AN INDOOR WIRELESS LAN LOCATION DETERMINATION SYSTEM

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information

Lanlan Song

Certificate of Approval:

Chungwei Lee Assistant Professor Computer Science and Software Engineering	Yu Wang, Chair Assistant Professor Computer Science and Software Engineering
Min-Te Sun Assistant Professor Computer Science and Software Engineering	Stephen L. McFarland Acting Dean Graduate School

AN INDOOR WIRELESS LAN LOCATION DETERMINATION SYSTEM

Lanlan Song

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

In Partial Fulfillment of the

Requirements of the

Degree of

Master of Science

Auburn, Alabama
December 16, 2005

AN INDOOR WIRELESS LAN LOCATION DETERMINATION SYSTEM

Lanlan Song

Signature of Author

Date of Graduation

THESIS ABSTRACT

AN INDOOR WIRELESS LAN LOCATION DETERMINATION SYSTEM

Lanlan Song

Master of Science, December 16, 2005
(B.S., Sichuan University, June 1999)

67 Typed Pages

Directed by Dr. Yu Wang

As pervasive computing becomes more popular, the need for context-aware applications increases. The context of an application refers to the information that is part of its operating environment, in which the physical location of the user is an important component. While traditional Global Positioning System (GPS) are mainly used to detect the location information in outdoor environment, precise, easy-to-build, and easy-to-use indoor location determination systems are still needed. So we focus our efforts on developing an Indoor Location Determination System.

In this thesis, a RF-based Indoor Location Determination System implemented in the context of IEEE 802.11b wireless LAN (WLAN) is presented. The system structure and implementation procedure are described. Five algorithms on determining the mobile client's location in the system are raised in this thesis and they are implemented. Real-world experiments are conducted to evaluate the reliability and performance of these algorithms under different parameter settings. The experiments show that the forth and fifth algorithm can achieve satisfying results for this system.

ACKNOWLEDGEMENTS

The author would like to express her deep gratitude to her advisor, Dr. Yu Wang, for her patient guidance, valuable advices, and encouragement throughout her studies. Sincere thanks are also due to her two graduate committee members, Dr. Chungwei Lee and Dr. Min-Te Sun, for their reviewing and advising efforts. Thanks also go to Adam Harder for his prior work and help during the course of this investigation process. The nice floor plan figure of the test bed in this thesis is generated by the tool-kit he developed. Thank Jeff Graves for helping create an Oracle account for this thesis research work. In addition, the author would like to thank her parents, Shufang Wang and Haichun Song, and her husband, Jie Che, for their consistent support.

Style manual or journal used

Computer software used Microsoft Word and Excel

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

As pervasive computing becomes more popular, the need for context-aware applications increases. The context of an application refers to the information that is part of its operating environment, in which the physical location of users is an important component. So we focus our efforts on developing an Indoor Location Determination System.

It is possible to obtain the physical location of a mobile station (MS) in two ways: by using a special infrastructure for positioning such as the global positioning system (GPS), or by enhancing the existing communication infrastructure to determine the location of users. GPS is not suitable for indoor areas because of the lack of coverage. And it is very expensive in terms of labor, spectrum and capital costs to implement a specialized infrastructure in indoor areas solely for position location. As such, it is preferable to employ the existing wireless communications infrastructure to determine the location of users within the network. [5]

In indoor areas, the wireless communications infrastructure is primarily based on wireless local area networks (WLANs), in particular the IEEE 802.11b standard that supports raw data rates of 11 Mbps [5]. Thus we choose to build a RF-based location determination system implemented in the context of IEEE 802.11b wireless LAN (WLAN).

Our system uses the signal strength observed from frames transmitted by the access points (APs) to infer the user location. Since the wireless cards measure the signal strength information of the received frames as part of their normal operation, this makes the system a software solution on top of the wireless network infrastructure.

There are two classes of WLAN location determination systems: client-based and infrastructure-based. Both have their own set of applications. Our system is currently implemented as a client-based system. A large class of applications, including location-sensitive content delivery and direction finding, can be built on top of the system.

WLAN location determination is an active research area. WLAN location determination systems usually work in two phases: a prerequisite off-line training phase to build up the system (hardware and database), and then the on-line location determination phase (real operation phase). For the client-based system, during the off-line phase, the system records the signal strength of frames received from the access points by the mobile host at selected training locations in the area, resulting in a database called radio map. During the on-line location determination phase, the system uses the signal strength vectors received from the access points to “search” the radio map to estimate the user location.

Five algorithms are presented and evaluated by experiments.

The remainder of this paper is organized as follows. Chapter2 describes the related work. In Chapter 3, research problems are discussed. The approach to build up the system is explained in Chapter 4 and five algorithms are presented in Chapter 5. Chapter 6 describes the implementation and experiments. Finally, Chapter 7 concludes the thesis and discusses the future work.

2. RELATED WORK

Many systems have tackled the problem of determining and tracking user location. Examples include the Global Positioning System (GPS), wide-area cellular-based systems, infrared-based systems, and magnetic tracking systems, various computer vision systems, physical contact systems, and radio frequency (RF) based systems [2]. Among these, the class of RF-based systems that use an underlying wireless data network [3], such as IEEE 802.11 wireless network, to estimate user location has gained attention, especially for indoor application. Unlike infrared-based systems, which are limited in range, RF-based techniques provide more ubiquitous coverage and do not require additional hardware for user location determination, thereby enhancing the value of the wireless data network.

RF-based techniques are purely software technique. They can be categorized into model based techniques and radio-map based techniques.

Model based techniques, use empirical models to capture the relation between the signal strength and distance. These models can capture the internal structure of the building, e.g. the number of walls between the access point and the receiver. However, they suffer from the noisy characteristic of the indoor wireless channel which makes the relation between the signal strength and distance a complex function that is difficult to capture using a general model.

To solve this problem, radio-map based techniques try to capture the signal strength received from the access points at selected locations in the area of interest. A radio map is a database about the radio signal strength information of different access points at each training location point. They work in two phases:

1. off-line phase, in which the signal strength received from the access points at each location is stored in some form in the radio map (database), and

2. on-line phase, in which the received signal strength vector, one entry for each access point, is compared to the stored radio map and the nearest match is returned as the estimated user location.

The representative is “The RADAR: An in-building RF-based user location and tracking system” [3, 4]. It uses the access points to record the signal strength from the mobile host. The mobile host broadcasts UDP packets and each access point used in the system records the Signal Strength (SS) measurement together with a synchronized timestamp t, i.e., it records tuples of the form (t, bs, ss) during the off-line phase and on-line phase. The system stores in the radio-map the average signal strength recorded by each base station for each training location during the off-line phase. During the on-line phase, the signal strength (ss) measurement is sent to a central server and the central server finds the location from the radio map that best fits the collected signal strength average information.

The TMI system proposed in [6] is based on triangulation, mapping and interpolation (TMI). In the TMI technique, the physical position of all the access points in the area needs to be known and a function is required to map signal strength onto distances.

The systems proposed in [33, 34] use a neural network approach to estimate the user location.

Probability theory can be used to infer the user location. For example, the Nibble location system [21] from UCLA uses a Bayesian network to infer a user location. Their Bayesian network model includes nodes for location, noise, and access points (sensors). The signal to noise ratio (SNR) observed from an access point at a given location is taken as an indication of that location. The system also quantizes the SNR into four levels: high, medium, low, and none.

The system in this thesis is implemented as a radio-map base system. The first algorithm in this thesis is similar to the RADAR system, but we use the client-based method to improve the privacy of the mobile host and reduce the continuous computational burden on the network. The second algorithm is based on the first algorithm with some change. Based on observation, the third, forth and fifth algorithms use the probability theory on the distribution of the signal strength.

3. RESEARCH PROBLEM STATEMENT

This chapter discusses Radio Frequency (RF) signal propagation in wireless LAN in 3.1, and describes the difference between the client- based or server-based method in 3.2. In 3.3, sampling experiments, a tool is chosen to sample the signal strength, and experiments are conducted to investigate the variation of signal strength as a function of time or physical position. They are the base of our approach.

3.1 Radio Frequency Signal Propagation in Wireless LAN

The IEEE 802.11b standard uses radio frequencies in the2.4GHz band, which is license-free around the world. The available adapters are based on spread spectrum radio technology. Accurate prediction of signal strength is a complex and difficult task. Due to reflection, refraction, scattering and absorption of radio waves by structures inside a building, the transmitted signal most often reaches the receiver by more than one path, resulting in a phenomenon known as multipath fading. Multipath causes fluctuations in the received signal envelope and phase, and the signal components arriving from indirect paths and the direct path, if this exists, combine to produce a distorted version of the transmitted signal. Multipath within buildings is strongly influenced by the layout of the building, the construction material used, and people in the building. [25]

As the number of people in the building varies, the propagation characteristics of RF signals change as well. This is because the human body is made up of water and water absorbs RF signals, and consequently people absorb signal. [25]

Although efforts have been made to model radio signal distribution in an indoor environment, experiments in different environments have arrived at different distributions and a general model remains unavailable.

The received signal varies according to time and the relative position of the transmitter and receiver. So the first step of our work is to test the variation of the signal strength, which will be discussed in 3.3.

3.2 Client- based or Server-based Method

There are two distinct approaches for recording signal strength: centralized client-based (client-centric, self-positioning) approach and infrastructure-based (server-based, server-centric) approach

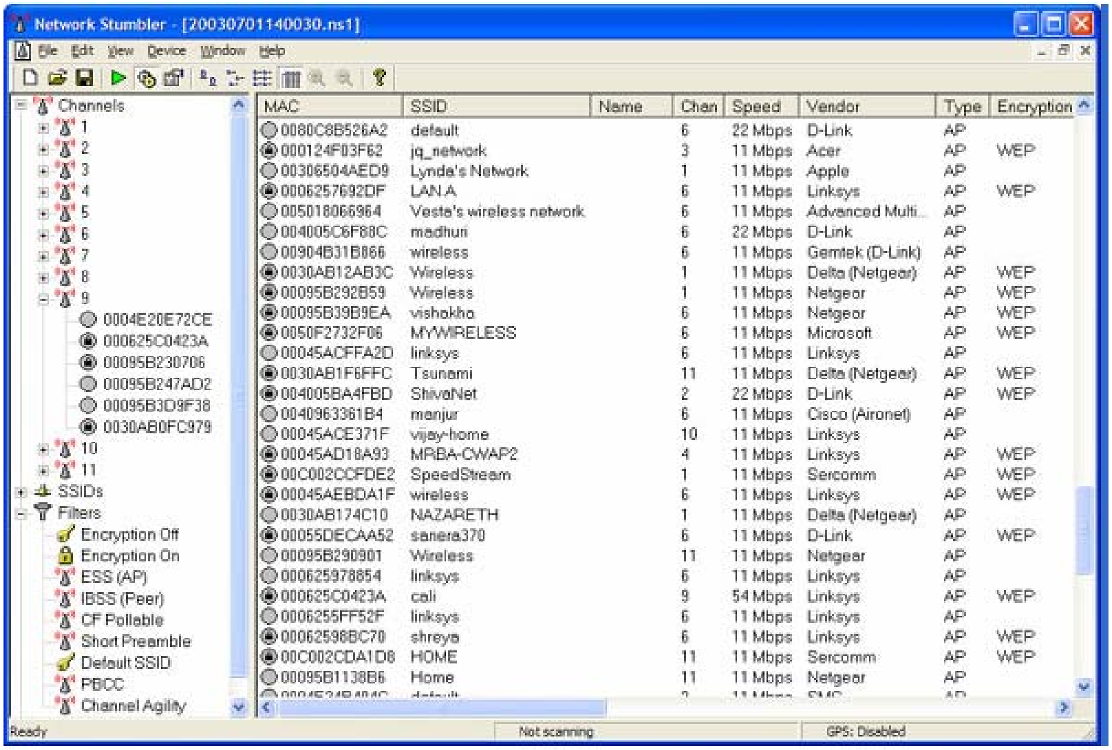
The centralized client-based approach means the mobile client is the active role to record signal strength received from access points. The client will send the collected information to the server when it requests location information and the central server will use certain algorithm to determine the corresponding location. Infrastructure-based (server-based) approach is that access points record the signal strength received from the mobile client and send it to the server, and the central server use certain algorithm determines and records the client’s location.

We choose the centralized client-based approach for our system. This will improve the privacy of the mobile station and reduce the continuous computational burden on the network.

3.3 Sampling Experiments

The signal strength data collection (sampling) phase is a key step in our research methodology. After investigation and experiments, we choose to install NetStumbler version 0.4.0 on the mobile host (laptop) to achieve this task.

NetStumbler [1] is a tool working on the Windows operating system, which allows you to detect Wireless Local Area Networks (WLANs) using 802.11b, 802.11a and 802.11g. The user interface is like below:



MAC	SSID	Name	Chan	Speed	Vendor	Type	Encryption
○ 0080C8B526A2	default		6	22 Mbps	D-Link	AP	
● 000124F03F62	jq_network		3	11 Mbps	Acer	AP	WEP
○ 00306504AED9	Lynda's Network		1	11 Mbps	Apple	AP	
● 0006257692DF	LAN A		6	11 Mbps	Linksys	AP	WEP
○ 005018066964	Veste's wireless network		6	11 Mbps	Advanced Multi...	AP	
○ 004005C6F88C	madhuri		6	22 Mbps	D-Link	AP	
○ 00904B31B866	wireless		6	11 Mbps	Gemtek (D-Link)	AP	
● 0030AB12AB3C	Wireless		1	11 Mbps	Delta (Netgear)	AP	WEP
● 00095B292B59	Wireless		1	11 Mbps	Netgear	AP	WEP
● 00095B39B9EA	vishakha		6	11 Mbps	Netgear	AP	WEP
● 0050F2732F06	MYWIRELESS		6	11 Mbps	Microsoft	AP	WEP
○ 00045ACFFA2D	linksys		6	11 Mbps	Linksys	AP	
● 0030AB1F6FFC	Tsunami		11	11 Mbps	Delta (Netgear)	AP	WEP
● 004005BA4FBD	ShivaNet		2	22 Mbps	D-Link	AP	WEP
○ 0040963361B4	manjur		6	11 Mbps	Cisco (Aironet)	AP	
○ 00045ACE371F	vijay-home		10	11 Mbps	Linksys	AP	
● 00045AD18A93	MRBA-CWAP2		4	11 Mbps	Linksys	AP	WEP
● 00C002CCFDE2	SpeedStream		1	11 Mbps	Sercomm	AP	WEP
● 00045AEBDA1F	wireless		6	11 Mbps	Linksys	AP	WEP
○ 0030AB174C10	NAZARETH		1	11 Mbps	Delta (Netgear)	AP	
● 00055DECAA52	sanera370		6	11 Mbps	D-Link	AP	WEP
○ 00095B290901	Wireless		11	11 Mbps	Netgear	AP	
○ 000625978854	linksys		6	11 Mbps	Linksys	AP	
● 000625C0423A	cali		9	54 Mbps	Linksys	AP	WEP
○ 0006255FF52F	linksys		6	11 Mbps	Linksys	AP	
● 00062598BC70	shreya		6	11 Mbps	Linksys	AP	WEP
● 00C002CDA1D8	HOME		11	11 Mbps	Sercomm	AP	WEP
○ 00095B1138B6	Home		11	11 Mbps	Netgear	AP	

Figure 1: NetStumbler User Interface

NetStumbler sends out a probe request from the mobile host, and reports the signal strength of responses from the detected access points. This is known as Active Scanning. Scan Speed can be adjust to control how frequently NetStumbler sends out probe requests for sampling. The time interval can be varied from 1.5 seconds (slow) to 0.5 seconds (fast). We use the default setting: once a second.

NetStumbler makes available the signal strength (SS). SS is reported in units of dBm. A signal strength of s Watts is equivalent to 10*log10(s/0.001) dBm.

Using NetStumbler, we could record the variation of signal strength as a function of time or physical position.

Figure 2 and 3 showing the variation of signal strength as a function of time. The signal strength is recorded by a laptop at a fixed position for about 24 hours, which generate about 86000 samples for each of the two access points detected. But excel can only show 32000 points (samples of 10 hours) at most in two dimensional figures. So this figure shows the variation of signal strength in 10 hours. We can see from the figure 2 about the access point named as “linksys”, which is in the same room with the laptop, that the signal strength is waved near 120. From the figure 3 of the access point named as “Apple”, which is in neighbor and father from the laptop than “linksys”, we can see that the signal strength is waved near 87. It can be seen from the figures that at the same timestamp, the signal strength received from access points, which are of different distances from the laptop, are different and they have their own “reliable” value range.

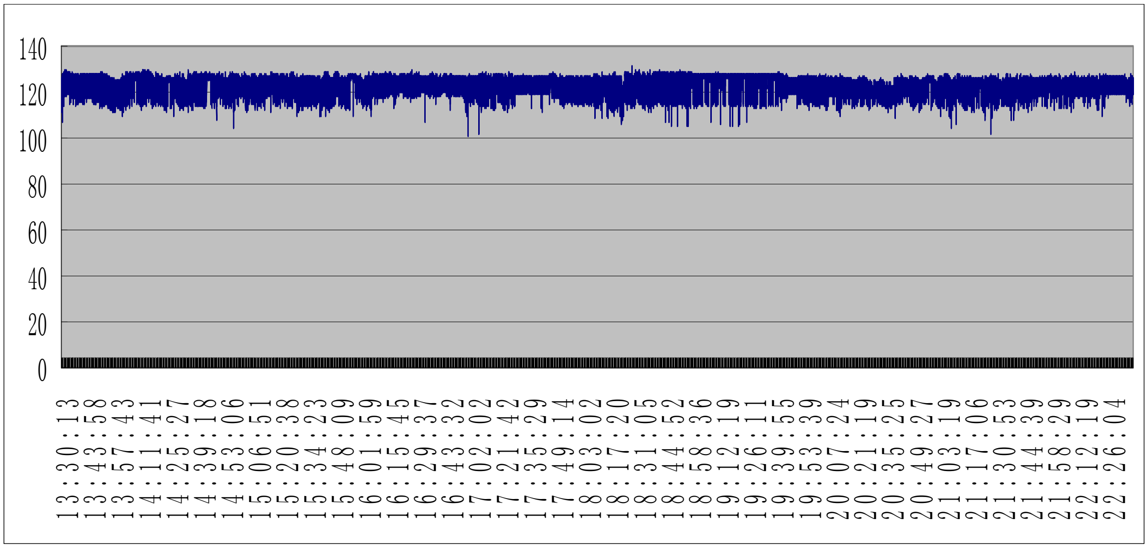


Figure 2: Signal strength received form the access point named "linksys" over time

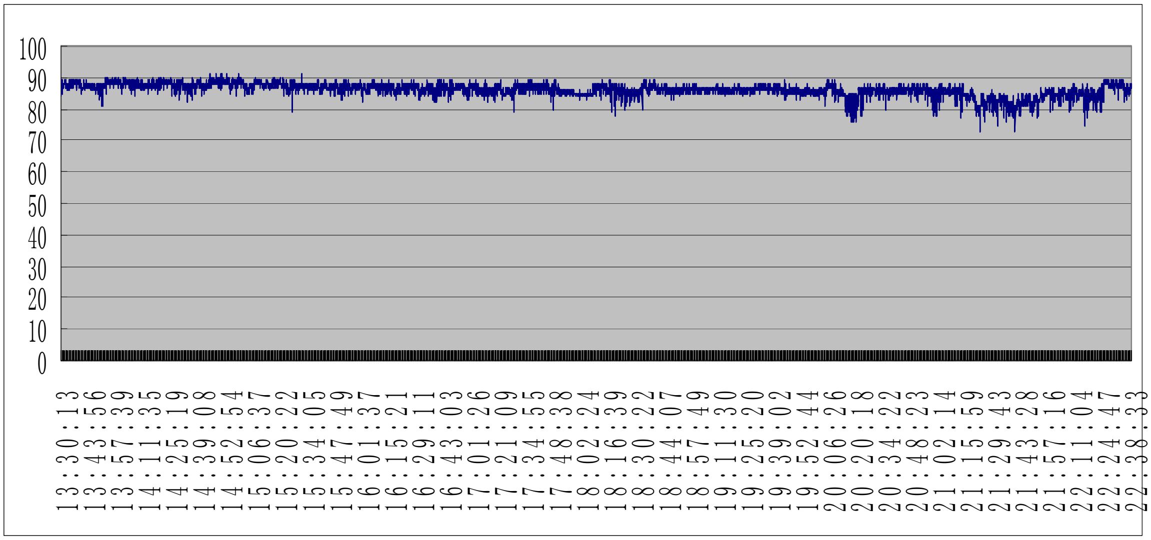


Figure 3: Signal strength received form the access point named "Apple" over time

Below are two figures showing the average signal strength as a function of the distance between the mobile host and the access point. There are five series of signal strength received from five different access points whose names are listed in the figure.

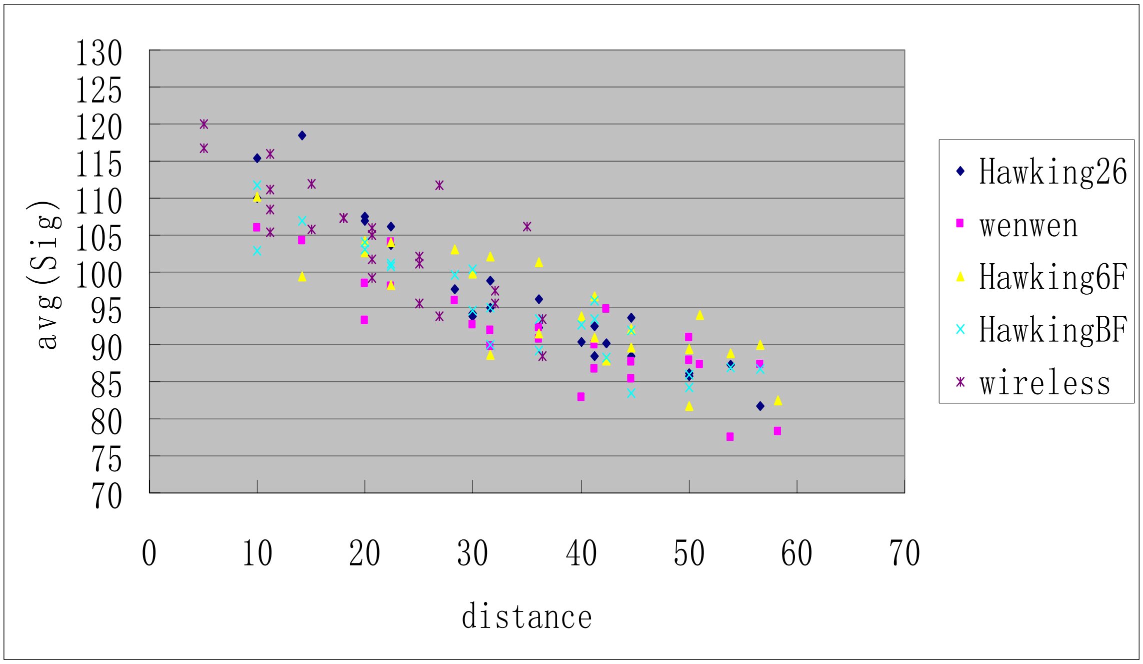


Figure 4: Average signal strength of five access points at 24 locations

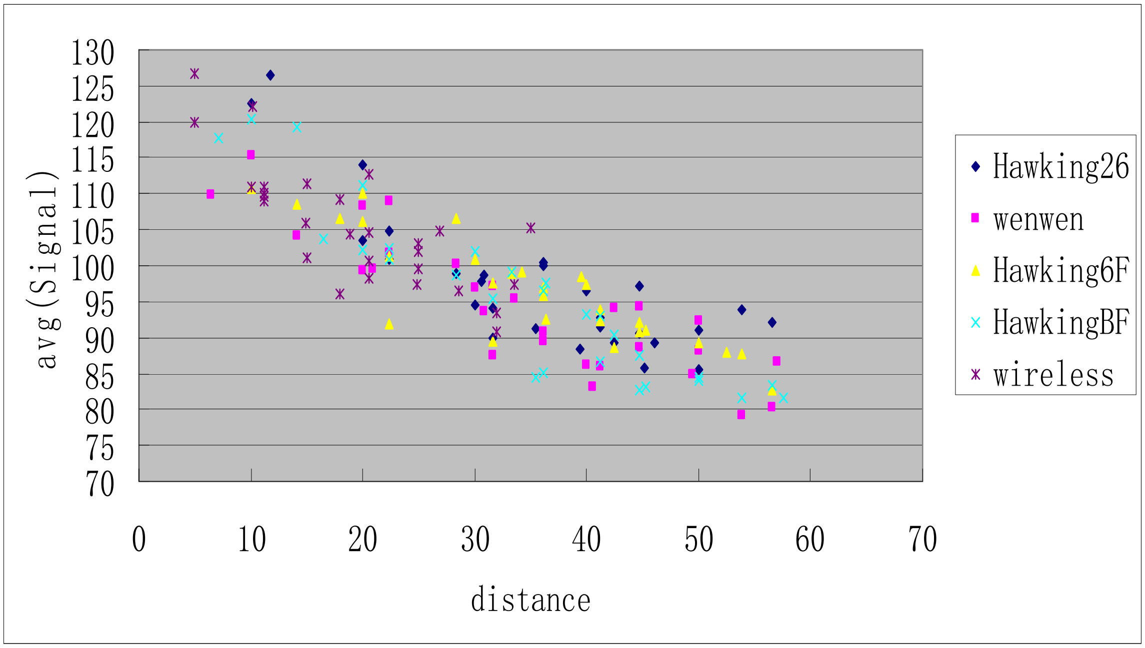


Figure 5: Average signal strength of five access points at 28 locations

Figure 4 shows average signal strength recorded by a laptop at 24 locations in the test bed described in section 5.1. At each location, it recorded for about 5 minutes.

Figure 5 shows average signal strength recorded by another laptop at 28 locations in the test bed described in section 5.1. At each location, it recorded for about 2 minutes.

We can see from these two figures that generally as the distance between the laptop and the access point increases, the received signal strength value at the laptop becomes smaller. But there may be some exceptions due to the complex environment impact. So we could infer that signal strength can be used to determine the mobile host's location, since at different location, there'll generally be a different signal strength value. But because of the environment variation, if we only use one signal strength series, there may be some wrong location determination result at some point. If we use several signal strength series together to infer the mobile host's location, the accuracy should be enhanced.

4. APPROACH

We use a radio-frequency based location determination approach. Our implementation is based on signal strength and access point information from the IEEE 802.11b wireless network. The fundamental idea is that in the wireless network, the signal strength or its distribution is a function of the mobile user's location.

We use a centralized client-based (client-centric, self-positioning) approach, which includes the off-line (training) phase and the on-line (operation) phase. In the prerequisite training phase, the system creates a radio map of the indoor area by at each training location designed, using the mobile host (laptop) to collect signal strength of the access points. These access points are set up at fixed location by us and can provide overlapping coverage in the area. They are called training access points in our system. Other access points detected by the laptop will not be used, because they are not the training access points of our system. A radio map is a database about the access points and their corresponding radio signal strength information at each training location, and it's stored on the central server. The detailed database structure will be determined by the specific algorithm used. In the real-time operation phase, the system operates as: when the mobile client requests location information, it will record signal strength using NetStumbler, and send the information to the central server. The server will select and process the signal strength of those training access points, and use certain algorithm to determine the client’s location and return the result to the client.

5 ALGORITHMS

Five different location determine algorithms as described in this chapter, will be used in this system. The first and the second algorithm use the average signal strength from each access point at each location during the training phase and the operation phase. The third, forth and fifth algorithm use the signal strength distribution of each access point at each location. They will be evaluated in section 6.3.

5.1 Algorithm 1: Finding the Best Match of the Averages

In the training phase, at each training location, the mobile host collects the signal strength information for certain time period and the average signal strength of each training access point will be calculated. So at each location, for each timestamp (each second) and each access point, there's a signal strength value collected. For each location and each timestamp, there's a sample vector of such signal strength value corresponding to each access point. Then for each location, there're a set of such sample vector corresponding to each timestamp. And for each location, there exists another vector of the average signal strength corresponding to each access point. All these information will be stored in the database (the radio map).

In the operation phase, at certain location, in order to request for location information, the mobile host will collect signal strength information for certain time period required by the system, and send the set of vectors like (location, timestamp, access point, signal strength), to the server. The server will then select the vectors of the signal strength of the training access points from it and calculate average signal strength for each training access point. Then a linear search is conducted in the database table. For each training location point, the average signal strength of each training point is found, and then a signal distance between the average signal strength of the client location and the stored training average will be calculated using Euclidean distance measure: sqrt ($\sum (SS_i - SS_i')^2$), in which SS_i is the i-th access point's average signal strength sending by the client, SS_i' is the corresponding access point's average signal strength in the database. The training location with the minimum signal distance will be the result returned to the client.

Different number of access points with the strongest signal strength will be used in the on-line phase to evaluate the algorithm's performance as described in section 6.3.1.

5.2 Algorithm 2: Finding the Mid-Point of the Two Best Match of the Averages

This algorithm is similar to Algorithm 1. The difference is: in the on-line phase, when the mobile host collects signal strength information for certain time period and sends it to the server to request for location information, the server will find a location with the minimum signal distance and another location with the second minimum signal distance. Then the midpoint of these two locations will be the result returned to the client.

Different number of access points with the strongest signal strength will also be used in the on-line phase to evaluate the algorithm's performance as described in section 6.3.2.

5.3 Algorithm 3: Signal Distribution Best Match I

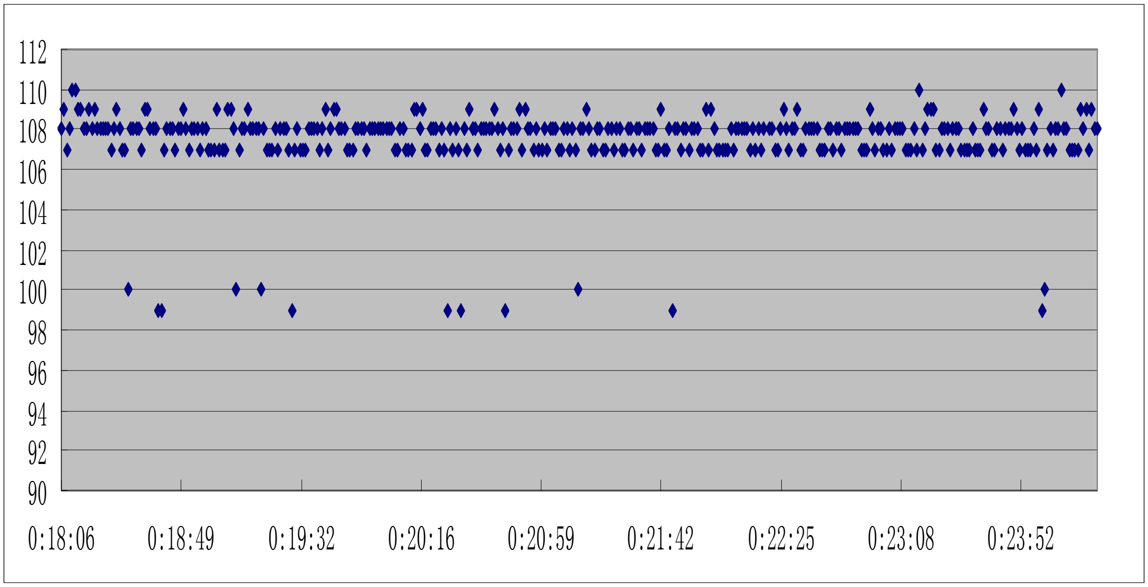


Figure 6: Signal strength of the access point named as “Hawking26” at location 0C over six minutes

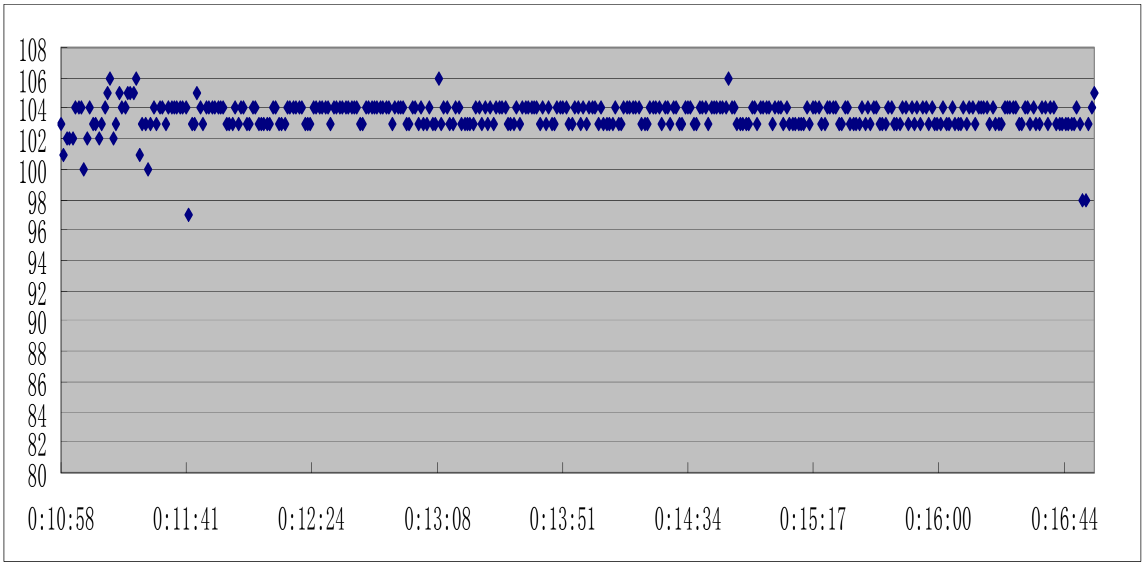


Figure 7: Signal strength of the access point named as “Hawking26” at location 1C over six minutes

The signal strength collected by NetStumbler is shown in a discrete integer value space. From Figure 6, Figure 7 and many experiment observations, it can be found that the signal strength from a specific access point at a specific location is in a specific narrow value range, and certain signal strength value may appear much more times than other strength values for this access point. For “Hawking26” access point, we can see in Figure 6 and 7, at location 0C (described in section 6.1. Experimental Test Bed), the signal strength value of 108 and 107 appear most; and at location 1C (described in section 6.1. Experimental Test Bed), the signal strength value of 104 and 103 appear most. Location 0C is nearer to “Hawking26” access point than location 1C, and most signal strength values of “Hawking26” at Location 0C are larger than at location 1C.

Therefore, the signal strength distribution at each training location (the appearing probability of each distinct signal strength value of each access point at each training location) seems reasonable to be used in an algorithm. Algorithm 3, 4 and 5 are based on this.

For Algorithm 3, in the training phase, at each training location, the mobile host collects signal strength information for certain time period. The probability of each appearing distinct signal strength value of each training access point at this location is calculated by: dividing the appearing times of the distinct signal strength value of this access point at this location by the total number of signal strength samples of the access point at this location. All the information, (location, access point, signal strength, probability) sets, are stored in the database (the radio map).

As described before, at each location, for each timestamp (each second) and each access point, there’s a signal strength value. For each location and each timestamp, there’s a sample vector of such signal strength value corresponding to each access point. Then for each location, there're a set of such sample vector corresponding to each timestamp. So in the on-line operation phase, at a certain location, the mobile client collects signal strength information for certain time period as required by the system, and sends it to the server to request for location information. The server will select the signal strength information of the training access points from it, and at each training location recorded in the database, looks for the probability of the signal strength value corresponding to each access point sent by the client, then multiplies the probability of each signal strength corresponding to each access point together to get a probability result for the sample vector at each timestamp (which means the signal strength of each access point in this sample vector are relative to each other), then the probability result of each sample vector are multiplied together to get the final probability result for the set of sample vectors at the operation location (which means the sample vectors in the set are relative to each other). The training location getting the maximum probability result on the operation sample vectors sent by the client will be the location result returned to the client.

Different number of signal strength sample vectors (different sample collecting time period) in the operation phase, eg. 1, 3, 5 or 10 sample vectors (or 1, 3, 5 or 10 seconds), will be used to evaluate the performance. Further discussion of this algorithm will be in section 6.3.3.

5.4 Algorithm 4: Signal Distribution Best Match II

This algorithm is similar to Algorithm 3. The difference is at the final step in the operation phase, the probability result of each sample vector is added together to get the final probability result for the operation location (which means the sample vectors in the set are independent to each other). Then the same, the training location getting the maximum probability result on the operation sample vectors sent by the client will be the location result returned to the client.

Different number of signal strength sample vectors (different sample collecting time period) in the operation phase, eg. 1, 3, 5 or 10 sample vectors (or 1, 3, 5 or 10 seconds), will also be used to evaluate the performance. Further discussion of this algorithm will be in section 6.3.4.

5.5 Algorithm 5: Signal Distribution Best Match III

This algorithm is also similar to Algorithm 3 and 4. The difference is in the operation phase, at each training location recorded in the database: after the server looks for the probability of each signal strength value of the corresponding access point sent by the client, the probabilities are added together to get a probability result for each sample vector (which means the signal strength of each access point in this sample vector are independent to each other), then the result of each sample vector are added together to get the final probability result for the training location (which means the sample vectors in the set are relative to each other). Then it’s the same that the training location with the maximum probability result on the operation sample vectors will be the result returned to the client.

Different number of signal strength sample vectors (different sample collecting time period) in the operation phase, eg. 1, 3, 5 or 10 sample vectors (or 1, 3, 5 or 10 seconds), will also be used to evaluate the performance. Further discussion of this algorithm will be in section 6.3.5.

6. EXPERIMENTS

Experimental test beds, experimental procedure, the implementation of the system, and experimental results of the five algorithms are presented in this chapter.

6.1 Experimental Test Beds

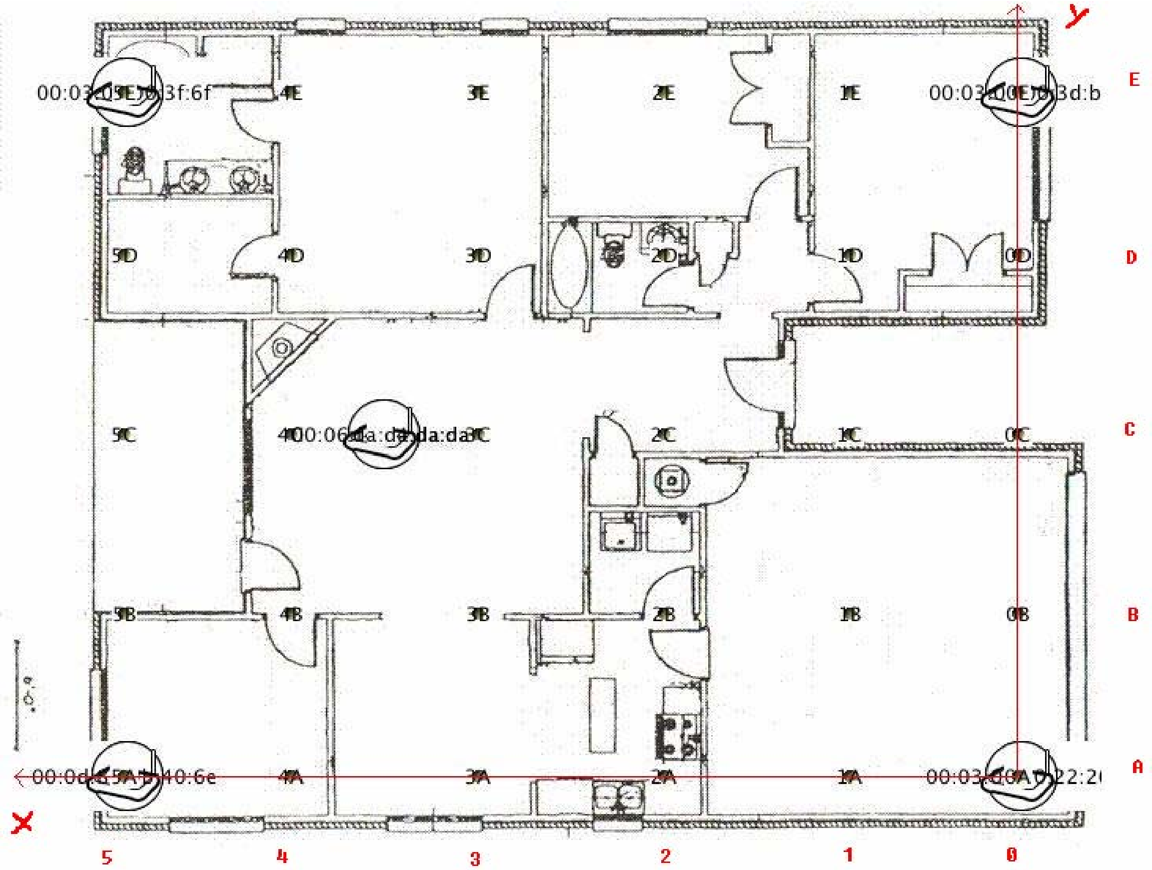


Figure 8: Test floor plan with coordinates

Our experimental test bed is a one-floor house of a rectangular area with the dimension of 55 by 45 feet. The training locations for the radio map (database) are set on a grid of points evenly spread with the spacing of 10 feet in the area. Four access points are set up at the four corners and the fifth is set in the middle of the house. The floor plan of the test bed is shown in Figure 8. A coordinate system was set with the original point at the lower-right grid point, and each grid point was given a coordinate name, from “0A” to “5E”. We can divide the area into 14 spaces including 12 rooms, a porch and a space of outside of the door as shown in Table 1. The coordinate unit is feet.

We performed two experiments in the test bed. In the first one, we used three laptops as mobile hosts. In the second one which was taken in another month, another laptop was used.

PointName	X	Y	Room/AP
0A	0	0	Hawking26
1A	10	0	Garage
2A	20	0	Kitchen
3A	30	0	Dining
4A	40	0	Study
5A	50	0	wenwen
0B	0	10	Garage
1B	10	10	Garage
2B	20	10	Kitchen
3B	30	10	Dining
4B	40	10	Study
5B	50	10	N
0C	0	20	OutDoor
1C	10	20	OutDoor
2C	20	20	Foyer
3C	30	20	Living
4C	40	20	Living
5C	50	20	Porch
0D	0	30	Bedroom3
1D	10	30	Bedroom3
2D	20	30	Bath2
3D	30	30	Bedroom1
4D	40	30	Bedroom1
5D	50	30	N
0E	0	40	HawkingBF
1E	10	40	Bedroom3
2E	20	40	Bedroom2
3E	30	40	Bedroom1
4E	40	40	Bedroom1
5E	50	40	Hawking6F
01DE	5	35	Bedroom3
1AB	10	6	Garage
12DE	16	36	Bedroom2
23BC	25	18	Living
3AB	30	6	Dining
34D	35	30	Bedroom1
34C	35	20	wireless
45AB	45	4	Study

Table 1: Location information

6.2 Experimental Procedure

Test bed is the environment of the location determination system. Access Points and the mobile host (laptop with Windows operating system) are the hardware components of this system. NetStumbler software, programs I wrote using Perl programming language and the location determination Oracle database I designed are the software components of this location determination system. The experiment procedure of the system is described in this section.

6.2.1 Experiment Test Bed and System Setting-Up

The first step to do the experiment is to set up the test bed and the hardware, which includes measuring the grid and mark training locations, setting up the access points at the designed locations to provide an overlapping coverage of this test area, and having the mobile host (the laptop) accessing to the network by one access point.

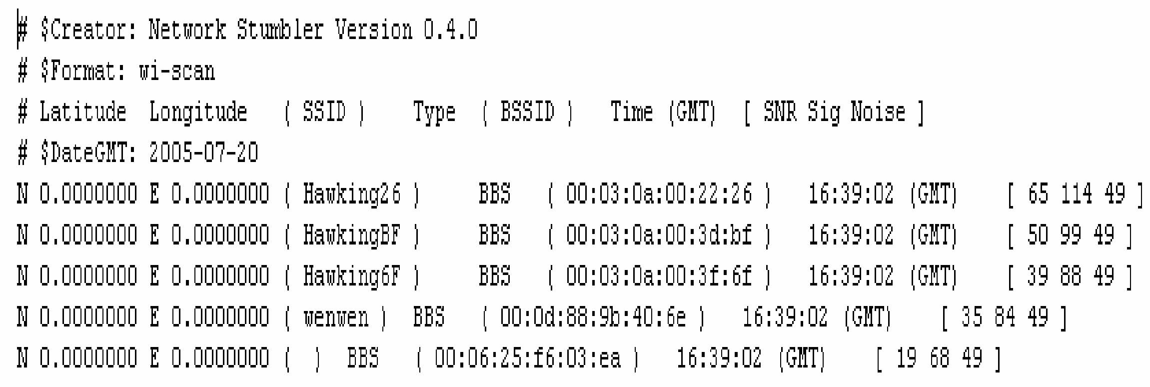
During the experiment, we found at different locations, except that all the five access points we set up in the house can be detected, there were different numbers of other access points of the neighbor detected. But we don't use signal strength information from those access points because only the five access points were under our control and they can provide a reliable and full coverage of the test bed with stronger signal strength.

6.2.2 Training Sampling

We performed this step twice. In the first sampling experiment, we used three laptops as mobile hosts, and took samples for about 5 minutes at each training location marked.

There're 24 points we used. In the second sampling experiment which was taken in another month, another laptop was used to take samples for about 2 minutes at each location. There're 28 points used in the second sampling collection process, 21 grid points the same as in the first experiment and another 7 were not on the grid.

We use NetStumbler tool on the laptops to collect the signal strength samples information and generate wi-scan file with the necessary signal strength information like the example in Figure 9



```
# $Creator: Network Stumbler Version 0.4.0
# $Format: wi-scan
# Latitude  Longitude   ( SSID )    Type  ( BSSID )   Time (GMT)  [ SNR Sig Noise ]
# $DateGMT: 2005-07-20
N 0.0000000 E 0.0000000 ( Hawking26 )     BBS   ( 00:03:0a:00:22:26 )   16:39:02 (GMT)    [ 65 114 49 ]
N 0.0000000 E 0.0000000 ( HawkingBF )     BBS   ( 00:03:0a:00:3d:bf )   16:39:02 (GMT)    [ 50 99 49 ]
N 0.0000000 E 0.0000000 ( Hawking6F )     BBS   ( 00:03:0a:00:3f:6f )   16:39:02 (GMT)    [ 39 88 49 ]
N 0.0000000 E 0.0000000 ( wenwen )  BBS   ( 00:0d:88:9b:40:6e )   16:39:02 (GMT)    [ 35 84 49 ]
N 0.0000000 E 0.0000000 (   )  BBS   ( 00:06:25:f6:03:ea )   16:39:02 (GMT)    [ 19 68 49 ]
```

Figure 9: NetStumbler wi-scan file example

6.2.3 Preprocess Training Samples for Analysis

The step is for signal strength analysis and is not used in the system.

Shell script and perl programming language are used. build.sh script will call preprocess.pl to select signal strength and time information of each training access point from the original wi-scan file and put the related information into a file for analysis.

6.2.4 Database Design and Building

Four steps are performed in this process.

6.2.4.1 Database initialization

Perl program init_db.pl will use the original wi-scan files (24 files corresponding to 24 training locations in the first sampling experiment and 28 files corresponding to 28 locations in the second sampling experiment) to generate init.sql file which contains all sql statements to create tables and insert data into them to build up the location determination database. Then init.sql file can be run in Oracle database to execute these sql statements.

There are 24 tables named as the training locations' names in the first sampling experiment and another 28 tables named as the training locations' names in the second sampling experiment. They have 3 columns: AP name, Timestamp and Signal Strength. These tables store signal strength information of each training access point at each sampling timestamp and at each of the training location.

Another AP_MAC table was generated to hold the AP name and MAC address information.

6.2.4.2 Querying the Initialized Database to Generate Summary Table

In this step, perl program summary.pl will access the initialized database built in last step to calculate the average, max, min and the standard deviation of signal strength of each access point at each location, and generate summary files grouped by AP or location for analysis. A sql file summary.sql will be also generated with sql statements to create a

Summary table and insert data into it. Then summary.sql file can be run in Oracle database to continue building the database.

6.2.4.3 Position Information Table Creation

A coordinate.txt file, like table , containing 4 columns information of Location Name, X coordinate, Y coordinate, and responding Room (or space), was used by pos_info.pl perl program to generate pos_info.sql which will contain sql statements to create POSITION INFO table and insert data into it. Then summary.sql file was run in Oracle database to execute these sql statements to continue building the database.

This POSITION_INFO table will be used in analysis and in the algorithm implementation to find the room of the on-line operation location and calculate the error distance.

6.2.4.4 Signal Strength Distribution Table

The distrb.pl perl program was run on database tables built for the signal strength information of each access points at each training location. It found the probability of each signal strength discrete value received from the specific access point at each location in the experiment (as described in section 4.3) and generate a sql file to create the signal strength distribution table, which will contain columns: Position, AP Name, discrete signal strength value and Percent of the SS value among all the corresponding AP SS values at this location. This table will be used in Algorithm 3, 4 and 5.

6.2.4.5 Algorithms implementation

Algorithms 1 to 5, as described in Chapter 5: algorithms, are implemented by perl programming language and the programs will query the Oracle database built in the prior steps.

6.2.4.6 On-line Sampling and Requesting Server for Result

The training and preparation steps have been finished before this step. Now it is the time of on-line operation phase. A mobile host can use NetStumbler sampling at certain location and send the wi-scan file generated to the server. The server will use the specific algorithm and query database to get the location result about it and return the result to the user.

6.3 Experimental Results of Five Algorithms

6.3.1 Algorithm 1: Finding the Best Match of the Averages

To evaluate this algorithm’s performance, the first sampling set which contains 24 location grid points is used as the training set to create the radio map database, and the second test set obtained in another month which contains 21 grid points and 7 points at other locations is used as the operation (on-line) set. The location and room results returned by the algorithm, and the real location and room are recorded and the error distance were calculated.

The algorithm is tested using the signal strength from the five training access points. And it is also tested using the signal strength from 4, 3, 2 or 1 training access points which provide the stronger signal strength at the operation location. The results are tabulated in Table 2. The unit of the Error Distance is feet and the unit of Signal Distance is dBm.

Table 3 to 7 show the results when signal strength from 5, 4, 3, 2 or 1 stronger access point at each operation location is used. Position column shows the operation location, next Result column shows the location determination result, then the Error Distance between them, the Minimum Signal Distance is the base to determine the result and shows the signal distance between the operation location and the location result, Room column lists the real room the operation location belongs, and the last Result column shows the room result which is determined by the location result. The shaded lines are corresponding to the 7 points not on the grid points.

Table 2 shows the Algorithm 1 result summary when using signal strength from different number of stronger access points at each operation location. The Correct Point Percentage of 21 column lists the 100% correct location determination result's percentage among the 21 training points on the grid; the avg(Error Distance) of 21 shows among the 21 training points the average error distance between the operation location and the result location; the avg(Error Distance) of 7 Other shows among the 7 operation points which are not on the training grid point, the average error distance between the operation location and the result location; the Correct Room Percentage of 21 column shows correct room guess result percentage among the 21 training points; and the Correct Room Percentage of 7 Other lists correct room guess result percentage among the 7 other operation points.

Number of APs	Correct Point Percentage of 21	avg(Error Distance) of 21	avg(Error Distance) of 7 Other	Correct Room Percentage of 21	Correct Room Percentage of 7 Other
5AP	52.4%	5.16	5.65	85.7%	100.0%
4AP	57.1%	5.07	9.89	85.7%	85.7%
3AP	61.9%	4.40	9.67	81.0%	85.7%
2AP	47.6%	6.81	9.67	66.7%	85.7%
1AP	19.0%	11.82	9.27	38.1%	85.7%

Table 2: Algorithm 1 results summary

We inferred in section 3.3: by using signal strength received from only one specific access point to infer the mobile host's location, there may be some wrong result at some point because of the environment variation. But if we use signal strength received from several access points altogether to infer the mobile host's location, the accuracy should be enhanced than using only one. This can be seen from the correct point result and the average error distance of the 21 training locations columns in the summary Table . When using one access point, the result is worst, and when more access points used together, the result is better. But if we use more than three access points, the result will not continue the accuracy increasing trend. So to use access points as many as possible is not a good idea, and the number of three got best result as in this test.

Position	Result	Error Distance	Minimum Signal Distance	Room	Result
0C	0C	0	6.92	OutDoor	OutDoor
1A	1B	10	5.52	Garage	Garage
1AB	1B	4	12.36	Garage	Garage
1C	1C	0	3.69	OutDoor	OutDoor
1D	1E	10	13.18	Bedroom3	Bedroom3
01DE	1E	7.07	7.04	Bedroom3	Bedroom3
1E	1E	0	12.75	Bedroom3	Bedroom3
2A	2A	0	5.64	Kitchen	Kitchen
2B	2A	10	9.23	Kitchen	Kitchen
2C	2C	0	4.65	Foyer	Foyer
2D	2C	10	6.28	Bath2	Foyer
2E	1E	10	10.31	Bedroom2	Bedroom3
3A	4A	10	5.95	Dining	Study
3AB	3A	6	11.74	Dining	Dining
3B	3B	0	9.67	Dining	Dining
3C	3C	0	7.99	Living	Living
3D	3D	0	4.37	Bedroom1	Bedroom1
3E	3D	10	9.74	Bedroom1	Bedroom1
4A	4A	0	12.16	Study	Study
4B	4B	0	4.64	Study	Study
4C	3B	14.14	11.15	Living	Dining
4D	3D	10	7.09	Bedroom1	Bedroom1
4E	3D	14.14	10.34	Bedroom1	Bedroom1
5C	5C	0	12.19	Porch	Porch
12DE	2E	5.66	3.44	Bedroom2	Bedroom2
23BC	3C	5.39	9.35	Living	Living
34D	3D	5	5.34	Bedroom1	Bedroom1
45AB	4A	6.4	9.59	Study	Study

Table 3: Algorithm 1 results (when using signal strength from 5 strongest access points at each operation location)

Position	Result	Error Distance	Minimum Signal Distance	Room	Result
0C	0C	0	6.74	OutDoor	OutDoor
1A	1B	10	5.4	Garage	Garage
1AB	1B	4	11.83	Garage	Garage
1C	1C	0	1.96	OutDoor	OutDoor
1D	1E	10	12.93	Bedroom3	Bedroom3
01DE	1E	7.07	7.01	Bedroom3	Bedroom3
1E	1E	0	10.59	Bedroom3	Bedroom3
2A	2A	0	5.55	Kitchen	Kitchen
2B	2A	10	9.12	Kitchen	Kitchen
2C	2C	0	4.65	Foyer	Foyer
2D	2C	10	5.46	Bath2	Foyer
2E	1E	10	10.27	Bedroom2	Bedroom3
3A	4A	10	5.23	Dining	Study
3AB	2D	26	8.09	Dining	Bath2
3B	3B	0	9.47	Dining	Dining
3C	3C	0	7.49	Living	Living
3D	3D	0	4.27	Bedroom1	Bedroom1
3E	2C	22.36	6.51	Bedroom1	Foyer
4A	4A	0	11.66	Study	Study
4B	4B	0	4.64	Study	Study
4C	4C	0	10.49	Living	Living
4D	3D	10	5.22	Bedroom1	Bedroom1
4E	3D	14.14	8.06	Bedroom1	Bedroom1
5C	5C	0	10.99	Porch	Porch
12DE	2E	5.66	1.71	Bedroom2	Bedroom2
23BC	4C	15.13	8.91	Living	Living
34D	3D	5	5.25	Bedroom1	Bedroom1
45AB	4A	6.4	8.11	Study	Study

Table 4: Algorithm 1 results (when using signal strength from 4 strongest access points at each operation location)

Position	Result	Error Distance	Minimum Signal Distance	Room	Result
0C	0C	0	6.63	OutDoor	OutDoor
1A	1B	10	5.37	Garage	Garage
1AB	1B	4	10.43	Garage	Garage
1C	1C	0	1.96	OutDoor	OutDoor
1D	1E	10	11.92	Bedroom3	Bedroom3
01DE	1E	7.07	6.37	Bedroom3	Bedroom3
1E	1E	0	9.36	Bedroom3	Bedroom3
2A	2A	0	5.55	Kitchen	Kitchen
2B	2A	10	9.07	Kitchen	Kitchen
2C	2C	0	1.65	Foyer	Foyer
2D	2C	10	4.17	Bath2	Foyer
2E	1E	10	10.09	Bedroom2	Bedroom3
3A	4A	10	5.23	Dining	Study
3AB	5C	24.41	4.37	Dining	Porch
3B	3B	0	8.3	Dining	Dining
3C	3C	0	6.42	Living	Living
3D	3D	0	3.13	Bedroom1	Bedroom1
3E	2C	22.36	5.54	Bedroom1	Foyer
4A	4A	0	11.28	Study	Study
4B	4B	0	4.59	Study	Study
4C	4C	0	8.43	Living	Living
4D	3D	10	5.14	Bedroom1	Bedroom1
4E	4E	0	3.13	Bedroom1	Bedroom1
5C	5C	0	8.85	Porch	Porch
12DE	2E	5.66	1.71	Bedroom2	Bedroom2
23BC	4C	15.13	7.5	Living	Living
34D	3D	5	4.33	Bedroom1	Bedroom1
45AB	4A	6.4	6.72	Study	Study

Table 5: Algorithm 1 results (when using signal strength from 3 strongest access points at each operation location)

Position	Result	Error Distance	Minimum Signal Distance	Room	Result
0C	0C	0	6.57	OutDoor	OutDoor
1A	1B	10	4.17	Garage	Garage
1AB	1B	4	8.42	Garage	Garage
1C	1C	0	1.96	OutDoor	OutDoor
1D	1E	10	11.82	Bedroom3	Bedroom3
01DE	1E	7.07	6.24	Bedroom3	Bedroom3
1E	1E	0	9.29	Bedroom3	Bedroom3
2A	1C	22.36	2.66	Kitchen	OutDoor
2B	3C	14.14	4.83	Kitchen	Living
2C	2C	0	1.39	Foyer	Foyer
2D	2D	0	1.95	Bath2	Bath2
2E	1E	10	7.42	Bedroom2	Bedroom3
3A	4A	10	3.04	Dining	Study
3AB	5C	24.41	1.19	Dining	Porch
3B	4B	10	4.79	Dining	Study
3C	3C	0	4.82	Living	Living
3D	2C	14.14	1.83	Bedroom1	Foyer
3E	3D	10	4.89	Bedroom1	Bedroom1
4A	4A	0	9.4	Study	Study
4B	4B	0	1.5	Study	Study
4C	4C	0	7.76	Living	Living
4D	3D	10	4.66	Bedroom1	Bedroom1
4E	4E	0	3.09	Bedroom1	Bedroom1
5C	4E	22.36	0.58	Porch	Bedroom1
12DE	2E	5.66	0.6	Bedroom2	Bedroom2
23BC	4C	15.13	7.44	Living	Living
34D	3D	5	3.57	Bedroom1	Bedroom1
45AB	4A	6.4	6.51	Study	Study

Table 6: Algorithm 1 results (when using signal strength from 2 strongest access points at each operation location)

Position	Result	Error Distance	Minimum Signal Distance	Room	Result
0C	0B	10	1.36	OutDoor	Garage
1A	1B	10	4.15	Garage	Garage
1AB	1B	4	7.98	Garage	Garage
1C	2B	14.14	1.18	OutDoor	Kitchen
1D	1E	10	7.55	Bedroom3	Bedroom3
01DE	1E	7.07	6.02	Bedroom3	Bedroom3
1E	1E	0	8.65	Bedroom3	Bedroom3
2A	1C	22.36	0.1	Kitchen	OutDoor
2B	3B	10	2.18	Kitchen	Dining
2C	4B	22.36	0.13	Foyer	Study
2D	3D	10	0.85	Bath2	Bedroom1
2E	1E	10	0.4	Bedroom2	Bedroom3
3A	4A	10	2.55	Dining	Study
3AB	3A	6	0.04	Dining	Dining
3B	4B	10	0.24	Dining	Study
3C	4C	10	0.09	Living	Living
3D	4B	22.36	1.17	Bedroom1	Study
3E	2C	22.36	1.03	Bedroom1	Foyer
4A	4A	0	9.37	Study	Study
4B	3D	22.36	1.23	Study	Bedroom1
4C	4C	0	6.78	Living	Living
4D	3D	10	0.58	Bedroom1	Bedroom1
4E	4E	0	0.52	Bedroom1	Bedroom1
5C	4E	22.36	0.02	Porch	Bedroom1
12DE	2E	5.66	0.21	Bedroom2	Bedroom2
23BC	4C	15.13	2.21	Living	Living
34D	4B	20.62	0.25	Bedroom1	Study
45AB	4A	6.4	3.94	Study	Study

Table 7: Algorithm 1 results (when using signal strength from 1 strongest access points at each operation location)

6.3.2 Algorithm 2: Finding the Mid-Point of the Two Best Match of the Averages

This algorithm is similar to Algorithm 1. The difference is: the result is determined by the mid-point the location with the minimum signal distance and the location with the second minimum signal distance.

To evaluate this algorithm's performance, like to evaluate Algorithm 1, the first test set which contains 24 location grid points is used as the training set to create the radio map database, and the second test set obtained in another month which contains 21 grid points from the 24 points in the first test and 7 points at other locations is used as the operation (on-line) set. The location results returned by the algorithm and the real locations and rooms are recorded and error distances were calculated.

The algorithm is tested using the signal strength from the five access points. And it is tested using the signal strength from 4, 3, 2 or 1 access points which provide the stronger signal strength at the operation location. The results are tabulated in tables. The unit of the Error Distance is feet and the unit of Signal Distance is dBm.

Table 9 to 13 show the results when signal strength from 3 stronger access points at each operation location is used. The shaded lines are the 7 points at other location.

Table 8 shows the summary results when using signal strength from different number of stronger access points at each operation location. The Correct Point Percentage of 21 column lists the 100% correct location determination result percentage among the 21 training point on the grid; the avg(Error Distance) of 21 column shows among the 21 training points the average error distance between the operation location and the result location; and the avg(Error Distance) of 7 Other column shows among the 7 operation points which are not on the training grid point, the average error distance between the operation location and the result location.

Number of APs	Correct Point Percentage of 21	avg(Error Distance) of 21	avg(Error Distance) of 7 Other
5AP	0.0%	7.75	9.07
4AP	0.0%	7.51	11.58
3AP	0.0%	7.49	9.97
2AP	0.0%	9.45	9.30
1AP	0.0%	10.71	8.59

Table 8: Algorithm 2 results summary

Like Algorithm 1 result, for the 21 training points, when using one access point, the result is worst, and when more access points used together, the result is better. But if we use more than three access points, the result will not continue the accuracy increasing trend. So using 3 stronger access points achieved the best result.

But Algorithm 2 doesn’t improve Algorithm 1 as inferred before the test, and Algorithm 1 got better error distance result than Algorithm 2. For the 21 training points, when using 5 APs, the error distance result of Algorithm 1 is 33.5% better than Algorithm 2; when using 4 APs, it’s 32.5% better; when using 3 APs, it’s 41.3% better; and when using 2 APs, it’s 27.9% better. Only when using 1 AP, the error distance result of Algorithm 2 is 9.4% better than Algorithm 1.This can be explained as: because of the complex signal environment, two nearest neighbors in the signal space are probably not two nearest neighbors in the locations space. So when averaging a location farther from the real location with a location nearer to the real, the result error distance will be larger.

Position	X Result	Y Result	Error Distance	Position 1	Position 2	Signal Distance 1	Signal Distance 2	Operation Room	Room 1	Room 2
0C	5	10	11.18	0C	1A	6.9	13.1	OutDoor	OutDoor	Garage
1A	5	10	11.18	1B	0B	5.5	10.8	Garage	Garage	Garage
1AB	5	10	6.4	1B	0B	12.4	16.8	Garage	Garage	Garage
1C	5	20	5	1C	0C	3.7	12.5	OutDoor	OutDoor	OutDoor
1D	10	35	5	1E	1D	13.2	15.7	Bedroom3	Bedroom3	Bedroom3
01DE	15	40	11.18	1E	2E	7	16.6	Bedroom3	Bedroom3	Bedroom2
1E	5	35	7.07	1E	0D	12.8	18.2	Bedroom3	Bedroom3	Bedroom3
2A	20	5	5	2A	2B	5.6	11.7	Kitchen	Kitchen	Kitchen
2B	25	0	11.18	2A	3A	9.2	12.8	Kitchen	Kitchen	Dining
2C	15	25	7.07	2C	1D	4.7	8.9	Foyer	Foyer	Bedroom3
2D	15	25	7.07	2C	1D	6.3	8.9	Bath2	Foyer	Bedroom3
2E	15	40	5	1E	2E	10.3	10.7	Bedroom2	Bedroom3	Bedroom2
3A	40	5	11.18	4A	4B	6	13	Dining	Study	Study
3AB	40	10	10.77	3A	5C	11.7	11.9	Dining	Dining	Porch
3B	35	10	5	3B	4B	9.7	11.9	Dining	Dining	Study
3C	35	20	5	3C	4C	8	11.6	Living	Living	Living
3D	35	30	5	3D	4D	4.4	5.8	Bedroom1	Bedroom1	Bedroom1
3E	25	30	11.18	3D	2D	9.7	9.9	Bedroom1	Bedroom1	Bath2
4A	40	5	5	4A	4B	12.2	19.5	Study	Study	Study
4B	35	10	5	4B	3B	4.6	9.7	Study	Study	Dining
4C	35	15	7.07	3B	4C	11.1	11.4	Living	Dining	Living
4D	30	35	11.18	3D	3E	7.1	10.9	Bedroom1	Bedroom1	Bedroom1
4E	30	35	11.18	3D	3E	10.3	11.6	Bedroom1	Bedroom1	Bedroom1
5C	40	25	11.18	5C	3D	12.2	18.2	Porch	Porch	Bedroom1
12DE	25	40	9.85	2E	3E	3.4	10.4	Bedroom2	Bedroom2	Bedroom1
23BC	35	20	10.2	3C	4C	9.4	9.9	Living	Living	Living
34D	25	30	10	3D	2D	5.3	7.3	Bedroom1	Bedroom1	Bath2
45AB	40	5	5.1	4A	4B	9.6	10.4	Study	Study	Study

Table 9: Algorithm 2 results (when using signal strength from 5 strongest access points at each operation location)

Position	X Result	Y Result	Error Distance	Position 1	Position 2	Signal Distance 1	Signal Distance 2	Operation Room	Room 1	Room 2
0C	5	20	5	0C	1C	6.7	11.8	OutDoor	OutDoor	OutDoor
1A	5	10	11.18	1B	0B	5.4	10.8	Garage	Garage	Garage
1AB	5	10	6.4	1B	0B	11.8	16.2	Garage	Garage	Garage
1C	5	20	5	1C	0C	2	5.9	OutDoor	OutDoor	OutDoor
1D	10	35	5	1E	1D	12.9	14.9	Bedroom3	Bedroom3	Bedroom3
01DE	15	40	11.18	1E	2E	7	16.6	Bedroom3	Bedroom3	Bedroom2
1E	15	40	5	1E	2E	10.6	17.6	Bedroom3	Bedroom3	Bedroom2
2A	20	5	5	2A	2B	5.6	9.6	Kitchen	Kitchen	Kitchen
2B	25	0	11.18	2A	3A	9.1	12.7	Kitchen	Kitchen	Dining
2C	15	25	7.07	2C	1D	4.7	8.9	Foyer	Foyer	Bedroom3
2D	15	25	7.07	2C	1D	5.5	8.4	Bath2	Foyer	Bedroom3
2E	15	40	5	1E	2E	10.3	10.7	Bedroom2	Bedroom3	Bedroom2
3A	40	5	11.18	4A	4B	5.2	13	Dining	Study	Study
3AB	15	30	28.3	2D	1D	8.1	10.7	Dining	Bath2	Bedroom3
3B	35	10	5	3B	4B	9.5	10.3	Dining	Dining	Study
3C	35	20	5	3C	4C	7.5	10.8	Living	Living	Living
3D	25	25	7.07	3D	2C	4.3	4.6	Bedroom1	Bedroom1	Foyer
3E	20	25	18.03	2C	2D	6.5	7	Bedroom1	Foyer	Bath2
4A	40	5	5	4A	4B	11.7	19.5	Study	Study	Study
4B	35	10	5	4B	3B	4.6	8.9	Study	Study	Dining
4C	35	15	7.07	4C	3B	10.5	11.1	Living	Living	Dining
4D	40	25	5	3D	5C	5.2	9.8	Bedroom1	Bedroom1	Porch
4E	35	35	7.07	3D	4E	8.1	10.9	Bedroom1	Bedroom1	Bedroom1
5C	35	25	15.81	5C	2D	11	11.3	Porch	Porch	Bath2
12DE	25	40	9.85	2E	3E	1.7	10	Bedroom2	Bedroom2	Bedroom1
23BC	35	20	10.2	4C	3C	8.9	9	Living	Living	Living
34D	25	30	10	3D	2D	5.3	6.6	Bedroom1	Bedroom1	Bath2
45AB	40	5	5.1	4A	4B	8.1	10.1	Study	Study	Study

Table 10: Algorithm 2 results (when using signal strength from 4 strongest access points at each operation location)

Position	X Result	Y Result	Error Distance	Position 1	Position 2	Signal Distance 1	Signal Distance 2	Operation Room	Room 1	Room 2
0C	5	20	5	0C	1C	6.6	11	OutDoor	OutDoor	OutDoor
1A	5	10	11.18	1B	0B	5.4	10.8	Garage	Garage	Garage
1AB	5	10	6.4	1B	0B	10.4	12.6	Garage	Garage	Garage
1C	5	20	5	1C	0C	2	4.9	OutDoor	OutDoor	OutDoor
1D	10	35	5	1E	1D	11.9	14.8	Bedroom3	Bedroom3	Bedroom3
01DE	15	40	11.18	1E	2E	6.4	14.1	Bedroom3	Bedroom3	Bedroom2
1E	10	35	5	1E	1D	9.4	14	Bedroom3	Bedroom3	Bedroom3
2A	20	5	5	2A	2B	5.5	9.5	Kitchen	Kitchen	Kitchen
2B	25	0	11.18	2A	3A	9.1	12.7	Kitchen	Kitchen	Dining
2C	20	25	5	2C	2D	1.7	7.7	Foyer	Foyer	Bath2
2D	15	20	11.18	2C	1C	4.2	7.8	Bath2	Foyer	OutDoor
2E	15	40	5	1E	2E	10.1	10.5	Bedroom2	Bedroom3	Bedroom2
3A	40	5	11.18	4A	4B	5.2	12.8	Dining	Study	Study
3AB	35	25	19.65	5C	2D	4.4	6	Dining	Porch	Bath2
3B	35	10	5	3B	4B	8.3	8.6	Dining	Dining	Study
3C	35	20	5	3C	4C	6.4	8.8	Living	Living	Living
3D	25	25	7.07	3D	2C	3.1	4.4	Bedroom1	Bedroom1	Foyer
3E	20	25	18.03	2C	2D	5.5	7	Bedroom1	Foyer	Bath2
4A	40	5	5	4A	4B	11.3	18.8	Study	Study	Study
4B	35	10	5	4B	3B	4.6	8.6	Study	Study	Dining
4C	35	15	7.07	4C	3B	8.4	11.1	Living	Living	Dining
4D	35	35	7.07	3D	4E	5.1	7.5	Bedroom1	Bedroom1	Bedroom1
4E	35	35	7.07	4E	3D	3.1	7.8	Bedroom1	Bedroom1	Bedroom1
5C	45	30	11.18	5C	4E	8.9	9.4	Porch	Porch	Bedroom1
12DE	15	30	6.08	2E	1C	1.7	8.9	Bedroom2	Bedroom2	OutDoor
23BC	35	20	10.2	4C	3C	7.5	8.7	Living	Living	Living
34D	25	25	11.18	3D	2C	4.3	5.8	Bedroom1	Bedroom1	Foyer
45AB	40	5	5.1	4A	4B	6.7	9.7	Study	Study	Study

Table 11: Algorithm 2 results (when using signal strength from 3 strongest access points at each operation location)

Position	X Result	Y Result	Error Distance	Position 1	Position 2	Signal Distance 1	Signal Distance 2	Operation Room	Room 1	Room 2
0C	10	10	14.14	0C	2A	6.6	8.2	OutDoor	OutDoor	Kitchen
1A	5	10	11.18	1B	0B	4.2	7.2	Garage	Garage	Garage
1AB	5	10	6.4	1B	0B	8.4	11.4	Garage	Garage	Garage
1C	5	20	5	1C	0C	2	2.7	OutDoor	OutDoor	OutDoor
1D	10	35	5	1E	1D	11.8	14.2	Bedroom3	Bedroom3	Bedroom3
01DE	15	40	11.18	1E	2E	6.2	13.9	Bedroom3	Bedroom3	Bedroom2
1E	10	35	5	1E	1D	9.3	13.5	Bedroom3	Bedroom3	Bedroom3
2A	15	10	11.18	1C	2A	2.7	3.7	Kitchen	OutDoor	Kitchen
2B	30	15	11.18	3C	3B	4.8	7.1	Kitchen	Living	Dining
2C	15	25	7.07	2C	1D	1.4	3.7	Foyer	Foyer	Bedroom3
2D	20	25	5	2D	2C	2	2.9	Bath2	Bath2	Foyer
2E	5	30	18.03	1E	0C	7.4	8.7	Bedroom2	Bedroom3	OutDoor
3A	40	5	11.18	4A	4B	3	11.3	Dining	Study	Study
3AB	35	25	19.65	5C	2D	1.2	4.9	Dining	Porch	Bath2
3B	35	10	5	4B	3B	4.8	7.1	Dining	Study	Dining
3C	35	20	5	3C	4C	4.8	8.5	Living	Living	Living
3D	20	25	11.18	2C	2D	1.8	2.7	Bedroom1	Foyer	Bath2
3E	30	25	15	3D	3C	4.9	5	Bedroom1	Bedroom1	Living
4A	40	5	5	4A	4B	9.4	17	Study	Study	Study
4B	35	10	5	4B	3B	1.5	6.3	Study	Study	Dining
4C	35	15	7.07	4C	3B	7.8	11	Living	Living	Dining
4D	30	35	11.18	3D	3E	4.7	6	Bedroom1	Bedroom1	Bedroom1
4E	35	40	5	4E	3E	3.1	6.5	Bedroom1	Bedroom1	Bedroom1
5C	35	40	25	4E	3E	0.6	7	Porch	Bedroom1	Bedroom1
12DE	15	35	1.41	2E	1D	0.6	4	Bedroom2	Bedroom2	Bedroom3
23BC	35	20	10.2	4C	3C	7.4	7.8	Living	Living	Living
34D	25	25	11.18	3D	2C	3.6	5.5	Bedroom1	Bedroom1	Foyer
45AB	40	5	5.1	4A	4B	6.5	8.9	Study	Study	Study

Table 12: Algorithm 2 results (when using signal strength from 2 strongest access points at each operation location)

Position	X Result	Y Result	Error Distance	Position 1	Position 2	Signal Distance 1	Signal Distance 2	Operation Room	Room 1	Room 2
0C	5	5	15.81	0B	1A	1.4	4	OutDoor	Garage	Garage
1A	5	10	11.18	1B	0B	4.1	7.2	Garage	Garage	Garage
1AB	5	10	6.4	1B	0B	8	11	Garage	Garage	Garage
1C	15	15	7.07	2B	1C	1.2	1.3	OutDoor	Kitchen	OutDoor
1D	10	35	5	1E	1D	7.5	12.4	Bedroom3	Bedroom3	Bedroom3
01DE	10	35	5	1E	1D	6	10.9	Bedroom3	Bedroom3	Bedroom3
1E	10	35	5	1E	1D	8.7	13.5	Bedroom3	Bedroom3	Bedroom3
2A	15	15	15.81	1C	2B	0.1	2.4	Kitchen	OutDoor	Kitchen
2B	20	15	5	3B	1C	2.2	2.6	Kitchen	Dining	OutDoor
2C	25	20	5	4B	1D	0.1	0.4	Foyer	Study	Bedroom3
2D	35	20	18.03	3D	4B	0.8	1.9	Bath2	Bedroom1	Study
2E	10	35	11.18	1E	1D	0.4	4.5	Bedroom2	Bedroom3	Bedroom3
3A	40	5	11.18	4A	4B	2.5	4.3	Dining	Study	Study
3AB	40	10	10.77	3A	5C	0	0.1	Dining	Dining	Porch
3B	25	20	11.18	4B	1D	0.2	0.8	Dining	Study	Bedroom3
3C	35	20	5	4C	3C	0.1	3.2	Living	Living	Living
3D	35	20	11.18	4B	3D	1.2	1.6	Bedroom1	Study	Bedroom1
3E	15	25	21.21	2C	1D	1	1.1	Bedroom1	Foyer	Bedroom3
4A	40	5	5	4A	4B	9.4	11.1	Study	Study	Study
4B	35	20	11.18	3D	4B	1.2	1.5	Study	Bedroom1	Study
4C	35	20	5	4C	3C	6.8	10.1	Living	Living	Living
4D	25	30	15	3D	2D	0.6	1.7	Bedroom1	Bedroom1	Bath2
4E	35	40	5	4E	3E	0.5	6.5	Bedroom1	Bedroom1	Bedroom1
5C	35	40	25	4E	3E	0	5.9	Porch	Bedroom1	Bedroom1
12DE	10	30	8.49	2E	0C	0.2	0.6	Bedroom2	Bedroom2	OutDoor
23BC	35	20	10.2	4C	3C	2.2	5.5	Living	Living	Living
34D	25	20	14.14	4B	1D	0.2	0.8	Bedroom1	Study	Bedroom3
45AB	40	5	5.1	4A	4B	3.9	5.7	Study	Study	Study

Table 13: Algorithm 2 results (when using signal strength from 1 strongest access points at each operation location)

Number of Samples	Correct Point Result of 28	Correct Point Percentage of 28	avg(Error Distance) of 28	Correct Room Result of 28	Correct Room Percentage of 28
1	15	53.6%	9.39	16	57.1%
3	16	57.1%	12.35	17	60.7%
5	12	42.9%	14.77	12	42.9%
10	7	25.0%	21.89	8	28.6%

Table 14: Algorithm 3 results summary

Number of Samples	Correct Point Result of 28	Correct Point Percentage of 28	avg(Error Distance) of 28	Correct Room Result of 28	Correct Room Percentage of 28
1	17	60.7%	9.12	20	71.4%
3	20	71.4%	8.58	20	71.4%
5	24	85.7%	4.25	24	85.7%
10	28	100.0%	0.00	28	100.0%

Table 15: Algorithm 4 results summary

Number of Samples	Correct Point Result of 28	Correct Point Percentage of 28	avg(Error Distance) of 28	Correct Room Result of 28	Correct Room Percentage of 28
1	18	64.3%	6.42	19	67.9%
3	21	75.0%	3.11	25	89.3%
5	22	78.6%	3.93	23	82.1%
10	26	92.9%	1.70	27	96.4%

Table 16: Algorithm 5 results summary

6.3.3 Algorithm 3: Signal Distribution Best Match I

At each operation location, the operation set of samples collected is composed of sample vectors at each timestamp (one sample vector per second). For each operation sample at each timestamp, Algorithm 3 sees the signal strength from different AP as relative components in the sample vector, so the probabilities of the signal strength for the corresponding AP at a location are multiplied together. And for different samples in the operation set collected at the operation location, their probability results calculated for a training location as described above are multiplied together again to get the probability result for the training location. This means the samples in the operation set are also seen as relative components in the operation set of the sample vectors. The training location with the maximum probability result will be returned as the location result.

The algorithm is tested on different numbers of operation sample vectors at each location. The numbers used here are 1, 3, 5 and 10. When more sample vectors is collected at the operation location and sent to the server, the accuracy should be better. But from Table 14: Algorithm 3 result summary for different number of operation samples, we can see this algorithm is not reliable, since when more sample vectors is used, the accuracy becomes worse.

Table 17 to 20 show the location determination results of Algorithm 3 for 1, 3, 5 or 10 samples used at each operation location.

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	0C	22.36	Bedroom3	OutDoor
2A	0C	28.28	Kitchen	OutDoor
2B	2B	0	Kitchen	Kitchen
2C	23BC	5.39	Foyer	Living
2D	0C	22.36	Bath2	OutDoor
2E	0C	28.28	Bedroom2	OutDoor
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	23BC	5.39	Living	Living
3D	23BC	13	Bedroom1	Living
3E	3E	0	Bedroom1	Bedroom1
4A	3E	41.23	Study	Bedroom1
4B	4B	0	Study	Study
4C	4C	0	Living	Living
4D	4C	10	Bedroom1	Living
4E	4C	20	Bedroom1	Living
5C	4C	10	Porch	Living
12DE	4C	28.84	Bedroom2	Living
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	34D	27.86	Study	Bedroom1

Table 17: Algorithm 3 results (when using 1 operation sample vector at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	0C	17.2	Garage	OutDoor
1C	0C	10	OutDoor	OutDoor
1D	0C	14.14	Bedroom3	OutDoor
01DE	01DE	0	Bedroom3	Bedroom3
1E	0C	22.36	Bedroom3	OutDoor
2A	0C	28.28	Kitchen	OutDoor
2B	2B	0	Kitchen	Kitchen
2C	0C	20	Foyer	OutDoor
2D	2D	0	Bath2	Bath2
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	0C	30	Living	OutDoor
3D	3D	0	Bedroom1	Bedroom1
3E	3E	0	Bedroom1	Bedroom1
4A	0C	44.72	Study	OutDoor
4B	0C	41.23	Study	OutDoor
4C	0C	40	Living	OutDoor
4D	4D	0	Bedroom1	Bedroom1
4E	4E	0	Bedroom1	Bedroom1
5C	0C	50	Porch	OutDoor
12DE	12DE	0	Bedroom2	Bedroom2
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	34D	27.86	Study	Bedroom1

Table 18: Algorithm 3 results (when using 3 operation sample vectors at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	0C	22.36	Garage	OutDoor
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	0C	15.81	Bedroom3	OutDoor
1E	0C	22.36	Bedroom3	OutDoor
2A	0C	28.28	Kitchen	OutDoor
2B	0C	22.36	Kitchen	OutDoor
2C	0C	20	Foyer	OutDoor
2D	0C	22.36	Bath2	OutDoor
2E	0C	28.28	Bedroom2	OutDoor
3A	0C	36.06	Dining	OutDoor
3AB	3AB	0	Dining	Dining
3B	0C	31.62	Dining	OutDoor
3C	3C	0	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	0C	36.06	Bedroom1	OutDoor
4A	4A	0	Study	Study
4B	4B	0	Study	Study
4C	4B	10	Living	Study
4D	4B	20	Bedroom1	Study
4E	4E	0	Bedroom1	Bedroom1
5C	0C	50	Porch	OutDoor
12DE	12DE	0	Bedroom2	Bedroom2
23BC	12DE	20.12	Living	Bedroom2
34D	34D	0	Bedroom1	Bedroom1
45AB	34D	27.86	Study	Bedroom1

Table 19: Algorithm 3 results (when using 5 operation sample vectors at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	0C	17.2	Garage	OutDoor
1C	0C	10	OutDoor	OutDoor
1D	0C	14.14	Bedroom3	OutDoor
01DE	0C	15.81	Bedroom3	OutDoor
1E	1E	0	Bedroom3	Bedroom3
2A	0C	28.28	Kitchen	OutDoor
2B	0C	22.36	Kitchen	OutDoor
2C	0C	20	Foyer	OutDoor
2D	0C	22.36	Bath2	OutDoor
2E	0C	28.28	Bedroom2	OutDoor
3A	3A	0	Dining	Dining
3AB	0C	33.11	Dining	OutDoor
3B	0C	31.62	Dining	OutDoor
3C	0C	30	Living	OutDoor
3D	3D	0	Bedroom1	Bedroom1
3E	0C	36.06	Bedroom1	OutDoor
4A	0C	44.72	Study	OutDoor
4B	0C	41.23	Study	OutDoor
4C	4C	0	Living	Living
4D	0C	41.23	Bedroom1	OutDoor
4E	0C	44.72	Bedroom1	OutDoor
5C	5C	0	Porch	Porch
12DE	0C	22.63	Bedroom2	OutDoor
23BC	0C	25.08	Living	OutDoor
34D	0C	36.4	Bedroom1	OutDoor
45AB	0C	47.76	Study	OutDoor

Table 20: Algorithm 3 results (when using 10 operation sample vectors at each operation location)

6.3.4 Algorithm 4: Signal Distribution Best Match II

Like Algorithm 3, for each operation sample at each timestamp, Algorithm 4 sees the signal strength from different AP as relative components in the sample vector, so the probabilities of the signal strength for the corresponding AP at a location are multiplied together. But different from Algorithm 3, Algorithm 4 sees the samples in the operation set as independent in the operation set. So for different samples in the operation set collected at the operation location, their probability results calculated for a training location as described above are added together to get the probability result for the training location. The training location with the maximum probability result will be returned as the location result.

Like Algorithm 3, the algorithm is tested on different numbers of operation sample vectors at each location. The numbers used here are 1, 3, 5 and 10. When more sample vectors is collected at the operation location and sent to the server, the accuracy should be better. From Table 15: Algorithm 4 result summary for different number of operation samples, we can see this algorithm is reliable, since when more sample vectors is used, the accuracy becomes better. Moreover, when 10 samples used and this algorithm gets the best result, the accuracy can be 100%.

Table 21 to 24 show the location determination results of Algorithm 4 for 1, 3, 5 or 10 samples used at each operation location.

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	1D	7.07	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	0C	28.28	Kitchen	OutDoor
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2C	10	Bath2	Foyer
2E	2E	0	Bedroom2	Bedroom2
3A	3AB	6	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	3C	0	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	3D	10	Bedroom1	Bedroom1
4A	0C	44.72	Study	OutDoor
4B	0C	41.23	Study	OutDoor
4C	4C	0	Living	Living
4D	4C	10	Bedroom1	Living
4E	4E	0	Bedroom1	Bedroom1
5C	0C	50	Porch	OutDoor
12DE	12DE	0	Bedroom2	Bedroom2
23BC	12DE	20.12	Living	Bedroom2
34D	34D	0	Bedroom1	Bedroom1
45AB	34D	27.86	Study	Bedroom1

Table 21: Algorithm 4 results (when using 1 operation sample vector at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2D	0	Bath2	Bath2
2E	2D	10	Bedroom2	Bath2
3A	45AB	15.52	Dining	Study
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	3C	0	Living	Living
3D	0C	31.62	Bedroom1	OutDoor
3E	3E	0	Bedroom1	Bedroom1
4A	0C	44.72	Study	OutDoor
4B	3A	14.14	Study	Dining
4C	0C	40	Living	OutDoor
4D	4D	0	Bedroom1	Bedroom1
4E	4E	0	Bedroom1	Bedroom1
5C	5C	0	Porch	Porch
12DE	12DE	0	Bedroom2	Bedroom2
23BC	23BC	0	Living	Living
34D	0C	36.4	Bedroom1	OutDoor
45AB	0C	47.76	Study	OutDoor

Table 22: Algorithm 4 results (when using 3 operation sample vectors at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2D	0	Bath2	Bath2
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	3C	0	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	3E	0	Bedroom1	Bedroom1
4A	3E	41.23	Study	Bedroom1
4B	4B	0	Study	Study
4C	4C	0	Living	Living
4D	4C	10	Bedroom1	Living
4E	4E	0	Bedroom1	Bedroom1
5C	5C	0	Porch	Porch
12DE	12DE	0	Bedroom2	Bedroom2
23BC	12DE	20.12	Living	Bedroom2
34D	34D	0	Bedroom1	Bedroom1
45AB	0C	47.76	Study	OutDoor

Table 23: Algorithm 4 results (when using 5 operation sample vectors at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2D	0	Bath2	Bath2
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	3C	0	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	3E	0	Bedroom1	Bedroom1
4A	4A	0	Study	Study
4B	4B	0	Study	Study
4C	4C	0	Living	Living
4D	4D	0	Bedroom1	Bedroom1
4E	4E	0	Bedroom1	Bedroom1
5C	5C	0	Porch	Porch
12DE	12DE	0	Bedroom2	Bedroom2
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	45AB	0	Study	Study

Table 24: Algorithm 4 results (when using 10 operation sample vectors at each operation location)

6.3.5 Algorithm 5: Signal Distribution Best Match III

Unlike Algorithm 3 and 4, for each operation sample at each timestamp, Algorithm 5 sees the signal strength from different AP as independent in the sample vector, so the probabilities of the signal strength for the corresponding AP at a location are added together. The same as Algorithm 4, Algorithm 5 sees the samples in the operation set as independent in the operation set. So for different samples in the operation set collected at the operation location, their probability results calculated for a training location as described above are added together to get the probability result for the training location. The training location with the maximum probability result will be returned as the location result.

Like Algorithm 3 and 4, the algorithm is tested on different numbers of operation sample vectors at each location. The numbers used here are 1, 3, 5 and 10. When more sample vectors is collected at the operation location and sent to the server, the accuracy should be better. From Table 16: Algorithm 5 result summary for different number of operation samples, we can see this algorithm is also reliable, since when more sample vectors is used, the accuracy becomes better. When 10 samples used, this algorithm can get the best result: 92.9% accurate location determination results and 1.70 feet average error distance.

Table 25 to 28 show the location determination results of Algorithm 5 for 1, 3, 5 or 10 samples used at each operation location.

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1AB	14	OutDoor	Garage
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	1D	10	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2E	10	Bath2	Bedroom2
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	4C	14.14	Dining	Living
3C	3C	0	Living	Living
3D	3C	10	Bedroom1	Living
3E	3E	0	Bedroom1	Bedroom1
4A	3E	41.23	Study	Bedroom1
4B	3E	31.62	Study	Bedroom1
4C	4C	0	Living	Living
4D	4D	0	Bedroom1	Bedroom1
4E	4D	10	Bedroom1	Bedroom1
5C	4D	14.14	Porch	Bedroom1
12DE	4D	24.74	Bedroom2	Bedroom1
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	45AB	0	Study	Study

Table 25: Algorithm 5 results (when using 1 operation sample vector at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2C	10	Bath2	Foyer
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	23BC	5.39	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	3D	10	Bedroom1	Bedroom1
4A	4A	0	Study	Study
4B	45AB	7.81	Study	Study
4C	4C	0	Living	Living
4D	4E	10	Bedroom1	Bedroom1
4E	4E	0	Bedroom1	Bedroom1
5C	4D	14.14	Porch	Bedroom1
12DE	12DE	0	Bedroom2	Bedroom2
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	2C	29.68	Study	Foyer

Table 26: Algorithm 5 results (when using 3 operation sample vectors at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	1D	7.07	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2E	10	Bath2	Bedroom2
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	3C	0	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	3E	0	Bedroom1	Bedroom1
4A	3A	10	Study	Dining
4B	4B	0	Study	Study
4C	4C	0	Living	Living
4D	4C	10	Bedroom1	Living
4E	1D	31.62	Bedroom1	Bedroom3
5C	1D	41.23	Porch	Bedroom3
12DE	12DE	0	Bedroom2	Bedroom2
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	45AB	0	Study	Study

Table 27: Algorithm 5 results (when using 5 operation sample vectors at each operation location)

Position	Result	Error Distance	Room	Result
0C	0C	0	OutDoor	OutDoor
1A	1A	0	Garage	Garage
1AB	1AB	0	Garage	Garage
1C	1C	0	OutDoor	OutDoor
1D	1D	0	Bedroom3	Bedroom3
01DE	01DE	0	Bedroom3	Bedroom3
1E	1E	0	Bedroom3	Bedroom3
2A	2A	0	Kitchen	Kitchen
2B	2B	0	Kitchen	Kitchen
2C	2C	0	Foyer	Foyer
2D	2D	0	Bath2	Bath2
2E	2E	0	Bedroom2	Bedroom2
3A	3A	0	Dining	Dining
3AB	3AB	0	Dining	Dining
3B	3B	0	Dining	Dining
3C	3C	0	Living	Living
3D	3D	0	Bedroom1	Bedroom1
3E	3E	0	Bedroom1	Bedroom1
4A	4A	0	Study	Study
4B	4B	0	Study	Study
4C	4C	0	Living	Living
4D	3D	10	Bedroom1	Bedroom1
4E	4E	0	Bedroom1	Bedroom1
5C	5C	0	Porch	Porch
12DE	5C	37.58	Bedroom2	Porch
23BC	23BC	0	Living	Living
34D	34D	0	Bedroom1	Bedroom1
45AB	45AB	0	Study	Study

Table 28: Algorithm 5 results (when using 10 operation sample vectors at each operation location)

7 CONCLUSION AND FUTURE WORK

In this thesis, the framework of an indoor location determination system using radio signal strength, the algorithm and the database design and implementation are presented. Real-world experiments in the area of 55 by 45 feet with 10-foot grid spacing were conducted to evaluate the performance of the system using five different algorithms.

Algorithm 1, finding the location with the best average signal strength match, is similar to the algorithm used in Radar system [2]. In our test bed, when using three strongest access points detected at the operation location, the system can get the best result of 61.9% correct location point result, 4.40 feet average error distance, and 81.0% correct room result from the 21 training location points.

Algorithm 2, finding the location which is the mid-point of the two locations with the minimum signal distances, is designed to get improvement of Algorithm 1. But due to the complex difference between the signal space and the location space, the results of using 5, 4, 3 or 2 strongest access points are worse than Algorithm 1. Only when using 1 strongest access point, the result of average error distance is 9.4% better than Algorithm 1.

Algorithm 3, 4 and 5 are designed to use the discrete signal strength value distribution, i.e. the percentage of the signal strength value’s appearance in the training database. Algorithm 3 regards the signal strength value from the different access point in the sample vector as relative, and regards the sample vectors of the different timestamp in the sample vectors' set as relative; Algorithm 4 regards them as relative and independent correspondingly; and Algorithm 5 regards them as independent and independent correspondingly. The experiments show that Algorithm 3 is not reliable as it has contrary results. But Algorithm 4 and 5 are reliable and can achieve high accuracy when more operation samples are collected and used. The best result of Algorithm 4 is: when 10 operation samples are used, the accuracy can even achieve 100%. The best result of Algorithm 5 is: when 10 operation samples are used, 92.9% location point results are correct and average error distance of 1.70 feet is achieved.

More experiments can be designed to evaluate the performance of the system using these algorithms in the future.

Another future work can be: to give the room result of point not in the training point set, another table in the location determination system database needs to build, which will provide the relationship between the coordinate and the room. The columns can be Room, corresponding min x coordinate and max x coordinate, corresponding min y coordinate and max y coordinate. This can give out the room result by given any coordinates in this area because all rooms are rectangular. Now the system can only give out the room guess at the training grid points

For Algorithm3, 4 and 5, now the computation in the operation phase is not fast when more samples are collected and used in the operation phase to achieve higher accuracy. So finding methods to reduce the computation cost will contribute to the efficiency.

For Algorithm 2 improvement, we can add weights determined in some way by the signal distances of between the operation location and the two training locations found by the system to infer the location.

Another future work can be combining the system design, algorithm and database design with the convenient user interface toolkit developed by our group.

REFERENCES

[1] http://www.stumbler.net/

[2] Jeffrey Hightower and Gaetano Borriello, “Location Systems for Ubiquitous Computing”, IEEE Computer, 2001

[3] P. Bahl and V. N. Padmanabhan. “RADAR: An in-building RF-based user location and tracking system”, In IEEE Infocom 2000.

[4] P. Bahl, V. N. Padmanabhan, and A. Balachandran, “Enhancements to the RADAR User Location and Tracking System,” Microsoft Research Technical Report: MSR-TR-00-12 (February 2000)

[5] P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, “On Indoor Position Location With Wireless LANs”, The 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2002), Lisbon, Portugal, September 2002.

[6] Asim Smailagic, Daniel P. Siewiorek, Joshua Anhalt, David Kogan, and Yang Wang, “Location Sensing and Privacy in a Context Aware Computing Environment,“ Pervasive Computing, 2001

[7] IEEE Standard 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications.

[8] WANT, R., HOPPER, A., FALCO, V., AND GIBBONS, J. “The Active Badge Location System”, ACM Transactions on Information Systems 10, 1 (January 1992), 91–102.

[9] H. Hashemi, “The Indoor Radio Propagation Channel,” Proceedings of the IEEE, Vol. 81, No. 7, pages 943-968 July 1993

[10] Stallings, W., “Wireless Communications and Networks”, first ed. Prentice Hall 2002

[11] Adam Harder, Lanlan Song, Yu Wang, “Towards an Indoor Location System Using RF Signal Strength in IEEE 802.11 Networks”, itcc, vol. 02, no. 2, pp. 228-233, International 2005

[12] Travis Calvert, “Wireless Location Determination Using Existing 802.11 Wireless Networks to Determine a User’s Location”, 2004

[13] Bjarte Stien Karlsen, “A Survey of Indoor Positioning Systems”, 2004.

[14] K Raman Kumar, “Location based Services in Wireless LANs”, 2004.

[15] Vasileios Zeimpekis, George M. Giaglis, George Lekakos, “A Taxonomy of Indoor and Outdoor Positioning Techniques for Mobile Location Services”, ACM SIGecom Exchanges, Volume 3 Issue 4, December 2002.

[16] Manapure, S.S.; Darabi, H.; Patel, V.; Banerjee, P., “A comparative study of radio frequency-based indoor location sensing systems”, Networking, Sensing and Control, 2004 IEEE International Conference on Volume 2, 2004 Page(s):1265 - 1270 Vol.2.

[17] Dhruv Pandya; Ravi Jain; Lupu, E.; “Indoor location estimation using multiple wireless technologies”, Personal, Indoor and Mobile Radio Communications,

2003. PIMRC 2003. 14th IEEE Proceedings on Volume 3, 7-10 Sept. 2003 Page(s):2208 - 2212 vol.3.

[18] Youngjune Gwon; Jain, R.; Kawahara, T.; “Robust indoor location estimation of stationary and mobile users”, INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies Volume 2, 7-11 March 2004 Page(s):1032 - 1043 vol.2

[19] Lionel M. Ni, Yunhao Liu1, Yiu Cho Lau and Abhishek P. Patil, “LANDMARC: Indoor Location Sensing Using Active RFID”, PerCom’03, 2003.

[20] Andreas Haeberlen, Eliot Flannery, Andrew M. Ladd, Algis Rudys, Dan S. Wallach, Lydia E. Kavraki, “Localization: Practical robust localization over large-scale 802.11 wireless networks”, September 2004 Proceedings of the 10th annual international conference on Mobile computing and networking

[21] Castro, P.; Munz, R.; “Managing context data for smart spaces”, Personal Communications, IEEE [see also IEEE Wireless Communications] Volume 7, Issue 5, Oct. 2000 Page(s):44 – 46

[22] P Castro, P Chiu, T Kremenek, RR Muntz, “A Probabilistic Room Location Service for Wireless Networked Environments”, Ubicomp, 2001

[23] Andrew M. Ladd , Kostas E. Bekris , Algis Rudys , Lydia E. Kavraki , Dan S. Wallach , Guillaume Marceau, “Robotics-based location sensing using wireless Ethernet”, Proceedings of the 8th annual international conference on Mobile computing and networking, September 23-28, 2002, Atlanta, Georgia, USA

[24] Ping Tao, Algis Rudys, Andrew M. Ladd, Dan S. Wallach, "Location: Wireless LAN location-sensing for security applications", September 2003 Proceedings of the 2003 ACM workshop on Wireless security

[25] Ladd, A.M., Bekris, K.E., Marceau, G., Rudys, A., Wallach, D.S., Kavraki, L.E., "Using wireless Ethernet for localization", Publication Date: 30 Sept.-5 Oct. 2002 Volume: 1 On page(s): 402 - 408 vol.1

[26] T. Roos, P.Myllymäki, and H.Tirri, "A Statistical Modeling Approach to Location Estimation," IEEE Transactions on Mobile Computing, Vol. 1, No. 1, January-March 2002, pp. 59-69.

[27] P. Myllymaki, T. Roos, H. Tirri, P. Misikangas, J. Sievanen, "A Probabilistic Approach to WLAN User Location Estimation", The Third IEEE Workshop on Wireless LANs, 2001.

[28] S Thrun, "Probabilistic algorithms in robotics", AI Magazine, 2000

[29] Moustafa Youssef and Ashok Agrawala, "On the Optimality of WLAN Location Determination Systems", Communication Networks and Distributed Systems Modeling and Simulation Conference, January 18-24 2004, San Diego, California

[30] Asim Smailagic, Daniel P. Siewiorek, Joshua Anhalt, David Kogan, and Yang Wang, "Location Sensing and Privacy in a Context Aware Computing Environment", Pervasive Computing, 2001.

[31] Gutmann, J.-S.; Burgard, W.; Fox, D.; Konolige, K.; "An experimental comparison of localization methods", Intelligent Robots and Systems, 1998. Proceedings., 1998 IEEE/RSJ International Conference on Volume 2, 13-17 Oct. 1998 Page(s):736 - 743 vol.2

[32] Gutmann, J.-S.; Fox, D., “An experimental comparison of localization methods continued”, Intelligent Robots and System, 2002. IEEE/RSJ International Conference on Volume 1, 30 Sept.-5 Oct. 2002 Page(s):454 - 459 vol.1

[33] S. Saha, K. Chaudhuri, D. Sanghi, and P. Bhagwat, “Location determination of a mobile device using IEEE 802.11b access point signals”. In IEEE Wireless Communications and Networking Conference 2003, March 2003.

[34] R. Battiti, T. L. Nhat, and A. Villani, “Location-aware computing: a neural network model for determining location in wireless LANs”, Technical Report Technical Report DIT-02-0083, 2002.