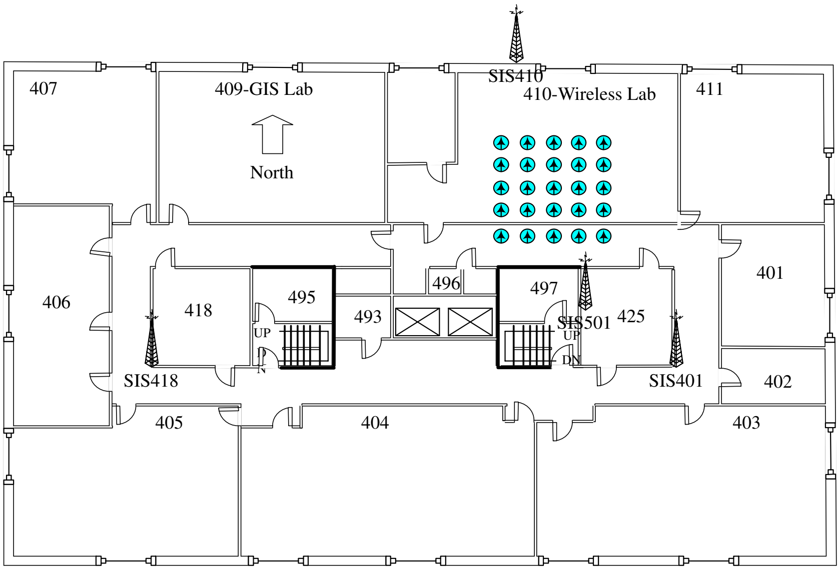

Fig. 1. Locations of access points and measurement points on the 4th floor of IS building.

The make of wireless cards is considered as the first factor in Section 4. Different makes of wireless cards could have different chipsets and hardware implementation. Some vendors have a better hardware design than others. Intuitively, the RSSI measured by different cards should have different results. Here, we try to uncover the major differences among available hardware that have an impact on the statistics of RSSI.

The second group of factors affecting the statistical properties of RSSI is discussed in Section 5. Because of changes in the environment over time (such as movement and varying number of people) around a measurement location, the second factor (time of measurement) is considered. This might indicate the need to collect location fingerprints at different times to reflect the time dependency. On the other hand, the third factor is considered in order to justify the minimum duration of measurement required to determine the location fingerprint at each position. This factor could affect indoor positioning accuracy and the total time required to collect location fingerprints. This study attempts to determine a suitable duration of measurement at each position and the number of samples per position that can provide sufficient statistics. The sources of interference in the fourth factor are limited to those signals using the same/different channels from other access points. This study determines the correlation of received signals from different access points.

The final factor in Table 1 is the building environment that we selected based on the available WLAN infrastructure at the University of Pittsburgh in the US and the National Electronics and Computer Technology Center (NECTEC) in Thailand. Three different areas were used in this study: a small office, a large hall, and a central corridor of a large office. The first studied area called *small office* is inside the School of Information Sciences (IS) building at the University of Pittsburgh. This building has eight floors. However, only one area on the fourth floor was used for our experiment. The floor plan with locations of APs in this environment is shown in Fig. 1.

As shown in Fig. 1, we defined a small area as a grid of 25 locations, of which 20 were placed inside a laboratory room and five were placed along the corridor. The minimum distance between two locations called grid spacing was fixed at one meter. Location fingerprints were collected at each location for a period of five minutes at a rate of four samples per second. This resulted in 25 locations $\times 3$ APs $= 75$ RSSI histograms. The user's orientation was limited to the north direction only. Additional factors for this test area labeled as Scenario 1 are listed in Table 2.

The second studied area called *large hall* is inside the Hillman library building at the University of Pittsburgh. Due to the limited space of this manuscript, we could not include the floor plan for this environment. Basically, this building has five floors including an underground floor. There are six APs located throughout the building. On the first floor, there is a large

Table 2
Experimental design and measurement factors.

Factors	Scenario 1	Scenario 2	Scenario 3
Campus	U. of Pittsburgh	U. of Pittsburgh	NECTEC
Building type	Small office	Large hall	Large office
Dimension	23 m × 37 m	60 m × 60 m	91 m × 40 m
Number of floors	8	5	6
Test area	Room in 4th Fl	Open space in 1st Fl	Corridor of 5th Fl
Proximity of laptop user	Presence	Presence	Presence
User and terminal orientation	North only	Dependent on location	Dependent on location
Time of measurement	Afternoon to evening	Afternoon to evening	Afternoon to evening
Span of measurement	1 day	32 days	1 day
Period of measurement per location	5 min	1 h	3 min
Sampling period	0.25 s	1 s	0.1 s
Max. no. of samples/AP/location	1200	3600	1800
Number of locations	25	71	16
Number of measured APs	3	6	1
Collectable histograms	75	299	80
Distance between locations	Uniform 1 m	Non-uniform, 2 m or more	Uniform 2 m
Client card vendor	Lucent Gold	Lucent Gold	SMC, D-Link, Hawking, 3COM, Intel
Access point vendor	Lucent WAVELAN	Enterasys roam about	Cisco Aironet 1100
Total number of APs	10	6	7
Software tool	Lucent's client manager	Lucent's client manager	WMI VBscript
AP's WLAN standards	802.11b	802.11b	802.11b/g

open space that shares the ceiling with the second floor thus forming a large hall. All measurements of the large hall were done inside this area of the first floor, from where signals from all six AP's could be detected. However, each AP did not completely cover the entire floor. This is why we could only obtain 299 histograms instead of 426 (6 APs × 71 locations) histograms.

Additional factors of this second area are listed in Table 2 under the column labeled Scenario 2. Note that the distance between different locations in Scenario 2 varied according to the locations of reading tables inside the library; therefore, we denoted it as non-uniform grid spacing. The sampling rate in this scenario was set at one sample/second to observe its effect on RSSI measurement. However, the measurement in this scenario was performed over a period of one hour causing a large span of measurement over several days due to the three-hour limitation of the laptop's battery. Because of limited control over the client manager software (data logging tool), the number of RSSI data points collected at each location in Scenarios 1 and 2 varied from one location to another.

The third studied area called *central corridor of a large office* is inside the NECTEC building. Once again due to limited space, we could not show the floor plan of the fifth floor of this six-story office building. There are a total of seven APs placed to serve most of the area. All APs are Cisco Aironet 1100 series which can support both IEEE 802.11b and 802.11g WLAN cards. A total of 16 locations along a part of the corridor on the fifth floor were selected as calibration points. Subsets of these 16 locations have direct line-of-sight with an AP on the fifth floor. Using customized MS Visual Basic Script (VBScript) based on [26], samples of RSSI were collected 0.1 s apart for a total of 1800 samples at each location per each WLAN card. For this area, we switched five different WLAN cards (SMC, D-Link, Hawking Technology, 3COM and Intel) over 16 locations and obtained 5 × 16 = 80 histograms. The rest of the factors for this test area called Scenario 3 are listed in Table 2.

4. Effects of hardware

In this section, we address three issues based on the hardware of WLAN cards that could affect the statistics of RSSI. First, we focus on differences of WLAN cards which influence the data collection of RSSI. Second, we briefly compare the RSSI collected from the same model of WLAN card. Finally, we point out the quantization effect of measuring the RSSI using WLAN cards. Note that the units of statistical parameters reported in subsequent tables and figures are as follows. The mean values are in dBm. The standard deviation is in dBm. The variance is in dBm^2.

4.1. Differences of WLAN cards

Comments from the authors of [7,31] suggested that location fingerprints with different makes of wireless cards could be different, perhaps substantially. In this section, we investigate in more detail whether different cards report significantly different RSSI values. It is said that some vendors implement better receivers for the IEEE 802.11b/g cards than others. In fact, it was reported that different vendors choose to measure RF energy differently [21].

Although the IEEE 802.11 standards define RF measurement value as a number between 0 and 255, the actual implementation of each vendor is limited to between 0 and a specific maximum RSSI value called "RSSI_Max" (not in dBm) [21]. For instance, Cisco's 802.11 card has a maximum RSSI value based on 100 levels, while the Atheros chipset has a maximum of 60 levels. These values are used internally by a microcode on the WLAN card and by a device driver [21] to

Table 3
List of available WLAN cards.

Vendor	Model	WLAN standards	Interface	Max (dBm)	Min (dBm)	Measurable Range
Lucent	Orinoco Gold	802.11b	PCMCIA 16-bit	−10	−102	92
Lucent	WaveLAN Silver	802.11b	PCMCIA 16-bit	−10	−94	84
Cisco	Aironet 350 series	802.11b	PCMCIA 16-bit	−10	−117	107
Proxim	Orinico Gold	802.11a/b/g	Cardbus 32-bit	−11	−93	82
SMC	EZ connect SMC2635W	802.11b	Cardbus 32-bit	−14	−82	68
D-Link	AirPlus DWL−650+	802.11b	Cardbus 32-bit	−50	−100	50
Hawking	HWC54G rev.R	802.11g	Cardbus 32-bit	0	−75	75
3COM	3CRUSB10075	802.11b/g	USB	+10	−94	104
Intel	PRO/wireless 2200BG	802.11b/g	mini-PCI	−10	−84	74

report the quality of the signal. Each vendor has its own RF measurement accuracy, granularity, range for the actual power in dBm, and range of RSSI values (0 to RSSI_Max) [21]. This clearly creates complications in using different WLAN cards for an indoor positioning system.

Moreover, some vendors report the RSSI in percentages using the RSSI range, but the RSSI level can be mapped to a value in dBm using a table. The mapping between the actual RF energy and the range of RSSI values is different and may not be publicly disclosed for each vendor. For location fingerprinting purposes, a wireless card with a wider range of RSSI values or good granularity is better since it allows a positioning system to better differentiate between two locations.

In this section, different wireless cards were tested using customized RSSI collecting software based on WMI in Windows XP [26] as discussed in Section 3. Fortunately, the query requested to Windows XP's network driver interface specification (NDIS) API reported back the received power in dBm and our code did not have to perform any data translation. Table 3 lists the vendors, models, WLAN standards, interface type, maximum RSSI, minimum RSSI, and range of RSSI for all the WLAN cards used in comparison of this section. Note that we divided the WLAN cards into two groups where the second group, which is the last three WLAN cards in the table, was acquired later in our research study.

Because the mapping between the actual RF energy and the RSSI range can vary from one vendor to another, the choice of WLAN cards can affect the performance of indoor positioning systems. Since the range and the measurement of RSSI depend on the WLAN card, we recommend the use of the same model of wireless card for collecting the location fingerprints and determining the location.

For communication purposes, a WLAN card that has a higher average received signal level at the same location is considered to be a better card. However, a better report of received signal strength level may not necessarily be important for location fingerprinting and positioning. For positioning purposes, we argue that the range of RSSI and the standard deviation of RSSI are the two most important properties of WLAN cards that could influence location determination performance.

First, the widest range of RSSI allows positioning systems to identify more locations by differentiating the RSSI values at such locations. Table 3 compares the range of each card. These ranges are based on our investigation by measuring the RSSI values at locations where the AP's antenna touches the WLAN card all the way up to those locations where the signal from the AP is very low or disappearing. An observation from the measurement results of the Cisco card indicates that some RSSI values in dBm will never be reported, which is also pointed out by Bardwell [21]. The mapping of RSSI to dBm is non-linear [21]. The data reported in Table 3 showed that the D-Link card had the shortest range of all, while Cisco card had the widest range. Based on this observation and our argument above, we could say that the Cisco card is more suitable for use in indoor positioning systems than the D-Link card. Note that the ranges reported in Table 3 may not correspond to the actual range between 0 and the RSSI_Max of each vendor.

Second, although signal fluctuations normally occur for any wireless communications due to changes in environment, we argue that a receiver which reports RSSI with smaller standard deviation can be useful for location fingerprinting. The main reason is that it is less likely to show a different value of the RSSI at a particular location. Clearly, a smaller standard deviation causes less confusion with nearby location fingerprints. Our first cut analytical model presented in [11] also confirmed this argument in which a smaller standard deviation of RSSI led to a better localization performance.

Figs. 2 and 3 report the means of RSSI values and the standard deviations measured by different WLAN cards at a location inside a room in the small office of Scenario 1. Each card collected 300 samples of RSSI over a period of 5 min (1 sample/Second) on each day and the average RSSIs and standard deviations were calculated. Although the means of RSSI seemed to vary a lot both with the same card and different cards over different measurement dates, the overall averaged RSSIs reported by the last bar of each group are around −64 dBm to −66 dBm for this particular location. Proxim, D-Link, Lucent Gold, and SMC cards reported approximately 1–2 dB better RSSI on average than Cisco and Lucent Silver cards. On the other hand, Cisco's Aironet card had the largest average standard deviation of 3.02 dBm, while the D-Link card had the smallest average standard deviation of 1.08 dBm. Note that we did not perform measurements on Hawking, 3COM, and Intel at this location; therefore, no results of these cards were included.

The reason for the difference in the standard deviation of these cards can be explained by the difference in mapping from the actual RF energy to the internal range of RSSI values. Since Cisco's Aironet card has the widest range, it can measure the signal with higher resolution and see more variation of signal. On the other hand, the D-Link's AirPlus card has the shortest range; therefore, a number of actual measured signal levels may be mapped into the same RSSI value and it sees less signal variation. This effect can be called the *quantization effect* of each WLAN card.

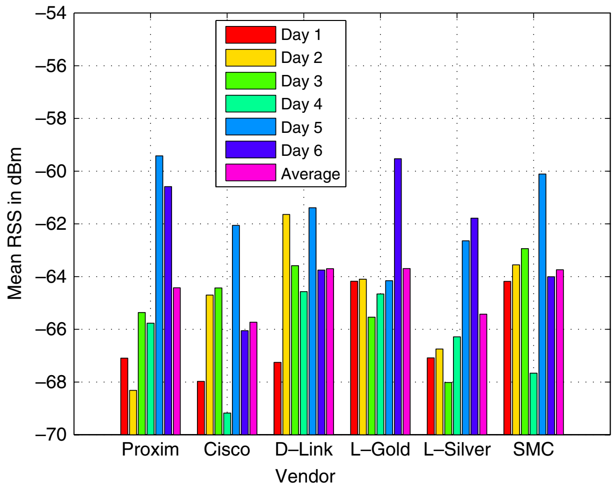

Fig. 2. Comparing mean RSSI of different vendors.

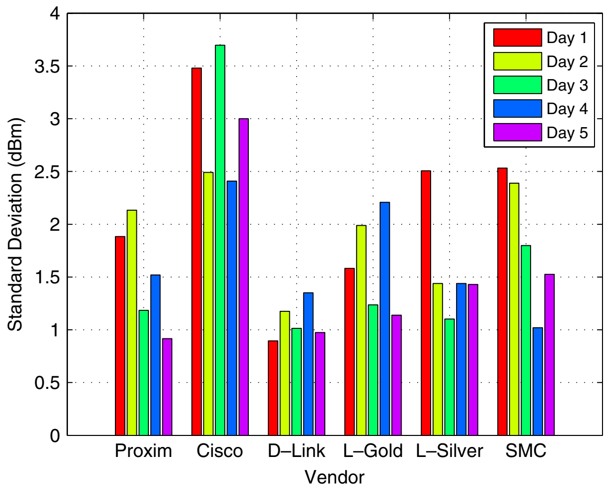

Fig. 3. Comparing standard deviation of different vendors.

However, there is yet another factor that limits the data collection of RSSI, which is the support of software tools to access the RSSI on the WLAN card. Although there are multiple alternatives for obtaining the RSSI as discussed in Section 3, without the knowledge of the proprietary mapping between the actual RF energy and the range of RSSI values for each card, the measurement of RSSI based on customized code or open source software could be an approximation at best. Thus far, only the Lucent Gold and Silver cards provide the RSSI logging capability in their proprietary client manager software. Fortunately, the Lucent gold card seems to have a good RSSI measurable range and average standard deviation; therefore, we selected this card for most of our measurement experiments because we believed that it could give the correct mapping of RSSI in dBm from the card.

4.2. WLAN cards from same vendor

A comparison of wireless cards from the same vendor and model is shown in Table 4 where two Lucent Gold cards were used. Each card collected 300 samples of RSSI over a period of five minutes (1 sample/second) to calculate the summary statistics in the table. Note that the measurement was done consecutively at an identical location inside the environment in Scenario 1. Although the means were different by about 1 dB, the range, the standard deviation (variance), and the skewness were very similar. This simple comparison confirmed that the WLAN cards form the same vendor should provide similar RSSI collection from the same location in indoor positioning systems.

4.3. Impact of quantization of RSSI values

Although the RSSI measurement reported by the software device driver is only in the quantization step of 1 dBm, the RF signals which are represented by RSSI are actually real and continuous values. Fig. 4 depicts an example of quantized

Table 4
Statistics of RSSI measured from two Lucent Gold cards.

Statistics	Lucent Gold 1	Lucent Gold 2
Mean	−62.41	−61.28
Median	−62	−61
Mode	−62	−61
Standard deviation	1.61	1.70
Sample variance	2.59	2.89
Skewness	−0.87	−0.48
Range	11	11
Minimum	−69	−68
Maximum	−58	−57

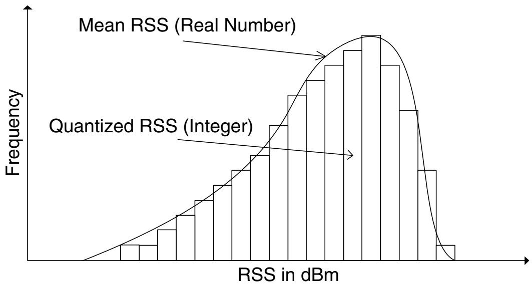

Fig. 4. Quantization of RSSI.

RSSI values which are reported by typical wireless cards. These quantization bins are represented by all possible values reported by RSSI of each WLAN card. Clearly, the larger the number of quantization of RSS in the graph is, the better the RSS represents the real WLAN signal. A WLAN card with more quantization steps should be a better card for indoor positioning systems. If only the integer number was used for the location fingerprints, the chance of any two location fingerprints having identical fingerprints will increase and degrade the performance of the fingerprinting technique. This argument is supported by the results in [9] where location estimation performance of the Gaussian method, which is non-integer, was better than the histogram method, which is integer. Typically, most of the research works in indoor positioning systems calculate the *average* values of RSSI and record them as real numbers. This in effect reduces the problem of the quantization effect.

5. Statistical properties of RSSI due to environmental effects

Changes to the RF signal due to indoor environment are difficult to predict because of the dense multipath environment and propagation effects such as reflection, diffraction, and scattering [32]. The multipath fading effect, which is the result of either a constructive or destructive combination of multiple signal copies at the receiver, causes the received signal to fluctuate around a mean value at a particular location. The received signal is usually modeled by the combined effects of large-scale fading and small-scale fading [6]. The large-scale-fading component describes the signal attenuation as the signal travels over a large distance and is absorbed by material such as walls and floors along the way to the receiver. This component predicts the *mean* of the RSSI and usually has a log-normal distribution [6].

On the other hand, the small-scale-fading component explains the dramatic fluctuation of the signal due to multipath fading. If there is no line-of-sight (LOS) component, the small-scale fading is often modeled with a Rayleigh distribution [6]. Such scenarios are commonly called non-line-of-sight (NLOS). If there is a line-of-sight component, the small-scale fading is typically modeled by a Rician distribution [6]. Note that when the RSSI is measured, the measurement averages out the small-scale-fading effects. Although these radio propagation models have been studied extensively in the literature, they are focused on their impact on receiver design and coverage. There is still a lack of the necessary understanding of the properties of RSSI from the perspective of indoor positioning systems.

5.1. Distribution of the RSSI

Traditionally, the average RSSI is believed to be log-normally distributed according the large-scale fading model [6]. The mean value is generally predictable and believed to follow one of several standardized path loss models discussed in [32]. However, there are some conflicting conclusions regarding the RSSI distribution measured at the software level by the wireless network interface card for indoor radio propagation in [12,33]. Moreover, the standard deviation and the stationarity of the RSSI are not understood very well.

The results in [12] were based on a five second sampling period over a long duration of five hours, 20 h, and one month. They concluded that the RSSI was log-normally distributed due to the similarity of the median, the mean, and the mode.

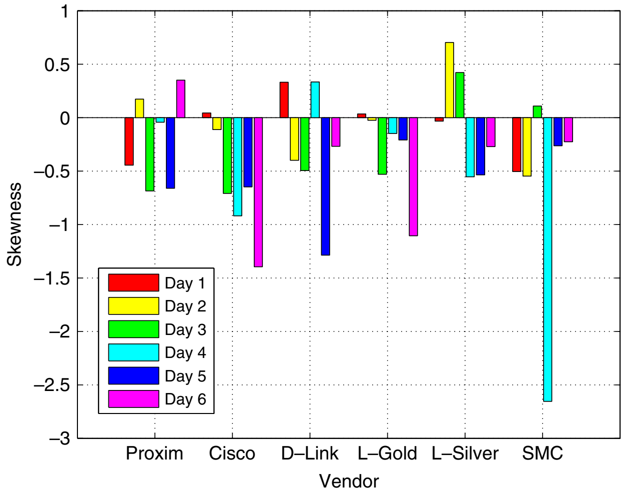

Fig. 5. Comparing skewness of different vendors.

However, they did not indicate whether the user was present all the time during the measurements. Thus, we suspect that the distribution of the RSSI in dBm that could be observed in reality may not be normal as described in [12]. A recent study of a 45-s measurement period with the user's presence in [33] pointed out that the RSSI distribution was non-Gaussian and asymmetric. Moreover, the histograms in [33] 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. These results did not however consider an in-depth analysis of the RSSI distribution.

Results from Scenario 1, Scenario 2, and Scenario 3 described in Section 3 yielded 75, 299, and 80 sets of histograms, respectively. Depending on the duration of measurement and availability of signals, Scenario 1 had samples of data anywhere between 441 and 1748 points, while Scenario 2 had samples of data anywhere between 2962 and 3956 points. Scenario 3 was conducted using customized code and we could collect 1800 sample points per card for each location. Observations from these 454 sets of histograms indicated that certain shapes of distribution often occurred at particular average values of the RSSI. Different shapes of distributions are caused by the upper and lower bounds of measurable RSSI at each location. Because the received signal is attenuated over distance, a received signal will never reach a value that is as high as the maximum transmitted power. The upper bound of the received signal in all of our measurement results showed no signal deviating higher than 10 dB above its average value. On the other hand, the lowest received signal is limited only by the receiver sensitivity. The signal deviation below the mean value can vary as much as the lowest receiver sensitivity.

Based on our observation, if the average RSSI is high (−80 dBm or above), the distribution or histogram of the RSSI will often have a long tail to the left which is called left-skewed distribution.[1] If the average RSSI is low (around −80 dBm), the distribution will not have a long tail but will be almost symmetric or appear to be a log-normal distribution (normal in dB).

Our measurement results showed different shapes of the distributions (based on plots of histogram) from three APs in Scenario 1, six APs in Scenario 2, and one AP in Scenario 3. In Scenario 1, we observed that signals from access points with stronger average RSSI and/or line-of-sight (LOS) have larger deviations, primarily to the left side, while distributions of signals from farther access points with weak average RSSI are almost symmetric. In Scenario 2 and Scenario 3, similar shapes with long left tails could also be observed. Note that histograms with multiple modes rarely occurred in any scenarios, which is in contrast to the results in [33]. A possible explanation is that multiple-mode histograms might occur with extended periods of measurement that are longer than one hour. Signals measured with different makes of card at the same location on different days also exhibited a majority of left-skewness as shown in Fig. 5.

To determine whether the data is significantly left-skewed, we compared the skewness of each data set with its standard error of skewness. The standard error of skewness can be estimated by $\sqrt{\frac{6}{N}}$ where N is the number of data points [35]. If an absolute value of skewness is larger than two standard errors of skewness, the data set is considered to be significantly skewed. The distributions tended to be left-skewed in most measurement results from all scenarios. In Scenario 1, we observed that 64 out of 75 histograms were significantly left-skewed, while 7 histograms were symmetric and 4 histograms were significantly right-skewed. In Scenario 2, we found that 191 out of 299 histograms were significantly left-skewed, while 51 of the data sets were symmetric and 57 histograms were significantly right-skewed. In Scenario 3, we found that 50 out of 80 histograms were significantly left-skewed, while 21 of data sets were significantly right-skewed and 9 of them

1 Skewness is a measure of symmetry of data. A probability density function (PDF) is said to be skewed to the left (tail on the left) when it has its mean less than its median which is less than its mode [34]. Skewness is reported by number in which a negative number reflects a left-skewed distribution. For univariate data $x_1, x_2, \ldots, x_N$ with mean $\bar{x}$ and standard deviation σ_x, Skewness = $\frac{\sum_{i=1}^{N}(x_i-\bar{x})^3}{(N-1)\sigma_x^3}$.

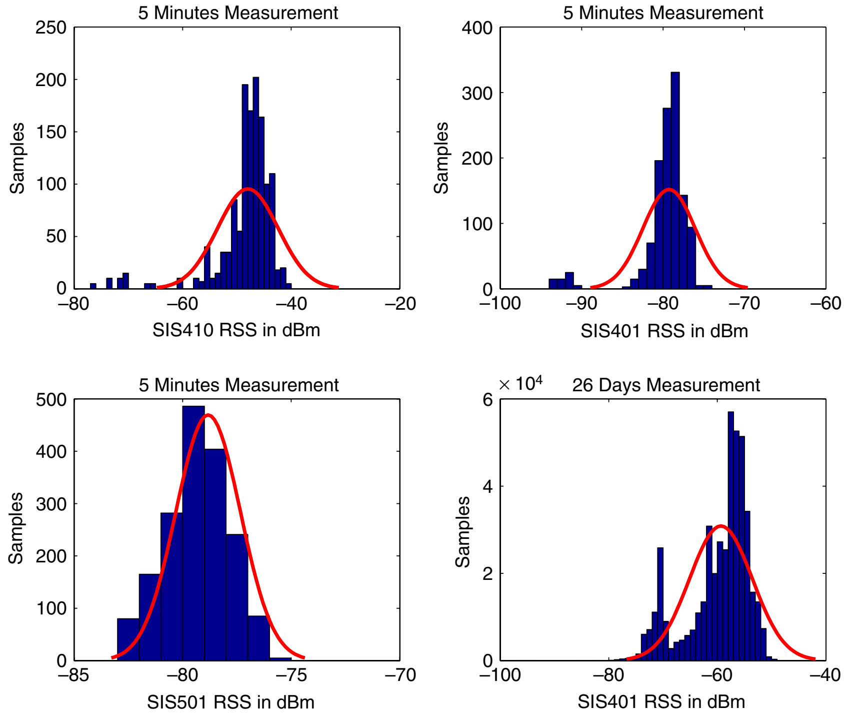

Fig. 6. Samples of RSSI distribution over five minutes and 26 days.

were symmetric. In the last scenario, we observed that WLAN cards from 3COM, Intel, Hawking technology exhibited a majority of left-skews while the SMC card had an almost equal number of left and right-skewed histograms. The only card that had majority of right-skewed histograms was D-Link. From our observations, the histograms that are significantly left-skewed are often the ones with strong average RSSI and/or when there is a line-of-sight between an AP and a MS. On average across all scenarios, the left-skewed histograms were 70% of the total while the right-skewed histograms were 20% and the symmetric histograms were 10%.

Four samples of RSSI histograms collected from three APs in Scenario 1 are plotted in Fig. 6. We compared three short measurements of five minutes with another long measurement of the signal from one of the AP in the same scenario. The long duration of measurement was collected over 26 days at a sampling interval of five seconds. The histogram of the long measurement is shown in the lower-right subgraph of Fig. 6. Notice that there are multiple modes in this measurement results corresponding to the 26 days. Fig. 6 illustrates examples of slightly left-skewed RSSI distribution measured from three APs.

The left-skewed distributions are prominent with both short (five minutes) durations and long (26 days) durations of measurement. The RSSI values are usually concentrated around the dominant modes. However, there are some histograms in our measurement that could be approximated by a log-normal distribution because they are slightly skewed or almost symmetric. Log-normal or Gaussian curves that can be fitted to data are superimposed on each histogram to compare the actual distribution with an ideal log-normal distribution. Slightly skewed distributions often occur when the RSSI level is low (the AP is far from the measurement location and/or there is no direct line-of-sight). These conditions are often valid for indoor environments. This observation could also explain why the measurements in [12] report a normal distribution where the measurement in that work was taken inside an office room with none-line-of-sight path.

Figs. 7 and 8 show the skewness of all histograms from Scenario 1 (small office) and Scenario 2 (large hall) in Table 2, respectively. In Fig. 7, the skewness of AP SIS501's histograms scatter around a zero value, while the other data from AP SIS410 and AP SIS401 have a rather large negative skewness. Note that the average RSSIs of AP SIS401 were found to have large negative skewness even though their levels are low (< -75 dBm). A possible explanation for this phenomenon is that the signal of AP SIS401 may reach the measurement area through the reflection effect along the corridor in Fig. 1 and it appears to be a line-of-sight signal. In Fig. 8, most of the data that have low RSSI values have skewness around the zero value, while the data that have higher RSSI values have larger negative skewness. Note that the measurement area in Scenario 2 did not have any direct line-of-sight to any of the APs.

The left-skewed distribution is difficult to model and does not fit to any well-known distribution. Most well-known distributions are right-skewed such as log-normal, Chi-square, Rayleigh, and Maxwell. We found a left-skewed distribution called Map–Airy distribution [36]. However, the expression is quite complex. A good representative distribution of the underlying RSSI process is needed to gain more understanding of location fingerprinting. If the RSSI distribution can be

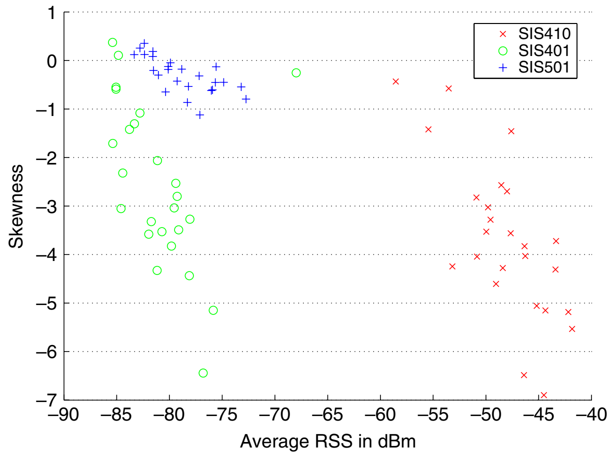

Fig. 7. Average RSSI vs. skewness from three APs in Scenario 1.

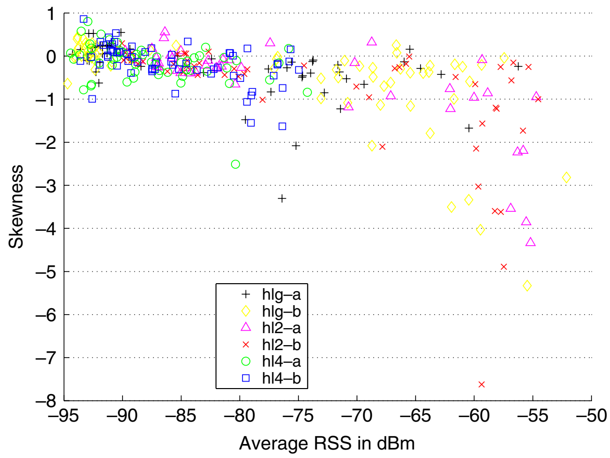

Fig. 8. Average RSSI vs. skewness from six APs in Scenario 2.

identified and modeled accurately, analytical models of location fingerprint and the indoor positioning system could be developed as demonstrated in [11] using Gaussian distribution. Although most of the histograms are not symmetric, an assumption of log-normal distribution can be used to partially describe the RSSI distributions. This is a reasonable choice but it is not the most accurate distribution for RSSI.

5.2. Standard deviation of the received signal strength

Using the same Lucent Gold card in both Scenario 1 and 2, this subsection reports sample standard deviations of all APs' signal at different locations. We found that the standard deviations vary from one location to another and from one AP to another. Sample standard deviations in Scenario 1 have values between 0.59 and 6.29 dBm, while sample standard deviations in Scenario 2 have values between 0.47 and 3.30 dBm. The major differences in the two environments are the distance and the existence or not of LOS between the APs and the receiver.

When we plot the mean RSSI versus the sample standard deviation, the results reveal that the farther the AP is from the MS, or the lower the received signal level is, the smaller the standard deviation is. On the other hand, the larger the mean RSSI is, the larger the standard deviation is. Fig. 9 depicts this relationship based on the results of Scenario 1. We observed that the standard deviation was large for high RSSI levels (−60 dBm – −40 dBm) when the signal from an AP had direct LOS to most of the locations in Scenario 1. Smaller standard deviations were observed in Scenario 2 in Fig. 10 when there was no LOS between any AP and each measurement location. From these results, it is clear that the WLAN signal will fluctuate less with low RSSI levels and with NLOS paths. Note that the measurement area in Scenario 2 is farther from AP's locations than in Scenario 1. This observation is similar to the skewness property discussed in the previous subsection that RSSI distribution is less skewed when RSSI levels are low. We argue that these phenomena are the combination results of the lower bound receiver sensitivity of WLAN cards and RF propagation effects in an indoor environment.

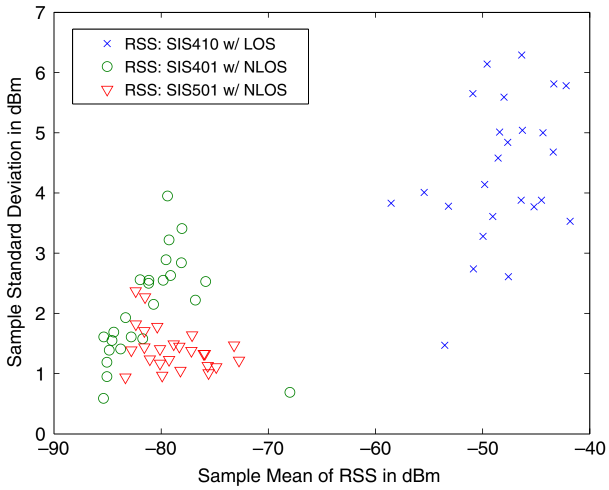

Fig. 9. Relationship between RSSI and its standard deviation from Scenario 1.

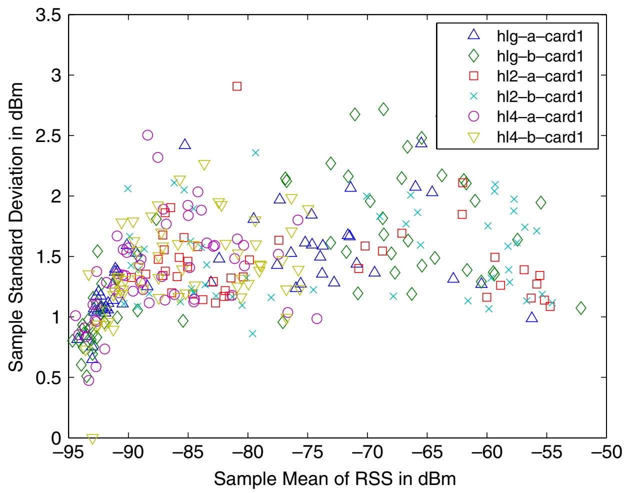

Fig. 10. Relationship between RSSI and its standard deviation from Scenario 2.

This property of the standard deviation suggests that the RSSI from two locations may be difficult to be separated or distinguished if both locations are close to the same AP due to the high signal level with a large degree of variation. On the other hand, two locations might be easily identified if both do not have LOS paths and are located far from the APs. Based on our preliminary model of indoor positioning in [11], the environment that has RSSI with lower standard deviation will have better accuracy and precision in location determination. This is rather counter-intuitive since the farther apart the WLAN user is from the AP, the worse the measurement accuracy should be as stated by Small et al. [12].

5.3. Stationarity of the received signal strength

Assuming that the Ergodic theorem[2] is applied according to Wiener's definition of stationarity [38], we analyze whether the RSSI is stationary by breaking the series of RSSI measurements into separate pieces over different time intervals. A random process is said to be *weakly stationary* when it meets the following two conditions [39]. First, its mean and variance remain the same over time. Second, its autocovariance function has the same shape for each separate time-series. Instead of plotting the autocovariance function, we plot the autocorrelation function versus time-lag, which is called correlogram, to test the second condition. Note that the autocorrelation function is the autocovariance function normalized to the zero-lag autocovariance. We investigated the stationary property over three time scales: pieces of 1 min and 15 s within five consecutive minutes, pieces of 15 min within the same hour, and pieces of one hour over five different hours. All of these measurements were taken at the same location within the area inside a room in the small office of Scenario 1.

2 Ergodic theorem states that the time average of a random process is equal to the space average of that process almost everywhere [37].

Table 5
Mean and variance of RSSI with user over short duration of 5 min.

	Mean				Variance			
Vendor	1st Period	2nd Period	3rd Period	4th Period	1st Period	2nd Period	3rd Period	4th Period
Cisco	−70.88	−68.37	−66.40	−71.05	9.86	6.72	4.16	19.48
D-Link	−64.57	−63.95	−64.49	−65.27	1.36	0.46	0.93	0.52
Lucent Gold	−64.55	−64.87	−64.32	−64.91	1.36	1.06	1.30	2.22
Lucent Silver	−66.65	−66.65	−65.40	−66.44	0.47	1.09	1.38	0.87
Proxim	−66.29	−66.13	−65.23	−65.43	0.94	2.14	1.02	0.73
SMC	−67.84	−67.11	−68.45	−67.25	4.22	3.31	1.20	3.19

Table 6
Mean and standard deviation of RSSI with user presence over 15 min.

Statistics	1st Qtr.	2nd Qtr.	3rd Qtr.	4th Qtr.
Mean	−71.71	−72.33	−71.82	−70.48
STD.	2.95	3.20	2.96	2.56
Var.	8.72	10.27	8.77	6.54

Table 7
Time dependency of RSSI (dBm) from SIS410 with user's presence.

Statistics	10 AM	12 AM	2 PM	8 PM	10 PM
Mean	−62.68	−60.02	−61.85	−63.12	−63.18
STD.	2.17	1.63	2.05	3.35	2.66
Var.	4.70	2.65	4.22	11.23	7.07

Using the data collected by different WLAN cards (except Hawking, 3COM, and Intel) in Section 4.1, we calculate the mean and the variance for four small periods of one minute and 15 s. Note that the sampling interval is one second over this small period resulting in 75 samples per period. Table 5 lists and compares the mean and the variance values of these cards across all four periods. By inspection, the means of the RSSI did not change by more than 1.5 dB for most cards. However, the mean RSSI of Cisco's card shifted quite a lot in its last period (−66.40 – −71.05 dBm or a change of 4.65 dB). The variance values also did not change much for most of the cards (less than 1.5 dB) except Cisco's card (4.16 up to 19.48 or a change of 15.32) and SMC's card (4.22 down to 1.2). The shift in means and variances at this time scale indicates that the RSSI is not a weakly stationary process. Since the first condition is not met at this time scale for all cards, we will not show the result of the correlogram for testing the second condition in this time scale.

In the second time scale, we divided a series of RSSI measurements at a location in Scenario 1 using the Lucent Gold card over a longer period of one hour into four groups of 15 min. Table 6 lists the summary statistics within each quarter of an hour. By inspection, the means and the variances across the four quarters were quite similar. These results suggest that the RSSI distribution could be stationary or time independent since the means are very close (less than 2 dB difference) and the sample variances of each quarter are in the same order. For this set of measurements, the first condition for stationarity is met.

Next, the correlograms were plotted to test the second condition for this second time scale. Fig. 11 depicts similar shapes for all quarters indicating that the second condition is also met for this time scale. Note that the fourth quarter plot in Fig. 11 had much smaller correlation coefficients at larger time lags which indicated that dependences of RSSI samples reduced faster in this last quarter. This implied a faster change of signal which might be a result of rapid change in the environment. The visual tests for this sample of measurement suggest the possibility of a weakly stationary process. Note that since the autocorrelations in the plots are significantly non-zero, the RSSI does have strong correlation between consecutive samples as pointed out by Youssef and Agrawala [10]. The correlation of RSSI samples was exploited in the design of Horus's indoor positioning algorithm. The authors of [10] found that autocorrelation of consecutive samples (lag = 1) could be as high as 0.9 [10].

Finally, the study on the third time scale consists of five periods of one hour over different times of day. The analysis was performed on a RSSI measurement at the same location in the small office of Scenario 1 using the Lucent Gold card. Table 7 shows a consistent mean, but inconsistent variance values of the RSSI as it shifted from 2.65 at 12 AM to 11.23 at 8 PM. Therefore, the test for the first condition for stationarity fails at this time scale. Based on all three time scales discussed in this section, we conclude that the RSSI random process is non-stationary. This is to be expected as the dynamic in the indoor environment changed over time. More observations of time dependency of RSSI will be discussed in detail in the next section.

5.4. Time dependency

The summary of statistics in Table 7 indicates that the RSSI is time-dependent. In this section, we performed a separate set of experiments inside the environment of Scenario 1. These experiments were aimed at determining time dependency

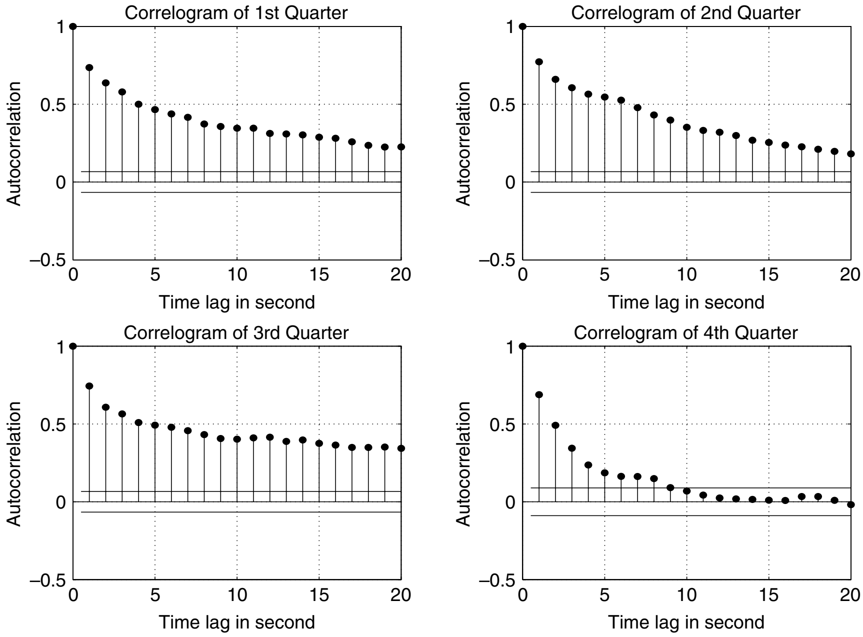

Fig. 11. Correlograms of RSSI within the same hour.

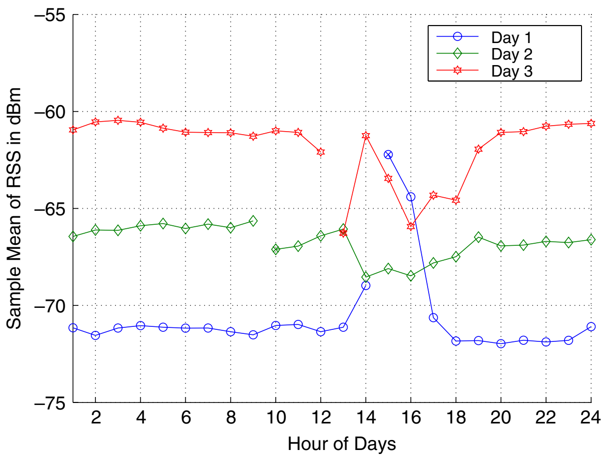

Fig. 12. Sample mean of RSSI over 24 h at three different days.

over hours of day and days of week. Only one location was considered in this analysis. To minimize the user's effect and focus on the time dependency effect only, we left a laptop on a desk in a room inside the small office of Scenario 1. No user operated this laptop after the measurement was started. However, the office was shared by other people; thus, within a vicinity of one meter or more, there could be other people sitting within the same room at any time. Moreover, the room had a door which was usually left open when a person entered this room. The presence of other people and the change of the door's position were uncontrollable in this experiment.

The measurement was performed over a continuous period of 24 h on three different days. The RSSI data were recorded once every second. The results of each day were divided into 24 series. The mean, variance, and skewness were calculated for each series to determine their time dependency. The results of mean and variance are plotted in Figs. 12 and 13, respectively. Note that the measurement of each day was started at a different hour of day denoted by a broken line in each figure. The measurement results were wrapped around for the time between the 24th and 1st h.

The observation of the mean values for three days in Fig. 12 shows stable signals during late night and early morning. The reason is that there were no other people around in the office and the office door remained closed during these periods. However, as we discovered in the previous section about the stationarity's property, the variance of the three signals in Fig. 13 shifted during the day over time between 10 AM and 8 PM. Considering the signal in the second hour after the

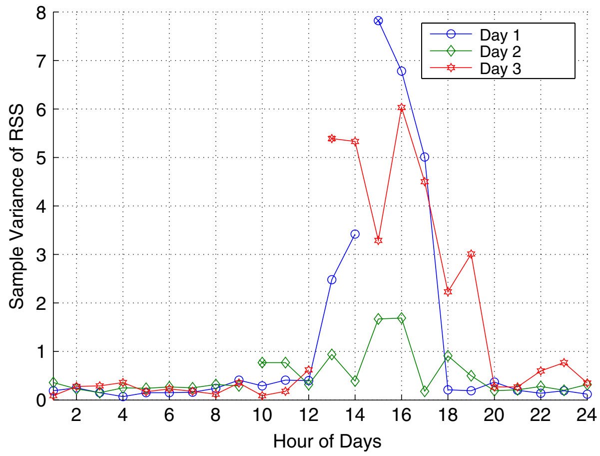

Fig. 13. Sample variance of RSSI over 24 h at three different days.

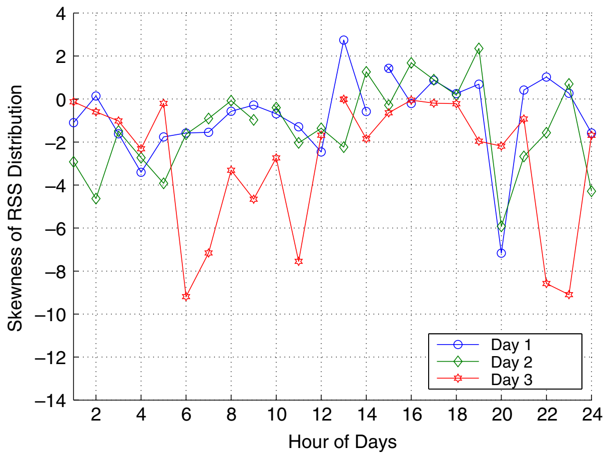

Fig. 14. Skewness of RSSI over 24 h at three different days.

starting of measurements each day, the signal variation occurred from changes in the environment when there were other people around and the office door was open.

Interestingly, the mean values on Day 1 during the day when people were present was higher than during the late night and early morning. Compared to the results on Day 2 and Day 3, the mean values during the day were lower than during the late night and early morning. This difference can be explained by the multipath effect, which might enhance or degrade the received signal level at the same location. In any case, the change in the RSSI's statistics by the environment is difficult to quantify because the exact same environment might not occur again.

Notice that the means of RSSI from the three different days were different with averages of −71, −66, and −61 dBm. This result suggests some dependency of RSSI over different days. Even though for each experiment the investigator tried to place the laptop with the same WLAN card at the same location, it was possible that a slight shift of the laptop's placement might occur and result in different means for each day. This may be interpreted as a problem for location fingerprinting where there were different means over different days. The results during the daytime when people were around indicated an average of RSSIs that was closer to a common value of −65 dBm. Therefore, it is possible that the human factor can dominate the average RSSI for the measurement at each location, which can result in similar average RSSIs over different days. Due to limited numbers of experiments, we cannot conclude that the same average RSSI will occur on each day of the week.

Fig. 14 plots the skewness over different times. The results in this figure show that the signals are often left-skewed during late night and early morning. The three signals shifted toward right-skewed distributions during the afternoon hours when there were more people and activity around the measurement point. These results indicate the dependency of the RSSI on changes in the environment. Therefore, these results suggest that the time dependency of the RSSI is in fact the dependency due to the environment that changes over time.

5.5. Interference and independence

This study quantifies the interference and the independence of multiple signals within each RSSI pattern by calculating the correlation coefficient between any two sequences of received signals collected at a location. The correlation coefficient is reported as a real number between 0 and 1. It is calculated by $R = \frac{\sigma_{i,j}^2}{\sqrt{\sigma_i^2 \sigma_j^2}}$ where $\sigma_{i,j}^2$ is the covariance between random variable i and j, and σ_i^2 and σ_j^2 are the variance of random variable i and j, respectively. If two signals have no effect on each other, the correlation coefficient will be 0. If two signals have a very strong relationship or depend on one another, the correlation coefficient will approach a value of 1. A guideline to classify the correlation coefficient of two random variables is that a value greater than 0.5 is high, 0.5–0.3 is medium, 0.3–0.1 is low, and anything smaller than 0.1 is trivial [40].

Samples of RSSI patterns collected in both Scenario 1 and 2 are used to verify the statistical independence of signals from different access points. Note that Scenario 3 only collected RSSI from one AP; therefore, it was not analyzed in this section. In Scenario 1 as described in Table 2, there were three signals that could be measured all the time from three APs in the environment. The correlation coefficient values of two of the three signals were calculated for all 25 locations which resulted in 73 correlation coefficients. Note that two correlation coefficients could not be calculated due to the unequal number of sample points. Most of the correlation values (51 of them) are lower than 0.1 which means that there is very small to trivial correlation. The rest of the values (21 of them) are between 0.1 and 0.3. Only one value is larger than 0.5 but less than 0.8. From these results, it is clear that WLAN signals do not have a very large correlation. Signals from different radio channels or signals from the same radio channel but different transmitters can be assumed to be uncorrelated. In the ideal case, it can be reasonably assumed that the RSSI from each AP is unrelated or independent in Scenario 1.

There are two APs in Scenario 1 that use the same frequency (channel number 6). One may think that the RSSIs from both APs might interfere with each other and cause difficulty in forming the location fingerprint. However, our measurement results indicate that both RSSIs have only a low to tiny correlation. The signals do not interfere with each other. Therefore, access points using the same radio frequency do not have any significant impact on the location fingerprint. This could be explained by the collision avoidance mechanism of the IEEE 802.11 [29], which enables a clear signal reception.

In Scenario 2, there were six signals available inside the large hall environment as described in Table 2. Because the area in this scenario was larger than the first one, not all signals could be measured at every location. Based on our analysis, only four correlation coefficients out of 269 values were between 0.5 and 0.3, while 90 of them were between 0.3 and 0.1, and 175 of them were less than 0.1. These results suggest that there is a moderate to very small correlation between any received signal pair. Once again, these results can lead to the reasonable assumption that the RSSI from multiple APs are uncorrelated or independent in Scenario 2.

5.6. Required number of samples

Collecting enough statistics for creating location fingerprints is the key to achieving good performance with any indoor positioning system. If the positioning system requires only the mean values to create fingerprints, a small number of samples is sufficient since the mean is relatively constant. Ideally, we would like to have as many samples as possible. In the literature, a small number in the order of 20 samples per location and orientation is used with acceptable location determination performance in RADAR [2]. Larger numbers of samples are required for the probabilistic approach to create accurate histograms. Youssef [41] and Xiang et al. [42] recorded 300 samples per location and orientation.

The durations of data collection in the literature are different due to the sampling period. For instance, RADAR [2] used a 0.25-s sampling period, while Xiang et al. [42] used a two-s sampling period. In Table 2, the sampling period was set differently in each scenario to observe any difference in statistics of measured data. However, results of all scenarios reflected the same trend of dominated left-skewed distributions. The sampling period is limited by either the software or the hardware. The software's limits depend on how often a device driver can be accessed and how often the BSSID scan list is updated. Some wireless cards have the capability to scan for APs' signals in the background [28]. The hardware's limits depend on how the vendor implements the scanning cycle and the amount of channel dwelling time. Clearly, it is difficult to obtain these limits because the information may be proprietary for each vendor. However, Microsoft specifies that a scanning query must be returned within two seconds after the query command is initiated [28]. Microsoft's design guidelines also provide an example of five seconds (for very frequently) as how often the scanning operation can be. In other words, Microsoft's specifications require at least six seconds between each scanning operation. Our own experiment with the SMC card with a one second scanning cycle showed that the scanning operation can be faster than six seconds as our BSSID scan list was updated every one second.

To study the required number of samples, we select a set of RSSI measurement data from a location (Location no. 47 of 71) and an AP called hlg-b-card1 in Scenario 2. Then, we calculate summary statistics of the AP with different numbers of samples ranging from 30, 50, 100, 150, 200 to 300, and the maximum amount of collected data as showed in Table 8. Assuming that the maximum amount of collected data represents the most accurate distribution of RSSI at that location, we compare the summary statistics of each number of samples. By inspection, it is true that only small numbers of samples (30 and 50) would be sufficient for a location fingerprint based on the mean values only. If we approximate the distribution of RSSI using log-normal distribution, the distribution can be completely described by the mean and the variance. Therefore,

Table 8
Summary Statistics of AP: hlg-b-card1's RSS at Location 47 in Scenario 2.

Number of samples	30	50	100	150	200	300	3545
Mean	−60.63	−60.96	−60.95	−61.16	−61.60	−61.59	−61.66
Standard error	0.18	0.15	0.11	0.10	0.11	0.08	0.02
Standard deviation	1.00	1.07	1.15	1.27	1.50	1.37	1.39
Sample variance	1.00	1.14	1.32	1.63	2.26	1.87	1.93
Skewness	0.06	0.13	−0.34	−0.29	−0.15	−0.09	−0.19
Range	3	4	5	5	6	6	9
Confidence level (95.0%)	0.37	0.30	0.23	0.21	0.21	0.16	0.05

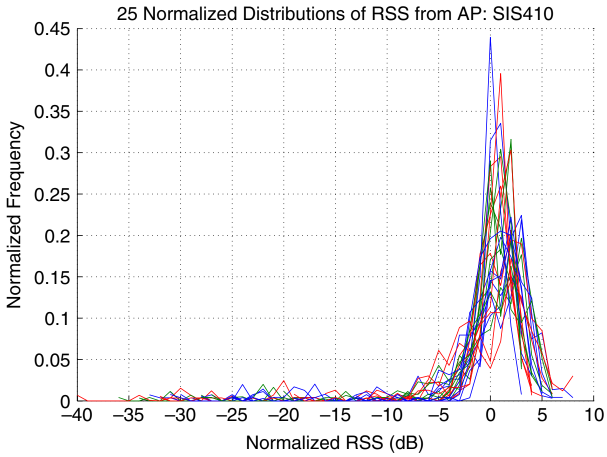

Fig. 15. Distribution of AP: SIS410.

the convergence of standard deviation can be used as a condition to stop collecting new samples. From our analysis, a number between 150 and 200 should be sufficient to reduce the standard error of the mean in the second row to about 0.1 and to obtain a good approximation of variance as shown in Table 8. Note that this conclusion did not take into account the effects of location determination algorithms or pre-processing techniques [15] which might reduce the required number of samples. The authors of [43] also suggested that a number of samples that was less than 30 might lead to in-accurate modeling of an RF environment for indoor positioning systems.

6. Causes of performance degradation in location determination

The randomness of RSSI is the major cause of error in any indoor positioning system that uses the RSSI for location inference. If there was no randomness in the RSSI, every indoor positioning system based on location fingerprinting techniques would have no location determination error (excellent accuracy and precision performance). In this section, we investigate the causes of error in identifying any indoor locations.

6.1. Randomness of received signal strength indication patterns

The randomness of RSSI patterns is clearly described by its probability density function (PDF) or its distribution. To understand the cause of the error, we need to understand the nature of the randomness of the RSSI. Based on our observation of all histograms collected during our statistical analysis, there were two major shapes of histograms which can be visualized as in Figs. 15 and 16. Both figures plot superpositioning graphs of normalized histograms (zero mean) of two APs with 25 locations in Scenario 1. The overall histograms of RSSI from AP SIS410 reflected a majority of left-skewed distributions where all locations had line-of-sight with an AP and strong average RSSI (−48 dBm) while histograms from AP SIS510 reflected symmetric distributions where all locations had no line-of-sight with an AP and weak average RSSI (−79 dBm).

Fig. 17 illustrates how the RSSI distribution (both mean and standard deviation) changes with the average RSSI. The stronger the mean RSSI, the larger the tail to the left and its variation. On the other hand, the weaker the mean RSSI, the more symmetric the distribution becomes. Based on our measurement results, symmetric distributions can be approximated by the log-normal distribution. Three different RSSI distributions are shown for comparison purposes in Fig. 17.

Since the level of RSSI sometimes also represents the distance between the AP and the mobile, this phenomenon suggests that the location detection algorithm will be more likely to make errors when a sampled RSSI falls in the tail of the left-skewed distribution. In this case, the algorithm will return incorrectly estimated location which is farther away from the

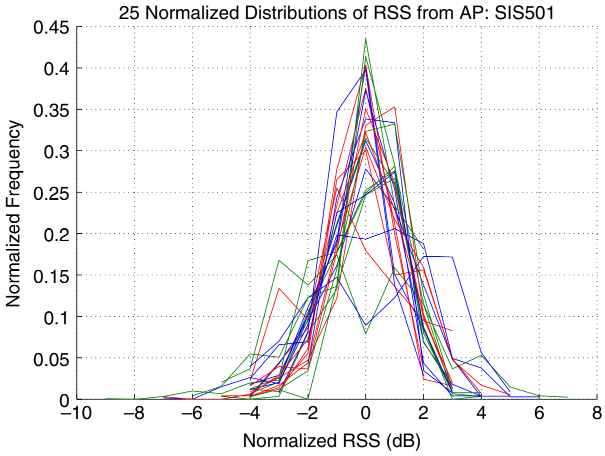

Fig. 16. Distribution of AP: SIS501.

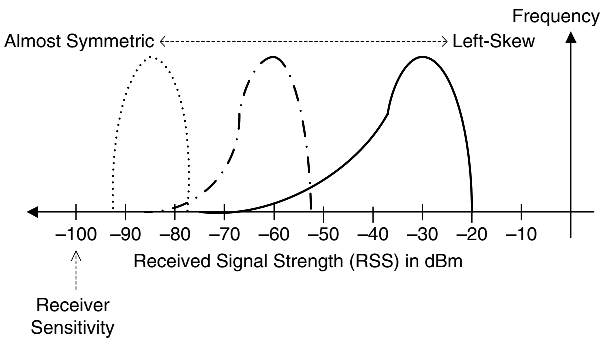

Fig. 17. Unique property of RSSI's distribution measured from typical WLAN card.

AP. On the other hand, the closer the actual location is (stronger RSSI) to an AP, the more difficult it is to identify the correct location among the high frequency of occurrence of RSSI. One possible reason behind the varying distribution is the non-linear mapping between the actual RF energy and the reported RSSI values of WLAN's card. Intuitively, the weaker the received signal level, the more difficult it should be for the WLAN card to be able to differentiate it. Ideally, if we have a perfect measurement tool that can measure small RF energy, we expect the distribution of RSSI to be left-skewed across all RSSI levels.

6.2. Separation of location fingerprints

The performance of indoor positioning systems depends greatly on the separation of location fingerprints. A location fingerprint corresponding to a location can be identified correctly if it is difficult to classify it (incorrectly) as another fingerprint by a pattern classifier. Note that we do not specify any pattern classifier at this point. Theoretically, a change in RSSI is proportional to the logarithm of a distance between a transmitter and a receiver. Therefore, two different locations with different distances from the same AP should have different average RSSI values. However, in practice, the RSSI is a random variable that has its value fluctuating around an average due to the dynamics in the environment. These fluctuating values from multiple APs can be grouped together as patterns of RSSI at a particular location.

From the square grid of 25 locations in Scenario 1, we found that the actual location fingerprints are not always nicely separated as shown in Fig. 18. Note that we plotted the average RSSI from two APs at each location and labeled each location with corresponding standard deviations (σ_{SIS410} and σ_{SIS501}). From the visual inspection of the fingerprints, we can see that they are almost random and sometimes not well separated. Some locations that should be closer together in physical distance may not be close together in RSSI signal space. Other cases that appear to be separated in physical space may be close in signal space. This phenomenon is difficult to model or characterize because the relationship between physical distance does not perfectly correspond to signal space distance.

However, by comparing Fig. 18 and the square grid of 25 locations in Fig. 19 closely, we can see that the neighboring structure of fingerprints is still apparent. For instance, location 13 is in the middle of the physical uniform grid and it is also in the middle of location fingerprints in Fig. 18. Based on the above visualization, two locations such as Location 13 and 19 were very close in RSSI signal space even though Location 19 was not the closest location in the grid to the Location 13.

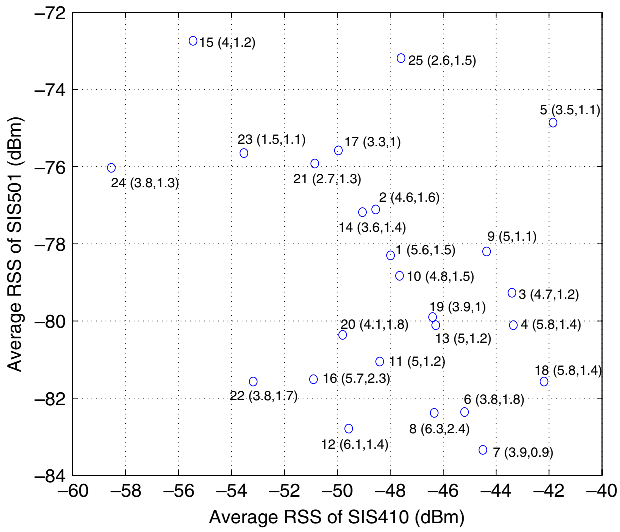

Fig. 18. Location fingerprints of 25 positions in Scenario 1.

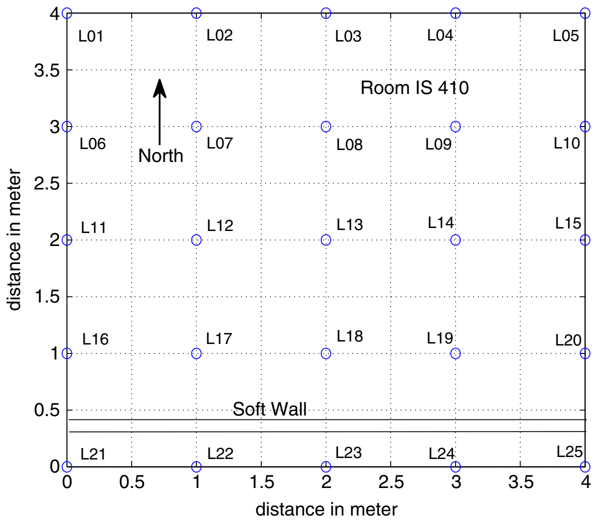

Fig. 19. Uniform grid space of 25 positions in Scenario 1.

Any two locations become difficult to identify if they are closer together and the distributions of their RSSI patterns significantly overlap.

To illustrate the separation problem, Figs. 20 and 21 show two-dimensional plots of RSSI patterns from four pairs of locations that are separated by one, two, three, and four meters in Scenario 1. Note that the pattern of RSSIs can be grouped together as clusters when removing some outliers. Highly overlapping patterns in the sub-plots of one-meter separation in both figures suggest that using two AP signals is insufficient to identify two locations separated by this distance. The degree of separation increased as the distance separation increased from one to four meters. However, the clusters of two different locations may not be nicely separated using only two AP signals.

Although Fig. 21 shows a similar problem at the same locations as Fig. 20, there is a minor difference between these two figures. That is that the AP SIS410 has a larger signal variation than the AP SIS401. Notice that the main cluster of SIS410's signal has a spread that is approximately 10 dBm on the abscissa, while the main cluster of SIS401's signal has a spread that is only approximately 5 dBm. Separation of RSSI patterns in Fig. 21 is much easier than in Fig. 20 for a separating distance of two meters or more. This observation suggests that signals with larger standard deviations (or variance) will make it more difficult to perform location identification based on location fingerprinting.

Fig. 22 plots the frequency of occurrence of two RSSI patterns called Location A and B which were sampled at 1 sample/second over one hour. The patterns near the center of each cluster do have very high frequencies of occurrence; therefore, the average or mean value of RSSI could represent the RSSI patterns very well. This visualization suggests that we may use the center of the cluster as a representative of the location fingerprint instead of the distributions of all

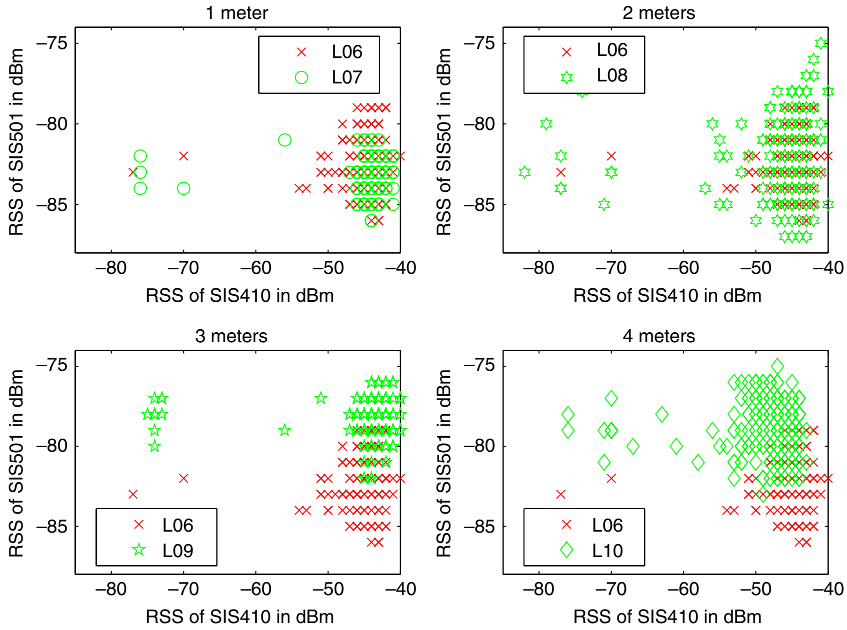

Fig. 20. Separation problem of two locations in Scenario 1 with SIS410 and SIS501.

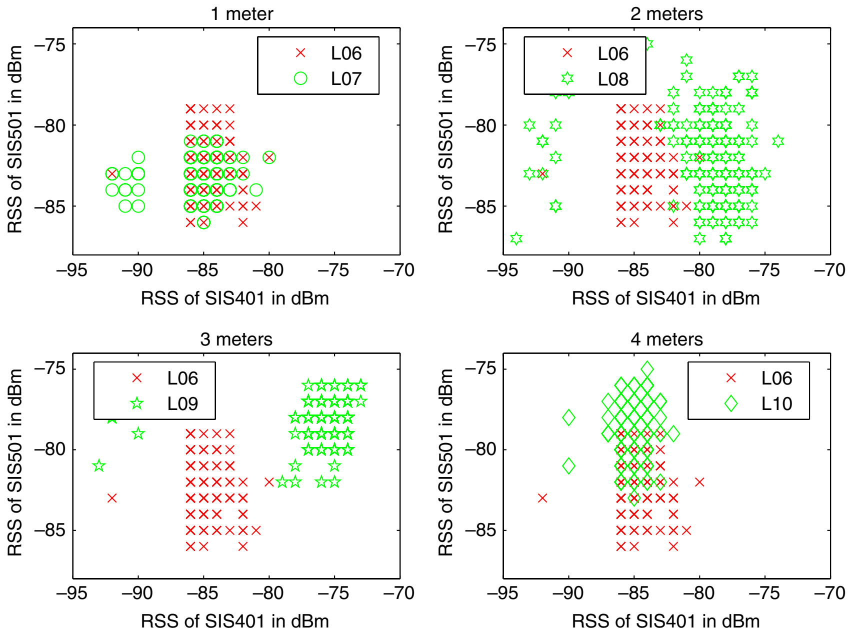

Fig. 21. Separation problem of two locations in Scenario 1 with SIS401 and SIS501.

RSSI features. The edge of a cluster could be drawn from the valley in between any two clusters. The edge between any two clusters can be used as the discriminant for pattern classification.

In real situations, the number of locations that need to be identified is much more than two (in the order of hundreds per floor). Increasing the number of access points is one way of separating two location fingerprints further. The overlap between the patterns becomes a lesser problem for the location discovery algorithm as the number of RSSI elements increases. This can be depicted in Fig. 23 when we plot the three dimensions for the RSSI patterns with the addition of the signal from the third AP. Clearly, the two RSSI patterns became easier to classify. Note that it is impossible to illustrate the frequency of each pattern in this case, but we may deduce from the previous figure that the highest frequency of occurrence will be at the center of each cluster.

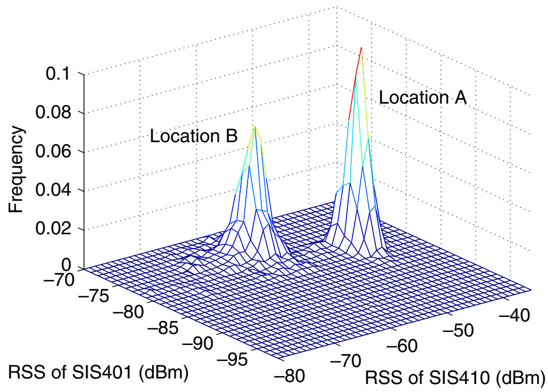

Fig. 22. Two clusters with frequency of RSSI fingerprints with two RSSI elements.

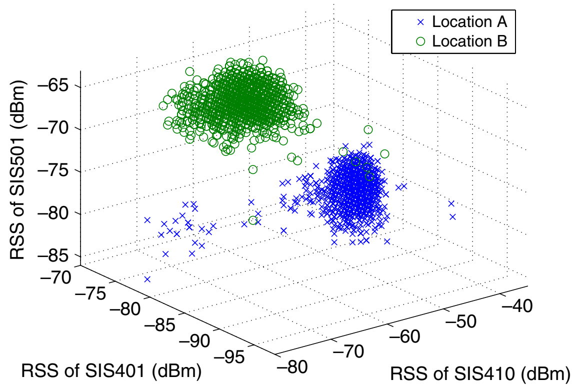

Fig. 23. RSSI fingerprints with three elements.

6.3. *Temporal received signal strength outage*

In real systems, access points are often installed opportunistically and mobile stations (MS) may not receive signals from all access points inside a building. In other words, the placement of APs might not be intended for use by positioning systems or there are some signal blind spots. A subset of access points may be present in the MS's view at a particular time. Our measurements indicate that some APs were never seen for a significant duration of time at a location, but suddenly became visible at other times because of the environmental changes, especially in multi-floor environments. Additionally, the effect of user orientation or the signal attenuation effect due to the human body can also cause a signal from certain APs to disappear at certain orientations. This intermittence of received signals can result in incomplete or censored RSSI patterns during the location detection phase of positioning systems. This clearly affects the performance of pattern classification algorithms when trying to match an incomplete RSSI pattern to a database of location fingerprints in indoor positioning systems.

For an indoor positioning system which has a large number of APs in a location fingerprint pattern, such as in [9], this kind of problem may not be noticeable. However, for a system with a small number of APs in a location fingerprint pattern, this could cause a degradation in performance of location determination algorithms. A possible solution to this problem is to carefully place all APs in such a way that every location in the positioning system can receive at least the minimum required number of signals to achieve a positioning performance goal. Our preliminary investigation in [11] indicated that the accuracy and precision of positioning systems depended on the number of APs. The research work of [44] also suggested that the performance depended on the placement of APs. Interesting results from access point placement experiments were reported in [45], where placing new APs did not always improve accuracy performance.

7. Summary of analysis results and implications to location fingerprinting

The analysis results in this study reveal characteristics of RSSI and its patterns beyond the general knowledge of the traditional wireless communications. The important characteristics and major implications to the design and the performance of indoor positioning systems are summarized as follows.

(a) The comparison of different WLAN cards confirmed that different hardware did not have the same standard in mapping a WLAN's RF signal to RSSI. The best WLAN card for indoor location fingerprinting should possess two unique properties.

First, it should have the widest range or largest quantization steps of RSSI. This is to allow better differentiation of indoor locations which are associated with different levels of RSSI. Second, it should report lowest variations or lowest fluctuations of RSSI during the measurement of the RSSI pattern. This is to reduce the confusion of RSSI patterns among neighboring locations. Both IEEE 802.11b and 802.11g WLAN cards reported similar results on the left-skewed property. The major difference is in their data rates, but not in location fingerprinting.

(b) The average RSSI is usually modeled by a log-normal distribution which is symmetric around a mean value [6], but our measurement results show that most distributions (70%) are often left-skewed. However, some distributions (10%) with a weak mean RSSI could be approximated by the log-normal distribution. Distributions of the signal from the same AP can have different shapes for different average values as shown by the skewness in Figs. 7 and 8. Signals with weak power and/or no line-of-sight often have symmetric histograms while signals with strong power and/or line-of-sight often have highly left-skewed histograms. This left-skewed effect is a result of receiver sensitivity which often has a wider range to the lower RSSI level (on the left) at a higher average value of RSSI. The implication of this discovery is that a better mathematical model or analytical model based on left-skewed distribution could provide better insight into the underlying mechanism of this type of positioning system as demonstrated by Gaussian distribution in [11]. Additionally, histogram based positioning algorithms such as in [10] could exploit the left-skewed properties instead of Gaussian distribution.

(c) The standard deviation or variance of RSSI can be different for signals from different APs within the same building. The standard deviation of signals from the same AP also varies with the location. The measurement results suggest that the signal from APs at locations with direct line-of-sight (LOS) often has a large standard deviation. On the other hand, the signal at non-line-of-sight (NLOS) locations often has smaller standard deviations. In large building environments such as in Scenario 2, the average of the standard deviation[3] is smaller than in small building environments such as in Scenario 1. We found that in Scenario 2 the average standard deviation of signals from all APs are close together, while in Scenario 1 the average standard deviations are quite different.

(d) When we consider the RSSI at a location as a random process, we found that the random process is typically non-stationary. Although the mean usually stays around the same value, the variance could shift over long periods of time such as different hours of the day. This varying of variance invalidates the assumption of stationary model of RSSI used in [10]. Changes in environment such as a human movement or furniture relocation could also change the mean RSSI. These non-stationary properties indicate the difficulty in modeling RSSI fingerprints for indoor positioning systems. Due to the time dependency property of RSSI, the location fingerprint collection process should be done at different periods of day for best results. This might provide another context to separate pattern of location fingerprints, which is that the radio map containing a database of location fingerprints should incorporate the difference of RSSI due to the change of time and environment.

(e) Signals from different APs within the range of reception can be considered as uncorrelated or independent because the correlation coefficient between any pair of signals is often small or trivial. The interference between the signals using the same radio frequency does not have any strong correlation, thus the interference by the co-channel signals may have little effect on the formation of the location fingerprint. The independence of signals is most likely a result of the collision avoidance mechanism of the medium access control protocol in 802.11. Beacons from APs using the same radio channels could be received because they did not encounter collisions.

(f) The most influential effect on the performance of location fingerprinting is the standard deviation of RSSI. It is the intrinsic performance limitation of indoor positioning systems based on location fingerprinting. The second most influential effect is the time dependency, which should be included in the design of positioning systems and when collecting the location fingerprints. The third most influential effect is the quality of the WLAN card used by the indoor positioning systems. The least influential effect on the performance is the co-channel interference or correlation among signals from multiple access points using the same radio frequency.

(g) The RSSI patterns for a given location usually cluster around one or more values near the center of patterns that have a high frequency of occurrence as shown in Fig. 22. This confirms that a vector of average values of RSSI can be used to represent a location fingerprint or a vector of means as done by the RADAR system [2]. However, the actual distance between two locations does not translate nicely into similar signal distance between two location fingerprints. Two simple approaches to improve the separation of RSSI patterns or location fingerprint are to increase the distance between two locations or to add additional access points into the pattern consideration to allow larger separation between clusters of RSSI.

(h) Based on different environments used in this work, we can identify that an ideal environment for indoor positioning systems should be an area with a large number of APs and no line-of-sight to most locations in order to provide better location determination performance. A set of recommended parameters for an indoor positioning system based on the assumptions of our work in [11] was that the standard deviation should be less than 4 dBm, the number of APs should be between three and five, and the grid spacing between two adjacent locations should be larger than 1.25 m.

[3] Although average of standard deviation should not be calculated as a summary statistics, this study does so in order to find a representative value to model the whole scenario.

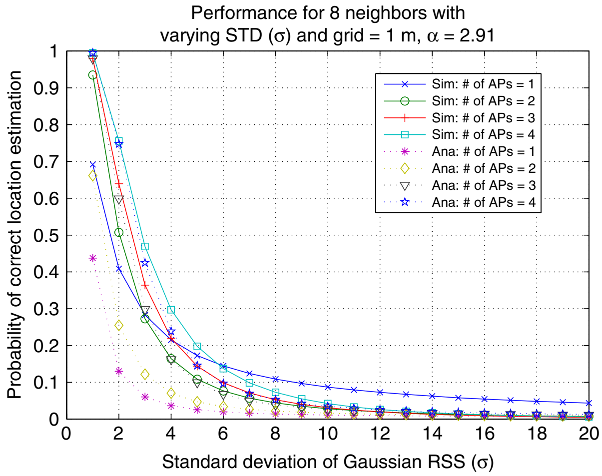

Fig. 24. Effect of RSS standard deviation on probability.

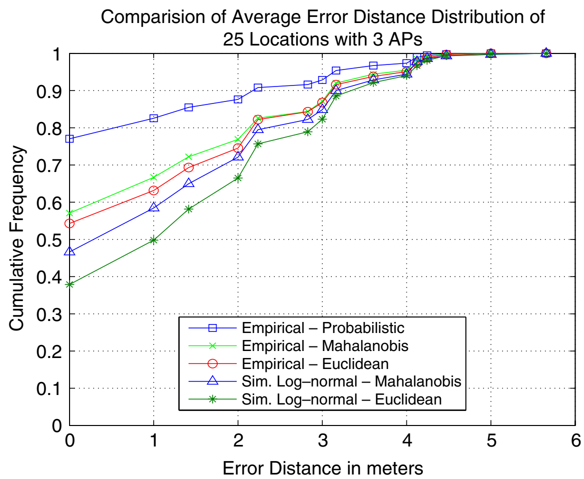

Fig. 25. Comparison of average cumulative distribution of error distance.

(i) Since the standard deviation or variance of RSSI is the most important factor which should be included when forming a location fingerprint beside the average of RSSI, a suitable pattern classification that should provide better location determination performance should include both mean and variance of RSSI into its consideration. Therefore, it is clear that a simple Euclidean distance technique will perform worse than Mahalanobis[4] distance or probabilistic based techniques. Fig. 24 shows the results from our preliminary simulation and analytical modeling of indoor positioning system presented in [11]. The effect of a higher standard deviation of RSSI at 10 dB would decrease the performance of in terms of probability of returning a correct location among the eight nearest neighbors to below 10%. Fig. 25 shows the results from another analysis in [47] where we compared the performance of three pattern classification techniques which were Euclidean, Mahalanobis, and probabilistic techniques. In that work, we confirmed that the probabilistic approach performs best compared to Mahalanobis and Euclidean based techniques. Note that we compared the performance of each technique using simulation data of log-normally distributed RSSI and real measurement data of RSSI called empirical data as labeled in the figure.

8. Conclusions

We presented a statistical data analysis of the RSSI values reported by IEEE 802.11b/g network interface cards that can be used in indoor location systems based on location fingerprinting. The major contributions of this work are the findings

[4] The Mahalanobis distance has three advantages over the Euclidean distance: automatically accounting for the scaling of the coordinate axes, correcting for correlation between different features, and enabling both non-linear and linear decision boundaries [46].

of important location-dependent mechanisms of RSSI, which are the left-skewed distribution, the standard deviation (or variance), the non-stationarity or the time dependency of RSSI, and the characteristics of the most suitable WLAN card for location fingerprinting.

The domination of left-skewed distributions among the histograms of RSSI has a major effect on how a pattern classification algorithm should decide when estimating a correct indoor location. If a closed form expression of left-skewed distribution can be employed a tractable analytical model can perhaps be developed to investigate the performance of indoor positioning systems based on location fingerprinting. On the other hand, if the left-skewed distribution is incorporated into the probabilistic-based pattern classification, the indoor positioning system should provide better location estimations. However, this argument has to be validated by a real experiment.

The errors in location estimation are proportional to the increase of the standard deviation of RSSI that we identified as the most influential parameter in indoor positioning systems. We found that the origins of RSSI fluctuations were from the intrinsic property of WLAN cards and the environment of a location. To improve the location determination performance, a designer must reduce the effect of standard deviation parameters. One possible approach is to select the best WLAN card that provides the widest range of RSSI and the most stable reception of WLAN signal. This finding is generic to the hardware of WLAN interface and the same selection is applicable for other mobile devices such as PDAs and smart phones.

To further improve the performance of the positioning system, the time dependency should be incorporated into the context of location fingerprints. That is, the location fingerprint may be changed with the time of day. However, we do recognize that this may result in more time being spent on collecting location fingerprints at different times of day, which is one of the disadvantages of this location fingerprinting technique.

Our investigation of RSSI also confirmed that RSSIs of a location have cluster-based patterns through visual inspection in which the means or the location fingerprint of RSSIs could represent the pattern. This result explains why the location fingerprinting technique based on the mean values of RSSIs alone can be used to estimate an indoor location. However, we pointed out that to achieve the best location determination performance a designer should include the standard deviation and the distribution of RSSI into consideration of pattern classification algorithms.

References

[1] M. Weiser, Some computer science issues in ubiquitous computing, Communications of the ACM 36 (7) (1993) 75–84.
[2] P. Bahl, V. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, in: Proc. INFOCOM'00: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, Tel Aviv, Israel, 2000, pp. 775–784.
[3] T. King, S. Kopf, T. Haenselmann, C. Lubberger, W. Effelsberg, COMPASS: a probabilistic indoor positioning system based on 802.11 and digital compasses, in: Proc. WiNTECH'06: First ACM International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization, Los Angeles, CA, USA, 2006, pp. 34–40.
[4] A. Papapostolou, H. Chaouchi, Orientation-based radio map extensions for improving positioning system accuracy, in: Proc. IWCMC'09: ACM International Conference on Wireless Communications and Mobile Computing, Leipzig, Germany, 2009, pp. 947–951.
[5] K. Kaemarungsi, P. Krishnamurthy, Properties of indoor received signal strength for WLAN location fingerprinting, in: Proc. MOBIQUITOUS'04: First Annual ACM International Conference on Mobile and Ubiquitous Systems: Networking and Services, Boston, MA, USA, 2004, pp. 14–23.
[6] B. Sklar, Rayleigh fading channels in mobile digital communication systems: I. characterization, IEEE Commun. Mag. 35 (7) (1997) 90–100.
[7] J. Krumm, E. Horvitz, LOCADIO: inferring motion and location from Wi-Fi signal strengths, in: Proc. MOBIQUITOUS'04: First Annual ACM International Conference on Mobile and Ubiquitous Systems: Networking and Services, Boston, MA, USA, 2004, pp. 4–13.
[8] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, J. Sievanen, A probabilistic approach to WLAN user location estimation, International Journal of Wireless Information Networks 9 (3) (2002) 155–164.
[9] A. Haeberlen, E. Flannery, A.M. Ladd, A. Rudys, D.S. Wallach, L.E. Kavraki, Practical robust localization over large-scale 802.11 wireless networks, in: Proc. MobiCom'04: Tenth Annual ACM International Conference on Mobile Computing and Networking, Philadelphia, PA, USA, 2004, pp. 70–84.
[10] M. Youssef, A. Agrawala, The HORUS location determination system, Wireless Networks 14 (3) (2008) 357–374.
[11] K. Kaemarungsi, P. Krishnamurthy, Modeling of indoor positioning systems based on location fingerprinting, in: Proc. INFOCOM'04: Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, Hong Kong, China, 2004, pp. 1012–1022.
[12] J. Small, A. Smailagic, D.P. Siewiorek, Determining user location for context aware computing through the use of a wireless LAN infrastructure, Online Dec. 2000. URL http://www-2.cs.cmu.edu/~aura/docdir/small00.pdf.
[13] U. Ahmad, A.V. Gavrilov, S. Lee, Y.-K. Lee, A modular classification model for received signal strength based location systems, Neurocomputing 71 (13–15) (2008) 2657–2669.
[14] H. Lim, L.-C. Kung, J.C. Hou, H. Luo, Zero-configuration indoor localization over IEEE 802.11 wireless infrastructure, Wireless Networks 16 (2) (2010) 405–420.
[15] H. Lemelson, T. King, W. Effelsberg, Pre-processing of fingerprints to improve the positioning accuracy of 802.11-based positioning systems, in: Proc. MELT'08: First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less environments, San Francisco, CA, USA, 2008, pp. 73–78.
[16] U. Grossmann, M. Schauch, S. Hakobyan, RSSI based WLAN indoor positioning with personal digital assistants, in: Proc. IDAACS'07: Fourth IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Dortmund, Germany, 2007, pp. 653–656.
[17] E. Elnahrawy, L. Xiaoyan, R.P. Martin, The limits of localization using signal strength: a comparative study, in: Proc. SECON'04: First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, Santa Clara, CA, USA, 2004, pp. 406–414.
[18] A.W. Tsui, Y.-H. Chuang, H.-H. Chu, Unsupervised learning for solving RSS hardware variance problem in WiFi localization, Mobile Networks and Applications 14 (5) (2009) 677–691.
[19] L. Mengual, O. Marban, S. Eibe, Clustering-based location in wireless networks, Expert Systems with Applications 37 (9) (2010) 6165–6175.
[20] C. Khauphung, P. Keeratiwintakorn, K. Kaemarungsi, On the effects of VLAN in WLAN-based positioning system, in: Proc. ECTI-CON'07, Chiang Rai, Thailand, 2007, pp. 1–4.
[21] J. Bardwell, Wildpackets' converting signal strength percentage to dBm values, White Paper Nov. 2002.
[22] Agere System Inc., Agere Product Brief: WaveLAN 802.11b chip set, Online Feb. 2003, URL http://www.agere.com/client/docs/PB03025.pdf.
[23] Vistumbler: a wireless network scanner for vista, Online Resource Aug. 2010. URL http://www.vistumbler.net.
[24] Microsoft's Windows XP Driver Development Kit, Software Development Kits 2003, URL http://www.microsoft.com/whdc/devtools/ddk/default.mspx.
[25] A. Balachandran, Wireless research API WRAPI, Online Resource Sep. 2002. URL http://ramp.ucsd.edu/pawn/wrapi/.
[26] C. Tunstall, G. Cole, Developing WMI Solutions: A Guide to Windows Management Instrumentation, Addison-Wesley Profession, Boston, MA, 2002.

[27] Microsoft's Native Wifi Windows, Online Resource Aug. 2010, URL http://msdn.microsoft.com/en-us/library/ms706556(VS.85).aspx.
[28] Microsoft's IEEE 802.11 Network Adapter Design Guideline for Windows XP, White Paper May 2003.
[29] IEEE 802.11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications 1997.
[30] The Microsoft Developer Network MSDN Library, WebPage Nov. 2004, URL http://msdn.microsoft.com/library/.
[31] M.A. Youssef, A. Agrawala, A.U. Shankar, WLAN location determination via clustering and probability distributions, in: Proc. PerCom'03: IEEE International Conference on Pervasive Computing and Communications, Dallas-Fort Worth, TX, USA, 2003, pp. 23–26.
[32] K. Pahlavan, P. Krishnamurthy, Principles of Wireless Networks: A Unified Approach, Prentice Hall PTR, Upper Saddle River, NJ, 2001.
[33] A.M. Ladd, K.E. Bekris, G. Marceau, A. Rudys, L.E. Kavraki, D.S. Wallach, Robotics-based location sensing using wireless Ethernet, in: Proc. MOBICOM'02: Eighth Annual ACM International Conference on Mobile Computing and Networking, Atlanta, GA, USA, 2002, pp. 227–238.
[34] R. Jain, The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling, John Wiley & Sons, New York, NY, 1991.
[35] B.G. Tabachnick, L.S. Fidell, Using Multivariate Statistics, third ed., Harper Collins, New York, NY, 1996.
[36] E.W. Weisstein, Map-Airy Distribution, Webpage, URL http://mathworld.wolfram.com/Map-AiryDistribution.html.
[37] K.S. Shanmugan, A.M. Breipohl, Random signals: detection, in: Estimation and Data Analysis, John Wiley & Sons, New York, NY, 1988.
[38] J.M. Gottman, Time-series Analysis: A Comprehensive Introduction for Social Scientists, Cambridge University Press, New York, NY, 1981.
[39] R.M. Grey, L.D. Davisson, An Introduction to Statistical Signal Processing, Cambridge University Press, Cambridge, UK, 2004.
[40] J. Cohen, Statistical power analysis for the behavioral sciences, in: Lawrence Erlbaum Associates, second ed., Hillsdale, NJ, 1988.
[41] M.A. Youssef, HORUS: a WLAN-based indoor location determination system, Ph.D. thesis, University of Maryland, College Park, MD 2004.
[42] Z. Xiang, S. Song, J. Chen, H. Wang, J. Huang, X. Gao, A wireless LAN-based indoor positioning technology, IBM Journal of Research and Development 48 (5/6) (2004) 617–626.
[43] J. Wierenga, P. Komisarczuk, SIMPLE: developing a LBS positioning solution, in: Proc. MUM'05: Fourth ACM International Conference on Mobile and Ubiquitous Multimedia, Christchurch, New Zealand, 2005, pp. 48–55.
[44] Y. Chen, H. Kobayashi, Signal strength based indoor geolocation, in: Proc. ICC'02 International Conference on Communications, New York, NY, 2002, pp. 436–439.
[45] O. Baala, Y. Zheng, A. Caminada, Toward environment indicators to evaluate WLAN-based indoor positioning system, in: Proc. AICCSA'09: IEEE/ACS International Conference on Computer Systems and Applications, Rabat, Morocco, 2009, pp. 243–250.
[46] J.T. Tou, R.C. Gonzalez, Pattern Recognition Principles, second ed., Addison-Wesley, Reading, MA, 1974.
[47] K. Kaemarungsi, Efficient design of indoor positioning systems based on location fingerprinting, in: Proc. IEEE International Conference on Wireless Networks, Communications and Mobile Computing, Maui, HI, USA, 2005, pp. 181–186.