2014

Post hoc Indoor Localization Based on Rss Fingerprint in Wlan

Hao Huang
University of Massachusetts Amherst

POST HOC INDOOR LOCALIZATION BASED ON RSS FINGERPRINT IN WLAN

A Thesis Presented

by

HAO HUANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

February 2014

Electrical and Computer Engineering

POST HOC INDOOR LOCALIZATION BASED ON RSS FINGERPRINT IN WLAN

A Thesis Presented

by

HAO HUANG

Approved as to style and content by:

Dennis L. Goeckel, Chair

Patrick A. Kelly , Member

Aura Ganz, Member

C.V.Hollot, Head
Electrical and Computer Engineering

"It is easier to invent the future than to predict it."

To my beloved parents

ACKNOWLEDGMENTS

Many thanks to Professor Dennis L. Goeckel

ABSTRACT

POST HOC INDOOR LOCALIZATION BASED ON RSS FINGERPRINT IN WLAN

FEBRUARY 2014

HAO HUANG

B.E., CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS, CHONGQING, CHINA

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Dennis L. Goeckel

In the investigation of crimes committed by wireless users, one of the key goals is to determine the location of the mobile device at the time of the crime. Since this happens during the investigative phase after the crime is committed, we term this the *post hoc geographical localization estimation* problem. In this thesis, we introduce the post hoc geographical localization estimation problem and present approaches for its solution based on radio frequency (RF) fingerprinting. Motivated by the goal of establishing a crime's location with enough accuracy to obtain a search warrant, our focus is on locating a criminal mobile device in indoor environments with roughly the granularity to distiguish between two adjacent rooms, without having the ability to enter those rooms or the building to gather input data for the RF fingerprinting algorithm. While empirical performance studies of instantaneous indoor positioning

systems based radio frequency (RF) fingerprinting have been presented in the literature, the core of this thesis is the first empirical study focused on the post hoc version of problem from the viewpoint of digital forensics. In this study, we set up experiments in a residential area and collect a large set of raw data in order to analyze and evaluate the algorithms, the best of which provides a mean error distance of roughly 1.4 meters. In addition, we consider enhancements to the baseline algorithms if knowledge of the blueprint of the building is available. In particular, we consider whether compensating the raw data for the attenuation caused by walls can improve algorithm performance.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation

The success of wireless networks and the ubiquitous usage of mobile devices have revolutionized certain types of cybercrime. According to a Norton Cybercrime Report [12], email account hacking, responded phishing attempt, identity theft and online credit card fraud are the four main types of cybercrime globally and domestically. And these crimes can be conducted anonymously by malicious hackers who exploit open wireless access points. In fact, identity theft is one of the fastest growing crimes in the United States and made up nineteen percent [1] of all fraud complaints in 2010. And there are nearly 12 million people affected by identity fraud each year [33]. One version exploits wireless access points through a technique known as *Wardriving* [4]. In this scenario, a hacker tries to establish a wireless connection through a non-protected access point into a personal or an organization's internal network and steal the confidential information. Afterwards, an investigative process will be triggered by law enforcement, and it is important that there is preliminary evidence to support a warrant providing further search rights.

As an exemplary scenario that we will use throughout this chapter to describe the method by which we pursue the goal, consider an attack that occurs in a hotel. A hacker could stay inside a hotel room and anonymously use the Internet through the motel's free Wi-Fi network. Malicious activities could take place like stealing a consumer's credit card, banking and other confidential business information. In 2008, at the luxury Thompson Hotel chain, a hacker captured personal and sensitive

emails sent by guests and staff members over its network and threatened to make them public [2].

Fortunately, the use of mobile devices by such an offender will typically result in digital evidence. Digital forensics is the study of techniques that allow the authorized investigation of an alleged crime or policy violation that involves digital data. One of the key goals in investigations of crimes committed by wireless users is determining the location of the mobile device at the time of the crime. As a prototypical example, consider a search warrant that might require locating a mobile device with fine enough granularity to distinguish between two adjacent rooms. We will term this *post hoc geographical localization*.

Assume that each access point (AP) stores the received signal strength (RSS) for every observable user and that the crime's RF fingerprint, which we define as the collection of RSS measurements across the APs for a given device at a given time, is then available to the investigator. Per above, the investigator's goal is to use this RF fingerprint to determine the location of the device at the time of the crime without having the ability to enter the building. The approach employed here is to build a radio frequency map after the fact. In particular, in this post hoc localization problem, we will place reference points around the building of interest where the crime occurred and gather the collection of RSS measurements across the APs (the RF fingerprint) for each reference point. Using this, we can then build a radio map that relates the RF fingerprints to position. The challenge is then to use such a radio map to determine the location of a device *within* the building, where we are unable to gather data, by considering the relation between the RF fingerprint from the crime and those collected from the reference points. Although this post hoc localization problem has been rarely considered in the past, previous efforts on studying instantaneous geographical localization can be used as a starting point.

1.2 Background

Per above, there has been extensive consideration at the localization of a wireless device at the current time to support location-based services (LBS) for both outdoor case and indoor scenarios. In the former case, global positioning system (GPS) [11] and cellular networks [27] have been widely exploited. GPS provides accurate results but consumes the battery power at a rapid rate, and does not return the location as quickly as the user wants. On the other hand, determining a user's location by employing cell towers can provide a fast response result but less accuracy. For the indoor case, several types of wireless technologies are used and can be categorized into GPS-based, RFID-based [18], cellular-based, UWB-based [17], WLAN-based [5, 17, 20, 30] and others. Tradeoffs have been considered according to different performance metrics, such as cost, complexity, robustness, scalability and security. Of these, the class of WLAN-based systems takes advantage of the ubiquitous coverage of wireless network infrastructure to estimate the mobile devices location by finding the subtle dependency between detected electromagnetic measurements and the user's location. Hence it has gained a lot of attention for indoor applications.

A widely adopted approach in WLAN indoor localization is fingerprint positioning [5, 20, 24, 31]. Fingerprint is a vector of received signal strength (RSS) measurements collected from detectable access points (APs). During an offline stage, fingerprints are collected at pre-determined positions, called reference points (RPs), to build a pre-mapping database or a table of RSS patterns around the area of interest [5, 10] (see Biaz and Ji for a survey of techniques [6]). This pre-mapping database is called a radio map [31], and it contains the coordinate location of the reference point and the related vector of RSS measurements. During the online stage, the mobile device moves into the area of interest and a set of sample measured RSS values can be collected from detectable APs and stored in the device itself. Then its location is estimated by calculating the similarity between the sample measurements and the

entries in the offline radio map. The computation can be done by the mobile device itself or by a remote server. The final step is to map the similarity to the coordinate location. The most challenging side of fingerprint based positioning is the design of an algorithm that measures the similarity between the online measurement and an entry in the database. Two algorithms from the literature will be reviewed in Chapter 2 and implemented in later experiments.

1.3 Problem Statement and Definitions

Traditionally, the aforementioned techniques and the consideration of performance tradeoffs are limited to the instantaneous indoor positioning problem, where a radio map is constructed in advance (i.e. before the target device gets into the area of interest). In this case there are two common drawbacks: (i) pre-mapping can be prohibitive for large geographical areas and private residences, and (ii) changes to the environment can render pre-mapped data inaccurate. Furthermore, many proposals rely on a regular grid, but it is unlikely that a regular grid will be available in a residential area where a crime might take place.

In this thesis, we are interested in locating mobile device from the forensics viewpoint. When the criminal commits the crime, the access points store the received signal strength for the criminal user. Now, an investigation is initiated, and it becomes important for search warrant purposes to distinguish between two rooms. In this forensics situation, the crime mobile device was used inside a building and we assume that we have no prior knowledge about the area of interest (such as the layout of the building). In order to estimate its position, we will proceed as follows by using the vector of signal strength measurements recorded from the access points.

- Access the vector fingerprint of observations of the signal strength of the criminal mobile device at the time of the crime.

- Dynamically and flexibly set reference points (RPs) around the area of interest, noting RPs cannot always be in optimal positions (e.g. only outside a private residence).

- At each reference point, a vector fingerprint of received signal strength is collected from the detectable APs.

- Each tuple containing the coordinate position of the RP and its associated fingerprint is used to build a radio map as a description of the RSS environment for the area of interest.

- A RSS-position dependency model will be employed to probabilistically map the vector fingerprint of RSS measurements of the criminal mobile as an estimated location.

There are five main assumptions that limit the scope of this work: (i) the observations of RSS stored in the criminal mobile device or at the access points at the time of the crime are accessible, (ii) there are enough access points in the area of interest, (iii) the target and the access points are located on the same geographic plate, (iv) the detectable access points are the same in terms of hardware, and (v) the mobile device is stationary.

However, this approach has three powerful advantages: (i) there is no requirement for real-time response; (ii) the measurement can be refined after observations take place. That is, forensics requires a reverse process for localization: observations (of a radio related to a crime) are taken in situ by a nearby, in-place wireless mesh or AP network, but mapping and localization occurs afterwards; (iii) we do not expect that in-place APs are ideally placed, but we do expect the ability to ideally choose new measurements to refine localization afterwards.

1.4 Approaches and Contribution

This thesis is centered around an experimental study of a post hoc fingerprint based indoor positioning system. First, we review the wireless-based positioning systems that provides for location-based services (LBS). Next, two types of RSS-position dependency models are considered in detail. In order to achieve a highly accurate estimation result, we adopt the Gaussian-based kernel model, first introduced for the instantaneous indoor localization problem, into the post hoc indoor localization problem. Next, we consider how to improves its performance. In particular, according to studies of indoor WLAN signal characteristics, the attenuation of signal strength is mainly due to obstacles like walls. More importantly, such obstacles significantly alter the RF fingerprint of devices located within and make positioning challenging, since it is such obstacles that break the dependence between signal strength and distance from the transmitter. Thus, in situations where we have an accurate blueprint of the structure, we consider incorporating a wall attenuation factor to compensate for the RSS attenuation caused by the building infrastructure. To support the incorporation of such a factor, we also present a method to find the average wall attenuation factor for the area of interest and evaluate it through experiment.

Our experiments are set up in a residential area. A large set of raw data is collected and analyzed in order to evaluate the algorithms presented in this work. First, we extensively analyze the location fingerprints, which helps develop an understanding of the signal's underlying features and the reasons why it can be used as an indication of location information. In particular, the distribution and the coverage of the RSS in the indoor environment are visualized.

Next, the critical topic of AP selection is considered. To ensure the success of an indoor positioning system based on location fingerprints, APs must be selected that can be used to distinguish different positions without introducing significant noise into the algorithm. Some APs can appear intermittently, which also makes

their employment challenging. In addition, constructing a RF fingerprint radio map requires a large amount of data from multiple APs; thus, subsets of detectable APs may report correlated measurements, leading to needless redundancy and possible biased estimation. This motivates the need to employ sophisticated algorithms (sucha as multiple discriminant analysis (MDA) [26]) to choose a subset of the detectable APs for use in positioning.

Currently, there are no clear guidelines on how to estimate the mobile device's position in the post hoc situation. Moreover, it is not clear how many reference points need to be set up to build the radio map for a given accuracy. The main goal is to study the accuracy and the precision performance metrics to identify a set of system parameters based on practical experiments for further guidance. The result of the system analysis can be applied to further study in post hoc indoor positioning problem, such as considering the influence of the area of interest's 3-D layout and the computational cost of algorithms chosen for locating mobile devices. The following is the list of contributions:

- A comparison of the current popular RSS-position dependency models theoretically and experimentally.
- Introduction of the study of the wall attenuation factor (WAF)s influence on indoor localization and tried to compensate its impact on current estimation algorithms.
- Guideline for future post hoc indoor localization studies and a discussion of a few technical challenges in this scenario.

1.5 Thesis Organization

The remainder of this thesis is organized as follows: Chapter 2 presents relevant background materials and provides an outline of existing methods applied in

a WLAN indoor instantaneous positioning system. Fingerprinting-based methods, such as the K-nearest neighbor technique and the Gaussian kernel based technique, are discussed and compared. Chapter 2 also presents AP selection methods used in this thesis. Chapter 3 motivates the WAF compensation and explores its adaptation to Gaussian kernel based methods in the post hoc problem. Chapter 4 provides a detailed description of this experimental setup, data sets, and testing scenarios used in evaluation. This chapter also describes the way to calculate the WAF and shows the result comparison based on different parameters chosen, such as kernel width, multiple discriminant analysis thresholds and the number of neighbors in KNN.

CHAPTER 2

BACKGROUND AND RELATED WORK

In this chapter, prior work and the necessary background for the tools employed in our post hoc localization algorithms and experiments are presented. Per Chapter 1, there has been extensive work in the determination of the current location of a user to support location based services (LBS). Since this is the closest application to our problem, these prior algorithms are discussed in detail in Section 2.1. Next, we turn to the description of the mathematical tools that will form key parts of our algorithms. First, an important aspect of the model is the mapping of the location to the set of received signal strengths (RSSs). Two methods for performing this mapping are presented, namely the radio propagation model and the fingerprint-based model. Next, we discuss two fingerprinting methods for determining a user's location, namely K-Nearest Neighbor (KNN) and Kernel-based, in Section 2.3.1 and Section 2.3.2, respectively. Finally, the underpinnings of multiple discriminant analysis, an additional technique that will be employed to improve the estimation performance, are described.

2.1 Wireless-based Positioning System Review

Wireless-based positioning systems are designed to provide accurate and reliable user location information in LBS applications.

2.1.1 GPS-Based And Cellular-Based

A well-known positioning system for the outdoor environment is the Global Positioning System (GPS). Unfortunately, GPS works poorly in many cities where tall

buildings form urban canyons to block a clear view of the GPS satellites [11]. Also GPS offers limited accuracy and coverage in indoor environments. There are several positioning systems build on GPS to overcome the drawbacks of conventional GPS [21]. For instance, assisted-GPS (AGPS) provides a GPS-based indoor technique with an average of 15 meters accuracy. It uses a location server with a reference GPS receiver that can detect the same satellites to help another GPS to find a weak GPS signal. Another approach exploits the mobile cellular network and is applied where the area of interest is covered by several base stations or one base station with a strong RSS received by the indoor mobile target. An example [27] of such implements the indoor localization on GSM mobile phones through the use of "wide" signal-strength fingerprints, where a wide fingerprint includes the six strongest GSM cells and readings of up to 29 additional GSM channels. The results in [27] show that the GSM-based indoor localization system can differentiate between floors and achieve median errors of around 4 to 5 meters. Although these methods can provide reasonable indoor position estimation, the extra requirement for cellular network infrastructure and the limited application on mobile phones make them not suitable for reliable and efficient LBS.

2.1.2 RFID-Based And UWB-Based

Besides the GPS and the cellular network solutions, different types of wireless technologies such as radio-frequency identification (RFID) and ultra-wideband (UWB) are employed. RFID is a means of storing and retrieving data through electromagnetic transmission to an RF compatible integrated circuit and a RFID system consists of a number of RFID readers, RFID tags and the communication between them. The reader is used to read the data from the tags. The tags can be categorized as either passive or active. Passive RFID tags receive the RF signal from a reader and add location information by the reflected signal. Active RFID tags act as transceivers,

which could actively transmit location information. For the positioning purpose, active RFID is generally more suitable because it has a smaller antenna and a much longer working range. SpotON [18] is a well-known location sensing system that uses an aggregation algorithm for 3-D location sensing based on radio signal strength analysis from many tags. Another application is called LANDMARC (indoor location sensing using active RFID) which uses extra fixed location reference tags to help location calibration and adopts the KNN method to calculate the location of the RFID tags.

Unlike conventional RFID systems, UWB is based on sending ultrashort pulses with a low duty cycle, and it utilizes a wide swath of the radio spectrum to transmit a signal for a much shorter duration. Some field experiments have been introduced in [17] where the Ubisense system is considered. It is a unidirectional UWB location platform with a conventional bidirectional time division multiple access control channel. It works by creating sensor cells throughout buildings or collections of buildings, with each cell including at least 4 sensors or readers. The readers receive location data from tags and send it to the Ubisense Smart Space software platform. Although these systems are able to localize users with reasonable accuracies, they require the installation of additional sensors, which are subjected to large-scale deployments.

2.1.3 WLAN-Based

Due to the widespread adoption of 802.11x wireless LAN as a common network infrastructure, Wi-Fi-based localization techniques have been explored in order to convert the detected electromagnetic signals into a measureable metric such as distance and angle for location determination [16]. For instance, [7] applied the Motley-Keenan propagation model, which is convenient to estimate the distance between neighboring sensors from received signal strength (RSS) measurements. Another type of measure-

ment is a propagation time measurement, such as time-difference-of-arrival (TDOA) and time of arrival (TOA).

The TOA or TDOA related methods rely on estimating the propagation time of signals between a transmitter and multiple receivers, such as one-way propagation time measurements and round-trip propagation time measurements. The one-way propagation measurements determine the time difference between the time the signal is sent from transmitter and the signal is received. The round-trip measurement finds the difference between the time the signal is sent from a sensor and the time of receiving the returned signal at the same sensor. Once TOA or TDOA measurements are collected, the distance between two sensors can be approximately calculated by considering the speed of signal propagation (i.e. the speed of light). Then there is a geometrical circle centered on each receiver with an estimated distance between transmitter and receiver. In the ideal case, propagation measurements at three receivers would define an intersection point of the circles as where the transmitter locates. Then based on the coordinates of the three receivers, the coordinates of the transmitter could be calculated. However, in order to get a satisfying position estimation, both approaches require perfect time synchronization, which would add cost to sensors by demanding highly accurate clocks or a sophisticated synchronization mechanism. Also, they are hampered by multipath propagation, non-line-of-sight (NLOS) signal paths, and other impairments.

RSS-based fingerprint positioning, [5, 20, 30], is the most widely used technique in the literature. This technique consists in having an RSS radio map that can be used to characterize the Wi-Fi radio coverage at different positions and then finding a target's position by comparing the target's RSS sample measurements with the radio map.

A traditional algorithm used to estimate the location computes the Euclidean distance between the vector of signal strength estimations of the mobile and each of

those for each entry in the radio map. The coordinates associated with the fingerprint that has the smallest distance are regarded as the estimate of the position. RADAR [5] is an in-building user location and tracking system, which uses a KNN method in order to find the closest match between the observed RSS vector and those in the radio map. The author in [5] also considered the signal propagation model and added the wall attenuation factor (WAF) and floor attenuation factor into account. The Horus system [31] adopts a joint clustering technique and probabilistic method for location estimation. Each reference coordinate is regarded as a class, and a location is chosen when its likelihood is highest and the error distance is minimized. The experiments in [31] also indicate that increasing the number of sample measurements at each reference point would improve the accuracy, because more data would improve the estimation for mean and standard deviation. A grid-based Bayesian location-sensing system is introduced in [30] which investigates an area of interest in their office building to achieve localization and tracking information with 1.5m accuracy over 50 percent of the time. Other advanced nonparametric algorithms such as kernel-based one have been introduced to determine the relationship between the sample RSS and the location fingerprints in the offline radio map. Several functions can be chosen for the kernel, with the most common one being the Gaussian kernel [20]. More details about the Gaussian kernel function can be found below.

2.2 RSS-Position Dependency Models

The key challenge in WLAN positioning is the determination of the dependency between the RSS and position, (i.e. relating the RSS radio vector from certain access points to a spatial position in 2D Cartesian coordinates). This is especially challenging in indoors due to multipath and shadowing conditions caused by the presence of walls, people, and other objects, which results in a different received signal strength at each position indoors that may be only loosely correlated with the inverse of the distance.

However, the overlapping coverage and the difference of received signal strength from multiple access points can still be used to describe the position information.

Basically there are two approaches to model the RSS-position dependency: radio propagation modeling [5, 14, 28] and fingerprint-based modeling [20, 29] .

2.2.1 Radio Propagation Modeling

In the indoor environment, radio propagation modeling approaches [5, 14, 28] capture the characteristics of signal propagation influenced by reflection, diffraction, and scattering of radio waves and they use simplified models relating the RSS measurements of the mobile device to the location information. One such model with the compensation of wall attenuation factor (WAF) can be found in [5]. It includes the effects of obstacles or walls between the access point (AP) and the mobile device. The received power can be obtained as:

$$P(d)[dBm] = P(d_0)[dBm] - 10n\log(\frac{d}{d_0}) - \begin{cases} n_W \cdot nWAF & n_W < C \\ C \cdot WAF & n_W \geq C \end{cases} \tag{2.1}$$

where d_0 indicates the reference distance, d is the mobile-AP separation distance and n is a coefficient characterizing the propagation in the environment. For example [16] in free path loss environment, we have $n = 2$. In indoor environments, this factor will be closer to 3. This parameter is rather important, as it may significantly change the estimation result. The parameter C is the maximum number of wall up to which the attenuation factor makes a difference and n_W is the number of walls between mobile and AP. In reality, the materials for walls and floors, the layout of rooms, and the location of objects have a significant effect on the path loss. Such uncertainty makes it difficult to find an explicit model applicable to general indoor environments. Also radio propagation modeling requires the exact knowledge of AP's location for

trilateration. This makes it impractical in large and ubiquitous deployments of Wi-Fi network.

2.3 Fingerprinting-Based Methods

Yet another localization technique is introduced as location fingerprinting [5, 19]. It works by constructing a form of radio map to characterize the RSS-position dependency for the area of interest in the offline phase and feeding new RSS measurements extracted from mobile device in the online phase to estimate the position.

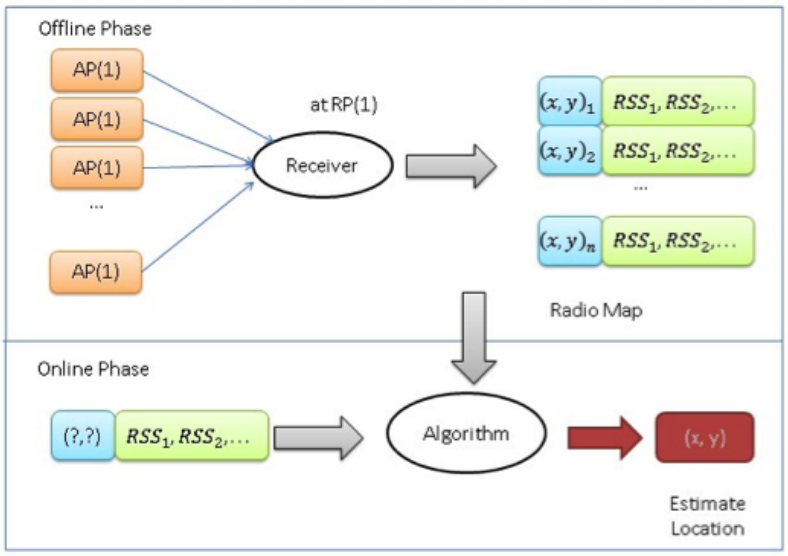


Figure 2.1: Two phases of fingerprinting: offline (training) phase and online (positioning) phase

A radio map R, as shown in Figure 2.2 , is a set of RSS measurements collected from different detectable APs at different reference points.

$$R \triangleq \{(p_i, F_{p_i})\}_{i=1,2,\ldots,L} \tag{2.2}$$

where p_i contains the coordinates of ith reference point, $F_{p_i} \triangleq [r_i(\vec{1}), ..., r_i(\vec{n})]'$ is a fingerprint matrix [20], L is the number of reference points and n is the number of offline phase time samples collected at each reference point. The vector $r_i(\vec{t}) \triangleq \left[r_i^1(t), ..., r_i^L(t)\right]$ contains the RSS measurements from L APs at time t at the reference point i located at the coordinates p_i. For instance, there is an area of interest as shown in Figure 2.2. After law enforcement placed a few reference points outside around the area, it can collect RSS measurements to form a fingerprint matrix F_{p_i} at each reference point, like at reference point P_1. And at the time stamp t, we will get a vector $r_i(\vec{t})$. Then during the time period n, we could get a radio map R as the yellow table shown in the Figure 2.2.

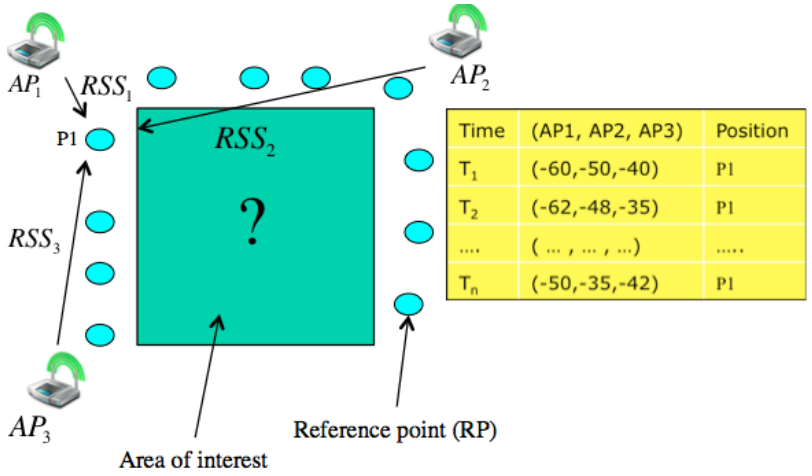


Figure 2.2: A radio map R collected from three detectable APs at reference point P_1; the yellow table shows the RSS measurements collected at different time stamp T_n

Note that n RSS samples collected at each reference point contain the information about the variation of RSS caused by time-varying multipath, shadowing and radio attenuation factors. However, there are often missing RSS measurements from some specific APs at some time samples. This will be further discussed in Chapter 4. Also in the real experiment, the number of detectable APs in the RSS measurements vector

$r_i(\vec{t})$ is determined by the number of APs detected during the following online phase. And there is a situation where the AP observed during the online phase may not show up during the offline phase at some reference point. This will be discussed in Chapter 4.

In the online phase, the mobile devices average RSS measurement vector $r(\vec{k}) \triangleq [\bar{r^1}(\vec{k}), ..., \bar{r^L}(\vec{k})]$ after a time period k is used to compute position estimation $\hat{p} = f(r(\vec{k}), R) = \sum_{i=1}^{L} p_i \omega(r(\vec{k}), F_{p_i})$, where $r(\vec{k})$ is the observed RSS measurement vector from the criminal device and the weighting function $\omega(.,.)$ provides the weight associated with the reference point, and the function $f(.,.)$ provides a mapping between the online observed RSS measurements and the radio map. In order to determine the weighting function and the mapping, there are two common and efficient approaches.

2.3.1 K-Nearest Neighbor Technique

The K-nearest neighbor (KNN) technique, which is the simpler of the two techniques we will discuss, was first introduced in [5]. In this approach, the mean value of the RSS measurements are used instead of the fingerprint matrix. And the distance between mobile device and the reference point is measured by Euclidean distance,

$$d^2_{KNN}(r(\vec{k}), F_{p_i}) = \| \bar{r(\vec{k})} - \bar{\vec{r_i}} \|^2 \tag{2.3}$$

where $r(\vec{k}) \triangleq [\bar{r^1}(\vec{k}), ..., \bar{r^L}(\vec{k})]$ is the target's average RSS measurement vector after a time period k and $\bar{\vec{r_i}} = \frac{1}{n} \sum_{\tau=1}^{n} r_i(\vec{\tau})$ is the average offline RSS measurement at p_i. Then the coordinates of RPs are listed such that $d^2_{KNN}(r(\vec{k}), F_{p_i})$ increases to yield the sequence $\{p(1), ..., p(L')\}$ and the $K < L'$ RPs are chosen to estimate the position of mobile device by averaging the coordinates of the K nearest RPs.

$$\hat{p} = \frac{1}{K} \sum_{i=1}^{K} p(i) \tag{2.4}$$

2.3.2 Fingerprinting-Based Technique

The second approach considers the minimum mean squared error (MMSE) [23] of position estimation.

$$\hat{p} = \underset{\hat{\boldsymbol{p}}}{\operatorname{argmin}} E\left\{(p-\hat{p})^T(p-\hat{p})\right\} \tag{2.5}$$

In this case, the MMSE estimate is conditioned on the RSS measurements, so

$$\hat{p} = E\left\{p \mid r\right\} = \int p f(p \mid r) dp \tag{2.6}$$

The key challenge in this estimation problem is that the posterior density $f(p \mid r)$ is unknown and must be estimated from the fingerprint. An ideal method to estimate posterior density is directly obtained from the fingerprint without the assumption of any prior statistical forms.

According to Bayes Theorem,

$$f(p \mid r) = \frac{f(r \mid p) f(p)}{\int f(r \mid p) f(p) dp} \tag{2.7}$$

The prior density $f(p)$ represents the knowledge of the environment conveyed by the reference points. Assuming the reference points are uniformly spread throughout the environment, a simple approximation is

$$f(p) \approx \frac{1}{L} \sum_{i=1}^{L} \sigma\left(\parallel p - p_i \parallel\right) \tag{2.8}$$

Where $\sigma(\cdot)$ is the Dirac delta function.

Subsituting (2.8) into (2.7), we obtain

$$\begin{aligned} f(p \mid r) &\approx \frac{f(r \mid p) \sum_{i=1}^{L} \sigma\left(\parallel p - p_i \parallel\right)}{\int f(r \mid p) \sum_{i=1}^{L} \sigma\left(\parallel p - p_i\right) dp} \\ &= \frac{\sum_{i=1}^{L} f(r \mid p_i) \sigma\left(\parallel p - p_i \parallel\right)}{\sum_{i=1}^{L} \int f(r \mid p_i) \sigma\left(\parallel p - p_i \parallel\right) dp} \\ &= \frac{\sum_{i=1}^{L} f(r \mid p_i) \sigma(\parallel p - p_i \parallel)}{\sum_{i=1}^{L} f(r \mid p_i)} \end{aligned} \tag{2.9}$$

Substituting (2.9) into (2.7),

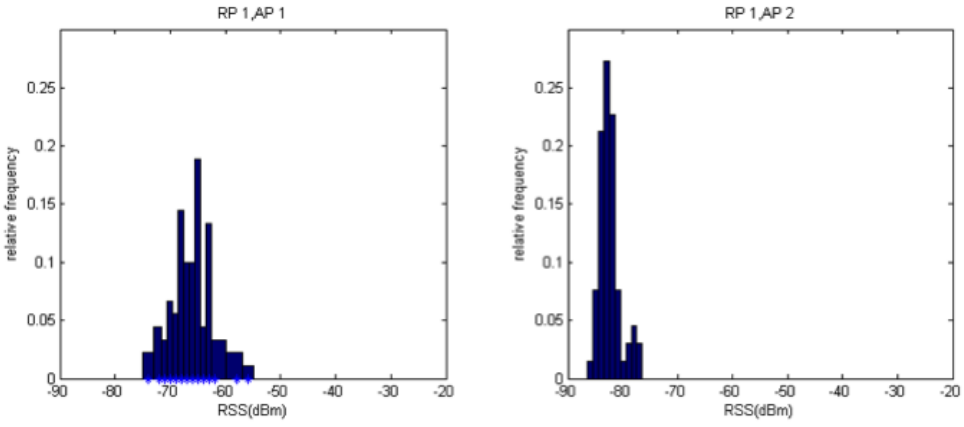
$$\begin{aligned}\hat{p} &\approx \int p \frac{\sum_{i=1}^{L} \sigma(\| p - p_i \|)}{\sum_{i=1}^{L} f(r \mid p_i)} dp \\ &= \frac{\sum_{i=1}^{L} \int p f(r \mid p_i) \sigma(\| p - p_i \|) dp}{\sum_{i=1}^{L} f(r \mid p_i)} \\ &= \frac{\sum_{i=1}^{L} p_i f(r \mid p_i)}{\sum_{i=1}^{L} f(r \mid p_i)} \end{aligned} \tag{2.10}$$

where we define the weight

$$\omega_i(r) \triangleq f(r \mid p_i) \tag{2.11}$$

Then, the MMSE estimation problem is reduced to estimating the likelihood density $f(r \mid p_i)$. Several methods are introduced to estimate the probabilistic weightings [8, 31], including a histogram-based one and a kernel-based one.

- The histogram-based method mainly relies on the structure present in the data. From Figure 2.3, we know the histogram based approach is not applicable to the likelihood estimation problem of this case because RSS distributions at fixed locations vary in timescale and do not generally follow parametric forms [9] .



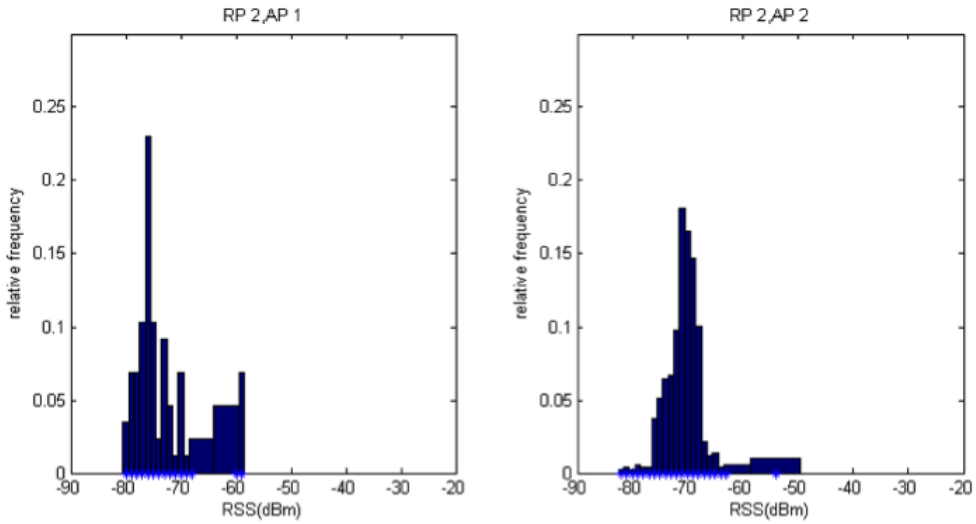


Figure 2.3: The variation of signal strength distribution from different access points at different reference points: the RSS measurements were collected across the access point AP1 and AP2 subsequently at reference point RP1 and RP2 for around 2 minutes

- As an alternative, the choice of a kernel-based method in [20] has been shown to provide more acceptable accuracy in the experiments. This method uses a kernel density estimator (KDE) to estimate the density: given a set of independent and identically distributed RSS samples $\left\{ r_i(\vec{\tau}) \mid \tau = 1, ..., n \right\}$ at reference point p_i, the kernel density estimate is given as

$$f(r \mid p) \approx \frac{1}{n\sigma} \sum_{\tau=1}^{n} K(\frac{r - r_i(\vec{\tau})}{\sigma}) \tag{2.12}$$

where $K(.)$ is the kernel, which is a symmetric function that integrates to one. Often a Gaussian kernel is employed. The σ is a smoothing parameter called the bandwidth, which controls the region of influence of each offline value. Large smoothing parameters capture the global structure of the density and when $\sigma \to \infty$ the density would be under fitted, whereas small values increase the detail and when $\sigma \to 0$ the density would be over fitted. In practice, this

parameter is determined based on the specific environment. And the τ is used as the index of time during the offline phase of the system.

Using a Gaussian kernel, the weighting function becomes

$$\omega_i(r) \triangleq \frac{1}{n}\sum_{\tau=1}^{n}\frac{1}{(2\pi)^{L/2} \mid \sum_{r_i} \mid^{1/2}} exp(-\frac{1}{2}(\vec{\bar{r}} - r_i\vec{(\tau)})^T \sum {}_{r_i}{}^{-1}(\vec{\bar{r}} - r_i\vec{(\tau)}))) \quad (2.13)$$

where $\vec{\bar{r}}$ stands for the vetor of the average RSS measurements of the criminal device, $\sum_{r_i}$ is the kernel bandwidth matrix and it is defined as $\sum_{r_i} = (\sigma^{\star}_{r_i})^2 I_{L^{\star}}$ where$I_{L^{\star}}$ is $L^{\star} \times L^{\star}$ identity matrix and $\sigma^{\star}_{r_i}$ is the bandwidth of the kernel. In practice, the choice of the kernel bandwidth parameter is not trivial and is generally data-dependent. For a Gaussian kernel, the ideal bandwidth minimizing the asymptotic mean integrated error between the estimated and true density is given in [20] as

$$\sigma^{\star}_{r_i} = (\frac{4}{2L^{\star}+1})^{\frac{1}{L^{\star}+4}} \hat{\sigma_{r_i}} n^{\frac{-1}{4+L^{\star}}} \quad (2.14)$$

where $\hat{\sigma_{r_i}} = \frac{1}{L}\sum_{l=1}^{L^{\star}}(\sigma^{l}_{r_i})^2$ is the average of the estimates of the marginal variances.

2.4 Access Points Selection

AP selection in fingerprinting-based location estimation can also be challenging. Estimation of a position in a two-dimensional space requires measurements from at least three APs. Due to the wide deployment of APs, the dimension of the measurements is generally greater than the minimum of three which increases the complexity of the WLAN positioning algorithm. Thus reducing the number of required APs wil not only improve the speed of positioning but also provide a better power efficiency and reduce the storage requirement. Also, a set of RSS measurements from detected AP is a description of the distance between the RP and the AP as well as the topology

of the environment in terms of the influence of obstacles and human activities. Therefore, subsets of the detected APs may report correlated and duplicated measurements. Clearly we want to select the APs with lowest correlation. These motivate the need for AP selection in the preprocessing step. Youssef et al. [31] used a joint clustering technique to reduce the computational requirements by clustering APs from which the receiver gets the strongest signal. Kushiki et al. [20] and Chen et al. [9] selected the most discriminative APs, defined as the ones that make most of the contribution to distinguish the places where the measurements be collected, and minimized the correlation between the detected APs in order to reduce the redundancy and maximize information gained from the APs. The method of AP selection used in this thesis is multiple discriminant analysis (MDA) with the goal of finding the optimum projection that can separate the RSS pattern among different locations [13, 15].

The MDA is the generalization of Fishers linear discriminant [13], which is used to transform the data that are best represented in the least squared sense [26]. The MDA is used when there are more than two classes. In our case, the classes can be regarded as the different reference locations and the objective of MDA is to find the collection of APs with enough entries that are useful to discriminate between the RSS at different locations.

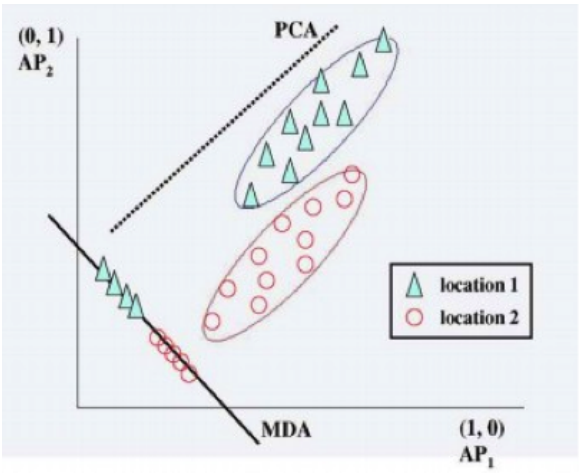


Figure 2.4: MDA based AP selection [15]. Two data sets respectively collected at location 1 and location 2 where there are two access points, AP1 and AP2.

A simple example is provided in Figure 2.4. The plot represents two different distributions of RSS measurements from two different APs at different two locations. If we want to decrease the dimensions, a generic AP selection method might just provide either the x axis oriented selection or the y axis oriented one. For both of these options, the original distribution would overlap on either the x axis or the y axis. But the two sets of data could be clearly classified while projecting original RSS measurements onto the line determined by MDA. Assuming $\vec{r}$ represents the original RSS data, the discriminant components (DCs) should be extracted via the MDA projection matrix A as follows,

$$Y = A^T \vec{r} \tag{2.15}$$

where $Y = [y_1, ..., y_D]$ is the projected data, $A \in \Re^{L' \times D}$ represents the MDA projection matrix, D refers to the number of DC chosen and $\vec{r} \in \Re^{L'}$ represents a measured RSS vector among L' APs.

In MDA, there are two scatter matrices called "between-class" (S_B) and "within-class" (S_W) matrices.

$$S_W = \sum_{r=1}^{L} \sum_{X \in l} (X - \mu_r)(X - \mu)^T \tag{2.16}$$

$$S_B = \sum_{r=1}^{L} n_r (\mu_r - \bar{\mu})(\mu_r - \bar{\mu})^T \tag{2.17}$$

where X stands for the RSS measurements collected during a period of time l, $\mu_r = \frac{1}{n_r} \sum_{X \in l} X$ this mean is the average of RSS from $L^\star$ APs at the rth location, $\bar{\mu} = \frac{1}{L^\star} \sum_{r=1}^{L^\star} \mu_r$ is the global mean, and n_r is the number of samples at the rth reference location. As defined in [26] , S_W measures the closeness of the samples within the locations, while S_B measures the separations between locations. In order to maximize the between-class measure and minimize the within-class measure, one way to maximize the ratio [13] :

$$\hat{A}_{MDA} = \underset{\boldsymbol{A}}{\operatorname{argmin}} \frac{\left|A^T S_B A\right|}{|A^T S_W A|} \Rightarrow (S_B - \lambda_i S_W) A_i^\star = 0 \tag{2.18}$$

where $A_i^\star$, the columns of $\hat{A_{MDA}}$, are the generalized eigenvectors of $S_W^{-1} S_B$ corresponding to the eigenvalues λ_i, which are ranked in decreasing order [15]. As a result, a satisfying matrix A would lead to DCs that have high spatial separation and low temporal variation.

A proper value of D, the number of DCs, could provide a graceful balance between accuracy and complexity. Actually we could decide the number of DCs by calculating the cumulative percentage of eigenvalues obtained in MDA as follows:

$$\frac{\sum_{i=1}^{D} \lambda_i}{\sum_{l=1}^{L^\star}} \geq \eta \tag{2.19}$$

where the eigenvalues $\{\lambda_1, ..., \lambda_{L^\star}\}$ are obtained from 2.18. Each λ_i represents the contribution of each y_i in the projected space. The greater value of λ_i is, the more

information that y_i would contain. Then we set threshold λ that would represent the most percentage of contribution that the determined DCs contain, and in the experiment it is set to be 0.92 . Once the projection A is determined, both online measurements $\vec{r}$ and $F(p_i)$ should be adapted to the same space, which are denoted as $\vec{\vec{y}}$ and $M(p_i)$. So 2.13 turns out to be:

$$\omega_i^\star(r) \triangleq \frac{1}{n}\sum_{\tau=1}^{n}\frac{1}{(2\pi)^{L/2}\left|\Sigma_{r_i^\star}\right|^{1/2}}exp(-\frac{1}{2}(\vec{\vec{y}}-\vec{r_i^\star(\tau)})^T\Sigma_{r_i^\star}^{-1}(\vec{\vec{y}}-\vec{r_i^\star(\tau)})) \tag{2.20}$$

2.5 Chapter Summery

This chapter gives a brief overview of wireless-based location systems serving for LBS. It also discusses two kinds of RSS-position dependency models that are used to describe the relationship between the received signal strength and coordinate position. Two fingerprinting methods, KNN and Gaussian kernel-based probabilistic techniques are described in details, as they are implemented in later experiments. Finally, MDA-based access point selection for improving estimation performance and reducing the computing complexity is discussed. This technique is used by the proposed positioning system for post hoc indoor localization.

CHAPTER 3

EXPERIMENTAL DEPLOYMENT AND ANALYSIS

In this chapter, in order to evaluate our fingerprinting-based method on the post hoc indoor localization problem, we set up an environment to test the procedure of determining the crime mobile device's location at the time of crime as described in Section 1.3. We select a residential area where four apartment buildings are adjacent to each other. Each building has three floors and contains nine apartments (3.1), making 36 apartments in total. In the experiments, two rooms in one family's apartment are selected as the area of interest with the goal of determing in which room the mobile device was located at the time of crime. Twelve testing "crime position" have been set in the two rooms as ground truth, which will be used to calculate the error distance in the estimated position.

Figure 3.1: Residential Environment (26 Brittany Manor Dr, Amherst, MA 01002)

Imagine that a cyber-crime has been committed in the area of interest, and law enforcement investigates. The experimental procedure to validate the investigative process described in Chapter 1 will proceed, as follows:

- Collect a vector of RSS measurements for each potential "crime position" in the area of interest, either in room A or room B. Each vector dataset is regarded as the one that law enforcement could extract either directly from the mobile device itself after the time of crime or from nearby access points.

- Set reference points around the area of interest; for each, collect a vector of RSS measurements to form the fingerprinting radio map post hoc. This step is to simulate the step that the law enforcement would take to collect data to determine the device location.

- For each reference point, apply the methods of Chapter 2 to calculate the weight that is used to describe the closeness between the RSS vector corresponding to

the reference point and the criminal device. If the value of the weight is high, it means that the target was close to the reference point.

- Sum up the weighted coordinates of each reference point to get the targets estimated location.

Per Section 1.3, we assume that the device used to collect the RSS measurements is located with the same altitude as all the APs are and that all of the detectable APs are the same in terms of hardware. We also only consider stationary wireless devices in our work, which means that all of the data is collected while the device was not in motion. Finally, the impact of the experimenter's human body on the signal attenuation while conducting the experiment is ignored.

3.1 General Setup

3.1.1 The Buildings

The experimental data sets used in this thesis are collected from an apartment complex in the southern part of Amherst where there are four buildings (e.g. Apt 121, Apt 119, Apt 115 and Apt 117) located closely to each other as shown in Figure 3.1. Each of the four buildings has nine families in total with an average of 3 people per family. In contrast to many evaluations in the existing literature, however, this thesis does not use an artificially constructed setup. Instead, the existing WLAN communication infrastructure in this area is utilized. Up to 85 access points can be observed in the experimental area and 28 of them can be detected in all four buildings. Hence, there is a sufficient number of access points in the experimental area.

Each family's apartment is around 750 square feet and has the room layout illustrated in Figure 3.2. Recalling that a key goal is to determine the room in which the device was located at the time of the crime, two bedrooms are selected in one

apartment and marked as A and B-referring to possible criminal scenes (as shown in Figure 3.1).

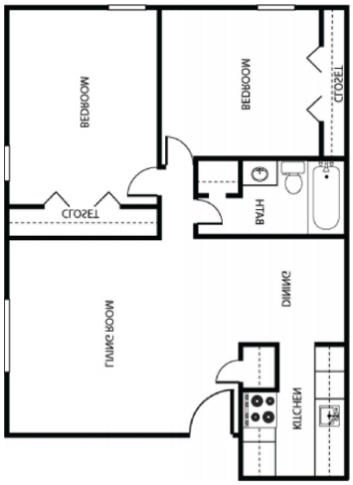


Figure 3.2: The apartment layout of the region of interest

3.1.2 Measurement Apparatus

RSS measurements used in this thesis were collected using a Dell Inspiron 1545 laptop with a Pentium processor, an external TP-Link TL-WN722N network adapter, an external portable GPS, and the Windows 7 operating system. RSS readings were obtained through publicly available wardriving software, Kismet-2011-03-R2 [3].

Kismet passively scans 802.11 channels, captures the wireless frames from the local interface (the TP-Link adapter in this case), which supports monitoring mode, measures the received signal strength, and reports GPS coordinates when integrated with a GPS device. It generates several log files including a .gpsxml which contains all of the information required. The RSS measurement is reported as integers in the range of (-100, 0) in unit of decibels relative to 1 milliwatt (dBm).

A sample set of measurements in Kismets .gpsxml output format are shown in Figure 3.3. This file includes the information about the detected access point's MAC address, the timestamp when capturing the measurement, the coordinate location of

the access point, the speed of the device's movement and the received signal strength. When conducting the experiment, there are around 5000 samples from different observable access points within two minutes in one output.

3.1.3 Testing/reference Point Placement

This thesis studies the stationary user scenario: we regard the position of the criminal mobile device as fixed during the commission of the crime and, naturally, assume that law enforcement would stand still to collect the reference point measurements.

t15-Kismet-20120504-12-27-33-1.gpsxml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE gps-run SYSTEM "http://kismetwireless.net/kismet-gps-2.9.1.dtd">

<gps-run gps-version="5" start-time="Fri May  4 12:27:33 2012">

    <network-file>/home/hao/Desktop/Kismet/Data/Kismet-20120504-12-27-33-1.netxml</network-file>

    <gps-point bssid="00:25:9C:5F:C0:64" source="00:25:9C:5F:C0:64" time-sec="1336148854" time-usec="341494" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-83" noise_dbm="0"/>
    <gps-point bssid="00:13:10:80:F4:CC" source="00:13:10:80:F4:CC" time-sec="1336148854" time-usec="362997" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-88" noise_dbm="0"/>
    <gps-point bssid="00:18:E7:E5:55:0E" source="00:18:E7:E5:55:0E" time-sec="1336148854" time-usec="366735" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-73" noise_dbm="0"/>
    <gps-point bssid="00:14:D1:35:AC:6A" source="00:14:D1:35:AC:6A" time-sec="1336148854" time-usec="381746" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-76" noise_dbm="0"/>
    <gps-point bssid="00:18:E7:E5:55:0E" source="00:18:E7:E5:55:0E" time-sec="1336148854" time-usec="398995" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-73" noise_dbm="0"/>
    <gps-point bssid="90:27:E4:8D:97:7E" source="90:27:E4:8D:97:7E" time-sec="1336148854" time-usec="421620" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-87" noise_dbm="0"/>
    <gps-point bssid="78:CA:39:43:CC:1D" source="78:CA:39:43:CC:1D" time-sec="1336148854" time-usec="564990" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-72" noise_dbm="0"/>
    <gps-point bssid="78:CA:39:43:CC:1D" source="78:CA:39:43:CC:1D" time-sec="1336148854" time-usec="665236" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-71" noise_dbm="0"/>
    <gps-point bssid="20:4E:7F:0C:59:62" source="20:4E:7F:0C:59:62" time-sec="1336148854" time-usec="720252" lat="42.347782" lon="
    -72.528915" spd="0.000000" heading="0.000000" fix="3" alt="53.551998" signal_dbm="-72" noise_dbm="0"/>
    <gps-point bssid="GP:SD:TR:AC:KL:OG" time-sec="1336148854" time-usec="767660" lat="42.347782" lon="-72.528915" spd="0.000000"
    heading="0.000000" fix="3" alt="53.551998"/>
```

Figure 3.3: A screenshot of .gpsxml, one of Kismet's output log files

A test point for a given trial corresponds to the spot where the criminal mobile device was used. Since these are inside the building we fail to record their location by the mounted GPS. Instead, the testing points for different trials are chosen to lie on a uniform grid, with each marked as "T" in Figure 3.4. Thus it is easy for us to associate a location with each test point in the building so as to establish the "ground truth" against which our measurement algorithm will be measured. As shown in Figure 3.4, six different testing points were placed in each room, with a distance between the two closest points of 1.5 meters.

The collection of reference points (RPs) used on each trial are also chosen from a uniform gird and are shown as 'R' in Figure 3.2. The distance between the two closest RPs is 3 meters.

Table 3.1: The Latitude/Longitude Coordinates of Each Testing Point

Testing Point	Latitude	Longitude	Testing Point	Latitude	Longitude
TP 1	42.3477222	-72.5289444	TP 2	42.3477360	-72.5289444
TP 3	42.3477460	-72.5289444	TP 4	42.3477222	-72.5289167
TP 5	42.3477360	-72.5289167	TP 6	42.3477460	-72.5289167
TP 7	42.3477222	-72.5288889	TP 8	42.3477360	-72.5288889
TP 9	42.3477460	-72.5288889	TP 10	42.3477222	-72.5288611
TP 11	42.3477360	-72.5288611	TP 12	42.3477460	-72.5288611

Table 3.2: The Latitude/Longitude Coordinates of Each Reference Point

Reference Point	Latitude	Longitude	Reference Point	Latitude	Longitude
RP 1	42.3476389	-72.5289722	RP 11	42.3477778	-72.5288333
RP 2	42.3476667	-72.5289722	RP 12	42.3477778	-72.5288056
RP 3	42.3476944	-72.5289722	RP 13	42.3477778	-72.5287778
RP 4	42.3477222	-72.5289722	RP 14	42.3477778	-72.52875
RP 5	42.34775	-72.5289722	RP 15	42.3477778	-72.5287222
RP 6	42.3477778	-72.5289722	RP 16	42.34775	-72.5287222
RP 7	42.3477778	-72.5289444	RP 17	42.3477222	-72.5287222
RP 8	42.3477778	-72.5289167	RP 18	42.3476944	-72.5287222
RP 9	42.3477778	-72.5288889	RP 19	42.3476667	-72.5287222
RP 10	42.3477778	-72.5288611	RP 20	42.3476389	-72.5287222

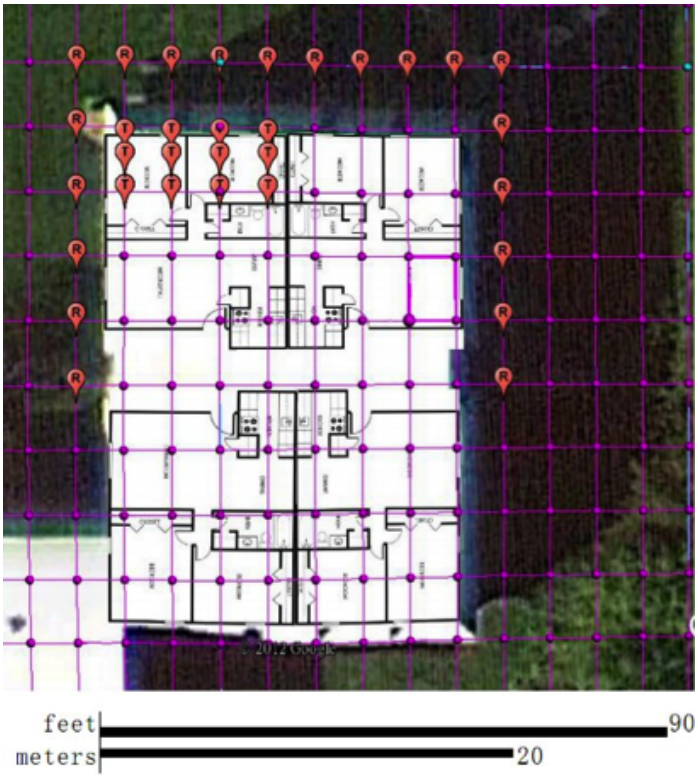


Figure 3.4: Experimental area: 'R' points represent RPs, 'T' points represent TPs.

3.2 Data Sets

3.2.1 Preprocessing

Before validating and comparing the fingerprinting-based algorithms based on the collected data, we need to consider how to convert the data into useful matrices, including translating the .gpsxml file into a MATLAB readable file, selecting the number of useful access points, and handling missing or multiple data at certain timestamp.

Per Section 3.1.2, the raw log file of the Kismets output is not human readable and not applicable to analysis and calculation. So the raw log is processed to produce a .txt file that just contain the BSSID (basic service set indentification), timestamp,

latitude, longitude and the received signal strength. Such a file is then easy for MATLAB to convert into a matrix.

A great challenge is processing the set of noisy and unreliable measurements to produce the proper set of pre-processed input for the algorithm.

For a particular AP, the received signal strength generally decreases as the distance to the AP increases. Thus, the RSS measurements collected at the testing points for nearby APs can offer more data sets used to describe the environment than measurements collected from more distant points; hence, these nearby will provide will provide fewer outages during measurement and hence provide greater utility.

Therefore, we filter the raw data by a certain percentage of RSS availability. This percentage of availability should be balanced according to the fact that a high percentage would result in fewer sets of measurements, and a low percentage would result in high redundancy. Both of the situations would negatively impact the estimation result. The useful set of RSS measurements in this experiment is set to the availability of RSS that has at least 65 percent of the possible time samples. Also, we found that sometimes there are more than one measurement at the same timestamp from the same AP, and we consider these should be individually added up when calculating the availability of RSS.

We need also to deal with the situation where there are some missing RSS measurements from spefic APs. As defined in Chapter 2, a fingerprint matrix is $F_{p_i} \triangleq \left[\vec{r_i}(1), ..., \vec{r_i}(n)\right]'$, where p_i stands for ith reference point, n is the number of timestamps recorded at the reference point, and $\vec{r_i}(t) \triangleq \left[r_i^1(t), r_i^2(t), ..., r_i^{L^\star}(t)\right]$ contains the RSS measurements from $L^\star$ APs at time t at the reference point p_i.

In the experiment, $L^\star$ is the cardinality of the set of APs observed at the *crime mobile device*'s point of view. Thus, for the fingerprint matrix collected at the reference point, sometimes there are missing RSS measurements at time t from the specific AP $l^\star$. It seems reasonable to set the missing data to be -99 dBm since the AP is

not detectable. But in order to avoid singularities while estimating the location, we set the missing one to be (-100+2*rand) dBm instead. And the rand represents the uniformly distributed pseudorandom numbers with a range of [0, 1]. Moreover, there are some timestamps when there are multiple RSS measurements collected from the same access point. In this case, we set the RSS value in the data to be processed to the average value of the measurements collected from the access point.

3.2.2 Data Collection

After processing the raw measurements, we obtain empirical estimates at different times from which we need to build the RSS value database F_{p_i}, $i = 1, 2, ..$ at position p_i. A sample set of data for a given F_{p_i} is shown in Table 3.1 for a given TP or RP.

Table 3.3: WLAN RSS Sample Database Profile

Timestamps	AP 1 [dBm]	AP 2 [dBm]	AP 3 [dBm]	AP 4 [dBm]	...
1	-98.69	-71	-78	-78	...
2	-98.69	-76.5	-77	-78	...
3	-98.69	-74.3	-78	-91.52	...
4	-98.69	-70	-65.5	-98.69	...
...	...	...	...	...	...
t	-66	-62.3	-79	-78	...

For the RSS data collected at a testing point, we build a 1 by $L^\star$ vector that contains the mean RSS value for each of the $L^\star$ detected access points. Associated with the coordinates of each reference point (partly shown in Table 3.2), a radio map $R \triangleq \{(p_i, F_{p_i})\}_{i=1}^{L^\star}$ is constructed.

Next, after pre-selecting access points and filtering with MDA (see Chapter 2), Table 3.3 shows the MAC addresses of the 5 APs that are most useful in estimating the location.

Table 3.4: AP MAC Adress List

AP NO.	MAC Adress
AP 1	F4:6D:04:8C:0C:38
AP 2	E0:91:F5:AF:78:F0
AP 3	C8:3A:35:3C:63:B8
AP 4	C4:3D:C7:8D:A6:7A
AP 5	C0:C1:C0:35:07:5D

3.3 Figure of Merit

In order to calculate the error, Haversine formula [2, 22] is used to calculate the distance between two latitude/longitude points. For two points on a sphere with lattitudes and longitudes, the distance between the two points is a spherical distance where the Haversine formula has been applied on .

3.4 Choose Experiment Parameters

3.4.1 The number of neighbors in KNN

As described in Section 2.3.1, KNN (K-Nearest Neighbor) will be used as a baseline for comparison with the more sophisticated fingerprinting algorithms. It is a deterministic approach that uses the average value of the RSS measurements from the radio map to estimate the criminal mobile device's position.

In order to find the proper value of K for the best performance of the KNN algorithm in this experiment, we pick four testing points (see Table 3.5) and estimate their location individually with the same radio map built from the measurements collected at all the reference points. For each testing point, we set the K to be from 3 to 20 with step of 1 and then get the average error distance depending on different values of K.

Figure 3.5 illustrates the effect of K on the average error distance. It shows that the average error distance can vary over a range of 7 meters if K is not chosen carefully, which is significant while locating a mobile device in the indoor environment for our forensics application. The performance improves as K increases to about 10. However, K should obviously not be set to a high value: this would imply that the algorithm takes so many reference points into account that the estimated location would lean to be the center of the area because the reference points are placed around the area of interest (and recall that the KNN algorithm does no weighting of the locations employed). So K is set to the optimal value of 10 for the following comparison of different algorithms.

Table 3.5: The Testing Points Used for Finding the Proper K in K-Nearest Neighbor

Testing Point	Latitude	Longitude	Testing Point	Latitude	Longitude
TP 1	42.3477222	-72.5289444	TP 2	42.3477360	-72.5289444
TP 3	42.3477460	-72.5289444	TP 4	42.3477222	-72.5289167

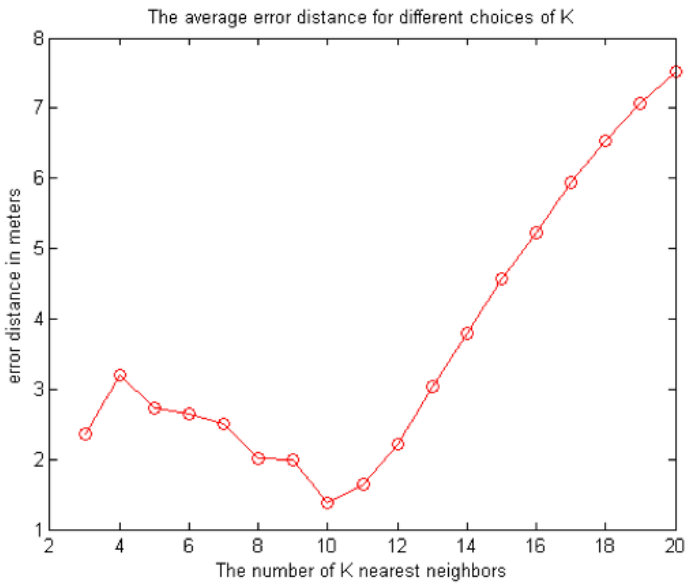


Figure 3.5: To Find the Proper K in KNN: the average error distance for different choices of K and K = 3,4,...,20

3.4.2 Kernel Width $\sigma^{\star}_{r_i}$

For the more sophisticated algorithm of Chapter 2 employing the Gaussian kernel, the bandwidth, also known as the window width, determines the width of the kernel or the region of influence of each reference RSS fingerprint matrix. A large window width captures the global structure of the density of the fingerprint matrix while a smaller one reveals more details. As $\sigma^{\star}_{r_i} \to 0$, the density estimate approaches the set of data centered at the set with a significant over-fitting. As $\sigma^{\star}_{r_i} \to \infty$, the density estimate approaches a uniform density with an under-fitting.

In pratice, the choice of the kernel bandwidth parameter is generally data-dependent, defined as $\sigma^{\star}_{r_i}$ in (2.14) according to [20] . In order to determine a proper width for our experimental environment, $\sigma^{\star}_{r_i}$ is replaced by $\alpha \times \sigma^{\star}_{r_i}$, and α in the interval [1,3] (with a step of 0.2) were considered.

Five sets of fingerprint matrices collected at the testing points (Table 3.6) have been used. Figure 3.6 illustrates error distance as a function of the kernel width for the testing data. Both too large kernel width and too small kernel width can impact the performance on estimation by inducing under-fitting or over-fitting of the original data. In practice, the experimenter can collect validation data first and then choose a proper kernel bandwidth for further use. In our case, the α is set to be 1.9.

Table 3.6: The Testing Points Used for Finding the Proper bandwidth for the Gaussian kernel

Testing Point	Latitude	Longitude	Testing Point	Latitude	Longitude
TP 2	42.3477360	-72.5289444	TP 1	42.3477222	-72.5289444
TP 3	42.3477460	-72.5289444	TP 4	42.3477222	-72.5289167
TP 7	42.3476944	-72.5288889			

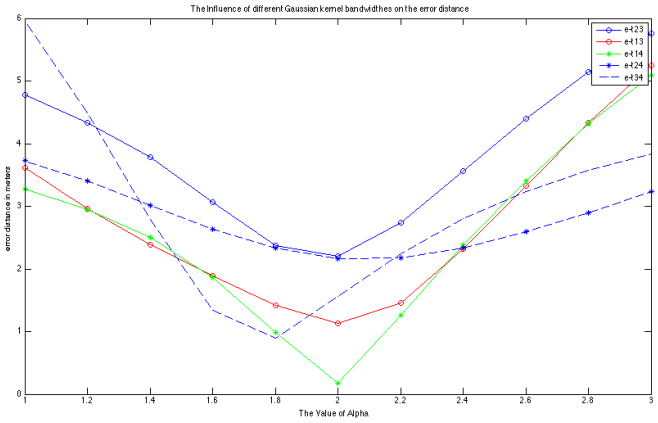


Figure 3.6: The influence of different Gaussian kernel bandwidth on error distance

3.4.3 MDA threshold η

Another parameter that is required to be determined experimentally is the MDA threshold η defined in (2.20), as it determines how many and which APs should be

used as the discriminant components (DCs) to carry enough information about the RSS environment. Therefore, the number of DCs is intelligently determined based on the characteristics of the RSS environment. In the experiment, we use three test points to consider the performance for different thresholds. The figure shows that the larger value of η is the larger error distance would be. And when the η is equal to one, the number of DC would be the number of detectable access points, in which case all the detectable access points would be used (which then, of course, negates the purpose of using MDA to refine the estimated result). So the threshold is set to be 92% for further use.

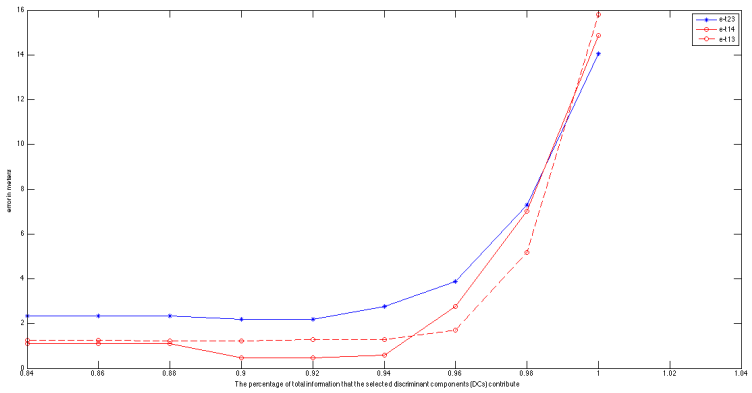


Figure 3.7: The MDA threshold η's influence on the performance: when η is .92 the error distance is minimized with around 1 meter and the variance around that value is flat and stable

3.5 Signal Propagation Analysis

The IEEE 802.11b/g/n WLAN standard uses radio frequencies in the 2.4 GHz band, which is license-free in most places around the world. However, it suffers from inherent disadvantages, such as the interference from microwave ovens, Bluetooth devices and other devices. Moreover, the indoor environment has unique properties that influence the radio signals used by the positioning systems and make the signal

propagation complicated. In particular, it may be weaken or obstructed by walls and human bodies and reach the receiver through multipath due to reflection, refraction, scattering and absorption of radio waves by inside structures. These prominent factors also cause the signal strength to vary over time. In addition [20] shows that the movement of human beings and other uncontrollable factors such as temperature and air movement create random effects of radio propagation inside the building. So positioning based on received signal strength patterns is difficult.

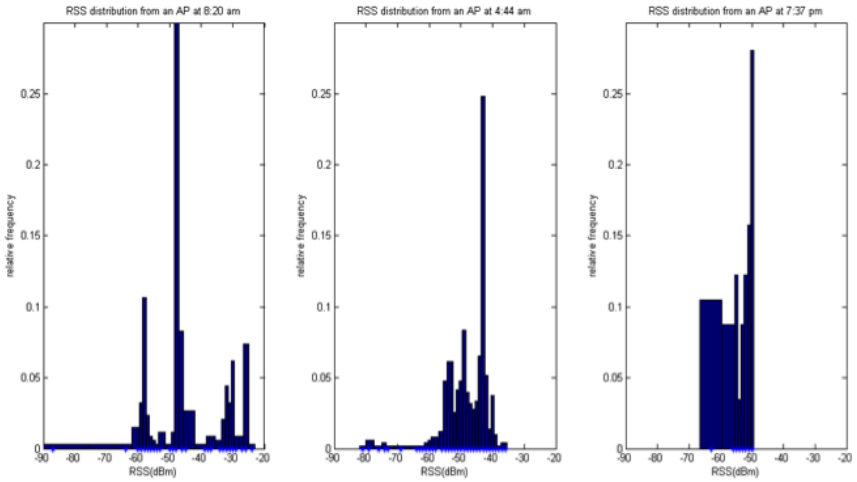


Figure 3.8: Examples of RSS distributions for the same AP over two minutes at the same location at 8:20 am, 4:44 am and 7:37 pm (from left to right)

Figure 3.8 shows the variation of the receive signal strength collected at different times. The samples for the histograms are produced by averaging approximately 2 minutes of RSS measurements collected from the same access point at the same location inside the building. As can be seen, the noisy signal can be as weak as -80dBm or as strong as -10dBm. From the receiver's point of view, such 70-dB range implies that there is a noisy environment indoors and the shape of the distribution is not predictable; that is, we cannot find a traditional distribution to describe it. Therefore,

the dynamic and unpredictable propagation characteristic of RSS in indoors means that an explicit formulation of the RSS-position relationship is impossible.

3.6 Compensation of Wall Attenuation

As shown in Figure 3.8, RSS measurements present quite unpredictable variation in the indoor environment. As described in Section 2.2.1, radio propagation models applied in the indoor environment can take into account the additional attenuation cased by walls. This motivates us to improve the estimation by taking into account the significant path loss caused by walls in indoor radio propagation. In the experiment, the wall attenuation factor (WAF) is used to adjust the signal measurements collected at the testing point, with the purpose of approximately compensating the signal strength loss caused by the exterior wall around the area of interest and thus aiming to improve the estimated results.

3.6.1 Calculate the WAF

We found the wall attenuation factor (WAF) [5, 25] by finding the average signal strength difference between the outside and the inside separated by the wall(s). In reality, the law enforcement could find a similar apartment with the same building infrastructure and measure the WAF. In our experiment, the average value of the difference is roughly 10 dB.

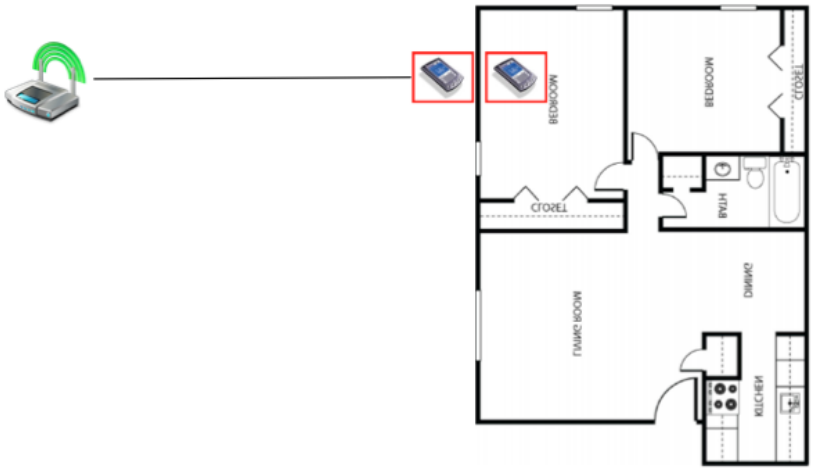


Figure 3.9: Method to calculate the WAF in an apartment

3.6.2 Pre-determine the Position of the Target and Modify the RSS measurements

For the criminal mobile device's viewpoint, the impact of the wall attenuation is different once the mobile device is located in a different room, because there may be different number of walls on the path of the radio propagation from the same access point. For example, there are two rooms illustrated in Figure 3.10. And the number of walls on the path of the radio propagation from the same AP1 to the criminal mobile device in Room B is one while the number in Room A is two. So the impact of the RSS attenuation at Room B would be less than the one in Room A from the same access point, AP1. Thus there would be different values of the WAF compensated in the two cases. However, there is no way for law enforcement to determine which room the criminal mobile device was in after the crime, and, in particular, this is one of the key goals of this work. So in order to make a correct compensation on the collected RSS measurements, it is necessary to pre-determine which room the target

has more probability to be in so that we can determine the number of walls between the outside and the inside.

The Mahalanobis distance has been used in this case to pre-determine the position. The Mahalanobis distance is defined as $D_M = (\vec{r} - \vec{\bar{F}})^T C_{-1}(\vec{r} - \vec{\bar{F}})$, where $\vec{r}$ represents the mean of the targets RSS data, $\vec{\bar{F}}$ represents the mean of the radio fingerprint collected at specific area, and C represents the covariance matrix of data in $\vec{F}$. According to its definition, we know that the Mahalanobis distance considers the probabilistic characteristic of the fingerprint map, such as its variation.

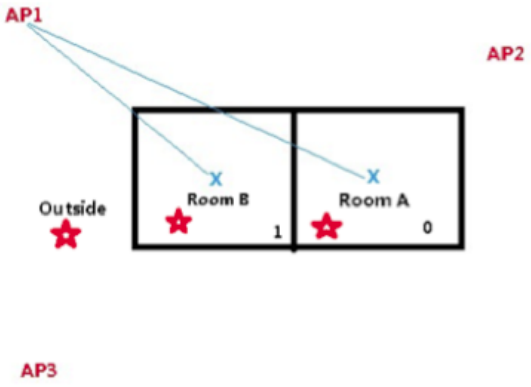


Figure 3.10: Pre-determine the position of the target

In reality, we may not get into the area of interest to collect the RSS measurements to describe each room's signal environment. So in the experiment, we choose a set of measurements collected close to but outside the room as the room's sample RSS measurements. Since it is a probabilistic way to estimate the position of the criminal device, there is a case where the values of the two distances are similar and we can not pre-determine the position of the criminal device based on a small amount of difference in probability. So we need to balance the two probabilities that the criminal device was in each of the two rooms. To do this, we define the admittedly ad hoc WAF coefficient as $\omega = -(D_{M_B} - D_{M_A})/(D_{M_B} + D_{M_A})$, where D_{M_B} (D_{M_A}) is the

Mahalanobis distance between the target and Room B ($RoomA$) and it is bigger than zero. This coefficient has following characteristics: (i) its absolute value is always less than one; (2) its absolute value is proportional to the probability that the criminal mobile device was in one of the rooms versus the other one.

For example, if the target happened to be in Room A, ω<0. Then the set of RSS value from AP2 should be modified by $w * WAF$ and the set of data from AP1 and AP3 should be modified by $w * WAF * 2$ (there are two walls on the path of propagation). Otherwise, ω >0, and the data from AP2 should be modified by $+w * WAF * 2$ and the set from AP1 and AP3 should be modified by $+w * WAF$.

To simplify the WAF compensation, we group the detectable access points in the neighborhood based on their locations into five groups, namely G_{115}, G_{117}, G_{119}, G_{121_L} and G_{121_R} (as shown in Figure 3.11). And we regard each group of access points as one and consider its relative location to the area of interest to find the number of walls in between the AP group and the suspect room.

After compensating for the effects of the walls and creating the "correct" data collected at the testing points, we will apply the fingerprinting-based technique on the modified data to estimate the position.

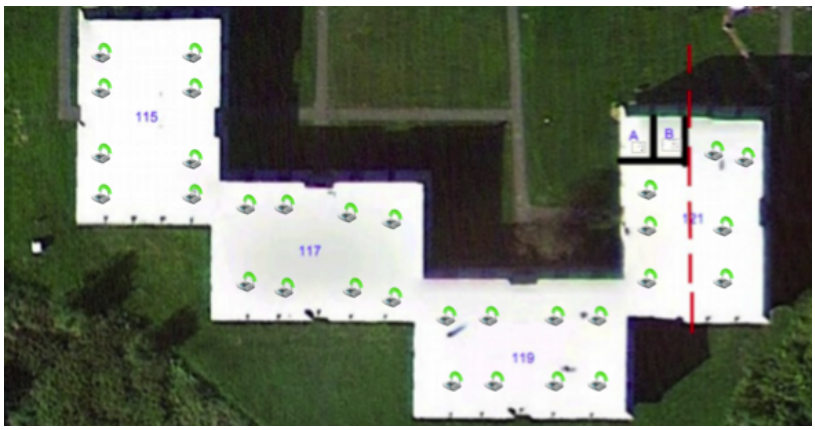


Figure 3.11: The groups of the detectable access points in the neigborhood

3.7 Performance Comparison

Throughout the above experiments, a set of optimal parameters that gives the best performance of KNN and Gaussian kernel-based techniques can be determined and is given in Table 3.4. The performance of the Gaussian kernel-based technique is then compared to the KNN technique in terms of the cumulative distribution function (CDF) of the position error.

Table 3.7: A set of optimal parameters for the Gaussian kernel-based technique applied

Number of APs pre-selected by RSS apperance	21
Number of discrimnant components	5
Kernel Width	$1.9^{*}\sigma^{\star}_{r_i}$
MDA threshold η	92%

The position errors of the four methods are compared in Figure 3.12 and Figure 3.13. In the experiments, there are 12 testing points involved in the comparison of the accuracy of the KNN algorithm and the Gaussian kernel-based algorithm. Clearly, the performance of the Gaussian kernel-based positioning algorithm outperforms the KNN algorithm; it improves significantly in terms of mean error (53%), maximum error (38%) and variance (66%). And according to Table 3.5 and Figure 3.13, the compensation of WAF slightly improves the performance of Kernel-based technique by decreasing the maximun error and the variance. Table 3.5 summarizes the statistics of the position errors of the algorithms.

Table 3.8: A set of optimal parameters for the Gaussian kernel-based technique applied

Method	Mean [m]	Max [m]	Variance [m^2]
KNN	2.99	5.36	2.53
MDA-projected Kernel method with noWAF	1.40	3.33	0.86
MDA-projected Kernel method with WAF=15	1.46	2.53	0.61
MDA-projected Kernel method with WAF=10	1.36	2.78	0.54

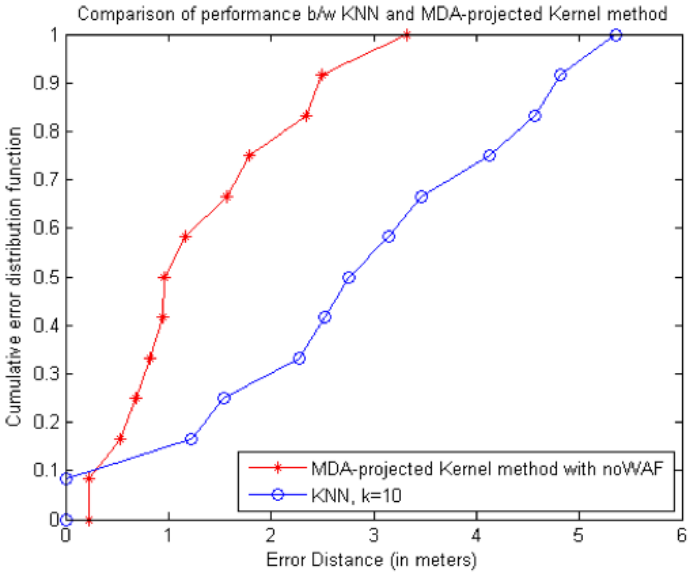


Figure 3.12: The comparison of performance between KNN and MDA-projected Kernel method

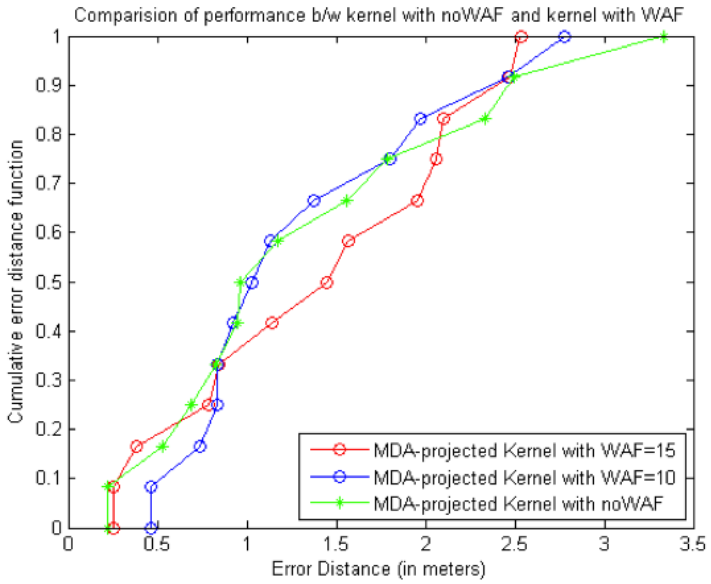


Figure 3.13: The comparison of performance between Kernel method with noWAF and the one with WAF compensation

3.8 Chapter Summary

This chapter documents experiments for determining the location of a crime mobile device to an accuracy which can distinguish between two adjacent rooms. In particular, the performances of two positioning algorithms - KNN and Gaussian kernel-based - are studied in the post hoc forensics scenario. In addition, the effect of the compensation of WAF in indoor environments has been considered. The experiments were carried out on the second floor of an apartment complex in the southern part of Amherst, MA.

The analysis of the noisy characteristics of the signal environment in indoors implies that the traditional signal propagation model is not sufficient to indicate the subtle dependency of the RSS measurements on the position. We next experimentally determine the optimal parameters used in the experiment to compare the two algorithms. As expected, the MDA threshold shows the trade-off of the number of

access points used in the Gaussian kernel-based approach, whereas the selection of K in KNN indicates the impact of the number of reference points on the performance of indoor positioning. Our comparison of the two positioning techniques shows that the Gaussian kernel-based positioning technique outperforms the KNN technique in terms of cumulative error distributions. We also observed a slight improvement of the performance due to the compensation of WAF.

CHAPTER 4

CONCLUSION AND FUTURE WORK

Modern cybercrimes leave behind digital evidence; hence, the ability to identify a perpetrator from such evidence - digital forensics - is of importance. In this thesis, we have studied the *post hoc fingerprint-based indoor positioning problem*. Using the motivating example of obtaining a search warrant, we consider determining the location of a criminal mobile device at the time of the crime with fine enough granularity to distinguish between two adjacent rooms. Our main contribution is an experimental evaluation of fingerprint-based methods in a residential area. Meanwhile, we also consider introducing a new factor due to wall attenuation and consider the effect of its incorporation on positioning the mobile device. While empirical performance studies of instantaneous indoor positioning systems based on fingerprinting have been widely presented in the literature, this study is the first to consider post hoc indoor prositioning problem from the forensics; in particular, the ability to determine the room in which a crime was committed using only the RF fingerprint of the device at the time of the cirme and RF fingerprints gathered from test points *outside* the building is the main contribution.

Unlike instantaneous geographical localization, which: (1) tries to locate the target while the target is in the area of interest, and (2) exploits a received signal strength (RSS) fingerprint matrixes at each of the reference points (RPs) taken *a priori* within the building of interest, our study considers placing reference points around the building of interest, where the target used to be. As in indoor instantaneous localization work, the RSS fingerprint matrixes are used to build a radio map, and, after the

radio map is done, an RSS-position dependency model is employed to map the vector fingerprint of RSS measurements of the criminal device to an estimated location. In this thesis, we have explored two RSS-position dependency models: the K-Nearest Neighbor (KNN) model and a Gaussian Kernel-based one.

In order to evaluate the performance of these two RSS-position dependency models in the post hoc indoor positioning problem, we set up an experiment to test the procedure of determining the crime mobile device's location at the time of crime. We selected a residential area where four apartment buildings are adjacent to each other and there are a large number of detectable access points. To simplify the experiment, we picked two adjacent bedrooms in one apartment as the area of interest. In each room, there were six testing points used to represent hypothesized crime mobile device locations in the room. The radio map was built for 20 reference points placed outside the apartment.

In our experiments, we found the indoor environment is very noisy, with around 70 dB of dynamic range and the shape of the received signal's distribution was not predictable enought to find a traditional distribution to describe it. Hence, the KNN model and a MDA-projected Gaussian kernel model are introduced and compared. We found better results from the kernel model, which demonstrated a mean error distance of 1.4 meters, which is precise enough to location the mobile device to a given room with sufficient certainty to establish the probable cause required to obtain a search warrant. Finally, we ameasured the wall attenuation factor (WAF) at roughly 10 dB per wall in our experiment. Since walls significantly impact the signal strength in indoor environments, we used the WAF to compensante the RSS measurements collected at the testing point with the goal of improving the estimation result. The experiment results showed that our current WAF compensation only slightly improves the performance of the MDA-projected Gaussian kernel model

4.1 Future Work

During the experiment, we found that the some reference points selected were far enough from the area of interest that the fingerprint data collected at such RPs did not aid in describing the RSS environment in the area of interest, thus adding unnecessary computation and an extra amount of RSS measurements. Therefore, a key feature of the *post hoc indoor localization problem* is the adaptive placement of reference points as data is collected and the area of interest gets smaller. It seems reasonable that one could use the current measurement information to guide the placement of successive reference points so as maximize the discrimination in answering a question of interest: for example, the apartment from which a wireless cybercrime was committed. As an example, consider our experiment. At the beginning we might hvae no idea about which apartment building the criminal device was located at the time of the crimbe, but instead might but just know it was in the four-building residential area. After some estimation, it would be easy to figure out in which building the crime was committed, which would then guide reference point placement to points around that building. This topic, which is one of the distinctive advantages for the investigator in the *post hoc indoor localization* problem versus standard instantaneous indoor positioning for location-based services (LBS), is of high interest for further study.

A second avenue for future research is the improvement of the algorithm to compensate the wall attenuation factor (WAF) on the crime mobile devices fingerprint database. The current way to compensate the WAF is based on the presumption of the mobile device's position. However, in the experiment, it worked especially poorly when the distance of the two positions to be distinguised were small. In that case, the WAF coefficient was so small as to result in little WAF compensation adjustment to the fingerprint vector. Another limited factor is that we assumed that there was only one floor in that area; that is, we ignored the significant impact of floor attenuation on the RSS measurements. Hence, we believe that there is potential for methods

which exploit detailed blueprints of buildings to perform detailed three dimensional attenuation compensation in the collection of RF fingerprints.

BIBLIOGRAPHY

[1] Computer Crime Section. http://www.ag.virginia.gov/CCSWeb/IDTheftFAQs.aspx.

[2] More Than 12 Million Identity Fraud Victims in 2012 According to Latest Javelin Strategy and Research Report . https://www.javelinstrategy.com/news/1387/92/1.

[3] Software: Kismet . http://www.kismetwireless.net/.

[4] Sources of identity theft: rogue access points and wardriving . http://biometricnews.typepad.com/biometric_news_and_inform/2010/12/sources-of-identity-theft-rogue-access-points-and-wardriving.html.

[5] Bahl, Paramvir, and Padmanabhan, Venkata N. Radar: an in-building rf-based user location and tracking system. pp. 775–784.

[6] Biaz, Saad, and Ji, Yiming. A survey and comparison on localisation algorithms for wireless ad hoc networks. *Int. J. Mob. Commun. 3*, 4 (May 2005), 374–410.

[7] Carlos Serodio, Luis Coutinho, Luis Reigoto, and Matias, Joao. A Lightweight Indoor Localization Model based on Motley-Keenan and COST .

[8] Castro, Paul, Chiu, Patrick, Kremenek, Ted, and Muntz, Richard R. A probabilistic room location service for wireless networked environments. In *Proceedings of the 3rd international conference on Ubiquitous Computing* (London, UK, UK, 2001), UbiComp '01, Springer-Verlag, pp. 18–34.

[9] Chen, Yiqiang, Yang, Qiang, Yin, Jie, and Chai, Xiaoyong. Power-efficient access-point selection for indoor location estimation. *IEEE Trans. Knowl. Data Eng. 18*, 7 (2006), 877–888.

[10] De Luca, Damiano, Mazzenga, Franco, Monti, Cristiano, and Vari, Marco. Performance evaluation of indoor localization techniques based on rf power measurements from active or passive devices. *EURASIP J. Appl. Signal Process. 2006* (Jan. 2006), 160–160.

[11] Djuknic, Goran M., and Richton, Robert E. Geolocation and assisted gps. *Computer 34*, 2 (Feb. 2001), 123–125.

[12] DRUGS, UNITED NATIONS OFFICE ON, and CRIME. Comprehensive Study on Cybercrime.

[13] Duda, Richard O., Stork, David G., and Hart, Peter E. *Pattern classification and scene analysis. Part 1, Pattern classification*, 2 ed. Wiley, Nov. 2000.

[14] Eisenbltter, Andreas, Geerdes, Hans-Florian, and Siomina, Iana. Integrated access point placement and channel assignment for wireless lans in an indoor office environment. In *WOWMOM* (2007), IEEE, pp. 1–10.

[15] Fang, Shih-Hau, and Lin, Tsung-Nan. Projection-Based Location System via Multiple Discriminant Analysis in Wireless Local Area Networks. *IEEE Transactions on Vehicular Technology 58*, 9 (Nov. 2009), 5009–5019.

[16] F.Evennou, F. Marx. Improving positioning capabilities for indoor environments with wifi.

[17] Fontana, R.J., Richley, E., and Barney, J. Commercialization of an ultra wideband precision asset location system. In *Ultra Wideband Systems and Technologies, 2003 IEEE Conference on* (2003), pp. 369–373.

[18] Hightower, J., Want, R., and Borriello, G. SpotON: An indoor 3D location sensing technology based on RF signal strength. *UW CSE 00-02-02,* (2000).

[19] Kjaergaard, Mikkel Baun. A taxonomy for radio location fingerprinting. In *Proceedings of the 3rd international conference on Location-and context-awareness* (Berlin, Heidelberg, 2007), LoCA'07, Springer-Verlag, pp. 139–156.

[20] Kushki, Azadeh, Plataniotis, Konstantinos N., and Venetsanopoulos, Anastasios N. Kernel-based positioning in wireless local area networks. *IEEE Trans. Mob. Comput. 6*, 6 (2007), 689–705.

[21] Liu, Hui, Darabi, H., Banerjee, P., and Liu, Jing. Survey of wireless indoor positioning techniques and systems. *Trans. Sys. Man Cyber Part C 37*, 6 (Nov. 2007), 1067–1080.

[22] Mao, Guoqiang, Fidan, Bar, and Anderson, Brian D.O. Wireless sensor network localization techniques. *Computer Networks 51*, 10 (2007), 2529 – 2553.

[23] Moon, T. K., and Stirling, W. C. *Mathematical Methods and Algorithms for Signal Pro- cessing.* Prentice Hall, 2000.

[24] Pan, Jeffery Junfeng, Kwok, James T., Yang, Qiang, and Chen, Yiqiang. Accurate and low-cost location estimation using kernels. In *Proceedings of the 19th international joint conference on Artificial intelligence* (San Francisco, CA, USA, 2005), IJCAI'05, Morgan Kaufmann Publishers Inc., pp. 1366–1371.

[25] Shreyas, Y., Seth, Shivam, and Agarwal, Rajat. Wireless network visualization and indoor empirical propagation model for a campus wi-fi network. *Journal of World Academy of Science 42* (2008), 730–734.

[26] Sundaram, R. Multiple discriminant analysis.

[27] Varshavsky, Alex, de Lara, Eyal, Hightower, Jeffrey, LaMarca, Anthony, and Otsason, Veljo. Gsm indoor localization. *Pervasive Mob. Comput. 3*, 6 (Dec. 2007), 698–720.

[28] Wang, Y., Jia, X., Lee, H. K., and Li, G. Y. An indoors wireless positioning system based on wireless local area network infrastructure. *Technology Including Mobile Positioning* (2003).

[29] Yin, Jie, Yang, Qiang, and Ni, Lionel M. Learning adaptive temporal radio maps for signal-strength-based location estimation. *IEEE Trans. Mob. Comput. 7*, 7 (2008), 869–883.

[30] Youssef, Moustafa, and Agrawala, Ashok. The horus wlan location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services* (New York, NY, USA, 2005), MobiSys '05, ACM, pp. 205–218.

[31] Youssef, Moustafa A., Agrawala, Ashok K., and Shankar, A. Udaya. Wlan location determination via clustering and probability distributions. In *PerCom* (2003), IEEE Computer Society, pp. 143–.