TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

ZHE PENG

CROWDSOURCING ERROR IMPACT ON INDOOR POSITIONING

Master of Science Thesis

Topic approved by:
Faculty Council of Computing and Electrical Engineering
Examiners:
Associate Professor Elena-Simona Lohan
Postdoctoral Researcher Helena Leppäkoski
Examiner and topic approved on 29.03.2017

ABSTRACT

Tampere University of technology
Master of Science Thesis, 44 pages
November 2017
Master's Degree Program in Electrical Engineering
Major: Wireless Communication
Examiners: Associate Professor, Elena-Simona Lohan
Postdoctoral Researcher, Helena Leppäkoski

Keywords: Crowdsourcing, Fingerprinting, Kullback-leibler Divergence (KLD), Access point (AP), Received Signal Strength (RSS)

Nowadays, with the rapid development of communication technology, plenty of new applications of 5G and IoT have appeared which requires high accuracy positioning skills. Wi-Fi based fingerprinting method is one of the most promising approaches for indoor positioning. Crowdsourcing is an appropriate fingerprint data collecting method on one hand. However, it is vulnerable to different kinds of crowdsourcing errors which add errors to the fingerprint database and can decrease the accuracy of positioning on another hand.

The main target of this thesis is to statistically analyze the behavior of the crowdsourcing data collected by different devices, and the effects of different kinds of intentionally or unintentionally added errors through MATLAB.

From the analysis results, it can be concluded that two different kinds of manually added errors perform complete differently. Data modified with all constant RSS values, out of author's expectation, achieves a decent accuracy similar to the original data. While data modified with only position error shows a behavior that the positioning accuracy drops with the increase of modified data proportion. Most of the distributions are closest to the Burr type XII distribution, which is particularly useful for modeling histograms.

PREFACE

The master thesis and the research have been carried out under the supervision of Associate Prof. Elena-Simona Lohan and Dr. Helena Leppäkoski in the laboratory of Electronics and Wireless communications at Tampere University of Technology, Tampere, Finland.

I would like to express my gratitude to my thesis supervisors, Associate Professor Dr. Elena Simona Lohan and Dr. Helena Leppäkoski, for their meticulous guidance and inspiration. I would especially thank Dr. Elena Simona Lohan for her considerate support throughout the whole period of the thesis work. I've learned a lot more than the thesis itself from Dr. Elena Simona Lohan.

I would also like to thank those who have devoted their effort in this thesis work, including some bachelor students of Tampere University of Technology who built the application for fingerprint data collecting, and those who have devoted time in the crowdsourcing data collecting work.

Finally, I would like to thank my family and friends, for their support to me so that I can complete my master's study smoothly.

Tampere, 4th Nov 2017

Zhe Peng

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

5G	5th generation mobile networks
AP	Access Point
AOA	Angle of Arrival
BLE	Bluetooth Low Energy
CDF	Cumulative Density Function
CDMA	Code Division Multiple Access
CFK	Cluster Filtered K-Nearest Neighbor
dB	Decibel
dBm	Decibel-milliwatts
FM	Frequency Modulation
GNSS	Global Navigation System
GPS	Global Positioning System
GEV	Generalized Extreme Value
GSM	Global System for Mobile Communication
HLF	Hyperbolic Location Fingerprinting
IoT	Internet of Thing
IR	Infrared Positioning
KLD	Kullback-Leibler Divergence
KNN	K-Nearest Neighbor
LOS	Line-Of-Sight
MAC	Media Access Control
MU	Mobile User
NB	Narrow Band
NLOS	Non-Line-Of-Sight

NN	Nearest Neighbor
PDF	Probability density function
PL	Pass Loss
RFID	Radio Frequency Identification
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RP	Reference Point
SS	Signal Strength
TOA	Time of Arrival
TDOA	Time Difference of Arrival
TUT	Tampere University of Technology
TSARS	Time and Space Attributes of Received Signal-Based Positioning Technology
UHF	Ultra-High Frequency
UWB	Ultra-Wide Band
UNB	Ultra-Narrow Band
US	Ultrasound Positioning
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network
WKNN	Weighted K-Nearest Neighbor

LIST OF SYMBOLS

D_{KL}	Kullback-Leiber Divergence value
P_{Δ}	RSS difference
a	shape parameter
b	scale parameter
K	shape parameter
t	time
A	RSS at reference point 1m from transmitter
D	distance parameter
P	received power/signal strength
W	noise parameter
c	velocity of light in free space
n	path-loss coefficient
r	RSS
u	sequence of training data
θ	shape parameter
μ	mean
σ	variance

1. INTRODUCTION

1.1 Introduction

Positioning is becoming a more and more significant part in wireless communication. The development of 5G and Internet of Things (IoT) in the near future has set new requirements, such as accuracy and reliability, for positioning technology. Mobile communication technology is rapidly developing as well as mobile devices, in which smartphones are especially pervasive to the whole world. Location-based services (LBS), at the same time, offer targeted services with geographic position, are also widely used in almost every field, and can provide extra value of exiting devices. Positioning with high accuracy is significant in 5G communication. Accuracy is required to be at one meter or even below [5]. Existing Global Navigation Satellite System (GNSS) and wireless fingerprinting positioning method can only achieve the accuracy of 3 to 4 meters [5]. With the development of Internet of things (IoT), a growing amount of applications which require location-based services have emerged [7]. Nowadays Global Navigation Satellite System is ubiquitous all around the world and is able to offer outdoor positioning services with good accuracy. However, it has a poor performance for indoor positioning, the accuracy of which is intensively affected by three main factors:

1. There is usually a large quantity of obstacles at indoor environment, such as doors, walls and floors, which causes serious blockage of the signal.

2. Multi-path effect is common for indoor environment, which causes large fluctuation of the signal.

3. The signal received from satellites in indoor scenarios is quite weak, thus the indoor operational carrier-to-noise ratio is fairly low.

Thus, other positioning methods should be considered when applying indoor positioning. There are some solutions, such as Infrared Positioning (IR) [16] and Ultrasound Positioning (US) [22], which are extremely accurate, but are not widely adopted and limited by the effective range. As mentioned in [33], about 70% of the positioning systems uses standard wireless network technologies, including Wi-Fi, Bluetooth, Radio Frequency Identification (RFID) and Ultra High Frequency (UHF). Among them, Wi-Fi is the one with the largest amount of existing infrastructure, and Wi-Fi fingerprinting-based approaches are the most popular solutions [23].

From the perspective of system topology, there are two types of positioning system as self-positioning and remote-positioning [20]. Mobile device acts as the measuring unit in

a self-positioning system. Some transmitters with known positions send signal to mobile device and the positioning is done through mobile device. On the other way around, in a remote-positioning system, the mobile device acts as the transmitter whereas some fixed measurement units receive the signal from the mobile device and calculate the position of the mobile device. There is always a requirement for the measurement unit, thus the advantage of a self-positioning system is with a cheap existing infrastructure and the advantage of a remote-positioning is with power efficient mobile device [20]. The priority of selection between these two systems depends on the real scenario in which cost may vary greatly.

In this thesis, Wi-Fi fingerprinting with RSS method is used for indoor localization. However, since it requires huge amount of fingerprint data to achieve high accuracy, the biggest challenge for Wi-Fi fingerprinting-based approach is to lower the cost and time of fingerprint data collecting.

Crowdsourcing, as a way to distribute the tasks to undefined crowd can be utilized to save labor cost and increase the data collecting efficiency [9]. During the process of crowdsourcing data collecting, erroneous data caused by different reasons intentionally or unintentionally will inevitably occur, which will decrease the accuracy of the positioning result and decrease the reliability of the positioning system. With appropriate quality-control of crowdsourcing data, the result can be greatly improved. The target of this thesis is to statistically analyze the behavior of different potential errors caused by crowdsourcing as well as the impact of erroneous data on the positioning system.

1.2 Thesis objectives

This thesis focuses on the crowdsourcing impact on indoor fingerprinting positioning accuracy. The specific objectives are as follows:

1. Collect fingerprint data of all floors in a building of TUT with different crowdsourcing devices by using an Android application.

2. Build a MATLAB project to simulate the Wi-Fi positioning of the building, and get Cumulative Density Function (CDF) of errors of the result as the positioning accuracy with different crowdsourcing dataset.

3. Manually add different types of errors to the dataset to simulate crowdsourcing errors.

4. Statistically analyze the performance of crowdsourcing data with and without errors by comparing CDF error curves, power maps and KL divergence result in MATLAB.

1.3 Author’s contribution

Author’s contributions are as follows:

- Author performed state-of-the-art review about RSS based indoor positioning methods and crowdsourcing.
- Author has collected some measurements to the fingerprint database, and used MATLAB code to convert the original downloaded json format file to sorted readable data.
- Author analyzed the fingerprint data by comparing positioning estimation results with different datasets as training data. The positioning result is shown in the figures of CDF error curves from Chapter 6.
- Author added synthetic error data to fingerprint data to analyze error impact on positioning result.
- Author utilized the Kullback-Leibler Divergence (KLD) to find the best distribution for histograms of different datasets and distribution for histograms of power map differences.
- Author has published two scientific papers [54][55] based on the measurements and analysis of this thesis.

1.4 Organization

Organization of the thesis is as follows:

Chapter 2 briefly introduces some available indoor positioning methods and metrics of positioning, and mainly focus on the explanation of basic principle of RSS fingerprinting-based approach.

Introduction of crowdsourcing is presented in Chapter 3. The content is about the basic meaning of crowdsourcing and how it is related to and used in location based service. Also, the main error sources in crowdsourcing for positioning are mentioned, as well as some scenarios of unintentional and intentional errors occurrence in crowdsourcing.

Chapter 4 is about the explanation of Wi-Fi positioning error calculation or how the accuracy of positioning is attained, and the algorithm used for positioning.

Chapter 5 explains the process of measurement campaign. It gives a brief introduction about the Android application used in data collecting and the cloud server used for data storing. Here, the procedure of erroneous data creation is also mentioned, and the further analysis is based on these data.

Then, the main analysis of data is presented in Chapter 6. Here, several different methods to analyze different crowdsourcing dataset and the results are shown.

Finally, Chapter 7 summarizes the thesis and presents the conclusions. The open challenges for this topic are also presented.

2. LOW-COST SCALABLE INDOOR POSITIONING METHODS

2.1 Approaches to indoor positioning

There are different available technologies for building an indoor positioning system as well as different methods for positioning estimation. Due to the limitation and complexity of indoor environment, the solution to build an indoor positioning system with high accuracy and stability remains open. This section presents a brief overview of indoor localization technologies and measurement techniques.

2.1.1 Technologies for indoor localization

There are a lot of wireless technologies that can be applied for indoor positioning and they can be sorted by the frequency they use and the transmit distance they can achieve. As long distance wireless technologies, Frequency Modulation (FM), Global System for Mobile Communication (GSM) and Code Division Multiple Access (CDMA) have been used for a long time.

FM is used in radio broadcasting and the frequency of the radio spectrum is usually from 87.5 to 108.0 MHz. FM signal has a good penetration ability and it can transmit through the wall easily, thus there is no complicated requirement for the receiver. But since FM signal has a long wavelength, signal strength does not vary drastically with the position change in short distance, thus it's not suitable for indoor positioning. There is one example in [2], the accuracy is only around 50 meters when the cumulative density function of error curve reaches 70%.

GSM/CDMA is used in cellular network communication. GSM is applied in Second Generation (2G) communication and CDMA is in Third Generation (3G) communication. The frequency GSM uses varies from 850 MHz to 1900 MHz and up to 2100 MHz in CDMA. Although the existing infrastructures of them fulfill the location based service requirement, the development of them in positioning area is limited by the heavy patent [3].

Wi-Fi, as one of the most ubiquitous wireless technology, is widely used in building to provide wireless network service. There are two license-exempt bands as 2.4 GHz and 5 GHz utilized in Wi-Fi [19]. Since Wi-Fi infrastructure exists in most building and the signal can cover most part of the whole building, and the mobile device such as mobile phone or laptop is available for everyone, indoor positioning with Wi-Fi technology can be implemented easily and without heavy cost. Thus, it has attracted plenty of research focus and it is one of the most promising method for indoor positioning.

ZigBee is a specification based on IEEE 802.15.4 protocol. It is used in short distance duplex transmission. It is characterized by low complexity, low power consumption, low cost and low transmission rate. It's usually used in automatic-control and remote- control area. Fang et al [4] has introduced a ZigBee indoor positioning method with good accuracy.

Bluetooth uses same band as Wi-Fi, and is a personal area network standard. Bluetooth low energy (BLE) is one technology which has lower power consumption and cost compared to classical Bluetooth. As mentioned in [10], propagation of BLE and WLAN signal are similar and positioning with BLE technology is completely feasible.

In Ultra-Wide-Band (UWB), pulses of very short duration are transmitted through high frequency band. The transmission of UWB does not interfere with other narrow band and carrier wave transmission [11].

Radio Frequency Identification (RFID) is a technology that has been widely used by companies in warehouse management for scanning and picking goods [12]. Also, it's used for identifying books in library. One problem in RFID-based positioning, which characterizes in fact most of the Received Signal Strength (RSS)-based positioning approaches, is that the RSS fluctuates easily with the dynamic variation of environment [15].

Narrow band IoT (NB-IoT) and Ultra-Narrow band IoT (UNB-IoT) are important brands of IoT and new technologies in IoT and 5G communication area as well. They are Low Power Wide Area Network (LPWAN) radio technology standards and have advantages as low power consumption requirement and can extend the battery life of devices [14]. The authors in [13] study the performance of UWB and Narrow Band (NB) propagation of indoor positioning. The result shows that both UWB and NB are promising technologies for indoor positioning. Some characteristics of the technologies mentioned above are listed in Table 1.

According to [19], there are two fields of indoor positioning methods: the first one is based on 2D model and the second one is based on 3D model. The previous one includes Bluetooth, ZigBee and Wi-Fi. They are some technologies network of which has already been widely distributed. The latter one includes infrared, UWB and ultrasonic.

Table 1. *Different technologies for indoor positioning*

Wireless Technology	Range	Dedicated Infrastructure	Power consumption	Disadvantages
FM	100 km	No	Low	Signal varies little in small distance
GSM/CDMA	100 m~10 km	No	Moderate	Highly patented
Wi-Fi	10-100 m	No (for most places)	High	High variance signal
ZigBee	10~100 m (line-of-sight)	Yes	Very low	Cover range is limited
Bluetooth	10 m	Generally, no	Moderate	Cover range is limited
UWB	4-20 m	Yes	Low	Cover range is limited
RFID	Usually 10 cm-1 m	Yes	Low	Cover range is limited

2.1.2 Measurement principles

According to [8][20], there are general four measurement principles for indoor positioning: Time of Arrival (ToA) or Time Difference of Arrival (TDoA), Angle of Arrival (AoA), Received Signal Strength (RSS) and hybrid techniques [49].

1. Time of Arrival (ToA) or Time difference of Arrival (TDoA)

ToA method measures signal's transmission time from the transmitter to the receiver. Then the distance between transmitter and receiver can be easily attained by simply multiplying transmission time by the speed of light.

$$Distance = c * ToA \quad (2.1)$$

c represents the speed of light in this equation. However, to get a high accuracy, wide bandwidth is required, which results in expensive hardware cost [2]. Instead of absolute time of arrival, TDoA method measures time difference between departing from a transmitter and arriving a receiver.

2. Angle of Arrival (AoA) or Angle Difference of Arrival (ADoA)

AoA method measures the transmission direction of received signal. Usually, it is implemented with an antenna array. By calculating the Angle Difference of arrival (ADoA) of

individual antennas, the incident angle of received signal can be estimated. But considering the impact of multi-path transmission in line of sight situation, it is still hard to get an accurate AoA result without other hardware device [2].

3. Received Signal Strength (RSS) and fingerprinting

RSS represents the power of received signal typically in dBm form. Basically, stronger RSS means a shorter transmission distance when the transmission power of transmitter is stable. From this aspect, RSS can be directly used as a distance parameter to estimate the distance, and then, trilateration method can be utilized to implement positioning. Trilateration is a conventional method for estimating position, which is used in GNSS. To achieve positioning, coordinates of three or more transmitters or Access Points (AP) and the distances between each AP and the mobile user (MU) are required [3]. The most important procedure is the measurement of the Signal Strength (SS), and convert it to responding distance with accuracy. For indoor positioning, because of multi-path fading and fluctuation of signal power, there is no stable linear relation between RSS and the transmission distance, thus, high accuracy cannot be typically achieved with trilateration. In general, TOA, AOA and RSS based trilateration methods are not available for non-line-of-sight (NLOS) environment [46]. To provide a better performance of indoor positioning, combinations of RSS and fingerprinting are proposed to offer better accuracy.

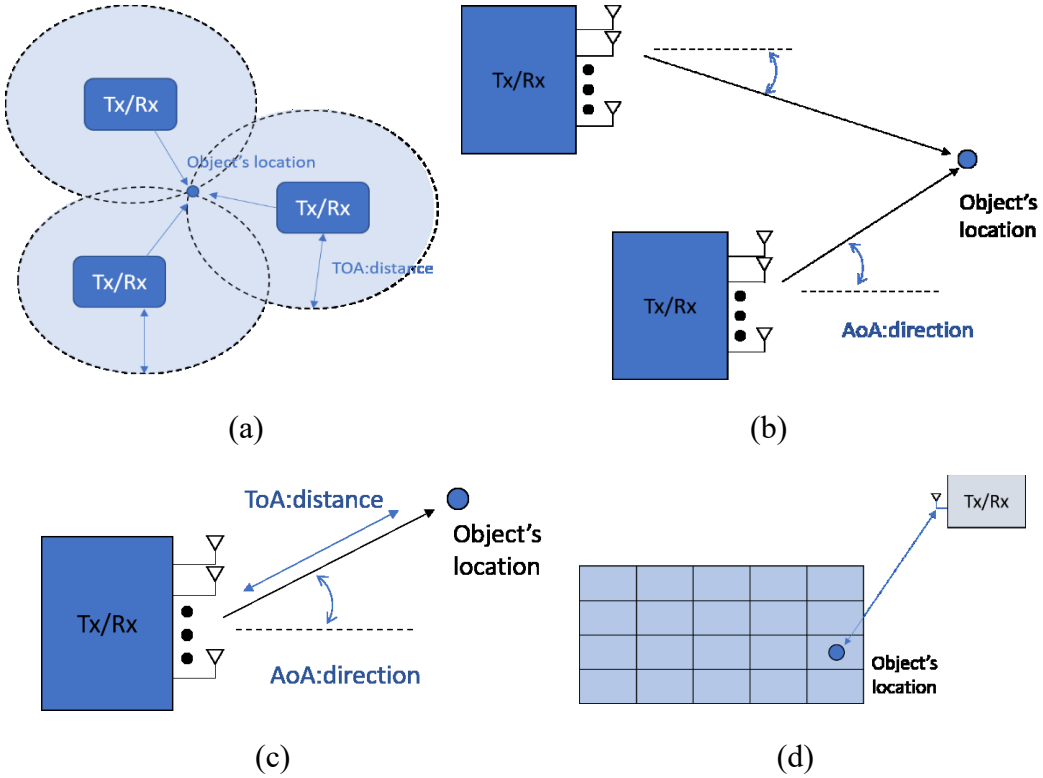

Figure 1. *Four basic approaches for indoor localization: a) Time of Arrival, b) Angle of Arrival, c) Hybrid ToA and AoA, d) Received signal strength and fingerprinting [8]*

4. Hybrid techniques

Hybrid techniques which combine ToA, AoA and RSS is possible. For example, hybrid ToA/AoA technique uses data from both ToA and AoA. It can reduce the requirement for nearby anchors [2] and positioning is possible with only one anchor. Authors in [7] have introduced a practical hybrid ToA/AoA appliance with only one anchor in an UWB positioning system. The above mentioned four basic measurement principles for indoor positioning are also shown in Figure 1.

Among these positioning methods, ToA and TDoA requires strict time synchronization and AoA requires access point which is equipped with special hardware to estimate the angle, while the hybrid method requires both. The distance between transmitter and receiver cannot be directly attained through RSS, and even if the location of the receiver keeps still, RSS can also vary for shadowing effects as shown in Figure 2. There are 8 RSSs in each subplot heard from an AP measured at different time but at the same measurement location. It is clear that the RSSs heard from all 4 APs fluctuate with time.

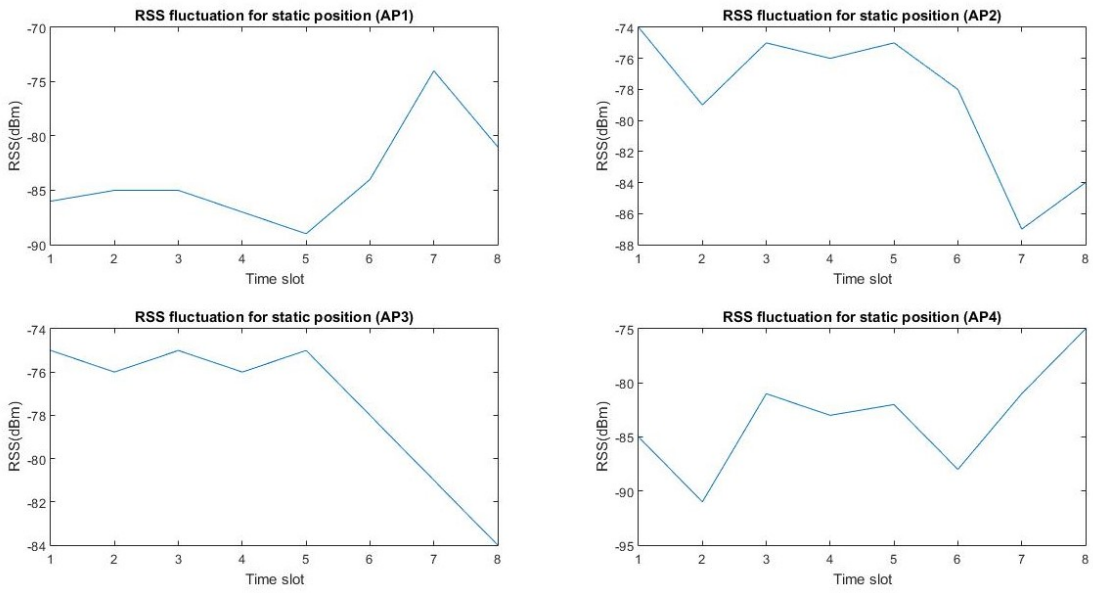

Figure 2. *Example of RSS fluctuation of static position with different time*

The above-mentioned effect may be also due to the non-stationary characteristic of the RSS value. Although RSS heard by each AP might vary with time, the mean value of RSSs of a same group of APs would not fluctuate as much as [43] mentioned.

Thus, to mitigate the error effects caused by RSS fluctuation, one available solution is to use a database with large amount of data as the fingerprint data, and RSS with fingerprinting is such another way to implement positioning. The idea is: if RSS of all locations are known, it is possible to create a power map of the building, which has each locations' RSS from all access points. For the estimation phase, by comparing the RSSs data collected by the user's device with the power map database, the data in which RSSs match

best can be used as the estimated result. RSS with fingerprinting is the cheapest method since it does not require other additional hardware than a smartphone. However, it cannot perform well in a dynamic environment since the fingerprint data changes with the environment. In addition, when the value of RSS does not vary considerably with the change of location, the accuracy will also be bad.

In addition, ToA/TDoA, AoA/ADoA, and hybrid ToA/AoA based technologies can be designated as Time and Space Attributes of Received Signal-Based Positioning Technology (TSARS) which is distinguished from RSS based positioning according to the classification done in [19]. The common feature of TSARS based positioning is using time and space attributes of received signal. In this way the classification of indoor positioning can be drawn as in Figure 3.

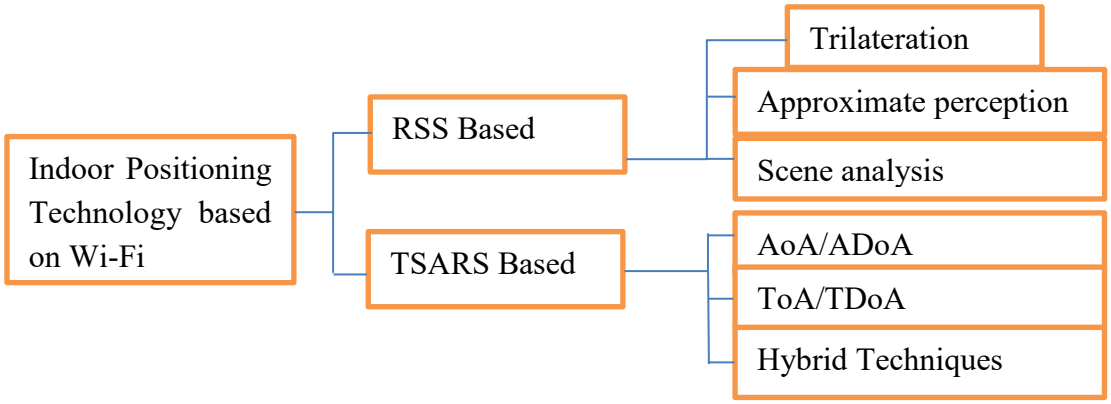

Figure 3. *Indoor positioning classification [19]*

From the perspective of [17], Wi-Fi indoor positioning algorithm can be sorted into three categories: proximity algorithm, triangulation algorithm and scene analysis algorithm. Triangulation algorithm is mentioned as ToA, AoA and hybrid ToA/AoA above. Proximity algorithm is similar with the RSS fingerprinting method, but it's more intuitive, which just uses the RP with highest RSS value as the estimated location. The third one, scene analysis algorithm is the data matching method and fingerprinting is one ubiquitous approach of it.

2.2 Access Point

An AP is a device which has the tasks of a centralized unit in a Wireless Local Area Network (WLAN), and it performs as a transmitter and receiver or simply called as transceiver in the WLAN. This transceiver connects a wired backbone LAN with wireless clients and provides wireless clients with wireless connections service.

According to [5][6], Multiple MAC addresses might come from the same location or the same AP since the AP can be with multiple antennas or a physical AP can support several MAC addresses. It is possible to remove some APs in training phase to mitigate the calculation complexity and at the same time provide good accuracy. In the measurement

campaign of this thesis, a large number of APs are heard, and some MACs are from the same AP. Besides, there might be some rogue APs, such as the hotspot of laptop or mobile device, which are also measured and can be one of the reason for having such large number of MACs in the building.

2.3 Fingerprinting

Fingerprinting-based positioning refers to the positioning approach using a database with collected data from known locations [18]. The collected data usually is the RSSs, but the devices used as transmitter and receiver varies with the communication technology utilized for positioning. No matter what technology is used, the basic process of fingerprinting is compatible for all.

There are two phases in fingerprinting method including offline training phase and online positioning phase [25]. The target of offline training phase is to build a fingerprint database which covers the positioning area. The fingerprint data is made up of coordinate of the location and the RSSs heard from all APs at this location as well as the Media Access Control (MAC) addresses of all available APs. Each location corresponds to a unique fingerprint data. Fingerprint data is collected at Reference Points (RPs), which are selected out from the indoor map and they are usually evenly distributed on the map to provide a good coverage of the positioning area.

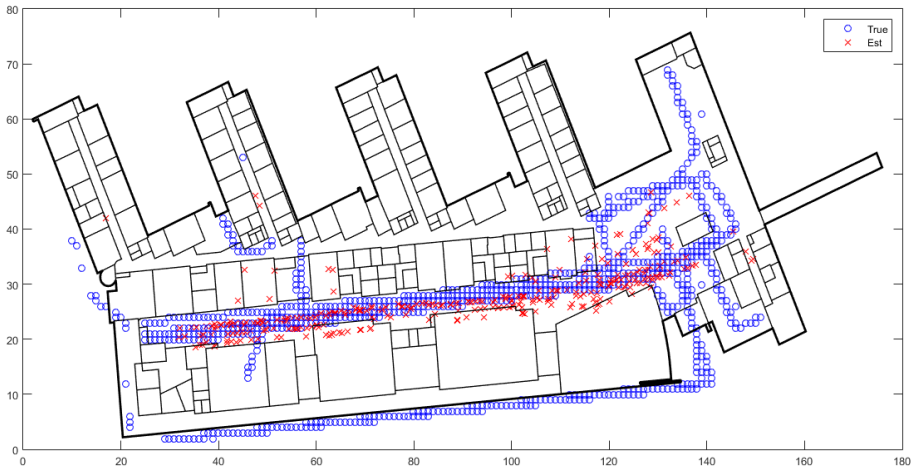

Figure 4. *Example figure of fingerprint data on floor map*

Then, the online positioning phase or estimation phase is conducted based on the fingerprint database. At this phase, user at a location collects RSSs heard from all available APs at that location, and the positioning is conducted by comparing the collected RSS measurements with the database with an algorithm [1]. Fingerprint data with closest RSSs will be selected out, and its coordinate is the estimated result. Large amounts of measurements and calculations are needed to guarantee good positioning accuracy [2]. Figure 4 shows

an example of the fingerprint data on the map from an overlook vision. Each blue circle represents for a fingerprint data collected in training phase, and it is point-wisely collected. Each red cross represents an estimated position. Figure 5 shows the training and estimation process of fingerprinting method.

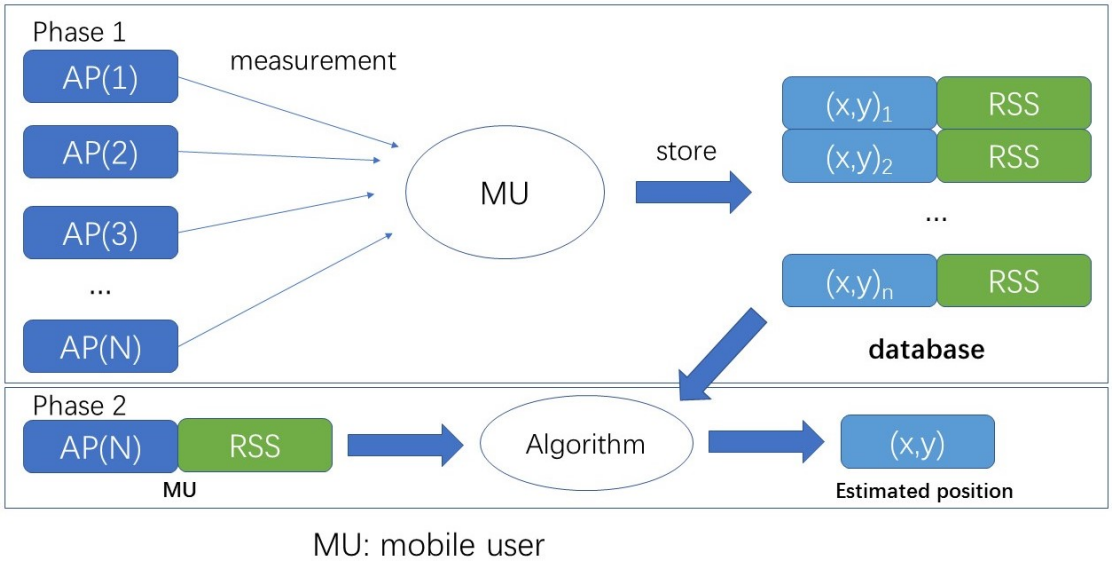

Figure 5. *Training and estimation process of fingerprinting method [1]*

In this thesis, the positioning estimation is done based on a log-Gaussian likelihood method [10]. Let's denote RSS_0 as one of the observed RSS values, u as the index of fingerprint data, and $RSS_{training}(u)$ as one RSS value from the training dataset.

$$F(u) = log\left(\frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{-\frac{\left(RSS_0 - RSS_{training}(u)\right)^2}{2\sigma^2}}\right) \quad (2.2)$$

The comparison is done under the premise that compared observed RSS and training data are from the same AP, which means with same MAC address. σ in the equation is a constant value representing the shadowing standard derivation. Here σ is a fixed value as 7 dB. After all values of RSS heard from APs are computed through this calculation, one final matrix of data can be attained by sum all log Gaussian likelihood values of one observed point.

$$F = \sum_{u} F(u) \quad (2.3)$$

By sorting the values in the matrix from highest to lowest, the first training point is selected out as the estimation result. To reduce the effect caused by noise, K Nearest Neighbor (KNN) method is used, which is widely used in data mining and machine learning [50]. Instead of simply using one best fit point as the estimation result, KNN takes the K best fit points out and exploits the average value of the K data as the estimation result. In

this thesis, 3 best estimation results are used, and the final estimation coordinate is the average of the 3 points coordinate.

RSS based fingerprinting method can offer a high accuracy for indoor positioning with existing infrastructure. Four RSS based fingerprinting methods are compared in [52], and all four methods can reach the accuracy of 2-3 meters for 90% of the estimation result. However, the attainable accuracy is to a great extend based on the amount of data in the training database. It's time consuming and laborious to collect data to build such a database which is also one of the biggest challenge for fingerprinting method. Thus, crowdsourcing, as a feasible solution to relieve the burden of site survey is selected here and utilized in the fingerprinting collecting.

2.4 Wi-Fi positioning metrics

Usually, accuracy is the main metric that we look at when evaluating a positioning system. Besides positioning accuracy, there are still some other benchmarks which are important to a positioning system. Thus, it's necessary to consider all metrics together when building a positioning system [20]. The metrics are as follows:

1. Accuracy/Measurement uncertainty

Still, accuracy is one of the most significant metric for a positioning system. It intuitively shows how well one positioning system performs. Accuracy is often represented by the mean distance error [19] which is the average Euclidean distance between true position and estimated location. An accuracy with smaller value of distance indicates better positioning result. Different systems have different requirements for accuracy, and the one with best accuracy may not be the best choice since all the facts should be considered. Measurement uncertainty is now sometimes used instead of accuracy, and it shows the quantification of a standard deviation [19].

2. Precision

CDF is usually used for measuring the precision of a positioning system. It tells about how well the accuracy a specific variable proportion of data can reach. The difference between accuracy and precision is that precision shows more detail about the positioning result, and the robustness of the system can be observed through precision rather than accuracy. Thus, in this thesis the analyze is mostly based on the precision of the system, but accuracy is still denoted.

3. Complexity

Complexity of a positioning system can be divided into three aspects: hardware complexity, software complexity and operation complexity [20]. Take Wi-Fi RSS based finger-

printing system for example, existing infrastructure greatly reduced the hardware complexity, and the Android or IOS based software also has low complexity. Usually the complexity is directly related to the cost of the system, which to a great extent decides whether this system is practical or implementable, thus it's also an important criterion.

4. Robustness

A highly robust system has the ability to function well even if error occurs. The robustness of RSS based system is mentioned in this thesis.

5. Scalability

The scalability of a positioning system represents its adaption to new environment, whether the system can resist the impact of space extension. For indoor environment, the further the distance between AP and mobile device, the worse it performs for positioning. Dimension of space is also the measurement of the scalability, usually with 2D and 3D spaces.

3. CROWDSOURCING FOR POSITIONING

This chapter will introduce the concept of crowdsourcing and about how it works in and is related with fingerprinting indoor positioning. The calibration issue and the existing challenge for crowdsourcing field are also referred.

3.1 Crowdsourcing

Crowdsourcing is a portmanteau of crowd and outsourcing. It refers to the process that tasks are outsourced to undefined crowd and solved through crowd's effort. As project's size expands and becomes increasingly complex, new paradigms and concepts including crowdsourcing are needed. Jeff Howe first introduced this concept in 2006: crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined network of people in the form of an open call (Howe, J., 2006) [4].

Since in this thesis, fingerprinting method is used in indoor positioning, there is a need for a database with large amount of data to ensure the accuracy of the positioning. Crowdsourcing is one efficient way to maintain the database, which saves time, reduces the workload and increases the efficiency of the data collecting job. Besides, since fingerprinting positioning accuracy can be drastically affected by the change of environment, data updating of fingerprint database becomes a significant task and crowdsourcing acts as one good solution to deal with data updating problem.

The biggest difference between crowdsourcing and outsourcing is that, in crowdsourcing, the work can be allocated to undefined public, whereas in outsourcing, tasks are distributed to experts or well-trained people [39]. Thus, as one of its biggest advantage, crowdsourcing is much cheaper and nevertheless the result could be as good as the outsourcing one if the result is appropriately filtered [39], and this is also the motivation of this thesis, to find methods to separate error data from the crowdsourced data.

There are two approaches for crowdsourcing as automated crowdsourcing and dedicated crowdsourcing. The difference between these two approaches is the way they report feedback data. In automated crowdsourcing, feedback is sent automatically through the device and without the aid of manual operation, while in dedicated crowdsourcing, feedback data is collected or supplemented with manual operation. In this thesis, dedicated crowdsourcing is used, and all crowdsourcing appears in the rest of the paper refers to dedicated crowdsourcing.

In crowdsourcing approach, Wi-Fi fingerprints are collected by multiple contributors, so each contributor only needs to collect a small amount of data which add up to total fingerprint data, and as the number of contributors increase, the effort each one needs to take will further decrease. In a positioning and data collection system used in this thesis, users also play the role of contributor. A mobile app can be used to collect fingerprints in an indoor positioning system. An Android app is used in this thesis and thus all mobile devices involved in this thesis are based on Android system. There is a feedback system in the android app which sends user feedback to the server in real-time. User can share fingerprint data after each positioning operation. With correct positioning result, user click the correct button and he or she records this measurement point as the correct one and save it to training data. If the result is incorrect, the user can click the correct position on the map, and the coordinate of this point on the map is recorded with the true RSS received from each available access point as one measurement point data. In this way, the training data will be continuously filled, and the positioning error will be decreased and get close to a threshold value. Since the work is distributed to unknown sources, the quality of work cannot be insured. Different errors may occur for various reasons.

First, different mobile devices report RSS differently because there is no strict range of RSS Indicator (RSSI) which is used for RSS measurements [34]. Thus, there is a scenario that the devices used to collect training data differ from the devices used as positioning devices, and it's hard to attain a good match when comparing the training data with the estimation data.

Secondly, the error may occur by operational accident. As explained previously, contributor needs to click the true floor and location on the map when it shows wrong estimated position. This manual operation error is inevitable but can be avoided as much as possible by improving the quality of the user interface (UI). The qualities for a great UI includes clarity, concision, familiarity, responsiveness, consistency, aesthetics, efficiency and forgiveness and all these are aimed at offering a user-friendly UI.

Besides, the error can be caused by the device itself when collecting data.

(1) With the distance from AP to mobile device increase, the received RSS from that AP will be weaker. Thus, far away APs can cause larger estimation errors whereas nearby APs can offer high accuracy estimation.
(2) Signal strength fluctuates for multipath effect and blockage caused by human body
(3) Usually the RSS is real-time measured, but if the received RSS is outdated due to the delay of scan, the error can occur [40].

3.2 Calibration in RSS-based positioning

The positioning accuracy with RSS-based localization system is affected by multipath and fading effects as well as temporal propagation dynamics such as temperature, humidity and movement of people [47], [48]. Thus, calibration for such system is needed to ensure the accuracy. Besides, when the device used for positioning differs from the one for training data collecting, calibration is also needed for the new device to be compatible with the existing radiomap.

3.2.1 Log-distance path model

As mentioned in [41], in RSS-based localization, log-distance path model is one of the mostly used PL models. The formula of this model is as:

$$P = A - 10n * \log_{10} D + W \quad (3.1)$$

In which P is the RSS value, D is the distance between transmitter and receiver, n is an environment coefficient and W is the noise parameter which includes natural noise, shadowing effects and RSS fluctuation. According to [41], A is the RP's RSS at 1m from the transmitter. Then it can be simply seen as the only parameter affected by the RSSI of mobile device. If the effects caused by A can be wiped out or at least decreased, then the diversity of different device can be mitigated.

3.2.2 Calibration-free methods

The idea of calibration-free method is to wipe out the effects of device dependent parameter A, and the simplest approach is to use difference of RSS instead of absolute values of RSS as the fingerprint data [34]. In this way, the new fingerprint data only consists 3 device-not-related parameters.

$$P_{\Delta} = -10n * \log_{10} D + W \quad (3.2)$$

However, the fact is that crowdsourcing devices have different value of A at same location, thus the parameter A cannot be calibrated between devices by simple subtraction. For crowdsourcing scenario, there is a requirement that the subtraction of RSS should be done between possible AP pairs heard by the same device. This has added more difficulty to data collecting process, and has set a minimum limit for data collected per device. Also, as the number of APs grows, the dimension of fingerprint data would be drastically increased [34]. One reference AP can be selected out to decrease the computing complexity. By subtracting the reference AP's RSS with all other RSS values, the fingerprint data size is narrowed down.

Besides above problems presented, because noise effect is amplified during subtraction, the differential fingerprinting will have a less accurate positioning result comparing to normal fingerprinting method regard less of device diversity [34].

As mentioned in [34], the Hyperbolic Location Fingerprinting (HLF) [26] and RSS ranking [27] method are other two methods aimed at reducing the device-dependent component, but both turn out to be not adoptable for some reason.

3.2.3 RSS data fitting methods

According to [34], the manual calibration and automatic calibration are two approaches in data fitting method. For manual calibration, no matter the relationships between the RSS of different devices are linear or not, there are various algorithms to create a mapping between different devices. But in all the algorithms, the user is required to collect some RSS data at some specific known location, which is not always feasible in real scenario, and is not suitable for a large number of devices. For automatic calibration, it's feasible to collect RSS data at unknown places but is with expensive computational fitting.

In this thesis, no calibration is adopted, thus the estimation result may be of larger error, averagely around 10 m's CDF error. But since the target of this thesis is to analyze the crowdsourcing impact, the comparison happens among all uncalibrated data and the result will not be greatly affected by calibration factor (possibility of influence caused by calibration is not excluded). Future work in the indoor positioning area could be to study the impact caused by calibration on crowdsourcing data.

3.3 Challenges for fingerprinting crowdsourcing-based indoor localization

Although crowdsourcing has relieved the burden of fingerprint data collecting, there are still some challenges for crowdsourcing based indoor positioning, and some are introduced by crowdsourcing itself. There are two main challenges as fingerprinting annotation and device diversity/heterogeneity [25].

The fingerprinting annotation is about how the coordinate information of the user is collected. There are two types of approaches as active fingerprinting crowdsourcing and passive fingerprinting crowdsourcing [25]. The active fingerprinting crowdsourcing is the traditional way of annotating fingerprints. The collector manually annotates the RP location with usually Cartesian coordinates, which is utilized in this thesis. One biggest problem is that it requires a precise floor/radio map to decrease the error of annotation made by the crowdsource contributor, and the accuracy of manual annotation is always limited. Another challenge is the intentional and unintentional mistakes made by the crowdsource contributor when reporting the coordinates. Passive fingerprinting crowdsourcing, as another annotation method, is implemented without user intervention. The movement track

of the user is recorded based on the sensors such as accelerometer and magnetometer on the mobile device. Compared to the active method, there is no requirement of an accurate map with high reliability. On the contrary, one physical map can be drawn with the combination of all measured trajectories [25][28][30]. There is an algorithm which automatically construct radio map based on crowdsourcing introduced in [29] and has presented a good accuracy performance. However, there is a privacy issue about passive fingerprinting crowdsourcing that the offline site survey process can cause some potential location privacy leakage [31].

Device diversity already exists without crowdsourcing method when fingerprints are collected by one device throughout the fingerprinting collecting process while users still use different devices for positioning. But with crowdsourcing, device diversity happens at the beginning of off-line measurement phase. Different mobile devices have different RSS measurement result of the same AP even if at the same location. Thus, calibration is needed to modify the RSSs received by different devices to a same range, and it increases the complexity of fingerprint database at the same time.

4. WI-FI POSITIONING ERRORS

Among the positioning metrics referred to in Chapter 2, positioning accuracy is normally the most important one. Error distance of positioning is used in this thesis as the accuracy. This Chapter will introduce the algorithm used for positioning and about calculation of positioning error distance.

4.1 Positioning algorithm

K-Nearest Neighbor (KNN) algorithm for indoor wireless local area network (WLAN) positioning is widely used [35]. The Euclidean distance can be calculated as follows:

$$D_i = \|r_i - r_u\| \quad (4.1)$$

r_i is the RSS of index i in the radio/power map, index i varies from 1 to the size of the radio map. r_u is the RSS from AP of u index as the estimation data. The idea of KNN algorithm is to find K fingerprints in the radio/power map database which offer K lowest value of D as D_{min}. After the K fingerprint data are determined, it's intuitive to choose the mean value of the coordinates of these K fingerprints as the positioning result:

$$\mathrm{C}(\mathrm{x},\mathrm{y},\mathrm{z}) = \frac{1}{K} \cdot \sum_{i}^{K} C_i(x,y,z) \quad (4.2)$$

$\mathrm{C}(\mathrm{x},\mathrm{y},\mathrm{z})$ is the coordinate of the result as the positioning location, and $C_i(x,y,z)$ is the ith KNN data.

Besides basic KNN, there are some improved algorithms such as Weighted KNN (WKNN) [36], Enhanced Weighted K-Nearest Neighbor (EWKNN) [36] and Cluster Filtered KNN (CFK) [36]. In WKNN, different neighbors have different weights and thus the result is not the simple mean value of all K neighbors. Some noisy fingerprint data might be presented with low weight value and in this way the effect of noise can be decreased. But it's hard and complex to assign corresponding weight to all the neighbors. When the fingerprint data grows significantly, it becomes even worse. Similar with WKNN, EWKNN mitigates noise effects by changing the value of K, which is to make the parameter K a variable. CFK is another advanced KNN algorithm. Instead of taking all the K nearest neighbors into calculation, it selects some neighbors from the K nearest neighbors and outperforms KNN [37].

All the advanced KNN algorithms mentioned above can offer better positioning accuracy than basic KNN algorithm. However, since the complexity raises with those algorithms, and the main focus of this thesis is on the comparison among data collected through crowdsourcing, which should not be affected by the practical accuracy the system can

achieve, it's reasonable to simply utilize KNN as the positioning algorithm. In this thesis 3NN is used through the analysis.

4.2 Positioning error calculation

With labeled training and estimation data (here, 3-D coordinate as the label), the positioning can be implemented without knowing the estimation data's coordinate. The positioning error is the Euclidean distance between estimated coordinate and reported true location's coordinate. It can be calculated as follows:

$$\mathrm{D}_{err} = \left\| \sqrt{(x_e - x_t)^2 + (y_e - y_t)^2 + (z_e - z_t)^2} \right\| \quad (4.3)$$

x_e, y_e and z_e are estimated 3-D coordinates result, and x_t, y_t, z_t are the true 3-D coordinates of fingerprint estimation data.

4.3 Probability distribution fitting

To statistically analyze different dataset's behavior, the RSS histograms of datasets are compared with 11 theoretical distributions including Gaussian, Exponential, Lognormal, Extreme value, Rayleigh, Gamma, Weibull, Logistic, Burr, Generalized pareto and Generalized extreme value. The comparison is based on Kullback-Leiber divergence (KLD) criterion which is also called relative entropy in mathematical statistics [56]. The value of KLD varies from 0 to infinity. When KLD gets close to 0, it indicates that the behavior of the two distributions are similar. When KLD increases, it indicates that two distributions are different. So, in this case, the distribution out of the 11 theoretical ones with smallest KLD value will be selected out as the best distribution.

The CDF of Gaussian distribution is also called Normal CDF (NCDF):

$$F(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt \quad (4.4)$$

μ is the mean of the distribution, σ is the standard deviation and σ^2 is the variance.

The Exponential CDF is:

$$F(x|\mu) = \int_{0}^{x} \frac{1}{\mu} e^{\frac{-t}{\mu}} dt = 1 - e^{\frac{-x}{\mu}} \quad (4.5)$$

Here, μ is the exponential factor.

The Lognormal CDF is:

$$F(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^x \frac{e^{\frac{-(\ln(t)-\mu)^2}{2\sigma^2}}}{t}dt \quad (4.6)$$

μ and σ are the mean and standard deviation, respectively.

The Extreme value CDF is:

$$F(x|\mu,\sigma) = 1 - e^{\left(-e^{\frac{x-\mu}{\sigma}}\right)} \quad (4.7)$$

μ and σ are the mean and standard deviation, respectively.

The Rayleigh CDF is:

$$F(x|b) = \int_0^x \frac{t}{b^2} e^{\left(\frac{-t^2}{2b^2}\right)} dt \quad (4.8)$$

b is the scale parameter of the distribution.

The Gamma CDF is:

$$F(x|a,b) = \frac{1}{b^a\Gamma(a)}\int_0^x t^{a-1}\mathrm{e}^{\frac{-1}{b}}\,\mathrm{d}t \quad (4.9)$$

a is a shape parameter and *b* is a scale parameter.

The Weibull CDF is:

$$F(x|a,b) = 1 - e^{-(x/a)^b} \text{ (x>0)} \quad (4.10)$$

Parameters of *a* and *b* are shape parameter and scale parameter, respectively.

The Logistic CDF is:

$$F(x|\mu,\sigma) = \frac{1}{1+e^{-\frac{x-\mu}{\sigma}}} \quad (4.11)$$

μ and σ are the mean and standard deviation, respectively.

The Burr CDF is:

$$F(x|a,\theta,k) = 1 - \frac{1}{\left(1+\left(\frac{x}{a}\right)^{\theta}\right)^k}, x>0, a>0, \theta>0, k>0 \quad (4.12)$$

θ and *k* are shape parameters and *a* is a scale parameter.

The Generalized pareto CDF is:

$$F = 1 - \left(1 + k * \frac{x - \mu}{\sigma}\right)^{-\frac{1}{k}} \quad (4.13)$$

μ and σ are location and scale parameters, k is the shape parameter.

The Generalized extreme value CDF is:

$$= F(x|a,b) = \begin{cases} e^{-\left(1+k*\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{k}}}, & k \neq 0 \\ e^{-e^{-(x-\mu)/\sigma}}, & k = 0 \end{cases} \quad (4.14)$$

μ and σ are location and scale parameters, and k is the shape parameter.

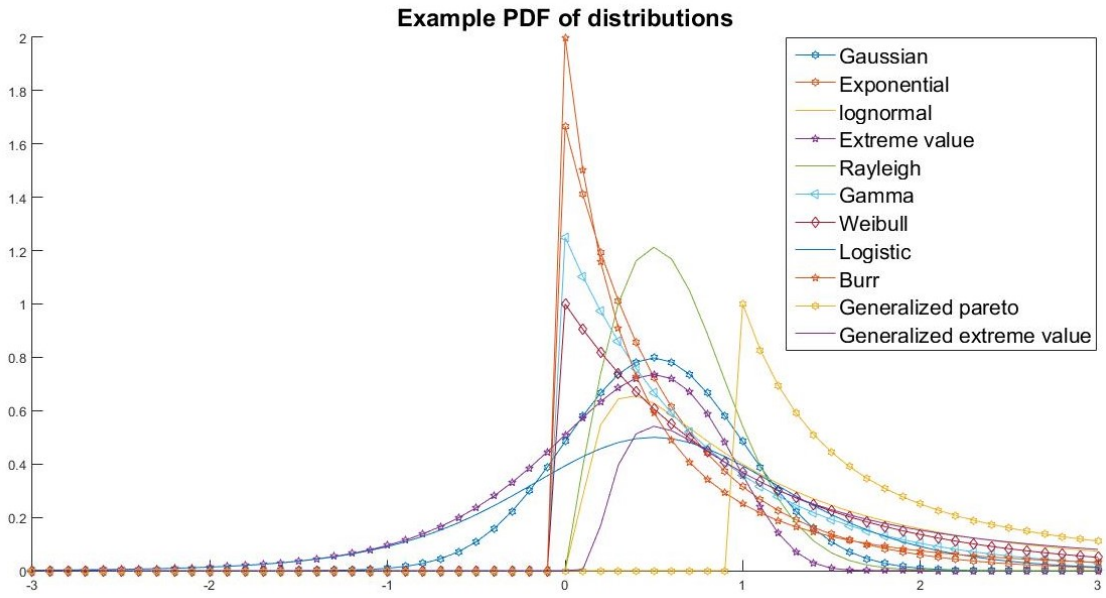

Figure 6. *Example PDF curves of some distributions*

PDFs of all the distributions mentioned above are shown in Figure 6. The parameters of the distributions are given in Table 2.

Table 2. *Distributions parameters for Figure 6*

Distribution	**Parameters**
Gaussian (Normal)	$\mu = 0.5$, $\sigma = 0.5$
Exponential	$\mu = 0.6$
Lognormal	$\mu = 0$, $\sigma = 1$
Extreme value	$\mu = 0.5$, $\sigma = 0.5$
Rayleigh	$b = 0.5$
Gamma	$a = 1$, $b = 0.8$
Weibull	$a = 1$, $b = 1$
Logistic	$\mu = 0.5$, $\sigma = 0.5$
Burr	$a = 1$, $\theta = 1$, $k = 2$
Generalized pareto	$k = 1$, $\mu = 1$, $\sigma = 1$
Generalized extreme value	$k = 1$, $\mu = 1$, $\sigma = 1$

5. DATA COLLECTION DURING THE MEASUREMENT CAMPAIGNS

The first process of fingerprinting positioning is to collect fingerprint data, and it is also mentioned as the offline training phase in Chapter 2. There is a measurement campaign during the research, and the analysis presented in this thesis in following chapter is based on measurements attained in this campaign. The processes of data collecting, storing and downloading are introduced in the following sections. The creation of synthetic erroneous data is also explained here.

5.1 Data collecting process

Two different types of fingerprint data collecting methods are utilized: pointwise collected crowdsourcing data and systematically collected data. Also, two different Android applications are used for these two methods.

5.1.1 Crowdsourcing data

The data collecting process is implemented through the Android application 'TUT Wi-Fi Positioning'. This application looks for all APs available and reads the MAC addresses and RSSs from all APs. There is already a fingerprint database in this application, so this application can offer position estimation function, which provides an initial reference position for the user feedback. In this application, each floor's map of one TUT building is available and the map of first floor is at the first sight of user's view. The user interface of the application can be seen in Figure 7. On the bottom side of this application interface, there are two function buttons 'ESTIMATE' and 'CENTER'. The estimation of user's position starts as soon as the 'ESTIMATE' button is clicked. After the mobile device scans for a while, the estimation result will be shown on the map as a small green circle. At the same time, a text box will appear on the bottom of the interface, above the two buttons mentioned before. If with correct result, user ought to click 'yes', and the data, including the coordinate of the position, the floor number and all the received RSS values as well as MAC addresses is reported. If the result position is not correct, user ought to click 'no', and then the application allows the user to freely click the correct position on the map (the chosen position will appear as a small pink circle) to report the data. All reported data will be instantly transmitted to a Google cloud server and stored in the cloud. The schematic of Google cloud server architecture is presented in Figure 8 below. User can also choose the floor number on the top of the interface when the result is with wrong estimated floor.

Figure 7. *Screenshots of application positioning process*

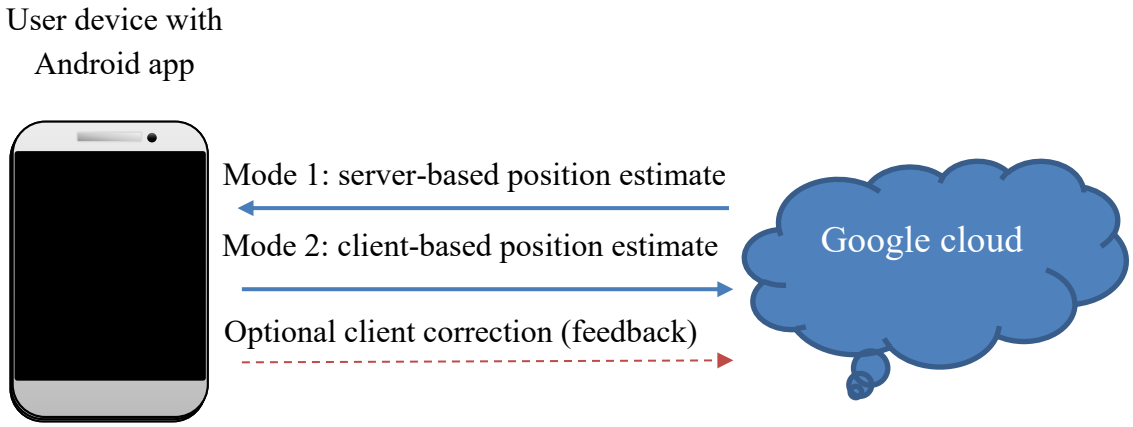

Figure 8. *Schematic client server architecture*

There are 4648 collected fingerprint data in total, and they are collected from 21 different mobile devices. The histogram which shows the numbers of measurements per device is shown in Figure 9. The data is plotted in descending order of numbers. 992 MACs in total are detected through the measurement. Multiple APs can be heard from the same location or transmitters, which results in such large number of MACs.

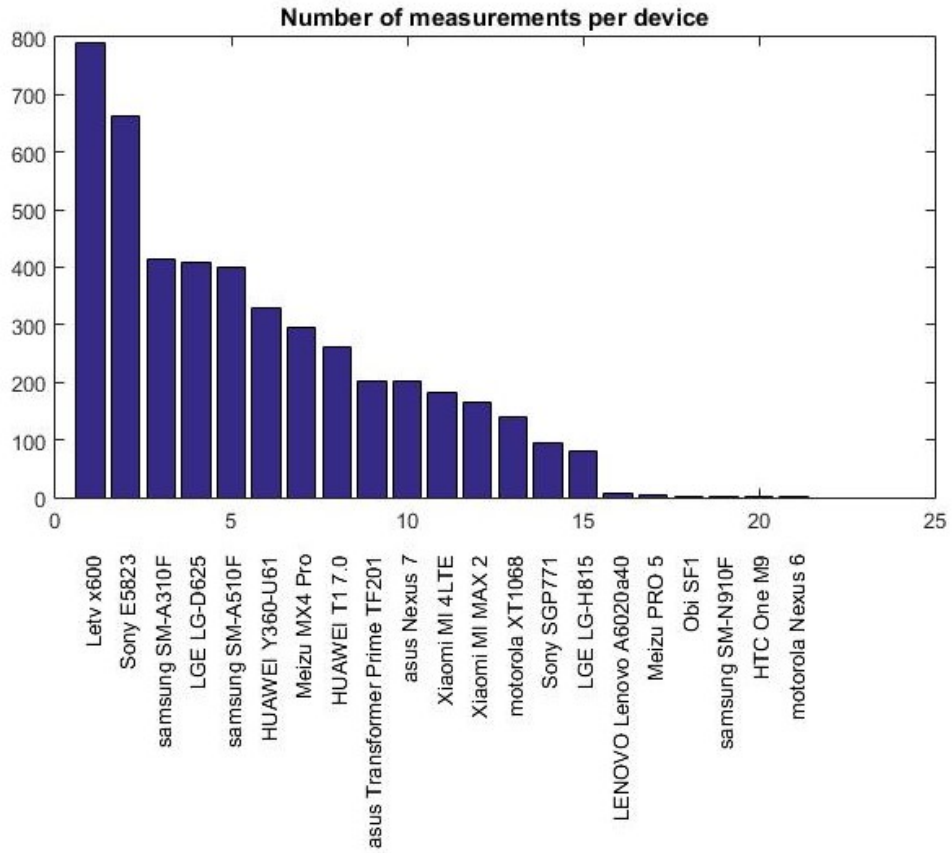

Figure 9. *Histogram of 21 devices fingerprint data*

5.1.2 Systematically collected data

Apart from the 4648 fingerprint data, there are 2220 fingerprints collected with another application by three different mobile devices including HuaweiT1 tablet, Huawei Y360 phone and Nexus tablet as three tracks, which are used as estimation data. The number of measurement per device is shown in Figure 10.

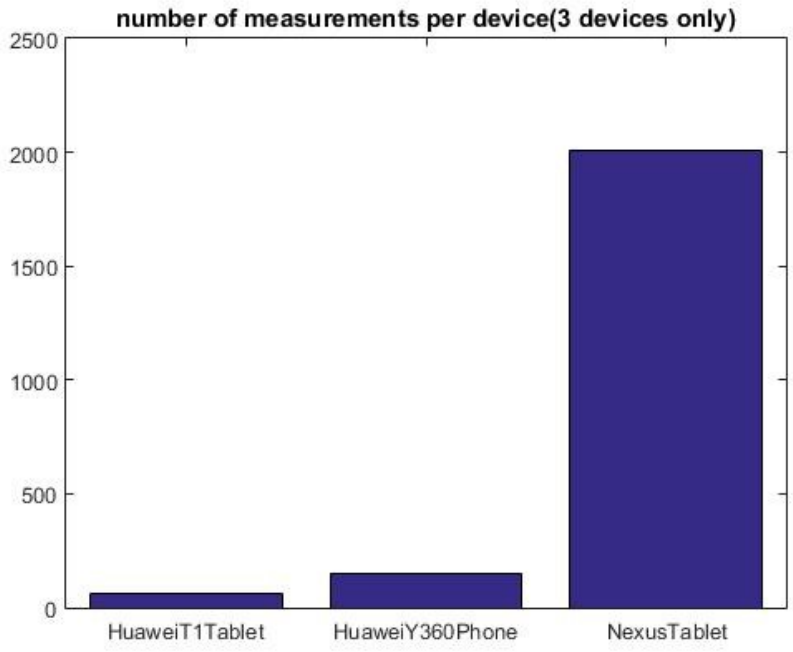

Figure 10. *Histogram of estimation fingerprint data*

These data are systematically collected with specific track. Most of these measurements are taken by Nexus tablet. To take full advantage of them, all three devices data together are utilized as the estimation data or the footprint track for testing.

5.1.3 Environment of the positioning area

Real environment of the building can be seen in Figure 11. There is an open space corridor of first and second floor as the two left pictures show which is linked to an entrance. Map of first floor can also be seen in Figure 4. Long office corridors take most of the space of upper floor. Wi-Fi signal covers most part of the building besides some small part of the office on upper floor which is vacant recently which might decrease the positioning accuracy but in an acceptable range.

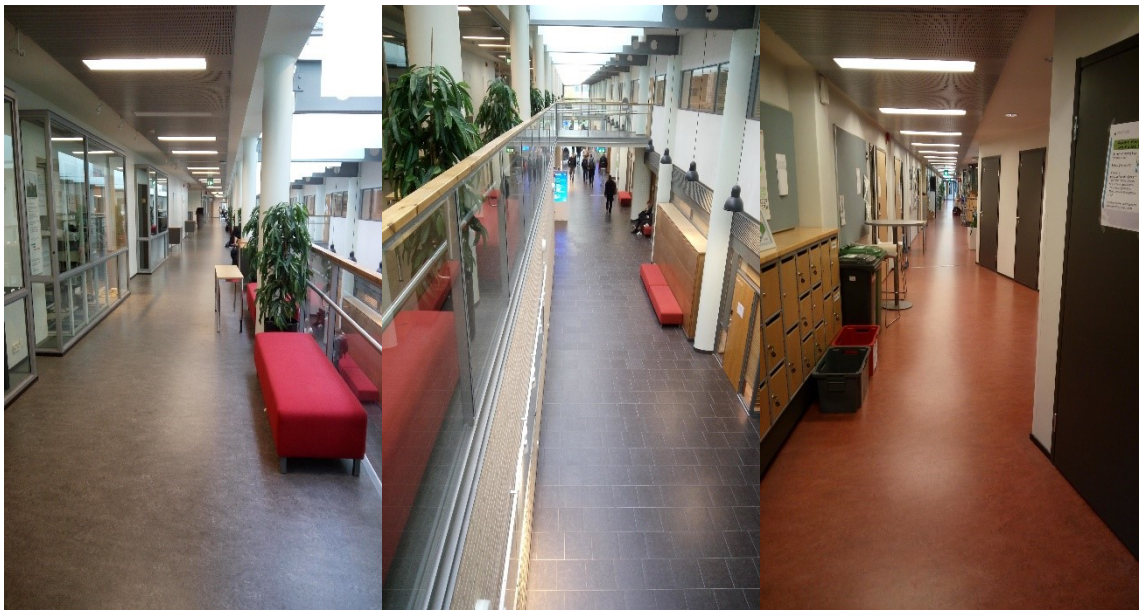

Figure 11. *Example pictures of environment, long corridors of first floor and office.*

5.2 Data downloading

After the data is stored in the cloud server, it can be accessed through a webpage as Figure 12 shows. The webpage is only accessible with administrator rights. On this webpage, the collected crowdsourcing data can be downloaded by clicking 'Download User Feedback Database'. The fingerprint database of the application can also be downloaded or uploaded if needed.

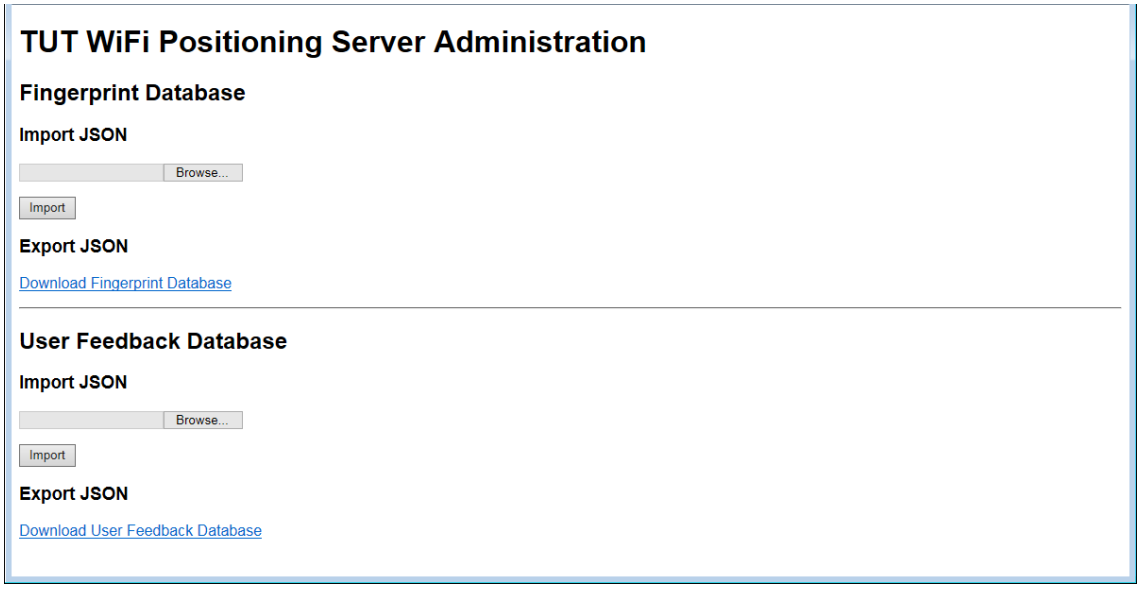

Figure 12. *The webpage interface of the data adminstration system*

All the data is saved in the form of .json file. MATLAB is used to read the data and extract the values of RSS and coordinates and sort the data in chronological order and with different mobile device models.

5.3 Creation of synthetic erroneous data

To statistically analyze the behavior of positioning accuracy when erroneous reporting happens, the author created errored data or malicious data, and modified the original fingerprint database with different proportion of errored data. Two types of error are considered in this thesis: first one is the malicious data with erroneous position, and the second one is data with incorrect RSSs reported. After erroneous data is constructed, the impact of the error is analyzed by comparing the positioning accuracy of using data with different proportion of error and without error.

5.3.1 Data with position error

Since the fingerprint data is collected through crowdsourcing, there are inevitable manual operating errors when using the Android application to report the data. The error may occur when user intentionally or unintentionally click the wrong position or more likely it happens when user click the position without choosing the floor number.

In this thesis, the position error data is modified in such a method as follows: First, according to the error proportion, a part of the data is chosen randomly from the database as the error data to be modified. The error proportions chosen here are 25%, 50%, 75% and 100%. To modify floor error, the floor number is changed to another one randomly. For example, if one data vector is obtained at floor number 2, then it will be changed to 1, 3, 4 or 5 (all the data measurement is done on floor 1 to 5 of this building). Then, to further modify the coordinate of the error data, the mid coordinate of x and y coordinates are computed, and the modified points are in symmetry to this midpoint.

The 3-D map with modified error points are shown in Figure 13, Figure 14 and Figure 15 for different percentage of position error, respectively. The red circles represent correct points and blue crosses represent modified error points. To make it clear for readers to see the relation between original correct points and modified error points, the data showed in these figures are just part of the complete database, since the full database with 4648 points will occupy most space of the map.

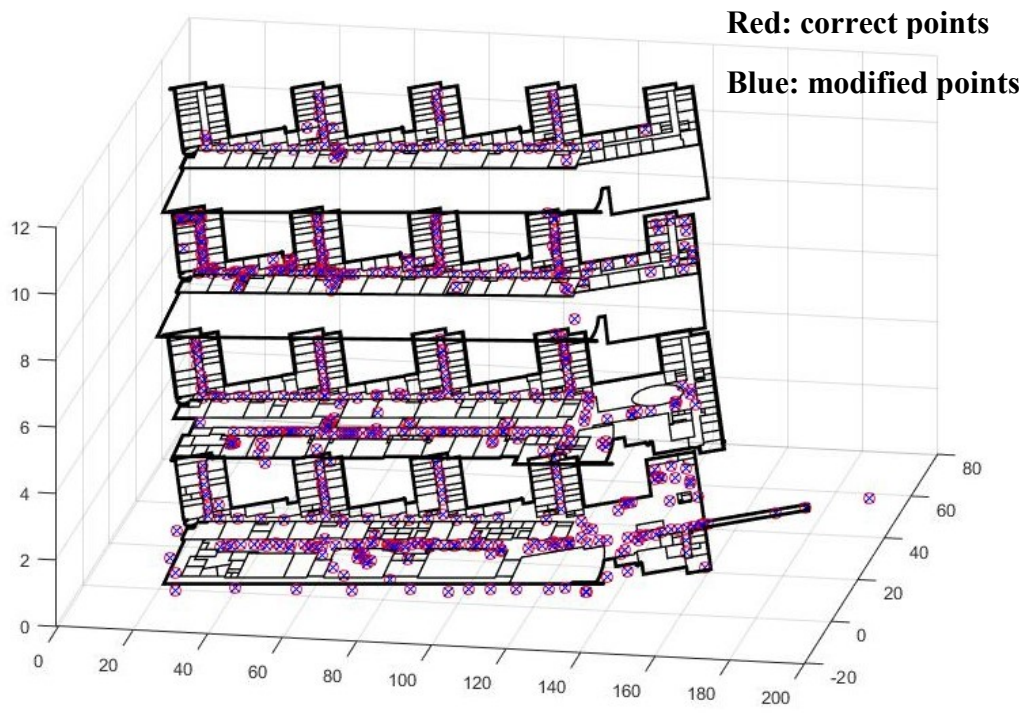

Figure 13. *Database without error*

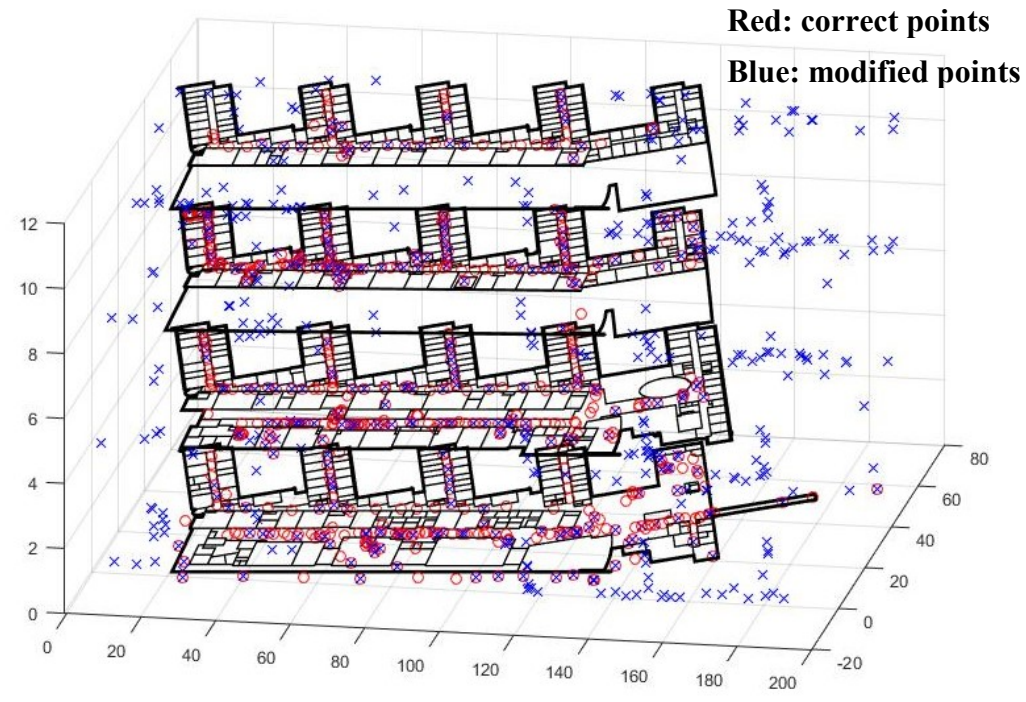

Figure 14. *Database with 50% position error*

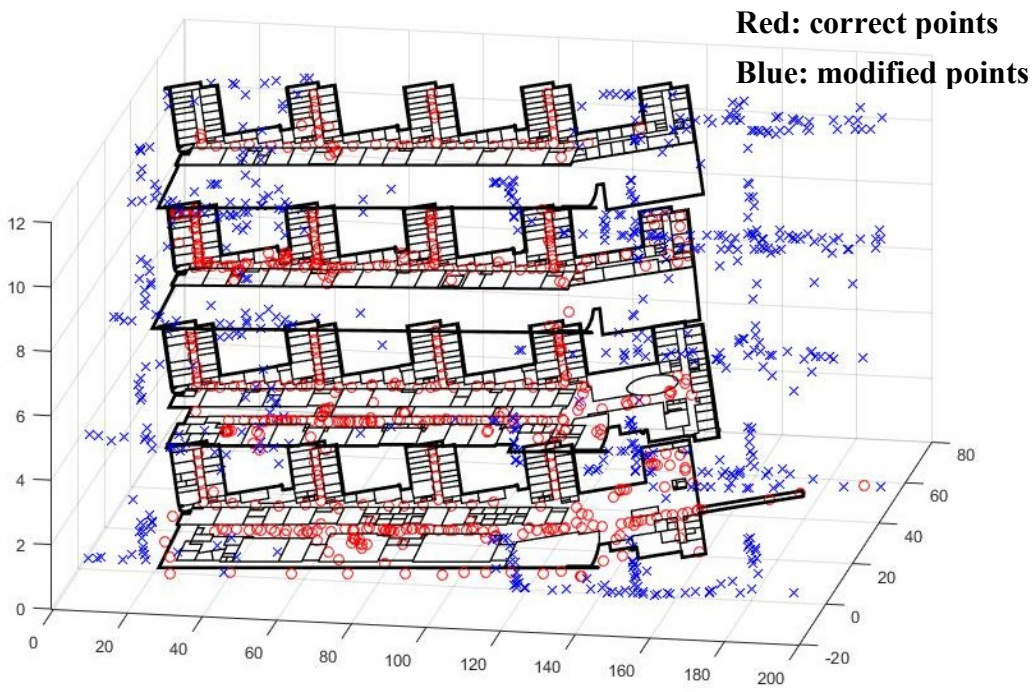

Figure 15. *Database with 100% position error*

After different proportions erroneous data positions are attained, each new dataset is used as a set of training data for the estimation process. Here, the systematically collected data are used as estimation data.

5.3.2 Data with incorrect RSS values

Besides incorrect position report, error may happen when the reported RSS values are incorrect. Because of human blockage and movement, multipath effect causes large RSS fluctuation [45]. Noise is another factor that can influence the RSS values [45]. In addition, faulty or malicious devices can report incorrect RSS data.

It's simple to modify error data with incorrect RSS, just by altering original data's RSS to desired new values. There are basically two schemes to alter the RSS values:

1. change the collected RSS values to new random values, the values should be within the limit of original data's RSS. For the 4648 fingerprint data, the maximum RSS value is -14dBm and minimum is -102 dBm.

2. change all values to constant values such as -70 dBm.

In this thesis, author adopted the second scheme, which is to set original RSS values to constant incorrect values. RSSs of -90 dBm, -65 dBm and -40 dBm are chosen as the modified values, among them, -65dBm is the value which is closest to the average RSS as shown in Figure 16.

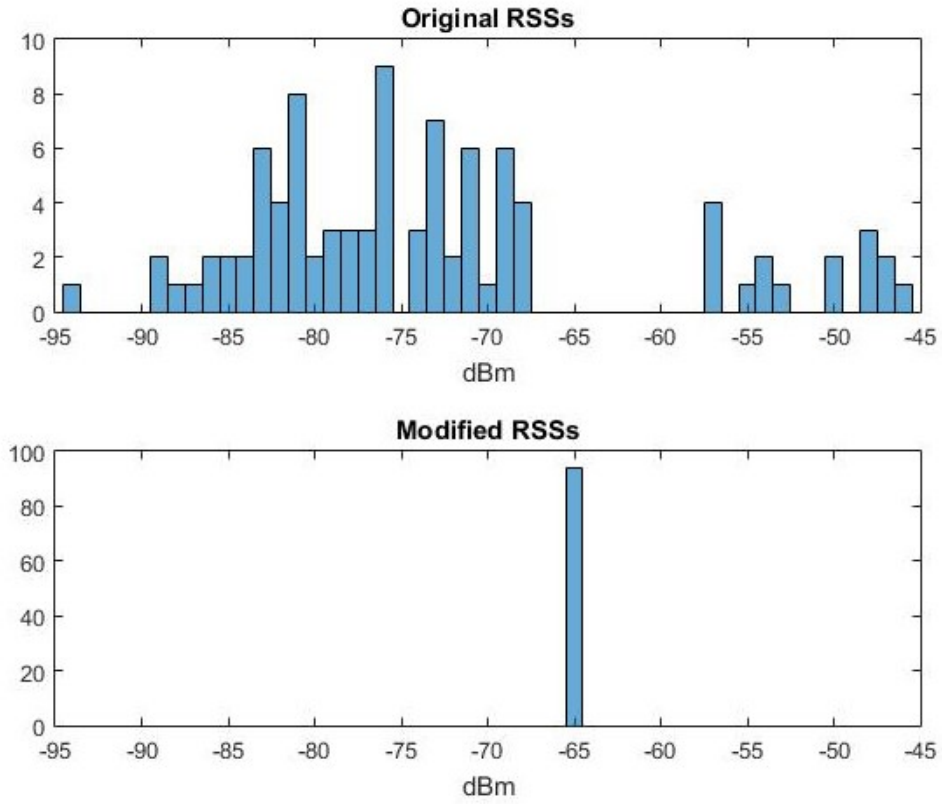

Figure 16. *Upper plot: original RSS values; lower plot: modified (incorrect) RSS values*

6. ANALYSIS OF DATA AND RESULTS

This Chapter presents the analysis of the crowdsourcing data. The analysis is conducted from several aspects. First, the positioning accuracy is analyzed by comparing the CDF of error curve of different dataset. Then, behavior of RSSs of different dataset is analyzed by comparing the best fitted distribution through KLD. Besides, the RSSs difference of different devices at the same AP is presented. Finally, the impact of two types of erroneous data is analyzed.

6.1 Analysis of crowdsourcing data

First analysis is based on the position estimation which is done by taking all 4648-crowdsourcing data as training data and the systematically collected data as estimation data, and the position is attained through 3-Nearest Neighbor (3NN) algorithm. The overall result is shown in CDF of error form in Figure 17. It can be observed from the blue curve that the positioning result is not so good, less than 70% of data can attain the accuracy of 10 m and up to 90% of data can get around 20 m's accuracy. From author's point of view, this is caused by multiple factors such as device heterogeneity, 2 different applications are used to collect training and estimation data, multipath effects, shadowing and fading, etc.

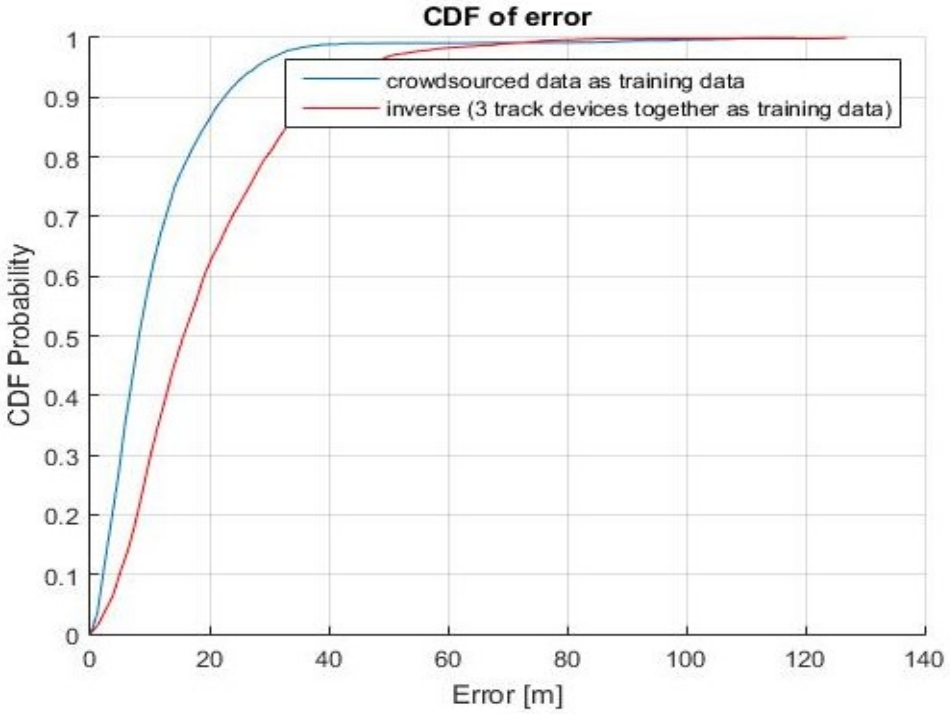

Figure 17. *CDF of error with overall crowdsourcing data*

Besides, the red curve which represent the inverse result with systematically collected data as training data and all crowdsourcing data as estimation data, presents a worse accuracy, which is reasonable since the number of training data has been declined from 4648 to 2220. To get an estimation result with high accuracy, a large quantity of training data is necessary from machine learning field's perspective [53]. This again clarifies the significance of crowdsourcing in data collecting.

The crowdsourcing data is collected with 21 different devices. It's feasible to analyze the dataset with different device separately. The estimation data keeps still, and each device's full data is selected out one by one as the training data. In Figure 18, CDF of error of different device is presented with various shape of line and with different color, and all CDFs are plotted in one figure.

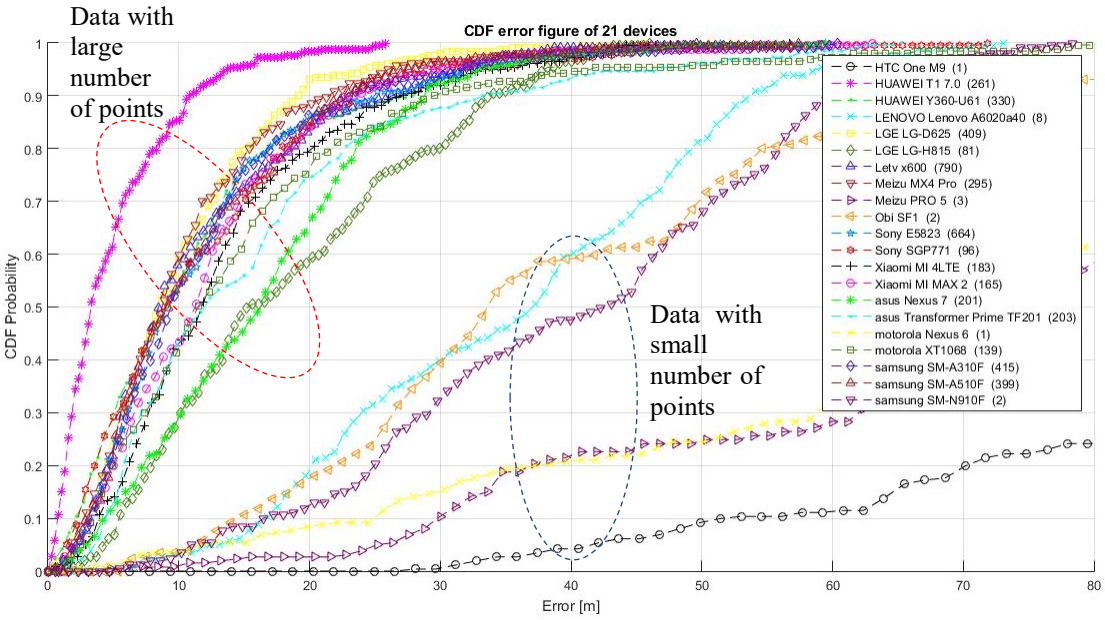

Figure 18. *CDF of error with all data sorted by device*

The number in the bracket after each device name represents the number of measurement points of corresponding device. There are 6 curves far away from the rest curves in this figure. These curves represent the datasets of devices with few measurement points. Thus, the focus is on the rest of the curves. The CDF error curve of HUAWEI T1 device which is drawn with purple asterisk shows the best accuracy. 70% of the positioning result is within 5m's accuracy and 90% can attain 10 m's accuracy. It performs much better than the overall crowdsourcing one as well as other devices positioning result. It's reasonable that device with a smaller amount of training data performs worse, but when the number of training points grows beyond a threshold value, for example 100, the positioning accuracy seems to be affected by other factors which of course include the diversity or heterogeneity of mobile devices. The device with the highest accuracy is HUAWEI T1 Tablet, whereas the one with largest number of measurement points is Letv-x600 device, and the measurement points of the latter one (790) are much more than the former one (261).

6.2 Power map and distribution

The RSSs and differences between the power maps are analyzed here by comparing the best fitted distribution of different datasets.

6.2.1 RSS distributions

First, the simply RSS histogram of dataset is compared with the 11 theoretical distributions with KLD. KLD calculation formula is:

$$D_{KL}(PDF_1 \| PDF_2) = \sum_{i}^{N} PDF_1(i) * \log\left(\frac{PDF_1(i)}{PDF_2(i)}\right) \quad (6.1)$$

PDF_1 is the PDF of analyzed fingerprinting RSSs and PDF_2 represents the theoretical distribution fitted to the fingerprinting RSSs. N is the segment number of histogram, and here 36 segments of histogram are used. $PDF_1(i)$ represents the probability of ith segment appearance and $PDF_2(i)$ is the probability of ith segment appearance of fitted distribution's curve. Since 0 value is not allowed for neither PDF, each 0 value of probability is replaced by a small value as 10^{-4}. When $PDF_1(i)$ equals to $PDF_2(i)$, the KLD value becomes 0 for this segment which shows the similarity of this segment.

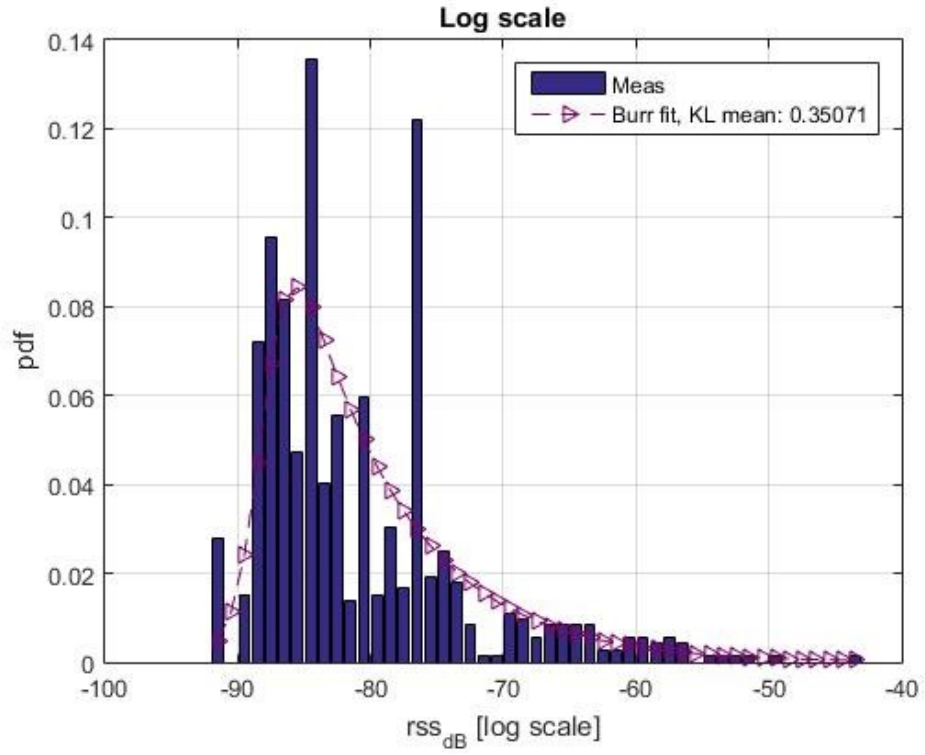

Figure 19. *Example of the RSS distribution for Letv-x600 device*

The best distribution among the 11 distributions is the Burr Type XII distribution. As shown in Figure 19, this is the RSS distribution of one random AP for Letv-x600 device

which is the one with most measurement points (790 measurement points out of total 4648 points). The distribution is not symmetric and is in skewed right shape (the right tail is much longer than the left tail). This shape fits Burr distribution, which usually also fits to real medical field data [51]. The cumulative distribution function (CDF) of Burr is:

$$F(x|a,\theta,k) = 1 - \frac{1}{\left(1+\left(\frac{x}{a}\right)^{\theta}\right)^{k}}, x > 0, a > 0, \theta > 0, k > 0 \quad (6.2)$$

The probability density function (PDF) is:

$$f(x|a,\theta,k) = \frac{\frac{k\theta}{a}\left(\frac{x}{a}\right)^{\theta-1}}{\left(1+\left(\frac{x}{a}\right)^{\theta}\right)^{k+1}}, x > 0, a > 0, \theta > 0, k > 0 \quad (6.3)$$

where θ and k are the shape parameters of the distribution and a is presented as the scale parameter. It is a very flexible distribution that it basically can express any distribution shapes and can fit a wide range of empirical data as shown in Figure 20. Here θ is replaced by c.

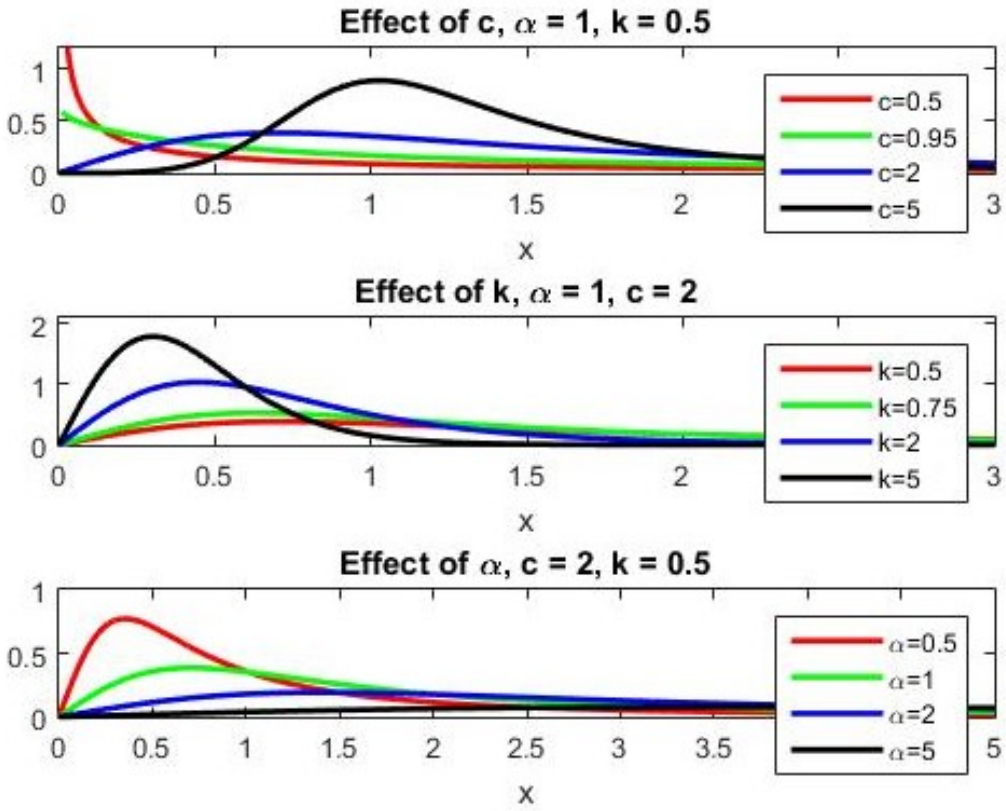

Figure 20. *Burr Type XII distribution examples, different parameters effects.*

Table 3 shows examples of best distribution and the KLD value of it and parameters of some datasets. Besides the best fit, the second-best fit belongs to Generalized extreme

value (GEV) distribution. Since the GEV distribution also consists of 3 parameters (Only GEV and Burr Type XII are made up of 3 parameters), it can fit a wide range of data as well.

Table 3. *Best distribution of RSS histograms for different dataset*

Dataset	**Best distribution and KLD value**	a	θ	k
crowdsensed dataset	Burr (0.5314)	2.4170e-07	4.0351	0.4201
estimation dataset	Burr (0.2740)	4.0024e-07	2.6310	0.9277
Huawei T1 only (from estimation dataset)	Burr (0.8581)	1.2806e-04	1.2724	2.3877
Nexus only (from estimation dataset)	Burr (0.2492)	4.1997e-07	3.2654	0.8205
Sony E523 only (from training dataset)	Burr (0.4452)	2.5648e-06	6.0779	0.2315
Letv-x600 only (from training dataset)	Burr (0.9422)	1.6750e-08	6.9034	0.6715

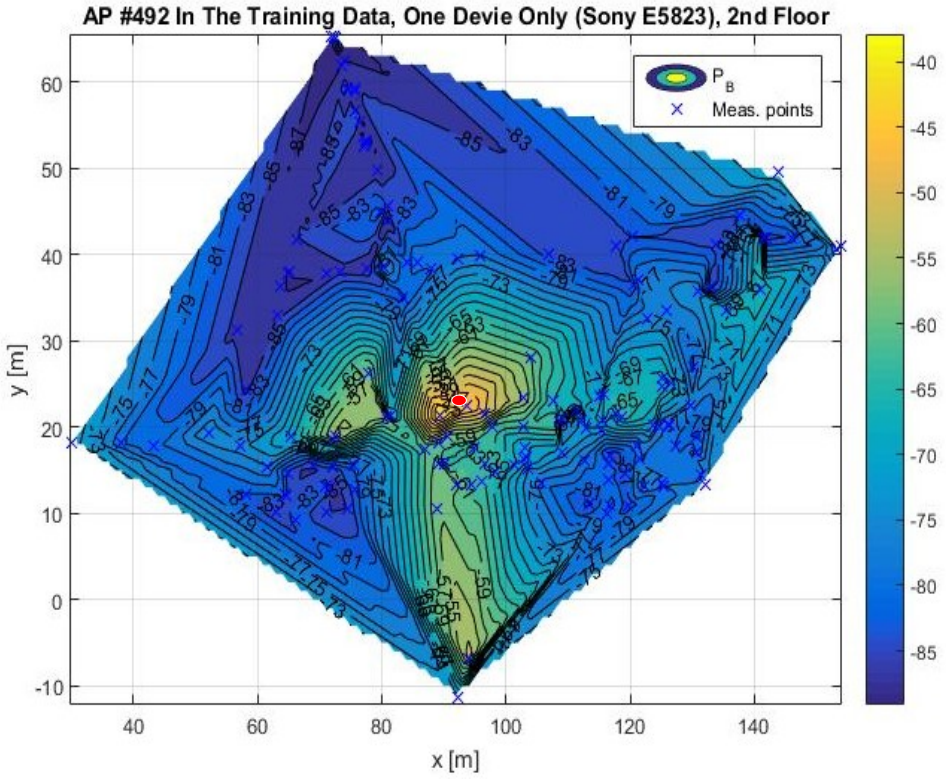

Figure 21. *Example power map of one AP for Sony E5823 (floor 2)*

Figure 21 and Figure 22 are the power maps of all fingerprinting RSSs heard from one same AP. Two figures respectively show the power map of two devices as Sony E523

and Letv-x600. As can be seen from the figures, the coordinate of the point with largest RSS value (shown in red circle on the figures) is the same for the two devices (x around 90 and y around 20). This should be the location of the AP, thus it's also feasible to locate all APs of the building with the RSS fingerprinting. Power maps of different devices are correlated but span over different space. The further comparison between these two power maps are presented in following section.

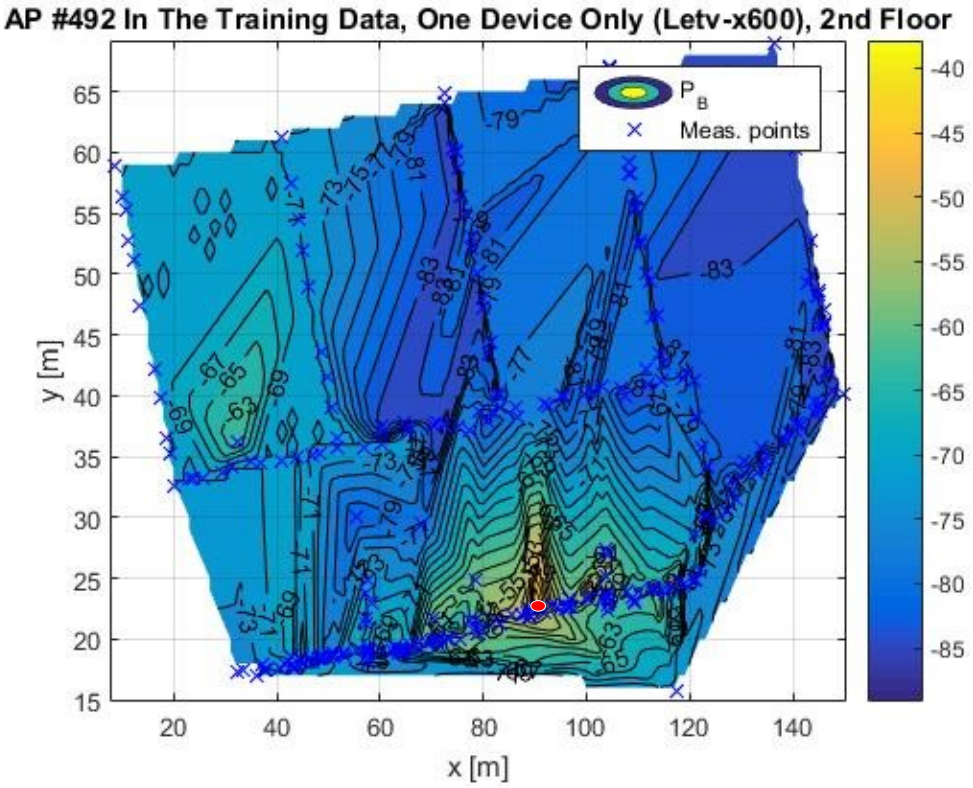

Figure 22. *Example power map of one AP for Letv-x600 device (floor 2)*

6.2.2 Distribution of power map difference

Next, the power map is compared by analyzing the distribution of power map differences. The power map difference is attained through 2 different power maps with same AP and at the same floor to build an interpolated and extrapolated power map. There is the same limit of the floor area computed for both power maps to make sure both power maps have the same spatial area and the subtraction between different power map can be smoothly processed. Then, the histogram of this difference is computed as the analysis object, in which the value is in dB form (addition and subtraction between dBm). The histogram is also compared with the 11 theoretical distributions in the same way as previously explained.

As shown in Table 4, the best distribution is still Burr distribution for all power map difference, and the KLD value is obviously much smaller than the ones presented in Table 3.

Table 4. *Best distribution of power map difference*

Dataset	Best distribution and KLD value	a	θ	k
comparing Huawei T1 estimation data with all crowdsensed data	Burr (0.1730)	4.6295	1.3756	1.5065
comparing Nexus estimation data with all crowdsensed data	Burr (0.0874)	13.4109	1.4326	1.8845
comparing Huawei T1 and Nexus data	Burr (0.1219)	2.0019	1.9125	2.3529
comparing Sony E523 and Letv-x600 data	Burr (0.0766)	9.5142	1.8192	1.4457

Figure 23 has shown an example histogram of power map difference between Letv-x600 device and Sony E5823 device. Although the Burr distribution fits best with lower than 0.1 KL value, the distribution curve is with good symmetry and the peak is located around 0 on x axis from observation.

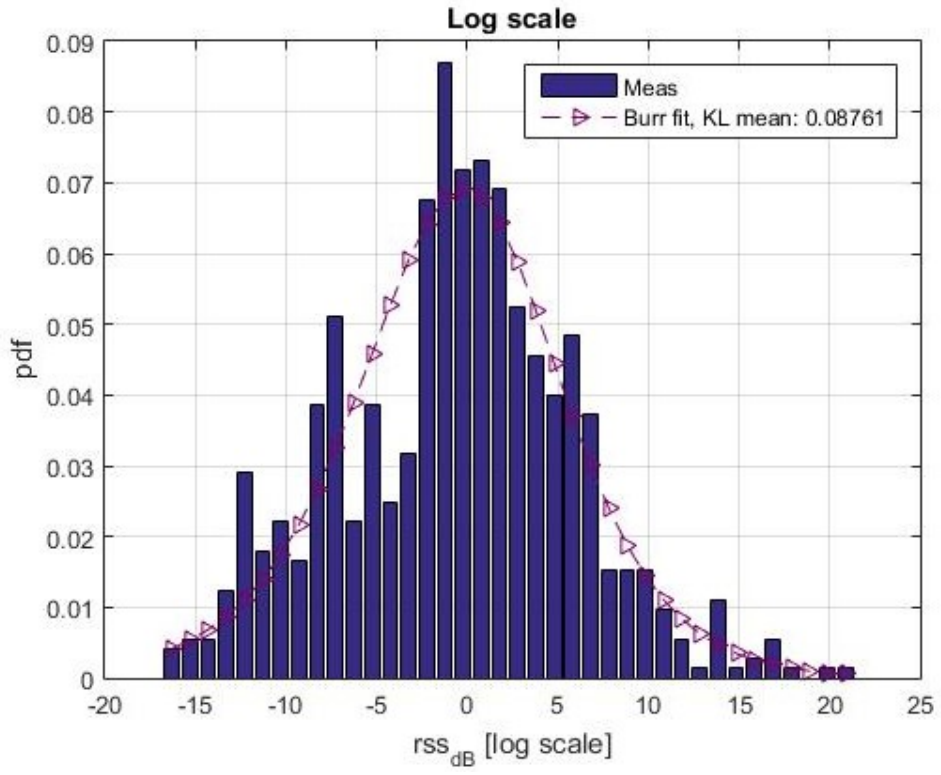

Figure 23. *Example of the distribution of power map difference (between Letv-x600 and Sony E5823 devices)*

From data presented in [52], there is a stable relation between RSS values attained with different device. In this way, the histogram of power map difference should be with constant values. However, in this thesis example, the power map difference value is varying from -15dB to 20dB, and it can be seen clearer from Figure 24, RSS difference varies with area. From author's point of view, measurements with a small number of APs and

with a smaller district like in [52] are not as affected by noise as in this thesis scenario. In real occasion, noise effects might be even bigger, thus the RSS difference value of diverse devices could be further away from constant.

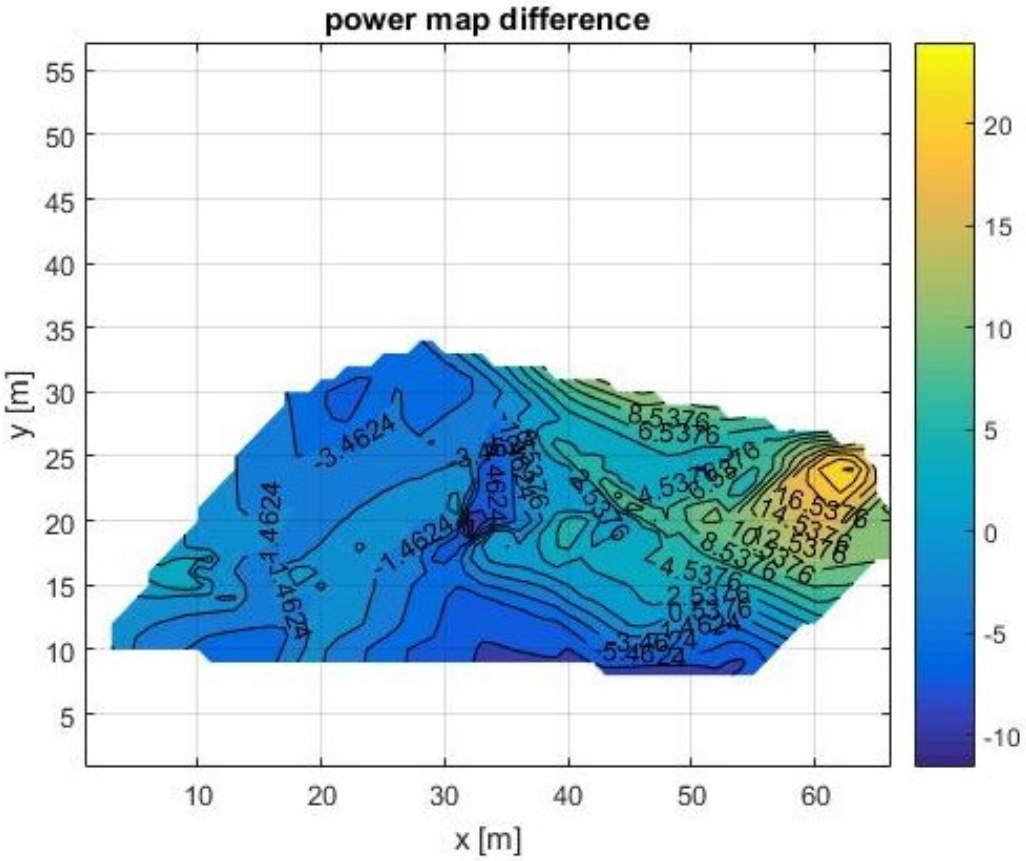

Figure 24. *Example Power map difference between Letv-x600 and Sony E5823 devices*

6.3 Analysis of erroneous data

To analyze the effect of error data on fingerprinting positioning, the most intuitive method is to use the errored data as training data and implement position estimate with it and compare the result with the one without error. The estimation result is shown in figure with CDF of error curves. Each curve shows how well the accuracy is attained with according set of data and higher curve indicates higher accuracy.

6.3.1 Data with incorrect position

As shown in Figure 25, the result with incorrect position data performs as expected. With higher proportion of error data, the estimation result looks worse. The dataset with no error obviously performs best in positioning among all the tested data. The modifying method of erroneous data is explained in chapter 4. The motivation to modify data with such huge error is to intuitively show the influence caused by 3-D position error. In real

scenario, big incorrect position error is usually caused by manual operation of selecting wrong floor and more often the error only varies within meters for imprecise clicking on the map.

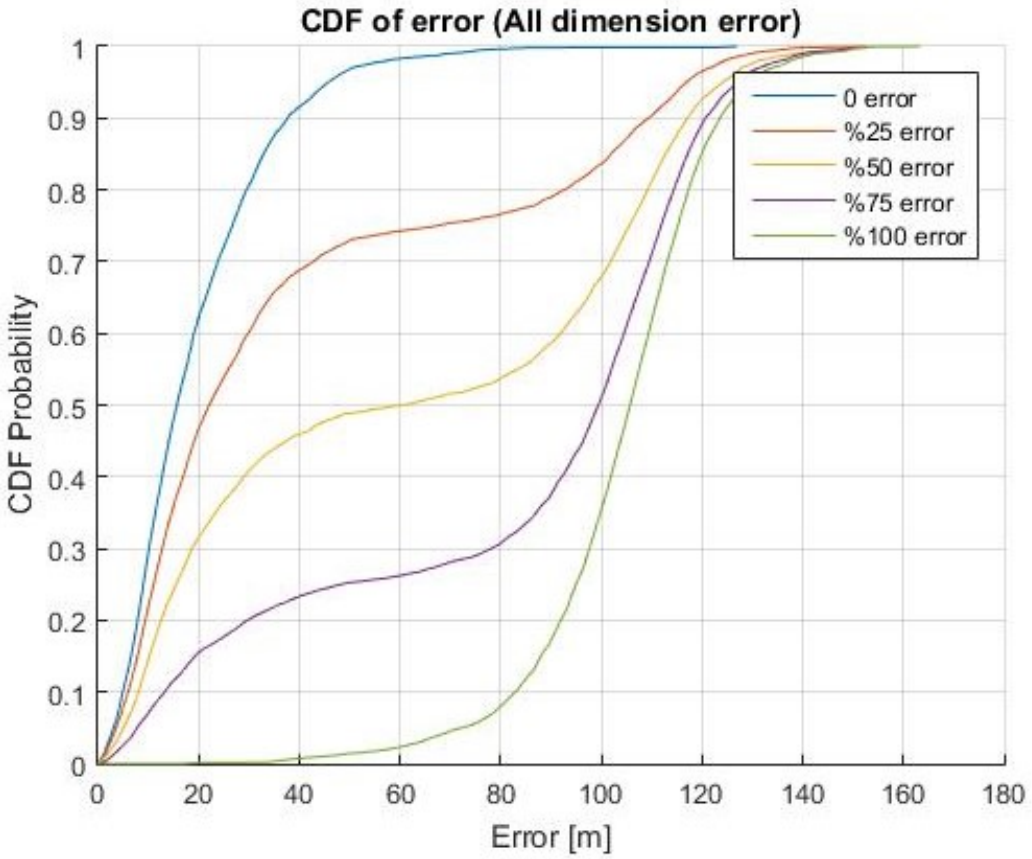

Figure 25. *CDF error figure of data with incorrect position*

6.3.2 Data with incorrect RSS values

Unlike the data with position error, to author's surprise, data with incorrect RSS values performs well in positioning, even similar with the original data without error. As Figure 28 shows, no matter how the proportion of modified data changes, the positioning accuracy just seems similar with the one without error. The curves have only little fluctuation even if the modified constant RSS value changes from -90 dBm to -40 dBm.

To make it clear to see, all the curves in Figure 26, Figure 27 and Figure 28 are drawn with different colors and markers. But since most parts of them are overlapped, it's still hard to distinguish them from each other. The overlapping happens most obviously for the first figure which is with -65 dBm RSS modified data. -65 dBm is the mean value for all the RSSs in the original fingerprint data, and -90 dBm and -40 dBm are the minimum and maximum value of the RSSs value of the whole fingerprint database. So, from this point of view, the impact of errored RSS data with different errored RSS value can be concluded as: errored RSSs close to mean value of original data keeps the accuracy at a

good level, while RSSs with big fluctuation value compared to the original data will decrease the positioning accuracy.

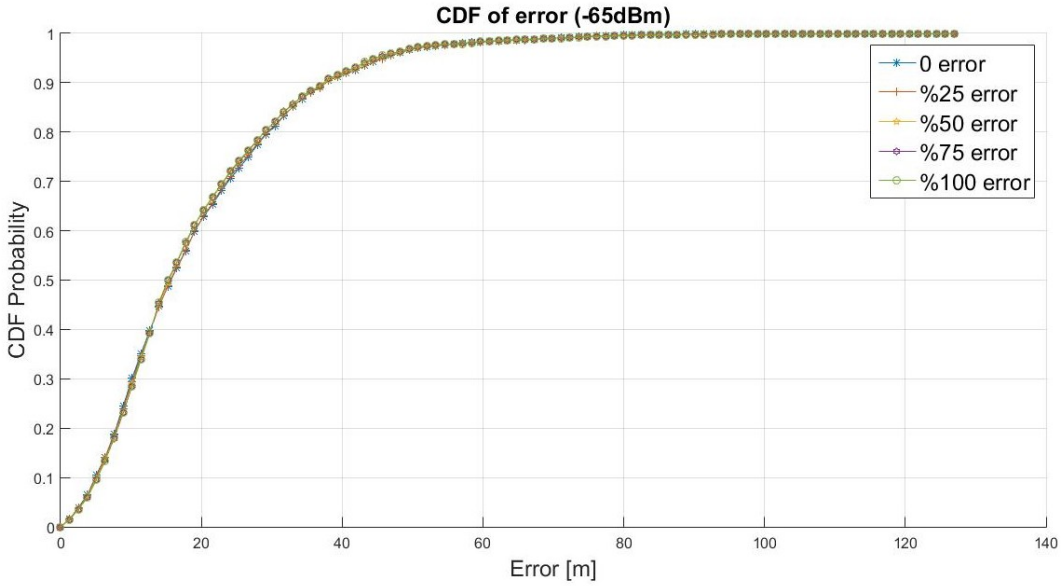

Figure 26. *CDF error figures of data with incorrect RSS values (-65 dBm)*

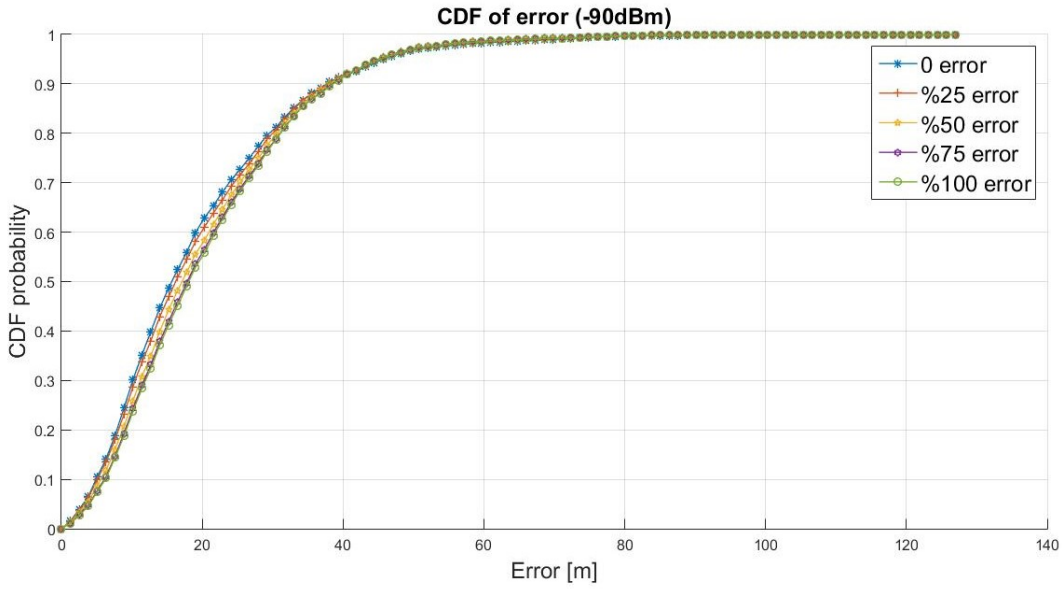

Figure 27. *CDF error figures of data with incorrect RSS values (-90 dBm)*

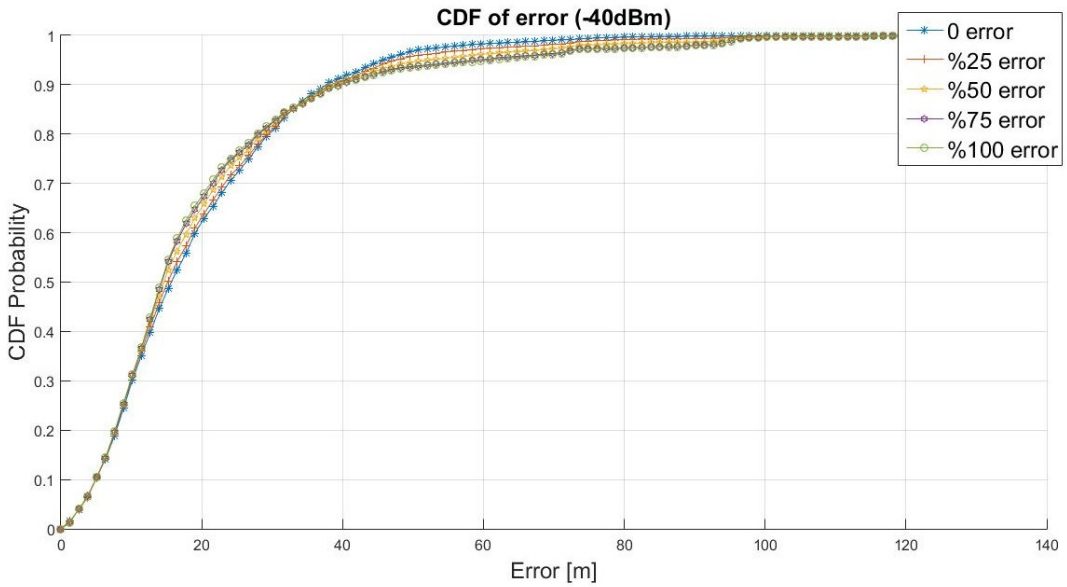

Figure 28. *CDF error figures of data with incorrect RSS values (-40 dBm)*

The reason why RSS values hardly affect the positioning accuracy is actually hidden in the estimate method. In fact, in the process of positioning estimation, the observed RSS is only compared with training data which is heard from the same access point. For those not matched points, the value in the log bracket is replaced by 10^{-6} to make sure that the calculation runs smoothly (the original value in the log bracket would be NaN if not replaced).

$$F(u) = log(\frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{-\frac{\left(RSS_0 - RSS_{training}(u)\right)^2}{2\sigma^2}}) \qquad (6.4)$$

When RSS is incorrect, but MAC address is correct and known, the Gaussian likelihood metric becomes close to rank-based metric and it is still able to estimate the user position only based on the number of commonly heard MAC addresses in the training and estimation phase. Thus, the estimation first compares the MAC addresses of the observed point with fingerprint data. The training data with most same MAC addresses would be selected out, and if multiple training data are chosen out, only then the Euclidean distance is needed for further comparing. Since the amount of APs or MAC addresses heard in this 4648 training data are as large as 992, the estimation result highly relies on the MAC addresses heard by the observed point instead of RSSs. It shows that RSS fingerprinting positioning system has good robustness. Besides, there is another conclusion that Gaussian likelihood metric might not be the most suitable metric to be used with crowdsourced data. Future research is needed to investigate the best positioning metrics with crowdsourcing data collection.

7. CONCLUSIONS

This thesis analyzed several different indoor positioning technologies and introduced some positioning measurement approaches. Wi-Fi based RSS fingerprinting is used as the positioning method and crowdsourcing is utilized in training phase to collect fingerprint data. Main target of this thesis is to analyze the impact of different crowdsourcing effects on Wi-Fi based indoor fingerprinting localization. The RSS fingerprints were collected through Android application with radio map of the building. Altogether 21 devices and 2 different applications as well as multiple participants were involved in the measurement campaign. In total, 4648 crowdsourcing fingerprint data samples were collected in the building through the campaign, and another 2220 systematically collected fingerprint data was used as the estimation data for positioning.

This thesis analyzed histogram distribution of RSSs in different dataset and the power map difference between data collected by different devices. Also, the thesis has analyzed the crowdsourcing impact by looking at the accuracy of positioning result with different dataset as training data. The positioning simulation is done through MATLAB. Different crowdsourcing errors were manually modified into the original database, and the behavior of errored data was observed through CDF figures. From the results it can be concluded that training data with higher proportion of 3-D position error has a worse positioning accuracy. However, fingerprint data modified with different proportion of constant RSS values can achieve almost similar positioning accuracy as the unmodified fingerprints. With enough IDs or MACs of AP correctly reported, RSS based fingerprinting localization can have a good positioning accuracy even if the RSSs fluctuate drastically.

Considering the non-stationarity of RSS, calibration in fingerprinting positioning can improve the positioning accuracy, thus, one of the future work is to analyze the crowdsourcing impact on calibrated fingerprint data, or the calibration impact on crowdsourcing fingerprinting positioning. Another future work is to investigate better positioning metrics with crowdsourcing data collection. Furthermore, in future 5G IoT standard, positioning with UNB RSS is more stable than with traditional Wi-Fi RSS, and the research can be continued on this new area.

8. REFERENCES

[1] B. Li, J. Salter, A. G. Dempster and C. Rizos, "Indoor positioning techniques based on wireless LAN" *first IEEE international conference on wireless broadband and ultra-wideband communications*, 2006.

[2] S. H. Fang, J. C. Chen, H. R. Huang and T. N. Lin, "Is FM a RF-Based Positioning Solution in a Metropolitan-Scale Environment? A Probabilistic Approach With Radio Measurements Analysis" in *IEEE Transactions on Broadcasting*, vol. 55, no. 3, pp. 577-588, September 2009.

[3] M. Ibrahim and M. Youssef, "CellSense: An Accurate Energy-Efficient GSM Positioning System" in *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 286-296, January 2012.

[4] S. H. Fang, C. H. Wang, T. Y. Huang, C. H. Yang and Y. S. Chen, "An Enhanced ZigBee Indoor Positioning System With an Ensemble Approach" in *IEEE Communications Letters*, vol. 16, no. 4, pp. 564-567, April 2012.

[5] E. Laitinen, E. S. Lohan, J. Talvitie and S. Shrestha, "Access point significance measures in WLAN-based location" *2012 9th Workshop on Positioning, Navigation and Communication, Dresden*, pp. 24-29, 2012.

[6] E. Laitinen, & E. S. Lohan, 2015, “Are all the Access Points necessary in WLAN-based indoor positioning?” in *Proceedings of 2015 International Conference on Localization and GNSS, (ICL-GNSS)*, 1 January 2015.

[7] R. Ding, Z. H. Qian and X. Wang, "UWB Positioning System Based on Joint TOA and DOA Estimation" in *Journal of Electronics & Information Technology*, 32(2), pp. 313-317, 2010.

[8] C. Yang and H. r. Shao, "WiFi-based indoor positioning" in *IEEE Communications Magazine*, vol. 53, no. 3, pp. 150-157, March 2015.

[9] S. H. Jung, S. Lee and D. Han, "A crowdsourcing-based global indoor positioning and navigation system" in *Pervasive and Mobile Computing*, vol. 31, pp. 94-106, September 2016.

[10] E. S. Lohan, J. Talvitie, R. Piche, P. Figueiredo e Silva, H. Nurminen and S. Ali-Löytty, "Received Signal Strength models for WLAN and BLE-based indoor positioning in multi-floor buildings" *International Conference on Localization and GNSS (ICL-GNSS)*, IEEE, 2015.

[11] A. A. Samuel A, C. J. Hill, and A. Isaac, "Evaluation of Ultra Wide-Band for Indoor positioning" *IIE Annual Conference. Proceedings*, 1-8, 2011.

[12] R. J. Kuo and J. W. Chang, "Intelligent RFID positioning system through immune-based feed-forward neural network" in *Journal of Intelligent Manufacturing*, vol. 26, pp. 755-767, August 2015.

[13] W. Vinicchayakul, R. Uppahad, P. Supanakoon and S. Promwong, "Study on performance of ultra wideband and narrow band propagation for an indoor positioning" *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA)*, Malacca City, pp. 194-198, 2016.

[14] H. Sallouha, A. Chiumento and S. Pollin, "Localization in long-range ultra narrow band IoT networks using RSSI" *2017 IEEE International Conference on Communications (ICC)*, Paris, pp. 1-6, 2017.

[15] H. Xu, Y. Ding, P. Li, R. Wang and Y. Li, "An RFID Indoor Positioning Algorithm Based on Bayesian Probability and K-Nearest Neighbor" in *Sensors*, vol. 17, (8), pp. 1806, 2017.

[16] E. Aitenbichler and M. Mhlhuser, "An ir local positioning system for smart items and devices" in *Proceedings of the 23rd IEEE International Conference on Distributed Computing Systems Workshops (IWSAWC03)*, pp. 334–339, 2003.

[17] R. Ma, Q. Guo, C. Hu and J. Xue, "An Improved WiFi Indoor Positioning Algorithm by Weighted Fusion" in *Sensors*, 15(9), 2015.

[18] P. Müller, M. Raitoharju, S. Ali-Löytty, L. Wirola and R. Piche, "A Survey of Parametric FingerprintingPositioning Methods" in *Gyroscopy and Navigation*, 7(2), pp. 107-127, 2016.

[19] S. Xia, Y. Liu, G. Yuan, M. Zhu and Z. Wang, "Indoor Fingerprinting Positioning Based on Wi-Fi: An Overview" in *ISPRS International Journal of Geo-Information*, 6(5), pp. 135, 2017.

[20] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems" in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067-1080, Nov 2007.

[21] M. Vossiek, L. Wiebking, P. Gulden, J. Weighardt and C. Hoffmann, "Wireless local positioning - concepts, solutions, applications" *Radio and Wireless Conference*, 2003. RAWCON '03. Proceedings, pp. 219-224, 2003.

[22] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system" in *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.

[23] Z. Jiang, J. Zhao, X. Li, W. Xi, K. Zhao, S. Tang and J. Han, "Communicating is Crowdsourcing: Wi-Fi Indoor Localization with CSI-Based Speed Estimation" in *Journal of Computer Science and Technology*, vol. 29, pp. 589-604, July 2014.

[24] K. V. N. Rajesh and K. V. N. Ramesh, "An introduction to crowdsourcing, " in *I-Manager's Journal on Information Technology*, vol. 4, no. 2, pp. 1, 2015.

[25] B. Wang, Q. Chen, L. T. Yang and H. C. Chao, "Indoor smartphone localization via fingerprinting crowdsourcing: challenges and approaches" in *IEEE Wireless Communications*, vol. 23, no. 3, pp. 82-89, June 2016.

[26] M. B. Kjaergaard, "Indoor location fingerprinting with heterogeneous clients" in *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 31–43, 2011.

[27] J. Machaj, P. Brida and R. Piche´, "Rank based fingerprinting algorithm for indoor positioning" in *Proc. IPIN* 2011, Guimarães,Portugal, pp. 1–6, September 2011.

[28] N. Boujnah and P. Korbel, "Crowdsourcing based terminal positioning using linear and non-linear interpolation techniques" *2016 Advances in Wireless and Optical Communications (RTUWO)*, Riga, pp. 101-106, 2016.

[29] N. Yu, C. Xiao, Y. Wu, and R. Feng, "A radio-map automatic construction algorithm based on crowdsourcing" *Sensors*, 16(4), pp. 504, 2016.

[30] P. Wilk, J. Karciarz and J. Swiatek, "Indoor radio map maintenance by automatic annotation of crowdsourced Wi-Fi fingerprintings" *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Banff, AB, pp. 1-8, 2015.

[31] S. Li, H. Li and L. Sun, "Privacy-preserving crowdsourced site survey in WiFi fingerprinting-based localization" in *EURASIP Journal on Wireless Communications and Networking*, December 2016.

[32] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen and M. Valkama, "High-Efficiency Device Positioning and Location-Aware Communications in Dense 5G Networks" in *IEEE Communications Magazine*, vol. 55, no. 8, pp. 188-195, 2017.

[33] C. Esposito and M. Ficco, "Deployment of RSS-Based Indoor Positioning Systems, " in *International Journal of Wireless Information Networks*, vol. 18, no. 4, pp. 224, 2011.

[34] C. Laoudias, R. Piche, and C. G. Panayiotou (2013), "Device self-calibration in location systems using signal strength histograms" in *Journal of Location Based Services*, vol. 7, no. 3, pp. 165-181, 2013.

[35] D. Zou, W. Meng and S. Han, "Euclidean distance based handoff algorithm for fingerprinting positioning of WLAN system" *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, pp. 1564-1568, 2013.

[36] X. Ge and Z. Qu, "Optimization WIFI indoor positioning KNN algorithm location-based fingerprinting" *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, pp. 135-137, 2016.

[37] J. Ma, X. Li, X. Tao and J. Lu, "Cluster filtered KNN: A WLAN-based indoor positioning scheme" *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, Newport Beach, CA, pp. 1-8, 2008.

[38] D. Zou, H. Sun, Y. Chen and N. Li, "A research on positioning technology based on internet of things" in *Applied Mechanics and Materials*, vols. 303-306, pp. 926, 2013.

[39] H. Li, B. Yu and D. Zhou, "Error Rate Bounds in Crowdsourcing Models" in *arXiv:1307.2674*, 2016.

[40] C. Wu, Z. Yang, Z. Zhou, Y. Liu and M. Liu, "Mitigating Large Errors in WiFi-Based Indoor Localization for Smartphones" in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6246-6257, July 2017.

[41] J. Talvitie, "Algorithms and Methods for Received Signal Strength Based Wireless Localization" in *Tampere University of Technology. Publication*, vol. 1365, Tampere University of Technology, 2016.

[42] M. Ponti, J. Kittler, M. Riva, T. D. Campos and C. Zor, "A decision cognizant Kullback–Leibler divergence" in *Pattern Recognition*, vol. 61, pp. 470-478, January 2017.

[43] B. Li, Y. Wang, H. K. Lee, A. Dempster and C. Rizos, "Method for yielding a database of location fingerprintings in WLAN" in *IEE Proceedings - Communications*, vol. 152, no. 5, pp. 580-586, 7 Oct 2005.

[44] S. He and S. H. G. Chan, "Wi-Fi Fingerprinting-Based Indoor Positioning: Recent Advances and Comparisons" in *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 466-490, First quarter 2016.

[45] S. H. Fang and T. N. Lin, "Accurate WLAN indoor localization based on RSS, fluctuations modeling" in *2009 IEEE International Symposium on Intelligent Signal Processing*, pp. 27-30, Budapest, 2009.

[46] S. H. Fang, Y. T. Hsu, B. C. Lu and W. H. Kuo, "A Calibration-Free RSS-Based Mobile Positioning System" in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, pp. 1-5, Yokohama, 2012.

[47] A. M. Bernardos, J. R. Casar and P. Tarrío, "Real time calibration for RSS indoor positioning systems" in *2010 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1-7, Zurich, 2010.

[48] B. J. Dil and P. J. M. Havinga, "On the calibration and performance of RSS-based localization methods" in *2010 Internet of Things (IOT)*, pp. 1-8, Tokyo, 2010.

[49] L. H. Chen, G. H. Chen, M. H. Jin and E. H. K. Wu, "A Novel RSS-Based Indoor Positioning Algorithm Using Mobility Prediction" in *2010 39th International Conference on Parallel Processing Workshops*, pp. 549-553, San Diego, CA, 2010.

[50] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for kNN Classification" in *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 43, April 2017.

[51] K. Maruo, T. Yamabe and Y. Yamaguchi, "Statistical simulation based on right skewed distributions" in *Computaional Statistics*, vol. 32, pp.889-907, September 2017.

[52] Z. Zheng, Y. Chen, T. He, F. Li and D. Chen, "Weight-RSS: A Calibration-Free and Robust Method for WLAN-Based Indoor Positioning" in *International Journal of Distributed Sensor Networks*, vol. 11, January 2015.

[53] X. Zhu, C. Vondrick, C. C. Fowlkes and D. Ramanan, "Do We Need More Taining Data?" in *International Journal of Computer Vision*, vol. 119, pp. 76-92, August 2016.

[54] E. S. Lohan, J. T. Sospedra, H. Leppakoski, P. Richter, Z. Peng and J. Huerta, "Wi-Fi Crowdsourced Fingerprint dataset for Indoor Positioning" in *Data*, vol. 2, October 2017.

[55] Z. Peng, P. Richter, H. Leppakoski and E. S. Lohan, "Analysis of crowdsensed WiFi fingerprintings for indoor localization" in *Fruct*, 2017.

[56] P. Réfrégier and F. Goudail, "Kullback relative entropy and characterization of partially polarized optical waves," in *Journal of the Optical Society of America A*, vol. 23, pp. 671-678, 2006.