TAMPERE UNIVERSITY OF TECHNOLOGY

MOHAMMADAMIR SHOKOUHIFARD
USER PRIVACY RISKS AND PROTECTION IN WLAN-BASED INDOOR POSITIONING

Master of Science thesis

Examiners: Assoc. Prof. Elena-Simona Lohan,
Prof. Robert Piche
Examiners and topic approved by the
Faculty Council of the Faculty of
Computing and Electrical Engineering
on 14.01.2015

ABSTRACT

MOHAMMADAMIR SHOKOUHIFARD: USER PRIVACY RISKS AND PROTECTION IN WLAN-BASED INDOOR POSITIONING
Tampere University of Technology
Master of Science thesis, 52 pages, 0 Appendix pages
October 2016
Master's Degree Programme in Information Technology
Major: Communication Systems and Networks
Examiners: Assoc. Prof. Elena-Simona Lohan and Prof. Robert Piche
Keywords: Indoor positioning, WLAN location privacy, Telecommunication security, Fingerprinting technique

Using location-based services (LBS) is the new trend for mobile users. LBS mostly exploit GPS and WLAN infrastructures for outdoor and indoor environments, respectively, in order to determine a user's location. After a location is known to a LBS, the network can provide location related contextual information such as nearby events, places, or navigation for the mobile users. Currently, LBS have been specifically growing rapidly in the domain of indoor positioning as more public places, e.g. schools, shopping centers, and airports are being equipped with WLAN networks.

The aforementioned situation leads to the fact that huge amount of tracking data gets possessed by a wide variety of different LBS and it poses the risk of location privacy violation of citizens. The problem is not only that this information reveals the places that a person has visited, but that it can also expose their behaviors and habits to the LBS and associated third parties. The conditions exacerbate as there are no appropriate regulations on how the tracking data is used by the LBS. In addition, the LBS data servers are under constant attacks by third parties who seek to access this kind of valuable data. Furthermore, the private sector has initiated the tracking of their customers in such places as shopping malls by means of simply collecting their MAC addresses.

The thesis is divided into two parts. In the literature part of this thesis, different indoor positioning techniques, location privacy leaks, and the solutions to tackle the problem will be explained. In the second part, we show practical implementation examples about how and at what extent a user may be positioned by the network, based simply on the mobile MAC address or using jointly MAC and signal strength information.

PREFACE

This master thesis has been conducted in the Faculty of Computing and Electrical Engineering in collaboration with Department of Automation Science and Engineering, Tampere University of Technology, Finland. I would like to express my sincere gratitude to my supervisors Associate Professor Elena-Simona Lohan and Professor Robert Piche for their support, guidance and patience during my work. Furthermore, I would like to thank Janne Redsven from IT help desk services who provided the TUT network data.

In addition, I would like to convey my thanks to my friends who were of great help and support throughout my whole education. Finally, I dedicate this thesis to my family without whom I would not be able to know myself.

I would also like to share a quote by Jean-Baptiste Alphonse Karr: plus ça change, plus c'est la même chose.

"The more it changes, the more the same thing it is"

Tampere, Finland
October, 2016
Mohammadamir Shokouhifard

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABE	Attribute-based encryption
AoA	Angle of Arrival
AP	Access Point
BSA	Basic Service Area
BSS	Basic Service Set
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
LAD	Localization Anomaly Detection
LBS	Location-based services
LCF	LB-SNS Certificate Fragment
LIA	LB-SNS Identity and Authentication
LPPMs	Location-Privacy Protection Mechanisms
MAC	Medium Access Control
mDNS	multicast Domain Name System
MS	Mobile Station
NLOS	Non-line-of-sight
NNM	K-nearest neighbor method
NNSS	Nearest Neighbor in Signal Space
OSI	Open Systems Interconnection
OUI	Organizationally Unique Identifier
PHY	Physical Layer
PPDM	Privacy Preserving Data Mining
QOS	Quality of Service
RSS	Received Signal Strength
SNR	Signal-to-noise ratio
STA	Wireless Station
TDoA	Time Difference of Arrival
ToA	Time of Arrival
UTAUT	Unified Theory of Acceptance and Use of Technology
VM	Verifiable multilateration
WLAN	Wireless Local Area Network
SNA	Secure Node Authentication
WSN	Wireless Sensor Network

LIST OF SYMBOLS

C	Centroid
C_f	Cost function
D	Euclidean distance
$E(...)$	Encryption function
E_e	Expected estimaion error
$Pr(...)$	Posterior distribution function
v	Set of vectors

1. INTRODUCTION

In this chapter, positioning systems will be briefly reviewed. In addition, the motivation, thesis objectives, and outline of the thesis will be presented.

1.1 Background and Motivation

Nowadays, the majority of people use different kind of communications devices thanks to ubiquitous computing. As a result, communications devices have become integrated in most aspects of our daily lives and the use of positioning systems have widely spread throughout societies. Global Positioning System (GPS) and mobile location estimation service, which is based on exploiting wireless communications networks, are the most common ways used for localizing a device. Location information is used by location-based services (LBS) to provide useful services, such as navigation or finding a nearby event for their clients.

Due to the growing number of users who become interested in using Location-Based Services, the LBS providers access to huge amount of information about attitudes of individuals as more people use these kinds of services. Although the users benefit from receiving the localization services, their whereabouts get exposed to the location service providers. It is important to point out that a user's location trace not only reveals their location on a map, but also the contextual data related to a trace shows a person's activities, interests, and habits.

With the growing number of cloud data centers in recent years, companies and governments have been able to store users tracking data for long-term periods for different reasons such as advertising, service providing, or national security. The downside of this trend is that the privacy of users is being put at risk due to the lack of effective privacy rules and potential adversary attacks to these data centers. Problems arise as these companies and the governments have no incentive in modifying their data collection and storage strategies to benefit the end users. Therefore, it is important to enhance the privacy of the users by means of privacy-preserving techniques, that brings about a trade-off between service quality and the level of privacy [34].

Indoor positioning can be considered as a new trend in the field of mobile positioning and it is creating new opportunities for businesses. As people spend so much time indoors, working, shopping, and so on, the LBS providers are trying to facilitate the ability for indoor positioning for their clients. One of the most practical systems to be used for indoor positioning is wireless local area network (WLAN), where a lot of effort has been put in order to increase the positioning accuracy. However, privacy related issues in positioning have been vastly neglected so far. Since every mobile device connecting to a wireless network is recognized by its unique Media Access Control (MAC) address, it is an unsolved challenge for the mobile users to be able to anonymously use the LBS [29].

1.2 Thesis objectives and author contributions

The objective of this thesis is to demonstrate the vulnerability of user location privacy in wireless local area networks. The main idea is to investigate how accurately a mobile device can be localized by its associated network administrator without the acknowledgment of the user. Furthermore, in order to provide a perspective on the subject of location privacy, a comprehensive literature review has been conducted about potential privacy leaks and privacy-preserving methods in different network layers. In addition, the procedure of data utilization and exploiting them by a network administrator or an adversary for user localization purposes has been simulated and analyzed in MATLAB environment.

1.3 Outline of the Thesis

In this thesis, the concept of privacy in indoor positioning will be described, and the fundamentals of indoor positioning will be explained. This thesis consists of six chapters. While this first chapter acts as a general introduction to the topic, chapter 2 explores WLAN-based indoor positioning, and elaborates on the key principles of positioning methods, i.e. proximity sensing, model-based approaches and fingerprinting-based approaches. Chapters 3 and 4 analyze privacy leaks and current solutions, respectively. Furthermore, chapter 5 and chapter 6 illustrate how indoor positioning can be performed in a terminal-based system and network-based system, respectively. In addition, we present examples with real data on how a network can locate the user with only basic information, such as MAC address or both MAC address and RSS. Finally, Chapter 7 summarizes the main points of this thesis.

2. WLAN-BASED INDOOR POSITIONING

IEEE 802.11 standard specifies the characteristics of the Medium Access Control (MAC) and Physical Layer (PHY) for Wireless Local Area Networks (WLANs) and has been modified periodically since its first release in 1997. As a result, a wide variety of amendments have been made to the original version of IEEE 802.11. Currently, the most widespread versions of WLAN implementations are: IEEE 802.11b functioning in the license-free 2.4 GHz frequency band and having an average throughput of 11 Mbps that provides a transmission range less than a few hundred of meters and IEEE 802.11g which benefits from a higher data rate of 54 Mbps. In addition, IEEE 802.11n is the most common standard that supports multiple-input multiple-output antennas over 2.4 GHz and 5 GHz frequency bands. By using multiple antennas, the network throughput increases to 600 Mbps. WLANs contain one or more Access Points (APs), each of which serve a group of Wireless Stations (STAs) that form a Basic Service Set (BSS). An AP covers a limited area called as Basic Service Area (BSA) [1].

Nowadays, WLAN services are widely available to the public, both in private and public premises such as hospitals, airports, shopping centers, university campuses, homes, etc. The current conditions make it promising for exploitation of WLAN equipment for demanding indoor positioning applications. In addition, a lot of effort has been put into utilizing the Global Navigation Satellite Systems (GNSSs) for indoor positioning without achieving the desired goals [28]. For instance, one well established positioning system is Global Positioning System (GPS) that has been open to the public since 2000. However, owing to the fact that GPS signal strength drops to about -160 dB indoors, it is hard for conventional GPS receivers to function correctly indoors [41].

The procedure of acquiring the location of a STA is called WLAN positioning. It can be achieved by measuring the signal characteristics that are position dependent such as Time of Arrival (ToA), Time Difference of Arrival (TDoA), Angle of Arrival (AoA), signal-to-noise ratio (SNR) and Received Signal Strength (RSS). In order to implement range measurement techniques like ToA, TDoA, and AoA, not only precise time synchronization and angle measurements are required, but also, due to non-line-of-sight (NLOS) propagation, shadowing, and multipath effect, additional short time delays occur that complicate the positioning process. Consequently, the three above-mentioned features are not widely used in the current WLAN-based indoor positioning systems. On the other hand, SNR and RSS values are calculated by capturing the IEEE 802.11 beacon frames that are broadcast by non-synchronized WLAN transceivers periodically. By using RSS information for positioning, there is no need for costly extra network equipment, thus the majority of WLAN positioning techniques are deployed or being developed based on SNR and RSS measurements [28].

Three main positioning techniques that exploit the beacon frame features will be discussed in this chapter as follows:

- Proximity sensing.
- Model-based approaches.
- Fingerprinting-based approaches.

2.1 Proximity sensing

Proximity sensing is a straightforward method in which a mobile device gets the MAC address of the access point with the highest signal strength. It is also known under the name of AP-ID, by analogy with the Cell-ID method in cellular networks. Afterwards, the MAC address is sent to a location provider, such as Google, Skyhook, HERE maps, etc., which can map the MAC address to a physical location in the real world. As a result, location of the user is exposed to the location provider [19].

Owing to the simplicity of this method, it fails to prepare a satisfactory level of accuracy. The position error can be even as high as hundred of meters, based on the AP ranges and on the number of APs installed in the premises. In an ideal system, in which APs are densely deployed, the accuracy increases significantly. Embedding room or corridor numbers into the beacon frames can also make the system more accurate [28] but this requires modification to the WLAN system.

2.2 Model-based approaches

Model-based approaches, also referred to path-loss-based approaches, use mathematical radio propagation models for calculating the position of a user by comparing the

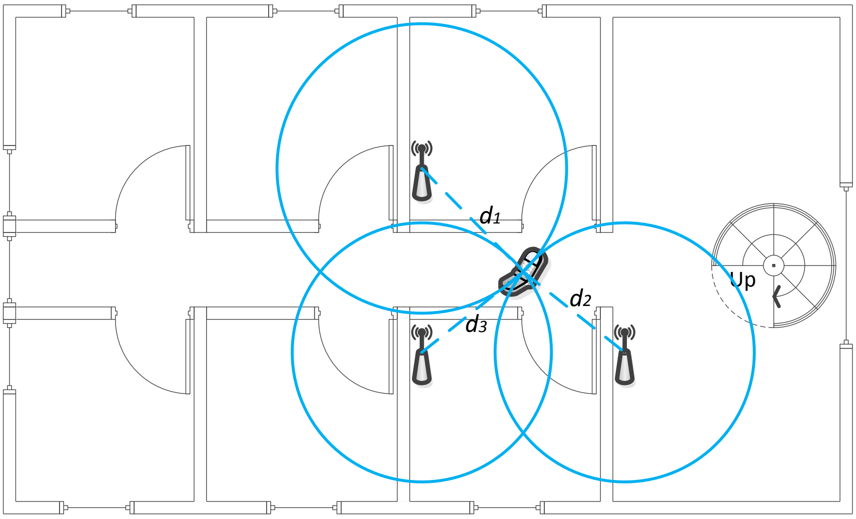


Figure 2.1 *Lateration technique with three APs.*

RSS with the transmitted signal strength from several access points. Since a lateration technique is used in this approach, the position of the access points must be known to the system as depicted in Figure 2.1. When creating the model, the impact of dynamic and chaotic conditions of the indoor environment on the transmitted signals must be taken into account because the signal attenuation occurs constantly due to free-space path loss, shadowing, and the multipath effects caused by receiving different versions of the same signal [29]. Absorption, scattering, diffraction and reflection caused by the walls, floors, furniture, and so on are extremely difficult to be predicted and modeled in a building. Therefore, the positioning process by using the mathematical models do not produce accurate results; furthermore, additional overheads are incurred because the signal strength at the transmitter must be known for calculating the path loss [28].

2.3 Fingerprinting-based approaches

Fingerprinting-based approaches comprise two phases. In the offline phase, also known as training or calibration, the access points' signal strength at predefined reference points, for example as shown in blue dots in Figure 2.2, are sampled and saved into a *radio map*. The data structure of the radio map consists of a set of vectors $v = (x, y, z, RSS_1, ..., RSS_n)$ in which (x, y, z) are the coordinates of a reference point and n is the number of access points 'heard' in that reference point. Therefore, the radio map contains spatial features of RSS patterns in a location [19],[29].

In the online phase, after the RSSs at a mobile station (MS) are measured and

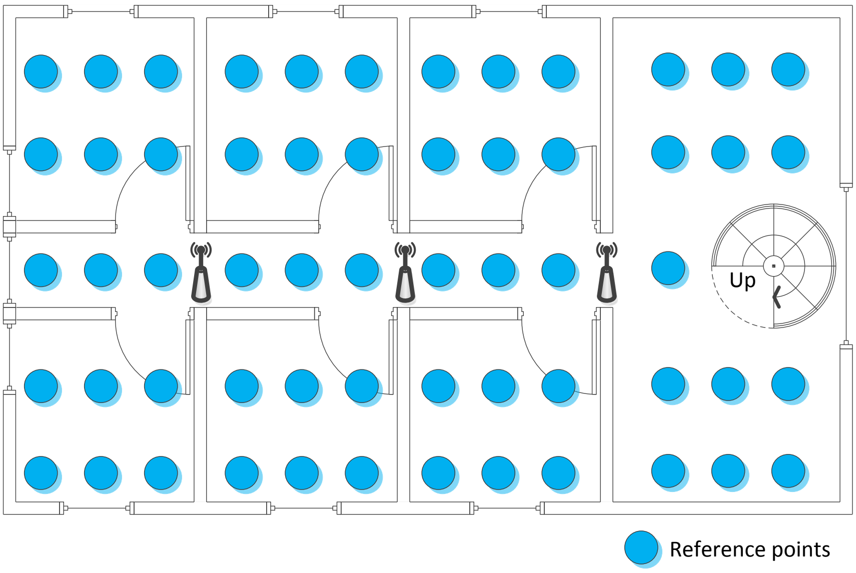


***Figure* 2.2** *Example of reference points.*

the best matching reference point in the radio map is found, the coordinates of that reference point is announced as the position of the MS. Therefore, the main challenge in creating a radio map is that it is a time consuming process to collect the data and the radio map has to be recreated regularly to account for changes in the topology of the access points. Forming a radio map in this way is also called as *empirical approach* [19].

Two other methods are used to deploy the empirical approach; deterministic and probabilistic. In the former, as proposed by Bahl and Padmanabhan in their positioning system *RADAR* in [5], first the mean value of several recorded RSSs is assigned to a reference point in the radio map. Then, in the online phase a metric introduced as *Nearest Neighbor in Signal Space (NNSS)* technique uses Euclidean distance to determine the location of the user:

$$D = \sum_{k=1}^{n} \sqrt{(RSS_{o,k} - RSS_{r,k})^2}. \tag{2.1}$$

where $RSS_{o,k}$ is the observed signal strength by the device and $RSS_{r,k}$ is the recorded one in the radio map. Having calculated the set of D, the reference point that has the minimum relevant D is reported as the location of the user. In a dynamic experimental location that had an area of 980 m^2, over 50 rooms, 70 reference points and 3 access points, RADAR achieved a median deviation of approximately 3 meters while reducing the number of reference points to 40 and performing 3 times of RSS measurements throughout the online phase [5]. Nevertheless, the accuracy

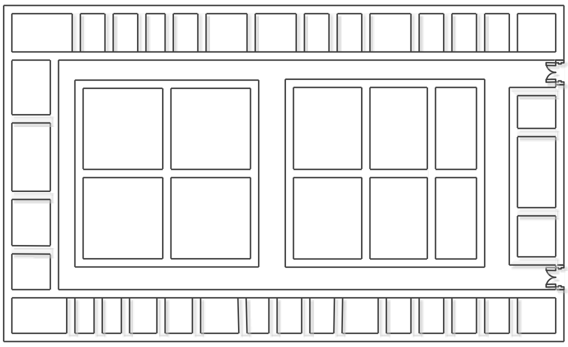
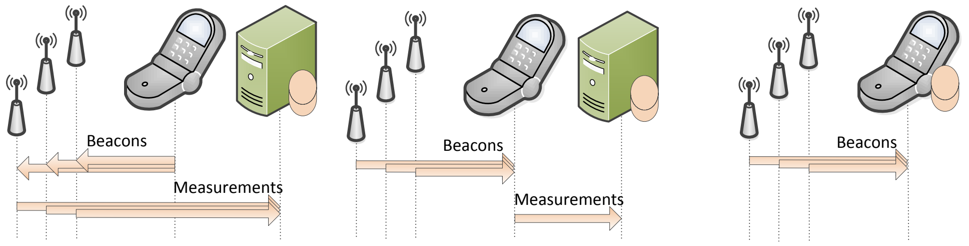


Figure 2.3 *Plan of the building in which the measurements were done in [42] . The dimension is 68 meters by 26 meters and the width of the corridor is 1.5 meters.*

of the system much depends on the orientation of the RSS and moving pace of the MS. The disadvantage of deterministic method is that NNSS must be run over each of the elements in the radio map to match the corresponding reference point.

Probabilistic fingerprinting is the other method that uses probability distributions of the signal strength values to alleviate noise effects during the online phase that cause deviation from the recorded RSSs. Furthermore, by using location clustering, data processing can be facilitated significantly. Using both the probability distributions of the signal strength and location clustering in a positioning system is called *Joint Clustering* and functions as follows. To create a cluster, the reference points that hear the same set of APs are grouped during the offline phase and saved in the radiomap. Afterwards, during the online phase, the cluster that has the most common APs with the the heard APs by a user's device is chosen. Finally, positioning is implemented by comparing the probability distributions of the RSSs in a set of APs in the chosen cluster and the heard APs [28].

Youssef et al. in [42] introduced a positioning system that exploits the joint clustering technique to reduce both the noise effect and computational costs. The proposed system was tested on an IEEE 802.11 WLAN framework in a university building having an area of 6096 square meters. In order to create the radio map, reference points were defined on a grid having cells located 1.2 meters from each other along the corridors . In addition, there were 110 reference points in each of which, around 4 access points were heard, and 300 samples were measured in each reference point at 1 second periods. As a result, positioning was performed on a terminal-based system effectively with an accuracy of 2 meters in 90% of cases. The plan of the building used in the experiment of [42] can be seen in Figure 2.3. By using joint clustering technique computational complexity reduced to the extent that positioning could be implemented on user's mobile device.



(a) Network-based (b) Terminal-assisted (c) Terminal-based

Figure 2.4 *Three configuration possibilities used in fingerprinting.*

In order to deploy the above mentioned system, during the offline phase a joint probability distribution model was formed, using Maximum Likelihood Estimation method, given the RSSs in each reference point. In addition, the location was grouped into clusters, each of which were built up of the reference points that were covered by the same APs. Then, during the online phase a cluster was chosen so that the most probable location according to the measured RSS would be searched in it. Afterwards, using Bayesian inference, the most probable reference point related to the observed RSSs was determined.

Fingerprinting can be developed based on three configurations. The different ways that tasks are assigned to three network entities (i.e. wireless terminals, access points, and servers) during positioning process, not only determine how the system functions but also affect scalability and privacy significantly. Figure 2.4 shows three possible configurations in an infrastructure-based positioning system as follows:

- Network-based.
- Terminal-assisted.
- Terminal-based.

First, there is the network-based case as it can be seen in Figure 2.4(a) in which the radio map is formed using signal strength of beacon frames broadcast by the terminals and measured by the access points in offline phase. Then, during the online phase, the location of the terminal is calculated by the servers and announced to the user [28]. The advantage of this system is that it is highly scalable, since there is no need to adjust the system for each new terminal and they can be automatically located as they enter a place that makes it convenient for asset tracking. However, it can turn into a privacy invasion tool as people can be tracked sporadically without being notified [19].

Second, there is the terminal-assisted case in which the terminal is responsible for measuring and sending RSSs to the servers and servers are responsible for maintaining the radio map and determining the location as depicted in Figure 2.4(b) [28]. This system is deployed when the current infrastructure cannot be modified in order to take on the whole workload, for functioning as a network-based positioning system. Although the users know about exposing their data to the system, there is no other choice if they intend to be located [19]. As a result, user location privacy in this case is also low.

Finally, there is the terminal-based case in which the radio map is saved in the wireless terminal by recording the RSS downlink measurements. Using the radio map and RSS observations, the terminal determines its location as shown in Figure 2.4(c) [28]. This method is assumed to be the most privacy-preserving method. However, in practice, it is challenging to be implemented due to the fact that there are constraints on transmission and maintenance of the bulk radio maps on a single device [19].

3. POTENTIAL PRIVACY RISKS IN WLAN-BASED INDOOR POSITIONING

Location-Based Services (LBS) provide the opportunity for mobile users to discover their surroundings and are proliferating in many dimensions of our everyday lives so that market for LBS could rocket from 2.8 to 10.3 billion dollars in a five-year period from 2010 through 2015 [46]. Social networks, such as Facebook or Twitter enable people to connect an activity to a location, tools such as Foursquare let people check in as they visit a place, and applications such as Yelp offer the opportunity to find proximate businesses or events. These are some examples that point out to the fact that LBS are being widely used in most aspects of our activities. However, a user may contribute to build up a tracking map history of all their visited locations unwillingly on a data server. Such tracking map history may be kept by quite a few of the LBS for an unlimited time. Moreover, privacy-related issues arise as on the one hand, existing privacy rules and regulations belong to decades ago, when LBS was not yet introduced, and on the other hand, the majority of them take sides against user privacy. All the above-mentioned matters make privacy subject of vital importance nowadays [34].

LBS usually exploit a vast variety of methods to collect customer information. For instance, Google implements its LBS based on the following three types of data: first, there is the *implicit location information*, in which the interest of a user in a specific location is determined, for example, if a user just looked up a place on the Internet. Second, there is the *Internet traffic information*, in which the IP address of a user is used to offer services to the end-users in an appropriate language and context according to the residing country. Third, there are the *device-based location services*, that require precise positioning of the device by using GPS signals, RF sensors, WLAN signals, and cell tower IDs [3]. This enables the LBS to integrate identity, location, and search terms of users, in order to produce a comprehensive data history for each of them; in addition, social networking sites can access their customers' photos, comments, friends list, gender, and etc. enabling them to create more detailed profile datasets [34].

To sum up, privacy violation of individuals or organizations can threaten civil liberties, which is an important indicator of the degree of democracy in a society. For instance, the U.S. governmental agencies account for the majority of requests to access citizens location information [34].

3.1 User privacy concerns based on survey results

Yun et al. applied Unified Theory of Acceptance and Use of Technology (UTAUT) in [43] on adoption and diffusion of LBS applications. The purpose was to ascertain the role of users' privacy expectations on the key constructs of UTAUT. The notion of privacy in LBS stems from the fact that of all cellphone users in North America and Europe, 33% and 20% of them use LBS respectively [33]. Yun et al. tried to determine how privacy concerns impact on spread of LBS applications. They found that subjects tend to respond to the technology as follows. First, privacy concerns slow down the spread of the technology among those who use the technology for the sake of performance expectancy, e.g. *productivity-oriented LBS applications* such as search engines, maps, and taxi services. Second, those who are more concerned about privacy seem to be the ones who use LBS more frequently than others. Therefore, users accept to take some risk due to the social influence, while they are still worried about their privacy. All in all, in order to increase the penetration rate of LBS for the first group it is necessary to satisfy both the performance and the information privacy expectations of the users. Meanwhile, for the second group persuasive marketing techniques to increase the social influence seems to be promising regarding the spread of LBS usage.

According to Ozer [34], there is already a high demand for accessing location tracking by law enforcement agencies and this poses a risk to users' privacy. A survey reveals that 55% of LBS users are already concerned about their privacy [2]. Consequently, LBS diffusion can be hindered if the consumers believe that LBS applications are not to their advantage, given privacy issues. The study in [40] shows that 34 percent of LBS providers lack privacy policies. Regarding the companies that do have privacy policies, some lack transparency in their policies on how they use the information, and some others place no restrictions on the amount of data they collect and how they use them. In addition, location service providers do not give any guarantee to stop them from handing the information to other companies or organizations. As a result, it is worth mentioning that users would feel more confident in dealing with LBS having concrete privacy practices.

In another study by Cheng et al. [8], public Wi-Fi users' information was gathered from 15 airports located in 4 countries. Analytics show that the exposure of transmitted data was approximately 70 percent while the public awareness about privacy vulnerability, caused by using public hotspots is quite low. The problem originates in the fact that there are no security methods implemented in this kind of networks, since they are meant to be as easily accessible as possible to travelers; on the other hand, based on the public networks policies, it's usually the users responsibility to take care of their privacy. Moreover, advertising businesses make money gathering consumers' private data.

Privacy leakage sources are categorized into three different areas in [8]. Firstly, there are users' devices that may use a multicast Domain Name System (mDNS) protocol that includes some human readable data, such as name resolution. Secondly, there are websites that may disclose a user's attributes using HTTP cookies, since many of Internet content providers track a user who surfs their website pages. Thirdly, there are online advertising companies that send advertisements to a user based on their browsing history, and this history can also be easily used to extract some other relevant information about the user. Overall, cellular networks offer higher levels of privacy preserving than public WiFi networks, but lead to high cost for subscribers.

3.2 Interception of user position by third parties

Currently, LBS are mainly provided by a means of communication between a user and a third party that results in the disclosure of user location to the third parties. According to [36], malicious LBS privacy issues can be categorized as *query privacy* and *location privacy*. In the former a user's query may be associated to their identity or an attribute of them by an LBS provider. For instance, regular queries about a political party's headquarters can reveal one's political trends. In the latter, a user's information is revealed according to their location. For example, if a user is located in a library it may infer that the user is interested in reading books. Location privacy and query privacy are closely related, so that locating a user can help to disclose a user's identity even though they use a location anonymizer. Accoriding to [22], a location anonymizer is a trusted party that functions as a medium between a user and a service provider. Assuming that the communication between the user and service provider is through an unsecured channel, an attacker could reidentify the users by accessing their locaiton queries. The task of the trusted party is to hide the exact location of the query issuer among $k-1$ other users' queries, in order to decrease the probabilty of reidentification to $1/k$.

As metioned in [21], mobile advertising is another concept that exploits context information of a user in order to offer tailored advertisements. For instance, as a user passes by a store, the relevant advertisements about that store pop up on their web browser. Although the benefits for both businesses and customers soar, privacy issues arise at the same time, since the advertiser accesses location data and activities of the users. To use the services, a user must trust the service provider; however, due to the vast amount of data and its significance, it is unrealistic to expect a company to be able to combat all the different adversarial attacks against their database systems. Also, an employee working with a *server-only* personalization system, e.g. Google and Yahoo, can easily access users identity. To address the problem, *client-only* systems can be employed so that sensitive data are stored on the user device. Systems such as Privad [20] and Repriv [16] offer a trade-off between server-only and client-only architectures.

According to [24], an Android mobile phone can be tracked by measuring the power consumed by its cellular radio. The idea is that there is a direct relationship between a cellular phone's distance from the associated cellular tower and its power consumption. Therefore, by creating a power consumption map for different geographical locations, one can find the location of a mobile user. However, this method is effective as long as the user is moving and the attacker has a prior knowledge of the places a user may visit. In addition, it is important to know the exact power consumption used only by the cellular radio antenna not by the applications. Still, a malicious application can track a user without asking for permission to access location data, as it only needs to access power data and network connectivity data.

3.3 Quantifying user privacy

Privacy preserving data mining (PPDM) is the practice of extracting information from datasets without exposing individual information about the users. Working with large amounts of data makes the process of privacy preserving difficult, due to the possible different ways in aggregating information about datasets. The lack of a comprehensive standard in the functionality of employed privacy preserving algorithms, in addition to the availability of many different types of them, poses serious privacy related problems when designing and using privacy preserving algorithms. In order to tackle the problem, a set of metrics are proposed in the literature that will be discussed in the following.

One of the novel researches done in the field of location privacy is the work done by Shokri et al. [38], in which quantifying different Location-Privacy Protection Mechanisms (LPPMs) is the aim of the proposed system. They claim that the lack of a standardized method for quantifying LPPMs' performance is a serious issue in location privacy as in PPDM. They believe that there is a lack of emphasis on modeling the adversary's attack conditions leading to inaccurate evaluation of privacy risks of the users. A framework is introduced so that takes into account both the prior knowledge and the attack methods of the adversary. The metric used in their quantification method is based on statistical methods, which is the estimation error of the attacker, presented as a tool: *Location-Privacy Meter* [37]. The location-privacy framework comprises of six different entities as shown in Table 3.1.

Table 3.1 *Location-privacy framework, from [38].*

U	A set of users
A	A set of actual traces of the users
LPPM	Location-privacy preserving mechanism
O	A set of traces, which is produced by LPPM, and observed by adversary
ADV	Adversary
METRIC	Performance of the adversary

Users produce a set of traces **A**, which are obfuscated traces produced probabilistically by the LPPM as **O** and the adversary tries to infer to **A** by having access to **O**s. X is defined as a set of values that the function of the attacker's objective can take for an attack. Since the attacker uses probabilistic methods, it can only extract an estimate of the real traces from the obfuscated location in the form of a posterior distribution function $Pr(x|O), x \in X$.

Adversary performance is the core of the location privacy quantification and can be evaluated according to its accuracy, certainty and correctness. However, the correctness of the attack is the metric used to quantify the location privacy. Equation 3.1 is the expected estimation error of the adversary where $\widehat{Pr}(x|O)$ is an estimate of the posterior distribution function $Pr(x|O)$, x_c is the real distance from the user, and x is the observed track by the adversary.

$$E_e(x_c, O) = \sum_x \widehat{Pr}(x|O)\|x - x_c\|. \tag{3.1}$$

According to Zhang et al. in [44], there are already many LBS Privacy Protection Mechanisms (LPPMs) proposed in the literature by researchers. Nevertheless, due to the lack of a generic adversarial model, it is problematic to evaluate the performance of the current LPPMs. Zhang et al. proposed a privacy quantification model to create a general adversarial model based on the adversary's estimation error. As a result, the privacy quantification model allows interpretation and comparison of different LBS privacy metrics based on a common perspective. The evaluation has been done based on an adversarial observation and the Bayes conditional risk was used as the privacy metric.

4. CURRENT SOLUTIONS FOR PRIVACY-PRESERVING INDOOR LOCALIZATION

In this chapter different methods and concepts for preserving location privacy are explained in order to represent a comprehensive perspective about addressing the issue.

4.1 Location cloaking

Location cloaking is usually the practice of adjusting the loaction information resolution according to specified anonymity needs of a user. This adjustment is done by reducing the accuracy in terms of space (*spatial cloaking*) or time (*temporal cloaking*) [18].

As explained in [10], by using spatial cloaking techniques, a client's precise location gets degraded into a spatial area for preserving their privacy given the clients' privacy expectations. In addition, k-anonymity and minimum spatial area are introduced as the most common techniques used currently. Mrorover, A spatial cloaking algorithm is implemented whether on the client device or a trusted third party to blur their location before connecting to a location-based database server.

4.1.1 k-anonymity

In order to use location-based services, the user has to report their location to the service that lets the service identify the user. k-anonymity tackles this problem by assigning at least $k-1$ other users in the cloaked location and exposing this location to the service. According to [10], there are two different ways to cloak a location by this means. Firstly, using a trusted central server in order to deploy k-anonymity algorithms. Here, the server connects to the location-based service instead of the user and the server knows the location of all its contacting users. Secondly, the users themselves determine the cloaked location by trusting in each other.

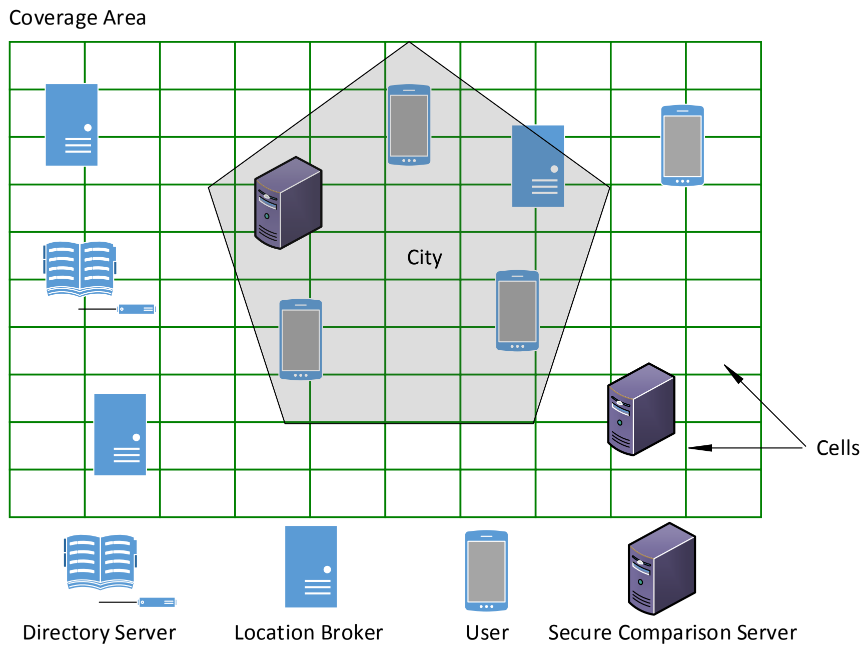


Figure 4.1 *Distributed k-anonymity system model. The directory server provides the contact information of a location broker for the user to register their location (cell). Afterwards, the user can contact the location brokers and a secure comparison server to get to know if there exist at least k users in their cell, from [45].*

Zhong and Hengartner in [45], propose a distributed k-anonymity system in which a set of servers are deployed to calculate the cloaking areas without the need to trust a single server or all the other users. The system works in two phases: first, by using a distributed k-anonymity protocol, a user determines if there are k other people in the area in collaboration with a set of different servers and a third party. Second, another protocol stops the association of a user to more than one server.

The distributed k-anonymity system introduced in [45] comprises 4 different parts as shown in Figure 4.1:

- A location broker.
- Users.
- A secure comparison server.
- The directory server.

First, there are multiple location brokers each of which belong to a different organization and maintain the information about a subset of users in the current cell throughout the coverage area.

Second, there are users that have the freedom to report their current location to one of the location brokers.

Third, there is a secure comparison server that covers the whole coverage area and deals with the users to acknowledge them if there are at least k other users in the current cell. In addition, each secure comparison server is maintained by a different organization.

Fourth, there is the directory server that broadcasts the contact information of the coverage area for both the location brokers and the secure comparison servers.

k-anonymity is prone to two kinds of attacks, namely *homogeneity attack* and *background knowledge attack*. Homogeneity Attack is the practice of extracting a user attribute from a dataset, since all the attribute values are the same in one equivalence group [35]. Background knowledge attack is the practice of associating certain attributes in a dataset to a user, due to the access of an adversary to background knowledge about a user's attributes [13].

In order to illustrate how k-anonymity can be defeated by an adversary, an example from [32] will be preseneted in the following. In this example user attributes are categorized as sensitive and non-sensitive. Sensitive data includes private information e.g. health details. Non-sensitive data consists of information that is usually publicly accessible e.g. zip code, age, and nationality. Moreover, it is the sensitive data that is not supposed to be exposed to an attacker. This example includes medical records of a fictitious hospital that can be seen in Table 4.1 as a 4-anonymous table.

In a homogeneity attack, suppose an attacker has access to the table and knows that a specific patient is 31 years old and lives in a neighborhood with the zip code 13053. Firstly, they can see that the patient must belong to one of the two groups having a zip code 130**. Secondly, considering the age column it is obvious that the patient must belong to the group with age assigned as 3*. As a result, it can be concluded that the patient has cancer due to lack of diversity in the sensitive attributes.

In a background knowledge attack, suppose that the attacker knows that a patient is a 21-year old Japanese person living in zip code 13068. Therefore, the patient must belong to the group having zip code 130** and age of < 30. Consequently, the patient may have either a heart disease or viral infection. However, the adversary has the background knowledge that Japanese tend to have low incidence of heart diseases. So the attacker concludes that the person has a viral infection.

Table 4.1 *Example of data division as in a 4-anonymous table given the non-sensitive data.*

	Non-Sensitive			**Sensitive**
	Zip Code	**Age**	**Nationality**	**Condition**
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	$3*$	*	Cancer
10	130**	$3*$	*	Cancer
11	130**	$3*$	*	Cancer
12	130**	$3*$	*	Cancer

4.1.2 L-diversity

According to Machanavajjhala et al. in [32], l-diversity is similar to k-anonymity, so that the same algorithms are used in both of them. However, l-diversity provides diversity among the sensitive attributes of the anonymized group by preventing sensitive attributes to be disclosed to attackers by making a user's attributes indistinguishable from l-1 other attributes.

A 3-diverse table has been depicted in Table 4.2. Given the abovementioned scenario for k-anonymity, the homogeneity attack is defeated as the attacker cannot figure out if a 31-year old patient living in zip code 13053 has either heart disease, viral infection, or cancer. In the background knowledge attack, a 21-year old Japanese patient living in zip code 13068 given the attacker's background knowledge, may have either cancer or a viral infection. It is worth mentioning that l-diversity method gets more challenging to implement when there exist multiple sensitive attributes in the table. In addition, diversity in k-anonymity is achieved according to the non-sensitive data, whereas, in l-diversity, diversity is achieved according to the sensitive data.

4.1.3 Onion routing

In order to preserve privacy in a network, both the content of a message and the communicating parties must be hidden from adversaries. Cryptography helps to hide the content of a message, however, the sender and receiver can still be identified

Table 4.2 *Example of data division as in a 3-diverse table given the sensitive data.*

	Non-Sensitive			**Sensitive**
	Zip Code	**Age**	**Nationality**	**Condition**
1	1305*	≤ 40	*	Heart Disease
4	1305*	≤ 40	*	Viral Infection
9	1305*	≤ 40	*	Cancer
10	1305*	≤ 40	*	Cancer
5	1485*	> 40	*	Cancer
6	1485*	> 40	*	Heart Disease
7	1485*	> 40	*	Viral Infection
8	1485*	> 40	*	Viral Infection
2	1306*	≤ 40	*	Heart Disease
3	1306*	≤ 40	*	Viral Infection
11	1306*	≤ 40	*	Cancer
12	1306*	≤ 40	*	Cancer

through traffic analysis and eavesdropping. Onion routing implements anonymous connections in packet-switching networks that is applicable to both connectionless and connection-based protocols in bidirectional communications networks. Onion routing runs by dynamically creating anonymous connections throughout a series of onion routers. An onion routing protocol is implemented in three different stages; connection setup, data movement, and connection teardown as follows [17]:

In the first stage an onion is built by the initiator, in which a path is created throughout the network. The characteristics of the route are then added repeatedly to the payload of the message; then by the use of the onion public key, the data gets encrypted. Afterwards, the message is sent and one layer of encryption is removed by each onion router through the path until the message arrives at the receiver as a plaintext. Finally, connection teardown occurs at the end or at any point along the path when needed. As a result, it is difficult to trace a packet along the network. However, due to the generated data overhead, an onion routing network consumes high amount of power for its operation.

4.1.4 Location uncertainty via resolution reduction

Another method for preserving location privacy is proposed by Cheng et al. in [9] by which the location information in space and time is degraded in order to lower the resolution of the location data that is accessible to the location-based services. The framework cloaks the location information of a user by reporting a larger area as the current location of the user.

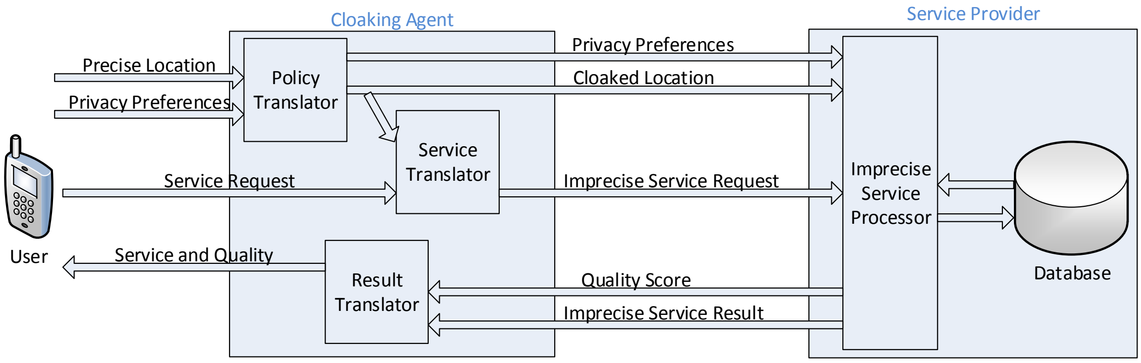


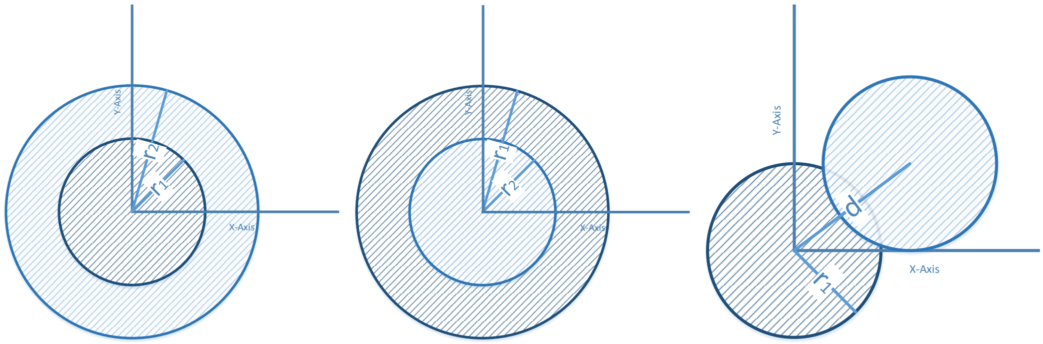
Figure 4.2 Managing Privacy and Service Quality with the Cloaking Agent, from [9].

Therefore, a user gains a higher level of location privacy by reducing the fitness of their location information granularity. On the other hand, the quality of service provided by the LBS also reduces. In order to provide a trade-off between the quality of service and the privacy of the users, adjusting the accuracy of the location information can be an option for the users. For instance, a social networking service that provides the information about some specific event in a city does not need to know the precise location of a user within the city. The proposed system is shown in Figure 4.2 where it lets the user to choose their location, service request, and privacy preferences and report it to the cloaking agent. The cloaked location and the imprecise service request is then generated by the cloaking agent. Finally the LBS responds to the user with the appropriate information. [9]

4.1.5 Dummy messages

As described by Kido et al. in [27], in this technique, a user produces several bogus location data (dummies) and sends them to the LBS along with their true location. Then the LBS respond to all these location queries without knowing which of them really belongs to the user while the user knows which information response to rely on.

In addition, the LBS should not be able to infer the true location of the user. For this, the produced dummy messages must not be distinguishable from the true data. For example, such a true datum can be the distance traveled by a car in a navigation system. This traveled distance is limited during a certain period of time. Consequently, a set of randomly generated dummy messages makes it easy for the LBS to determine if it is false data. In the algorithm in [27] the location of the first dummy message is generated randomly and the next dummy message gets generated in the neighborhood of the location of the current dummy message.



(a) Enlargement (b) Reduction (c) Shifting

Figure 4.3 *Three obfuscation methods. The dark blue circle shows the measured area of user location and light blue circle shows the obfuscated area of user location.*

4.2 Obfuscation

As explained by Ardagna et al. [4], obfuscation is the act of perturbing users' measured location information, in order to preserve their location privacy. The key features of location obfuscation proposed in their paper are: firstly, to let users choose their privacy preferences and secondly, to implement the privacy preferences against adversary attacks. Since the precision of user position is prone to measurement errors, which are caused by sensing technologies, the users' position is most often illustrated as *planar circular areas.* In [4], location obfuscation is implemented by reducing location accuracy by means of three obfuscation operators, as illustrated in Figure 4.3:

- *Enlargement:* Increasing the radius of the measured area from r_1 to r_2 with $r_1 < r_2$, in order to decrease the probability of the real location of a user to be in the obfuscated area.
- *Reduction:* Although reducing the radius of the measured area from r_1 to r_2 with $r_1 > r_2$, seems to be in conflict with the obfuscation concept, it reduces the probability of the real position to be in the obfuscated area.
- *Shifting:* Shifting the center of the measured area from c_1 to c_2 so that given d to be the distance between the centers of the areas then $d < 2r_1$.

4.3 Connection-free RSS measurement

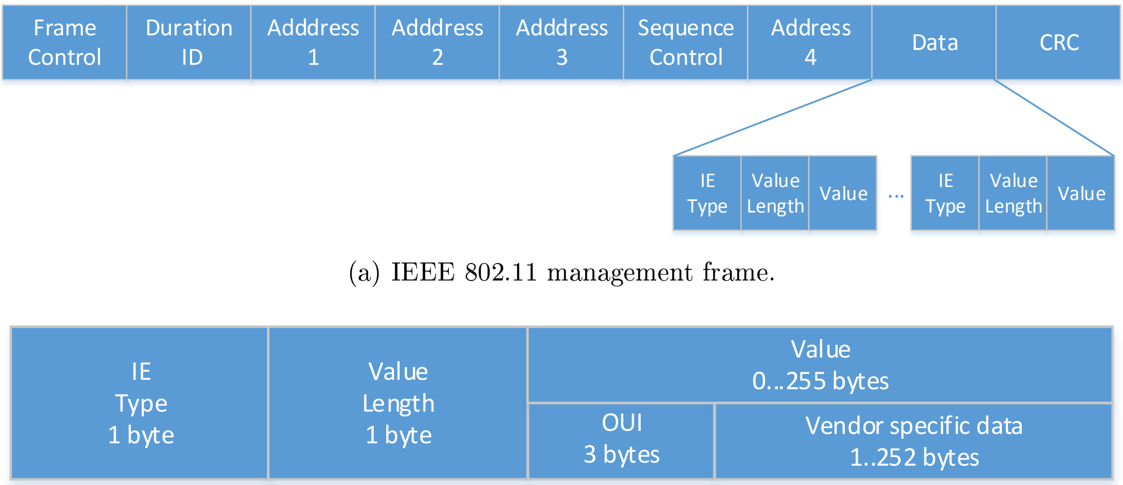
Privacy issues in location based services (LBS) arise, as WLAN positioning systems usually need both the received signal strength and the location of access points in order to calculate the location of a user. The information about access points is not usually available explicitly by a user. Moreover, network providers are usually the ones who have access to this information that makes it impossible for the users to calculate their position locally [19].

According to [19], by handing the procedure of position calculations to users' own devices, a network is only in charge of providing the necessary data. Therefore, the desired level of privacy will be attained. In addition, users take the liberty of whether keeping the data privately or sharing it with whoever they want.

There are three main infrastructures in positioning methods: infrastructure based, terminal assisted, and terminal based. It was mentioned before that the last one is the most privacy friendly method in which signal strength and position calculation are implemented in users' devices. However, it has its own drawback. As the device has to acquire some positioning related data from the infrastructure making it possible for the LBS provider to track the person.

In order to tackle the stated problems, Gschwandtner and Schindhelm [19] proposed a new positioning method for guaranteeing privacy preserving of the users, which is done by modifying the WLAN such that the locations of access points are inserted in WLAN 802.11 MAC management frames. Different parts of a management frame are shown in Figure 4.4(a). A management frame is broadcast by an access point to inform the users about the network features. IEEE 802.11 protocol makes it possible to feed in extra data, such as Information Elements (IEs), as shown in Figure 4.4(b), in the data field of a management frame. Information Elements of type 221 are not reserved by IEEE 802.11, therefore, they can be used to feed in vendor specific payload. In order to allow the users to be able to distinguish between different vendors, the data field includes a 3-byte OUI (Organizationally Unique Identifier), which identifies the vendor. By inserting the coordinates of the access points in the data field (vendor specific data) of a management frame, a user can perform the positioning process locally on their own device without the need of sending its location information or MAC ID to a server.

The proposed system works well if a proximity sensing technique is used for positioning, since the transmitted data per beacon will be around 13 bytes. However, in case of a fingerprinting system, more complex adjustments must be done so that the radio map can be transferred to user's device.



(a) IEEE 802.11 management frame.

(b) Vendor specific Information Element

Figure 4.4 *IEEE 802.11 Management Frame with Information Elements (IEs). The IEs reside within the Data field, which restricts the overall length of the IEs to a maximum of 2312 bytes.*

4.4 Other methods

In this section further ideas and concepts will be briefly introduced about both privacy and security subjects in user positioning area.

4.4.1 Preserving Location Privacy in Wireless LANs

In order to increase privacy protection in LBS, Jiang et al. [26] proposed to obfuscate some of the sensitive data, i.e. user signal strength, identity and time of transmission, transmitted by a mobile entity. Based on the results in [26], a user will not be traceable among N other mobile subscribers at the same location (with N=1000 in [26]) by any potential attacker. The downside of this method is that it degrades the service quality and users will experience delay in their communications due to the silent periods in applications such as Voice-over-IP. In addition, the designed system allows the users to choose the level of privacy. The system works as follows, in three stages.

First, the system uses anonymity to deter an adversary from using one's identity as a tool for tracking them. It is important to know that MAC and IP addresses of a device are assigned to a user as their identity in 802.11 WLAN systems. Although the frequent alteration of the pseudonyms can preserve a user's identity, it can also lead to some design problems. As for the MAC addresses, there is the probability of collision if they are randomly selected. In the proposed system, there is a local

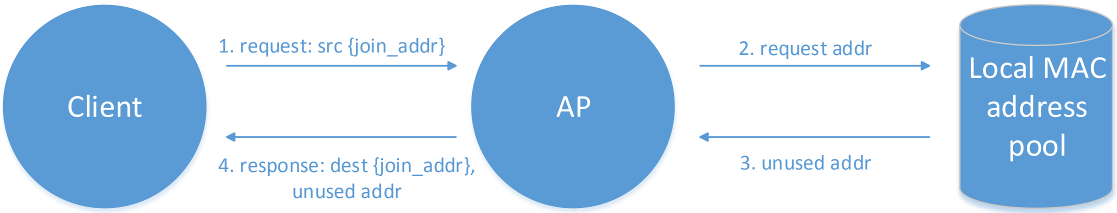

Client
1. request: src {join_addr}
4. response: dest {join_addr}, unused addr
AP
2. request addr
3. unused addr
Local MAC address pool

Figure 4.5 *The process of selecting the MAC address.*

MAC address pool from which, a user is assigned a MAC address called the *joint address* when trying to connect to an access point as shown in Figure 4.5.

Second, the system employs a silent period in order to stop adversaries from correlating the collected pseudonyms to one specific user. By employing a silent period, the user stops transmitting any data and it lets the sender to mix in with the other nodes. However, the efficiency of this approach is dependent on the density of the users. For instance, if a user is staying at home, changing the pseudonyms and using the silent period cannot really protect the user against the attacks. It is worth mentioning that privacy preserving methods have the best results in public places. In addition, using real-time communications can be problematic if the silent period is applied during data transmission. As a result, an *opportunistic* silent period is used to tackle this problem by which, the system changes the pseudonyms when the transmission has stopped for a longer period than the silent period.

Third, transmission power control (TPC) is applied to ensure a decline in positioning accuracy. Research shows that the more access points in reach of the user, the more precisely it can be located. As a result, the system reduces its transmission power so that a minimum number of APs are heard by the user's device, ensuring at least one AP to be accessible to the device all the time.

4.4.2 Secure LBS in Mobile Cloud Computing

Another proposal presented by Zhu et al. [46] tries to address both the security and the privacy issues regarding mobile could computing, as more users choose cloud services for data storage and computation purposes. Applying location-based fine-grained access control mechanism has been proposed in [46] for securing user data against unauthorized attacks and also for privacy protection against LBS providers. The designed mobile cloud system comprises three entities as shown in Figure 4.6.

- Mobile Users.

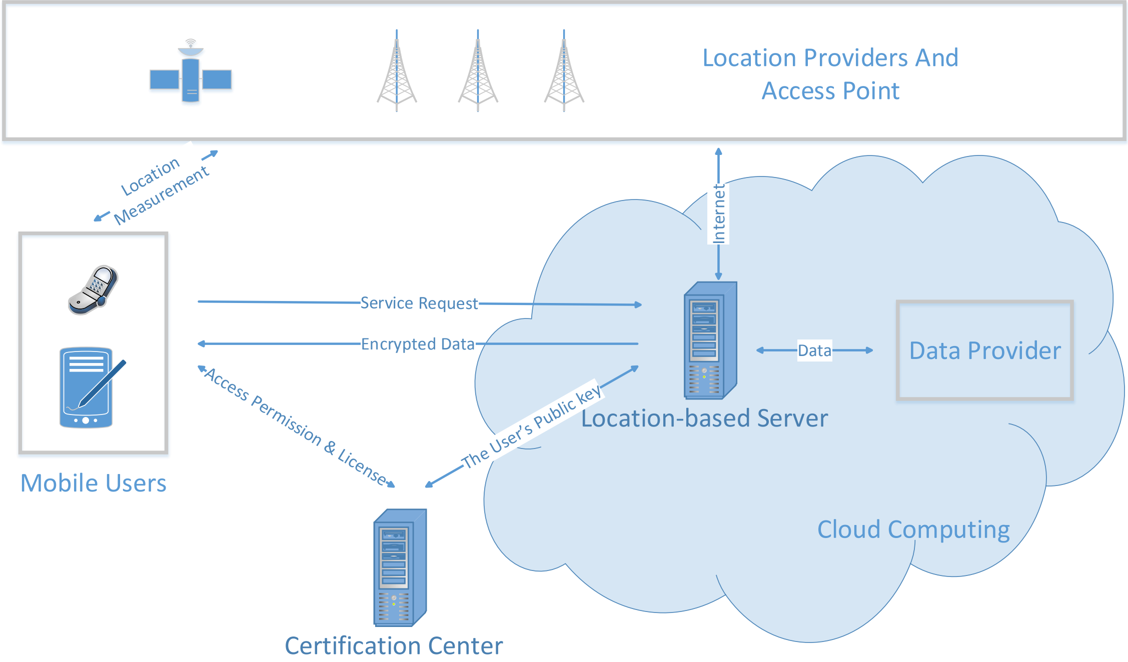


Figure 4.6 *Location-based service architecture.*

- Location-based Serivce Provider.
- Certification Center.

First, there are mobile users requesting the LBS data. Therefore, they are assigned a service certificate with specific access rights. The user can access the positioning data using GPS or a location information provider.

Second, there is Location-based Service Provider that offers the LBS for a user, given their location and request.

Third, there is the certification center which functions as a trusted third party and issues certificates for the users and produces the essential public parameter information to ensure secure LBSs.

By switching the service authentication to the client side, it becomes impossible for the LBS providers to access the user information. In this system, the certificate center is responsible for providing the service certificate. As the LBS provider gets the user's request and their location, it authenticates the access privilege and receives the LBS appeal securely. Then, the LBS provider processes the related data from data provider given the location of the user. Afterwards, the data is encrypted by the LBS provider and the user receives the ciphertext. Finally, it is the authorized user that is able to employ decryption to convert the ciphertext and extract the data using their private key.

Therefore, by implementing attribute-based encryption (ABE) and using integer comparison methods, only the mobile entity has the authority to access the location data. The attributes presented in this system are stored on the client's mobile device and called *ServiceName*, *Period-of-Validity*, *ServiceArea*, *Category*, and *Quality of Service (QoS)* as shown in table 4.3. The two latter attributes are integers defined by the service providers.

***Table** 4.3 Attribute lists for data user's service certificate (access privileges)*

ServiceName	Period-of-Validity	ServiceArea	Category	QoS
Wireless	2010/01/10-/06/15	(Phoenix)	=2(FamilyPlan)	$\geq$ 3(LowRate)

4.4.3 Cloud-Based Positioning

Homomorphic encryption is an encryption method that enables certain types of calculations to be performed on ciphertext without having access to the decryption key. Homomorphic encryption is a way to provide user defense against cyber adversaries' attacks that may happen during transmission of data to cloud servers that handle the calculations. As a result, a server makes the calculations on encrypted data and the mobile device is in charge of decryption. In contrast with the common encryption methods, the aforementioned algorithm not only is more secure, but also less complex. Moreover, it does not affect the accuracy of positioning results [15].

By using normal encryption methods, a cloud server will not be able to perform computations on the ciphertext. Hence, the decryption key has to be shared with the cloud server. In contrast, homomorphic encryption allows arbitrary functions to be performed on the decrypted data. For example, it is possible to add two encrypted numbers using this method without exposing either of the numbers to the adder.

Assuming a and b to be two integers with $\oplus$ and $\otimes$ as addition and multiplication signs subsequently, and considering $E()$ as an Encryption function, Equations 4.1 and 4.2 show that $E(a + b)$ and $E(a \times b)$ can be calculated based on the encrypted values of a and b.

$$E(a+b) = \oplus\{E(a), E(b)\} \tag{4.1}$$

$$E(a \times b) = \otimes\{E(a), E(b)\} \tag{4.2}$$

The proposed system is illustrated in Figure 4.7 and it works as follows: first,

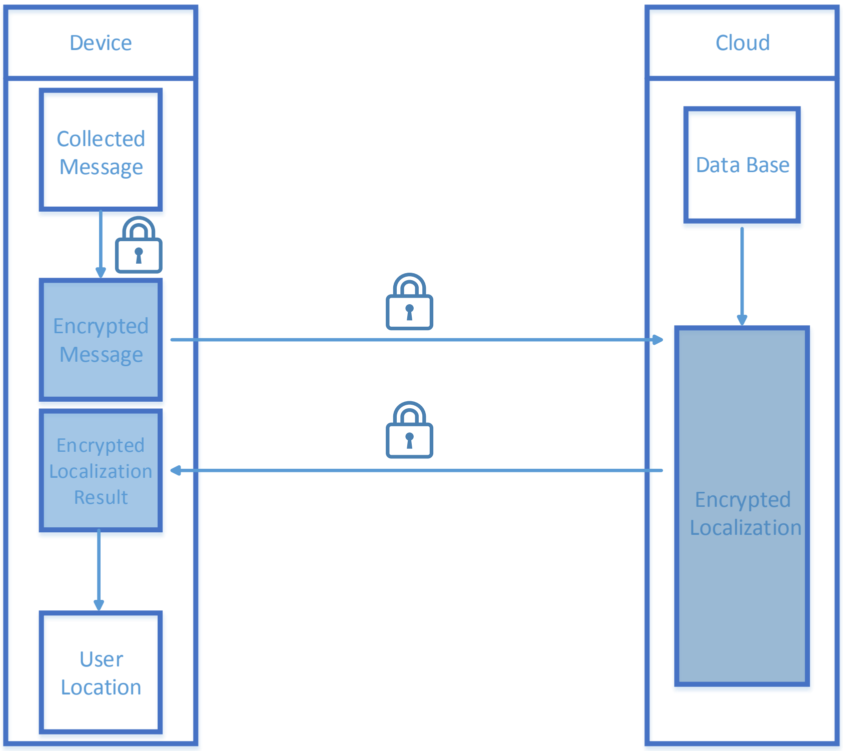


Figure 4.7 *Architecture of a secure cloud-based positioning system based on homomorphic encryption, from [15].*

the device gathers the received signal strength measurements and the data gets encrypted. Second, the encrypted message is transferred to the untrusted cloud server. Consequently, the cloud server calculates the position of the device using encrypted data. As a result, the data does not get exposed to the cloud server. Lastly, the encrypted data is sent to the device and it extracts the data.

4.4.4 Secure Localization in Wireless Sensor Networks

A wireless sensor network (WSN) faces a large variety of attacks, such as sinkhole, wormhole and Sybil attacks. All these can compromise the positioning process in two ways as follows [25].

First, an adversary may counterfeit the identity of another node that brings about the necessity for a secure node authentication (SNA). It contains the following categories:

- *Secure Localization for Unknown Nodes:*
    1. *Range-based Secure Localization:* Verifiable multilateration (VM) is a method proposed by Capkun and Hubaux [6] to assure secure positioning

based on distance-bounding protocols. VM is based on measuring the time-of-flight of the transmitted signals, by using a minimum of three, not necessarily synchronized, access points for verifying the location of the nodes.

2. *Range-free Secure Localization:* SeRLoc is a distributed range-free localization algorithm proposed by Lazos and Poovendran [30] that functions without the need for communication between the nodes. In SeRLoc there are some higher-power sectored antennas regarded as trusted locators, transmission range of which is higher than unknown nodes. The system works as the locators broadcast their position and sectors, afterwards the nodes calculate their position as the center of gravity of the overlapping locators' signal.

- *Secure Localization for Anchor Nodes:*

  LAD (Localization Anomaly Detection) is an abnormal anchor node detection scheme proposed by [12], which exploits the fact that sensors are mostly implemented in groups. Consequently, a deployment model is introduced so that nodes are located on a grid. Furthermore, the model follows a two-dimensional Gaussian distribution centering in the deploying point of the group. Having installed the nodes, each node has a recognized position. During the monitoring phase, LAD compares the calculated positions of the unknown nodes with the unknown deployment information and checks if they are consistent, given the following metrics: the Difference metric, Add-All metric and Probability metric.

4.4.5 Privacy in Social Network Services

Durr et al. proposed a new cross-layer protocol in [14] that is implemented as a secure and privacy-friendly framework for a peer-to-peer location-based social network called Vegas, which is personalized for reliable location-based advertising services. The design goals of the protocol are as follows. It must be able to: authenticate the advertised service, analyze the content of the advertisement and configure a reverse communication channel from the mobile entity to the access point for extra services. As a result, identification and data parts of the advertisement service are combined into the MAC frame and UDP packet respectively. Consequently, two new IEs are embedded into the beacon frame; one of which is *LB-SNS Certificate Fragment (LCF)* including AP certificates, and the other one is *LB-SNS Identity and Authentication (LIA)* including service identifiers and associated authentication information.

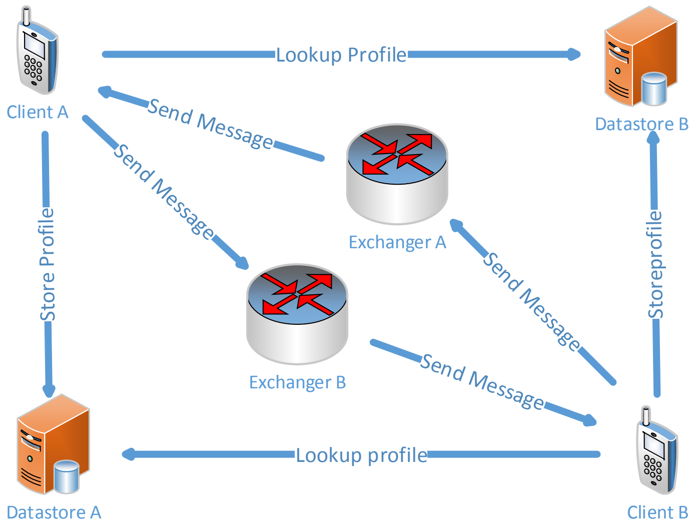


Figure 4.8 *Vegas model.*

Figure 4.8 shows the Vegas architecture in which a customer uses a mobile or stationary client to connect to a social network. An asynchronous message exchange is performed by Vegas depeneding on a common service such as email or SMS, by which *exchanger* unit is deployed. In order to keep the location privacy of the users, each pair of friends generates a public key pair to encrypt messages. As a result, a user's message can be identified and mapped to that user only by its friend that knows which public key belongs to which friend. The datastore is used to keep the profile for each customer and the exchanger is used for delivering messages between users. Furthermore, the address of the exchanger of each user is known to their friends in the system.

4.4.6 Attack Detection in Wireless Localization

The positioning infrastructure is prone to non-cryptographic attacks, such as signal attenuation and amplification, that cannot be tackled by common security services. As mentioned in [7], wireless devices, i.e. anchor/unknown nodes could be exposed to lower/physical layer attacks in a network. First, due to simplicity of wireless network architecture, it is unfeasible to implement traditional cryptographic methods for such systems. Second, wireless devices are vulnerable to physical attacks, for instance, signal attenuation, amplification, or reflection by an adversary. In order to address the aforementioned issues, Chen et al. proposed an attack detection method based on statistical significance testing that is tailored to suit multilateration and signal strength localization techniques. Practical trace-driven approaches evaluated the scenarios on both Wi-Fi and ZigBee networks and proved the reliability and accuracy of the algorithms [7].

5. IMPLEMENTATIONS OF USER TRACKING IN A TERMINAL-BASED SYSTEM

Positioning with means of wireless MAC addresses of a device will be implemented to illustrate how accurately and reliably a mobile device can be positioned indoors. And to try to prove better the significant motivation for studying the existing privacy-preserving location algorithms as it was done in Chapters 3 and 4 so far. Specifically, terminal-based positioning using only MAC addresses and RSS values of access points. In addition, terminal-based positioning is easy to implement on users' devices and it is the most privacy-preserving positioning method. As a result, it is important to know how a user device can be independently positioned on a terminal-based system.

5.1 Terminal-based positioning

This chapter compares two different positioning methods. First, a common positioning method is implemented using the collected MAC addresses and the measured RSS values. This method has long been used for positioning, e.g. in [5]. Second, another positioning method known as *rank-based positioning* [31] is implemented using only the observed MAC addresses of access points without using the measured RSS values.

5.1.1 Offline phase

The WLAN fingerprints were collected in 2011 by a measurement campaing in the Department of Electronics and Communications Engineering with a Windows tablet in 4 main floors of one of the buildings in Tampere University of Technology in Tampere, Finland [39]. During the offline phase, RSS values and MAC addresses of the heard access points were stored in a radio map in 1479 reference points. In addition, each reference point is a single point in the radio map that corresponds to a 5m-by-5m synthetic grid in the real map. The fingerprint dataset is floor-wise and the height of each floor is added as a z-dimension to the map. The MAC addresses

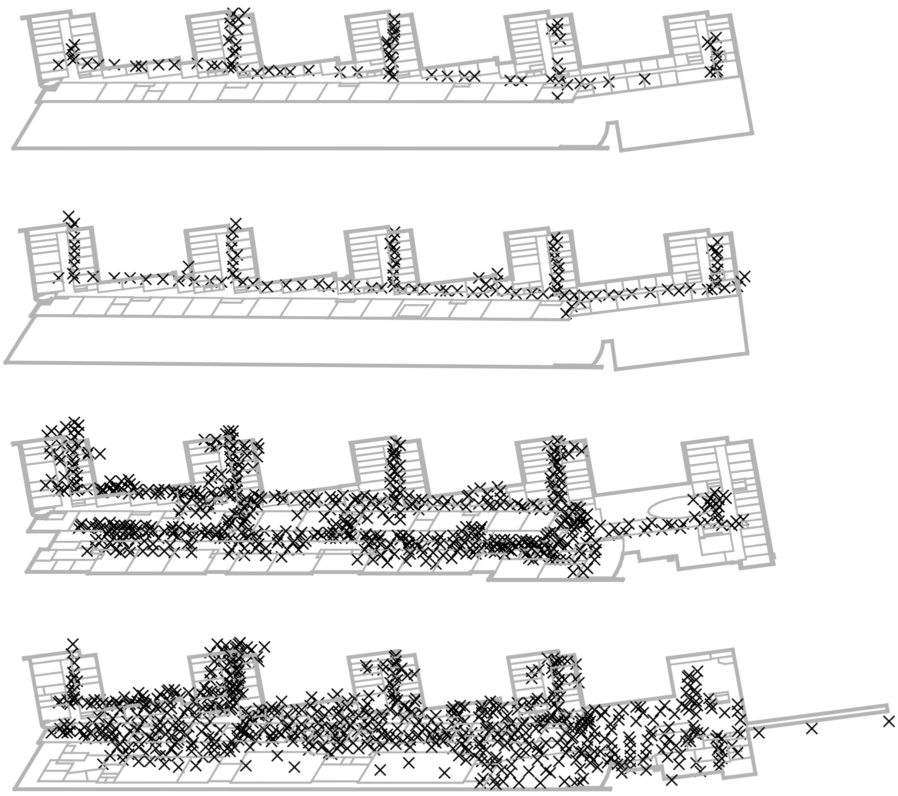


Figure 5.1 *The locations of 1479 fingerprinted reference points belonging to a 4-floor university building collected during offline phase, SCALE 1:1405 cm.*

are mapped to indices from 1 to 309. Since the access points support multiple BSSIDs, several MAC addresses may belong to the same access point. In addition, the locations of the access points are unknown. The 3D coordinates (x,y,z) map of the fingerprints are shown in Figure 5.1.

Figure 5.2 shows the distribution of all the measured RSS values as a histogram. X axis shows the RSS values, which are sorted into 20 equally spaced bins along the x-axis between -100 dB and -20 dB and the height of each rectangle indicates the number of RSS in the bin.

5.1.2 Online phase

During the online phase the same Windows tablet was used to measure the received signal strength in 490 points. The available dataset comprises the coordinates of the locations and the heard MAC addresses and their RSSs.

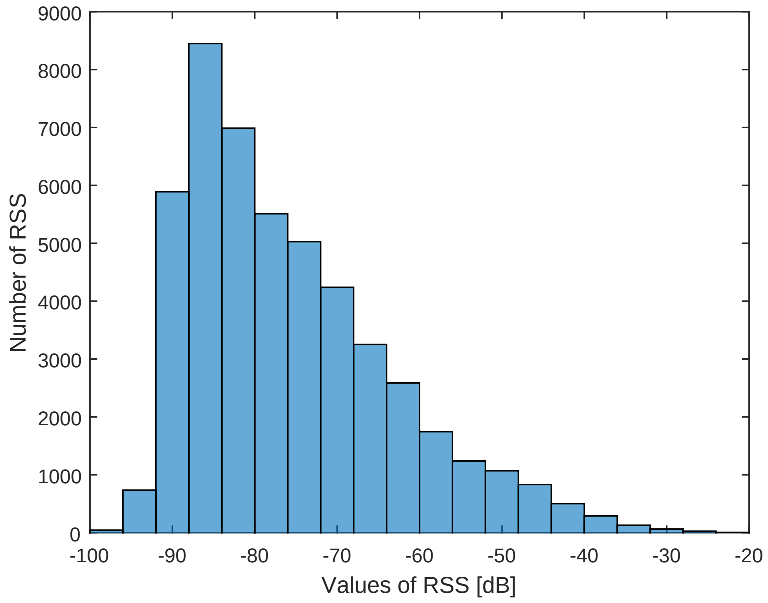


Figure 5.2 *Histogram of the RSS values measured during the offline stage in the fingerprinting process.*

5.1.3 Positioning using received signal strength and MAC addresses

In this subsection positioning will be conducted by exploiting both the MAC addresses and RSS values as follows. First, the common access points that are heard by the mobile station and the reference points at the tracking points are determined. Then the reference points that have the maximum number of common access points with a tracking point are chosen.

Finally, using k-nearest neighbor method (NNM) as shown in Equation 5.1 the minimum power difference between the observed RSS values by mobile station and fingerprinted RSS values is calculated and the location of one of the reference points that has the minimum power difference in RSS values with the tracking point is assigned as that point's location.

$$C_f = \frac{1}{N_{heard}} \sum_{AP_{heard}} (RSS_{f,AP} - RSS_{MS,AP})^2. \tag{5.1}$$

where C_f being the cost function per fingerprint, N_{heard} being the number of heard APs, AP_{heard} being the index of heard access points, $RSS_{f,AP}$ being fingerprinted RSS values of an AP, and $RSS_{MS,AP}$ being the observed RSS values by mobile station. Figure 5.3 shows the estimated locations of the mobile device that have

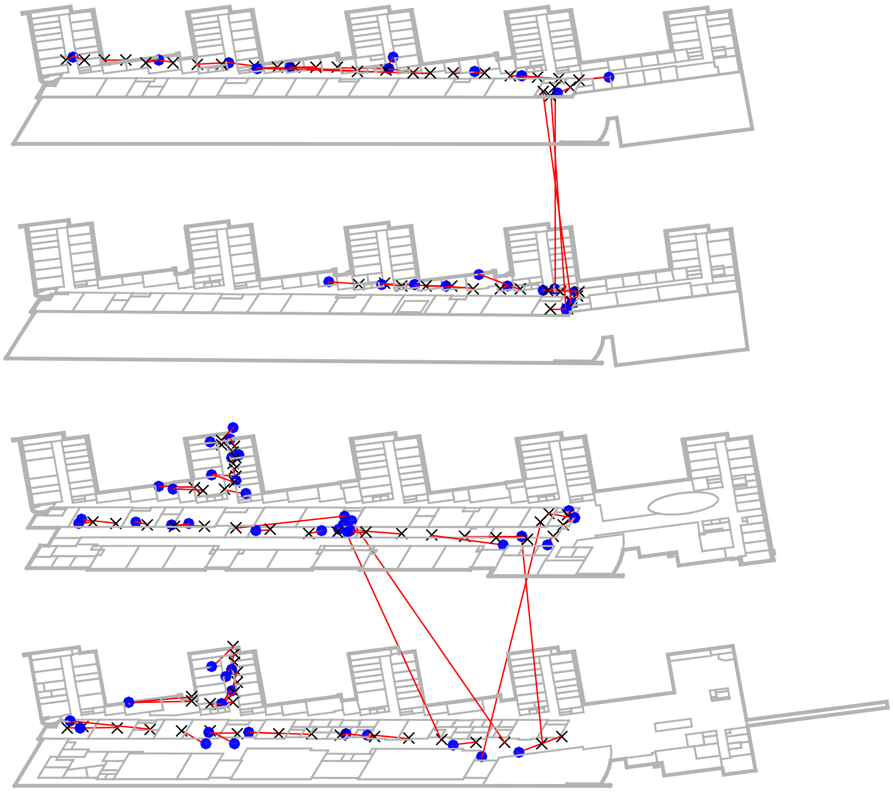


Figure 5.3 *The calculated locations of 20% of 490 tracking points using received signal strength and MAC addresses. Black crosses show the true locations and blue dots show the estimated locations that are connected by the red lines, SCALE 1:1405 cm.*

been calculated using this method.

The mean error of the positioning process was 6.25 meters. This method achieves better than 11 m accuracy in 90% of the cases and the correct floor detection probability was 88%.

5.1.4 Positioning using MAC addresses

In this subsection, positioning will be performed using the least amount of available information including the fingerprinting dataset and the observed MAC addresses of the access points by the mobile station. The goal is to evaluate the reliability of a simple positioning algorithm and to see how accurately it can reveal the location of the user. First, the common heard access points by the tracking points and the reference points are saved into a 490 by 1479 matrix. Afterwards, the maximum value in each row is chosen. This indicates which reference points have the highest number of access points in common with that particular tracking point. Then, by using the indices of these reference points the coordinates of the reference points are

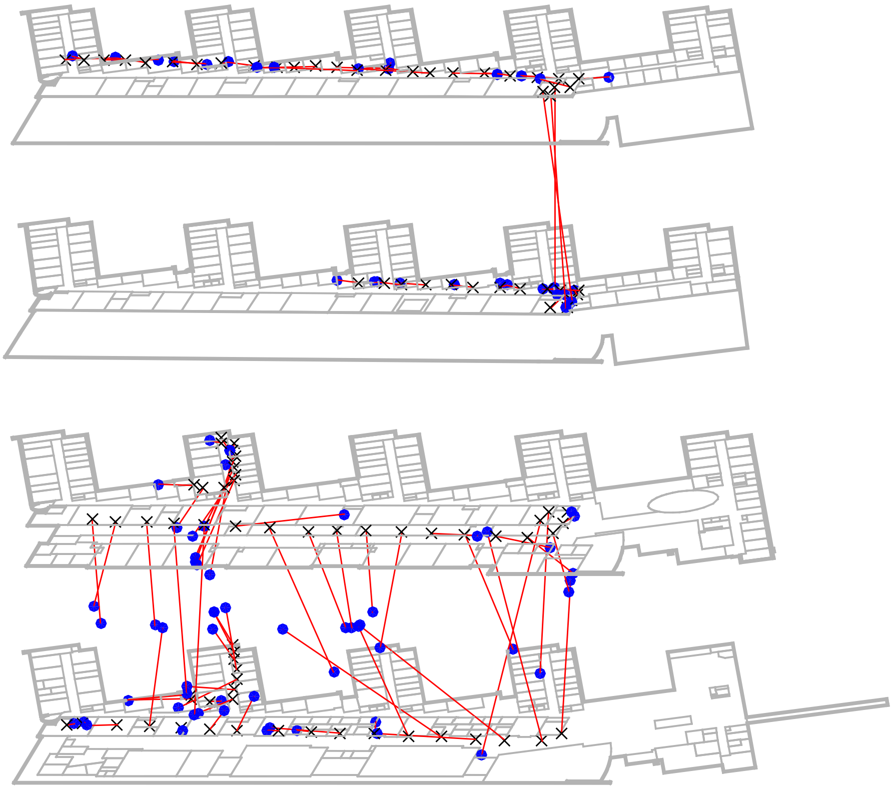


Figure 5.4 *The calculated locations of 20% of 490 tracking points using only MAC addresses. Black crosses show the true locations and blue dots show the estimated locations that are connected by the red lines, SCALE 1:1405 cm.*

fetched and in order to estimate the location of one tracking point, an averaging is performed on the coordinates of the reference points. The final positioning results can be seen in Figure 5.4. The mean error is about 7.40 meters and we can reach an accuracy below 30 m in 90% of the cases.

Since the location of the user is based on averaging the coordinates of a number of reference points, the calculated location may be in some place between two different floors. Therefore, in this next stage the estimated z coordinate will be mapped to its closest neighboring floor. The resulted map can be seen in Figure 5.5. During this stage, the correct floor detection probability was 46%.

5.1.5 Floor cluster-based positioning using MAC addresses

In order to find the location of the user; first, the common APs between reference points in each floor and all the tracking points are determined in four different vectors each of which represent one of the floors. Then, the vector that contains the maximum number of common APs is selected. Afterwards, all the reference points

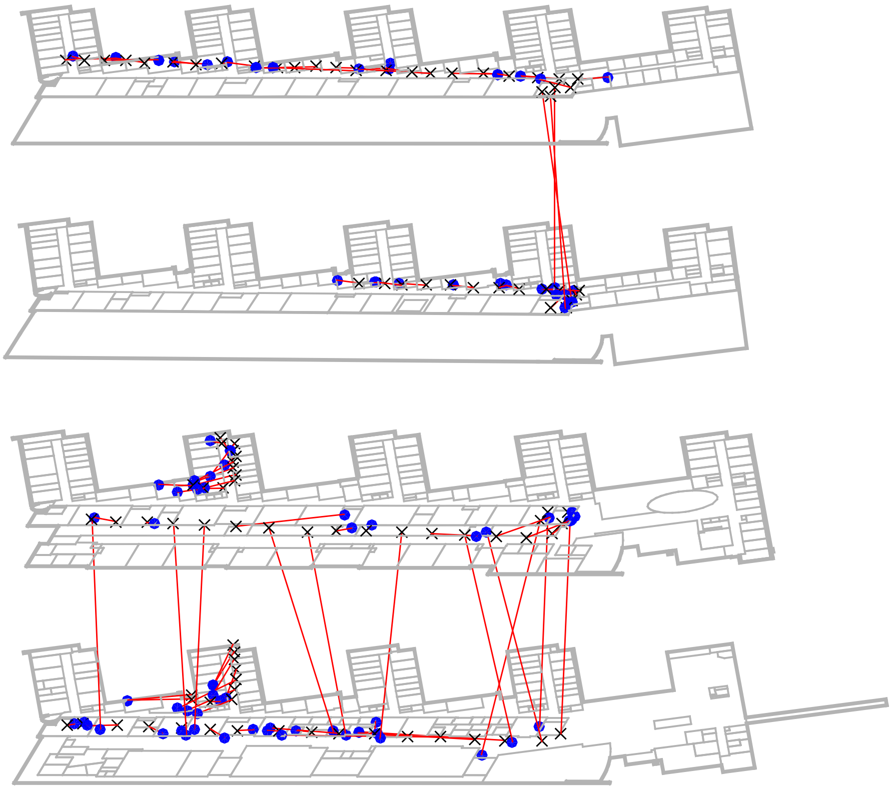


Figure 5.5 *The calulated locations of 20% of 490 tracking points using only MAC addresses with mapped z coordinates. Black crosses show the true locations and blue dots show the mapped locations that are connected by the red lines, SCALE 1:1405 cm.*

that have the maximum common access points in the vector are picked out. Finally, the location of the tracking point is estimated by averaging over the coordinates x and y of selected reference points in the previous step. It is worth mentioning that since the averaging is performed over reference points that all belong to one floor, no more processing is needed over the z axis. The estimated location of the tracking points are depicted in Figure 5.6.

The mean error of the positioning was 8.04 meters. We could achieve an accuracy below 28 m in 90% of the cases and the correct floor detection probability was 73%.

The error distance CDF for the three positioning methods can be seen in Figure 5.7. In addition, Table 5.1 represents a comparison among three methods in terms of floor detection probability, RMSE, and accuracy. As it can be seen, positioning using RSS is still the most accurate method and floor clustering did not increase the positioning accuracy significantly.

To sum up, it can be concluded that although using only MAC addresses for positioning is not as accurate as the traditional ways, it can still be implemented for

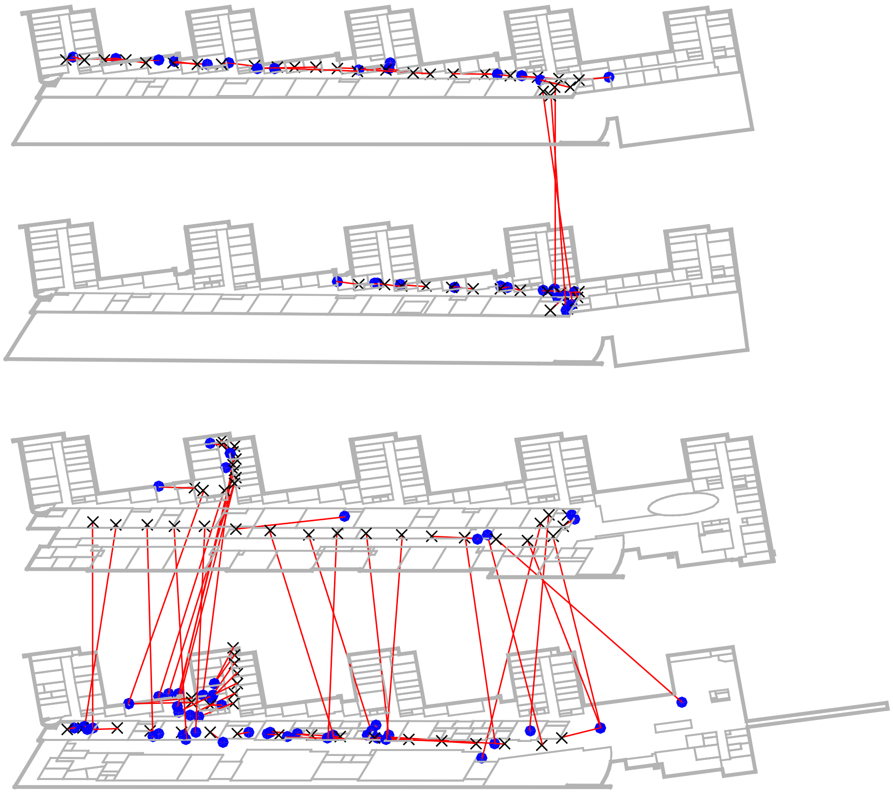


Figure 5.6 *The calculated locations of 20% of 490 tracking points by Floor cluster-based positioning using MAC addresses. Black crosses show the true locations and blue dots show the estimated locations that are connected by the red lines, SCALE 1:1405 cm.*

Table 5.1 *Positioning results based on floor detection, RMSE, and accuracy.*

Positioning Method	Floor detection probability	RMSE [m]	Accuracy < 10m probability
RSS+MAC	88%	6.25	86%
MAC	46%	7.40	40%
MAC+Clustering	73%	8.04	44%

positioning and tracking the mobile users. In addition, the results show that on the one hand by performing the positioning on a user's device a certain level of privacy can be achieved, but on the other hand an adversary can implement applications on a user's device to access data from the phone sensors and calculate the position of a user.

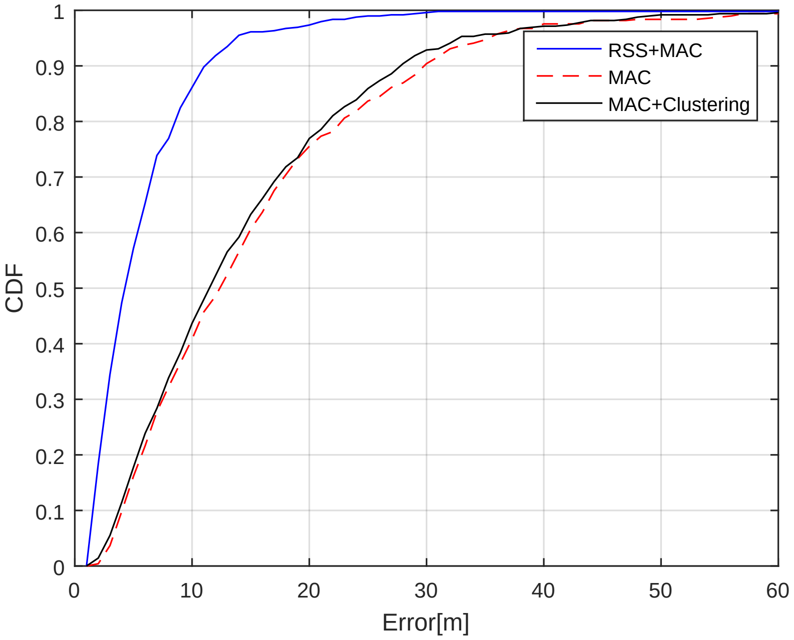


Figure 5.7 *Error distance CDF for all 3 methods used in this chapter.*

6. IMPLEMENTATIONS OF USER TRACKING IN A NETWORK-BASED SYSTEM

Many retailers have started collecting their customers' MAC addresses through their Wi-Fi connections for tracking the visitors. The tracking data is used to check whether a customer is a regular visitor or to find out which sites, shops or shelves in a shop they visited most frequently. For instance, Nordstrom, a retail chain in the United States, already tracks its customers' movements in its stores by means of sensors installed in a number of the stores [11]. The purpose of this tracking is to determine customer's behaviors in the stores for advertising and marketing. Consequently, customers' privacy gets violated without their acknowledgement [23].

This chapter focuses on some examples, based again on measurement data, of what kind of location information can be achieved by the network, a Location Based Server or a third party, from an unsuspecting user. The objective is to demonstrate that such localization can be in fact achieved with a relatively low amount of information about the user terminal, such as its MAC address and, possibly, some RSS values from or to other access nodes in range.

Two measurement databases are used:

- One of the database has been collected through a research conducted by Shweta Shretha, Elina Laitinen, Jukka Talvitie, and Simona Lohan at TUT during 2014 using a Nexus tablet and proprietary software with building maps.
- The second database was provided by TUT IT services, in conglomerate and anonymized form, and with hashed IDs. This was done in order to fully preserve the anonymity of the users.

The goal in using the first database is to show what kind of information can be achieved about the network by a possible attacker who has access to a fingerprints database or to devices to measure the radiomap of a building. This first aspect relates to the network security in positioning. We will show that even without network information, the Access Point location can be easily estimated from the radiomap of the building.

The goal in using the second database is to show to which extent users can be tracked based on the information that can be collected by the network, namely the user terminal MAC addresses and possibly their RSS values as measured at the network side.

6.1 Network-based positioning

In this section, two different approaches given security and privacy in Wireless LAN will be analyzed. In terms of security, a simulation will be run in order to investigate how feasible it is for an adversary to find the physical location of access points using the radiomap in a building. Positioning will be implemented using first both the RSS and MAC addresses and then only using the MAC addresses.

The data from the offline phase contains fingerprints collected with a Nexus tablet in two frequency bands, i.e. 2.4 GHz and 5 GHz, on four floors of a university building. The data is mapped to a 2m-by-2m synthetic grid. In addition, 422 MAC addresses were scanned in the offline phase and the number of online phase reference locations is 1803. Figure 6.1(a) shows the location of reference points on the building plan. Figure 6.1(b) illustrates how 64805 samples of the received signal are distributed based on their strength. The highest number of received signal strengths had a value between -81 dB and -80 dB with 3748 samples. Furthermore, the location of the access points are not available in the dataset.

It is worth mentioning that the histograms shown in Figure 5.2 and Figure 6.1(b) represent the distribution of RSS values in the same building, but after a big change in the WiFi infrastructure, meaning that the AP locations in the dataset used in Figure 5.2 have been different from the AP locations in the set used for Figure 6.1(b).

As a result of this change in the WiFi infrastructure, the percentage of RSS values in the range of -92 dB and -80 dB decreased from 44% to 33% in the radiomaps. This shows that a change in the network has occured.

6.2 Security-related example

Given security matters, it is worth knowing if an adversary can find the physical location of access points by scanning the MAC addresses and their related RSS in order to create a map of the access points in a building. By assuming that the created radiomap by an adversary is the same one as introduced in the previous section, the coordinates of each access point are estimated using weighted centroid that is formulated for each axis as follows:

$$\begin{aligned} C_x &= \frac{\sum C_{ix} RSS_i}{\sum RSS_i}, \\ C_y &= \frac{\sum C_{iy} RSS_i}{\sum RSS_i}, \\ C_z &= \frac{\sum C_{iz} RSS_i}{\sum RSS_i}. \end{aligned} \tag{6.1}$$

where i refers to the AP index and the three components (C_{ix},C_{iy},C_{iz}) specify the location coordinates in which AP_i was heard wtih RSS_i. (C_x,C_y,C_z) is the calculated coordinate of each access point.

The result of this calculation can be seen in Figure 6.2(a) as the real locations of the access points and their estimated locations are shown with red circles for the former and blue circles for the latter. After calculating the coordinates of the access points, the estimated values of the z axis were mapped to the nearest floor height of the building, i.e. 0, 3.7, 7.4, 11.1. As the error distance CDF diagram in Figure 6.2(b) illustrates, in around 80% of cased, the AP position can be estimated with less than 10 meters error.

I obtained the real coordinates of the access points in the following way: First, I found the physical location of each access point (each AP has a unique ID that is written on it) and marked it on the building map. Then, I mapped the marked locations on a digital map of the building to extract the corresponding coordinates of each location from this map.

6.3 Privacy-related example

The data in this phase was gathered by the IT administration of Tampere University of Technology during a one-week period in March, 2015. The collected data includes the received MAC address from a device and its RSS. The devices are defined as all mobile users who were in the building. In order to preserve their privacy, the MAC addresses were hashed by the IT administration. The raw data included 136487 collected MAC address samples. However, after the erroneous samples were omitted from the dataset, the number of samples dropped to 56520.

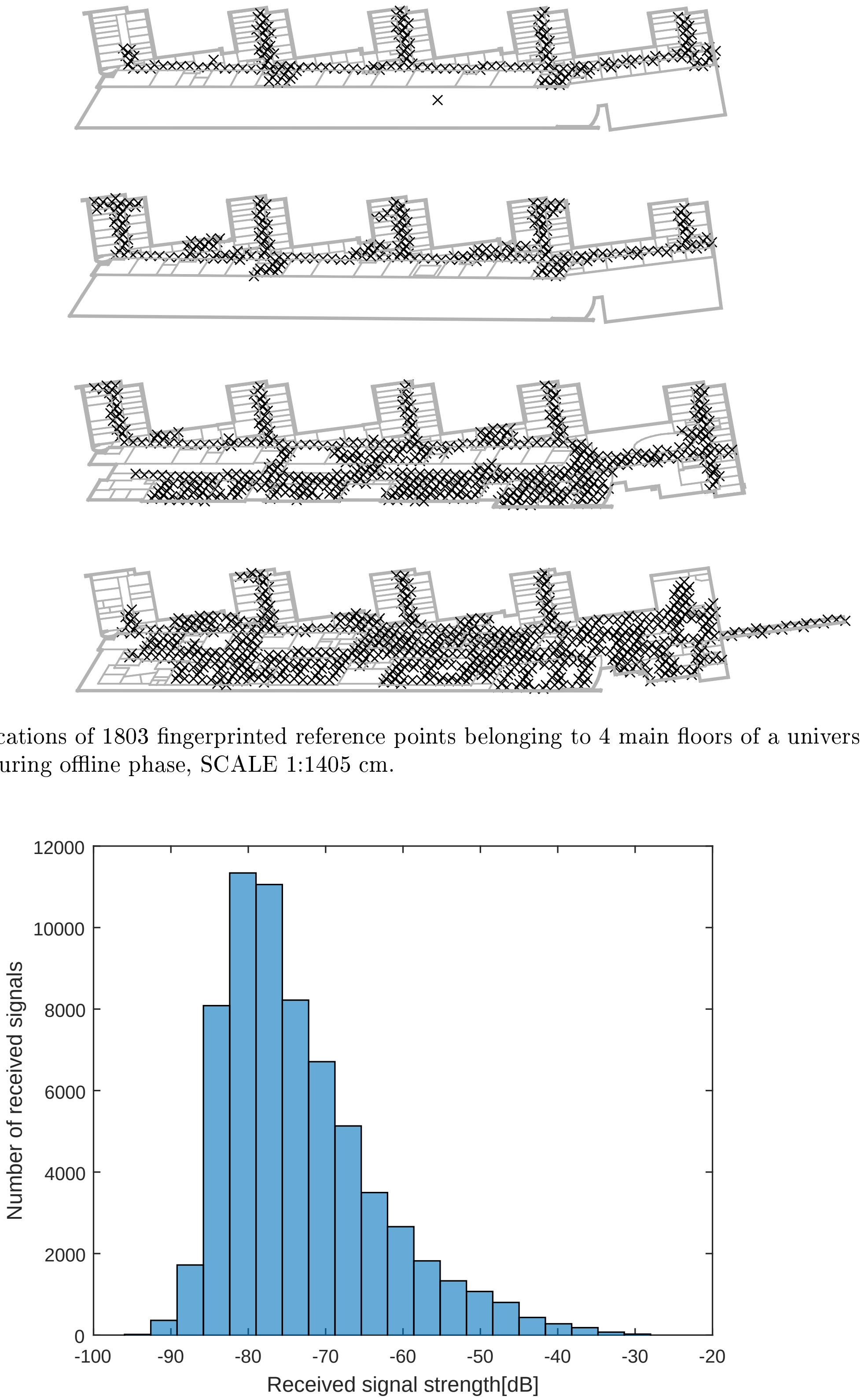
6.3.1 Positioning using received signal strength and MAC addresses

The process of positioning deployed in this stage is the same as used in subsection 5.1.3. The result of the positioning can be seen in Figure 6.3. Since the real locations of the users were unknown, it was impossible to provide comparison analysis as presented in the previous chapter for the positioning process. However, by contrasting Figure 6.1(a), which shows the location of the reference points, and Figure 6.3 it can be seen that the calculated locations of the users in first and second floors do not match to the locations of the reference points. On the other hand, the calculated locations resemble the locations of the reference points quite well for the third and fourth floor. This may occur due to the fact that the density of the reference points in first and second floor is much higher in comparison with the other two floors. As a result, it can be said that fingerprints which are based on samples from dense reference points do not increase the positioning accuracy necessarily and can lead to much longer processing time. Furthermore, the results show that user position distribution is mostly on the corridors and lecture rooms, which are situated in first and second floor. All in all, it can be concluded that user positions can be tracked by a network.

6.3.2 Positioning using MAC addresses

The positioning process in this stage has been done using the same algorithm used in subsection 5.1.4. The results are shown as diagrams in Figure 6.4, where Figure 6.4(a) illustrates the density of user positions per access points and Figure 6.4(b) shows the percentage of positioned users per floor. As it can be seen, most of the detected devices were located in the second and third floors. Since by using MAC addresses for positioning only the location of associated access points are detected, maps of user density per access points for each floor are provided in Figure 6.5. Where the blue dots illustrate the location of APs and the red circles indicate the density of located users per that access point. The larger the circle the higher the density of users per corresponding AP. In addition, Figure 6.5(b) and Figure 6.5(c) show higher user density per AP in 2nd and 3rd floors, in contrast with the other floors.

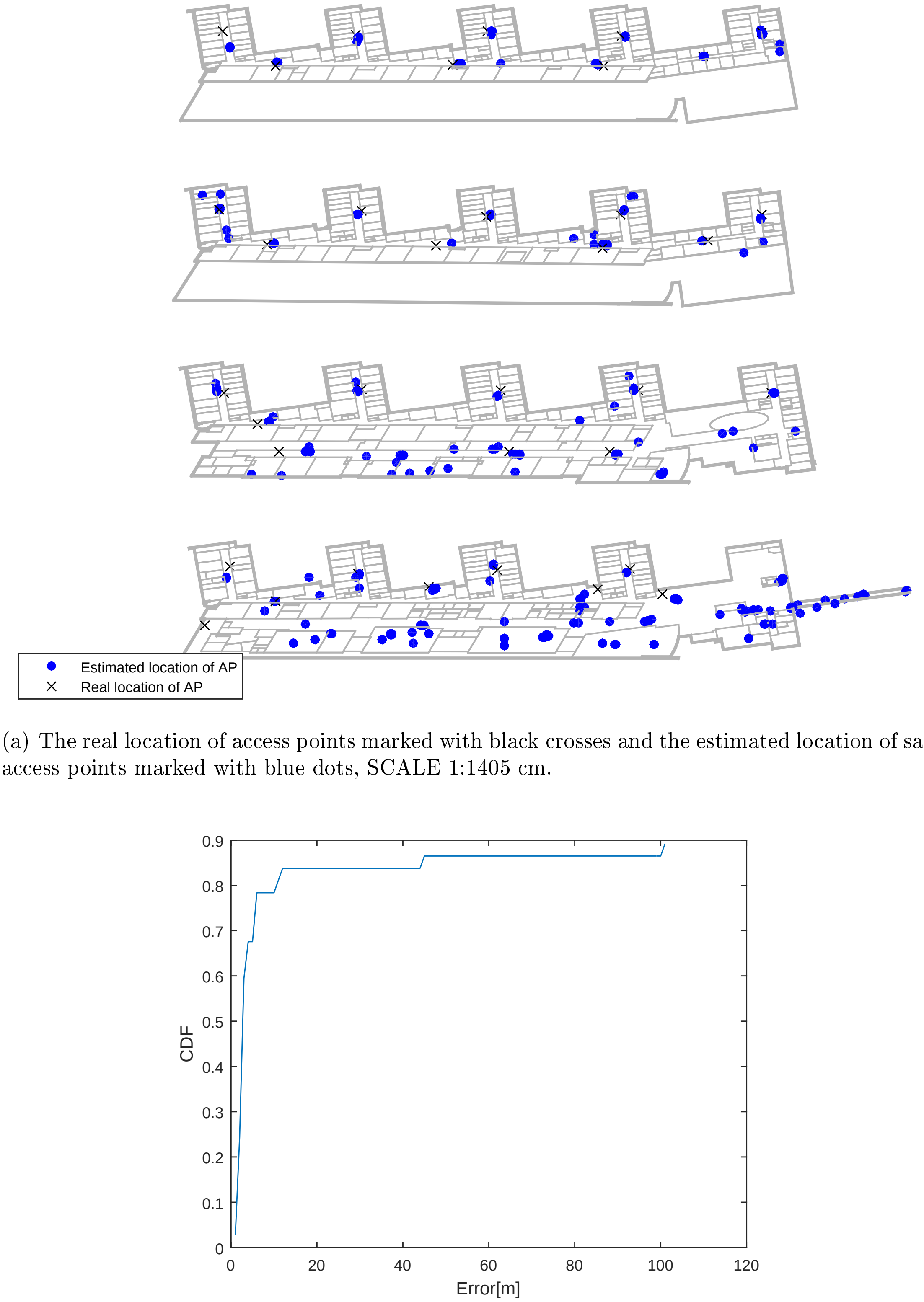
To sum up, a network can still discover the location of its users with an accuracy that has a direct relationship with the density of the access points installed in its buildings.



(a) The locations of 1803 fingerprinted reference points belonging to 4 main floors of a university collected during offline phase, SCALE 1:1405 cm.

(b) Distribution histogram of the receiced signal strength.

Figure 6.1 *The plan of the building used for fingerprinting and the distribution of the received signal strength in 2.4 GHz band.*



(a) The real location of access points marked with black crosses and the estimated location of same access points marked with blue dots, SCALE 1:1405 cm.

(b) Error distance CDF.

Figure 6.2 *Location of access points and the estimated location of them with their error distance CDF for z axis which is mapped to real heights of the building.*

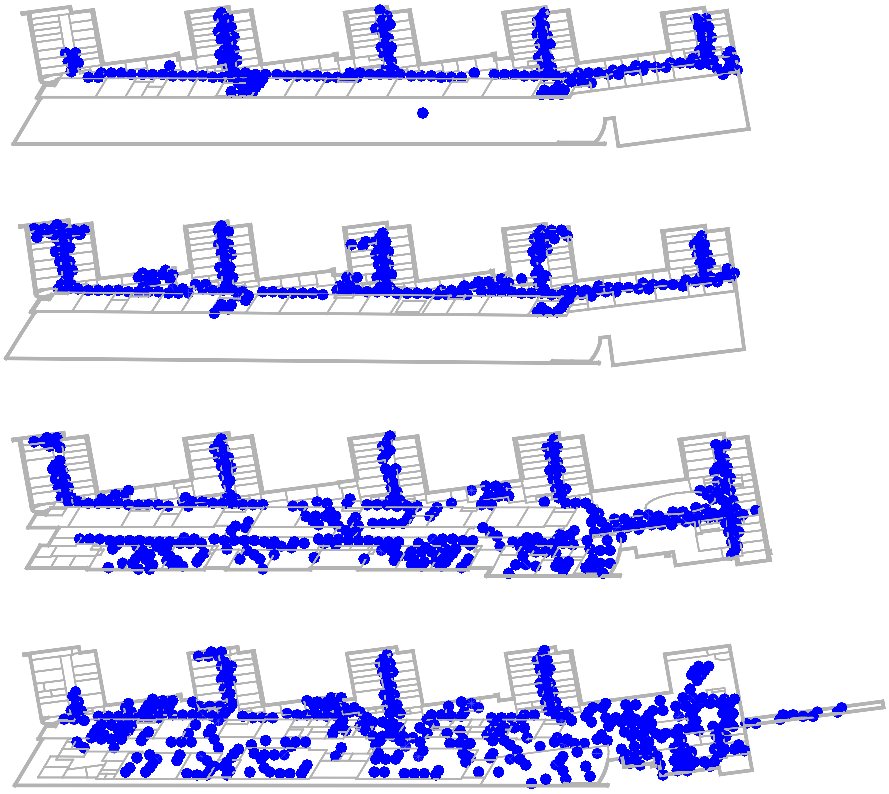
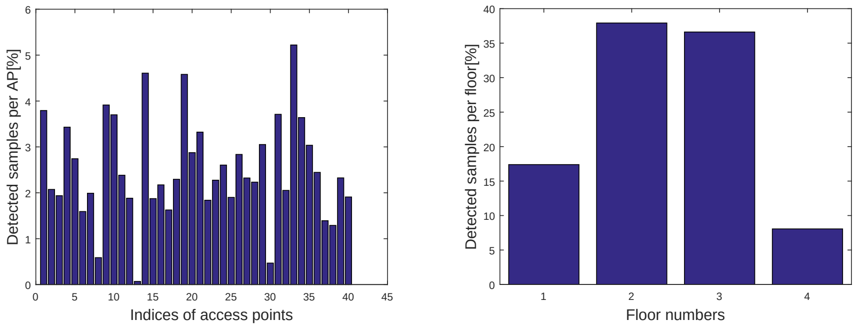


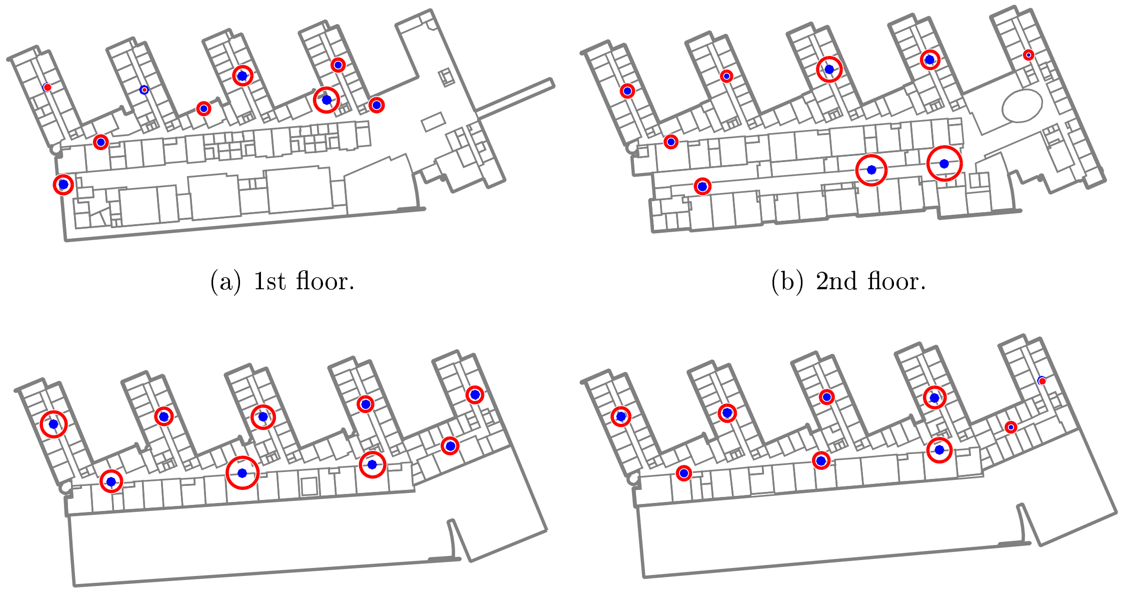
Figure 6.3 *The calculated locations of 56520 user positions using both MAC addresses and RSS values, SCALE 1:1405 cm.*



(a) Density of user positions per access point.

(b) Density of user positions per floor.

Figure 6.4 *Density of detected samples using only MAC addresses.*



(a) 1st floor.

(b) 2nd floor.

(c) 3rd floor.

(d) 4th floor.

Figure 6.5 *Blue dots and red circles show locations of APs and density of users respectively, SCALE 1:2522 cm.*

7. CONCLUSIONS

In this thesis, a wide variety of papers on the subject of location privacy have been reviewed. As it has been illustrated, different approaches to the matter already exist. Some try to hide user traces in the communications network and some use location cloaking to combine a set of locations to be introduced as the user location.

However, many companies and retailers have started collecting the MAC addresses of their customers' mobile devices for different purposes such as marketing or finding customers' positions on their premises. This offers a high motivation for looking for new privacy preserving methods. Based on this new habit of third parties in WLAN indoor positioning, it has been one of the goals of this thesis to implement several positioning methods using only MAC addresses of mobile devices and to compare the results with the common ways of indoor positioning which use both MAC address and RSS of a device.

Consequently, positioning was implemented in two infrastructure-based networks namely terminal-based and network-based configurations and we depicted how positioning can be done using simply the MAC addresses being broadcast on the air. The research area discussed in this thesis is quite new and it tackles the issue of privacy challenges of the uprising circumstances in indoor positioning. As a result, future related developments can be in the area of deriving more secure and privacy preserving location mechanisms for users who visit public places.

BIBLIOGRAPHY

[1] *IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Am*, pp. –.

[2] "Webroot survey finds geolocation apps prevalent amongst mobile device users, but 55 percent concerned about loss of privacy," 2010. [Online]. Available: http://www.webroot.com/ca/en/company/press-room/releases/social-networks-mobile-security

[3] Google, Inc.[online], http://www.google.com/policies/technologies/location-data/, 2015.

[4] C. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati, "An obfuscation-based approach for protecting location privacy," *Dependable and Secure Computing, IEEE Transactions on*, vol. 8, no. 1, pp. 13–27, Jan 2011.

[5] P. Bahl and V. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2000, pp. 775–784 vol.2.

[6] S. Capkun and J.-P. Hubaux, "Secure positioning in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 2, pp. 221–232, Feb 2006.

[7] Y. Chen, W. Trappe, and R. Martin, "Attack detection in wireless localization," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, May 2007, pp. 1964–1972.

[8] N. Cheng, X. Wang, W. Cheng, P. Mohapatra, and A. Seneviratne, "Characterizing privacy leakage of public wifi networks for users on travel," in *INFOCOM, 2013 Proceedings IEEE*, April 2013, pp. 2769–2777.

[9] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," in *Privacy Enhancing Technologies*, ser. Lecture Notes in Computer Science, G. Danezis and P. Golle, Eds. Springer Berlin Heidelberg, 2006, vol. 4258, pp. 393–412. [Online]. Available: http://dx.doi.org/10.1007/11957454_23

[10] C.-Y. Chow, “Cloaking algorithms for location privacy,” in *Encyclopedia of GIS*. Springer US, 2008, pp. 93–97. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-35973-1_136

[11] S. CLIFFORD and Q. HARDY, “Attention, shoppers: Store is tracking your cell,” 2013. [Online]. Available: http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html?pagewanted=all&_r=1

[12] W. Du, L. Fang, and P. Ning, “Lad: Localization anomaly detection forwireless sensor networks,” in *Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International*, April 2005, pp. 41a–41a.

[13] W. Du, Z. Teng, and Z. Zhu, “Privacy-maxent: Integrating background knowledge in privacy quantification,” in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD ’08. ACM, 2008, pp. 459–472. [Online]. Available: http://doi.acm.org/10.1145/1376616.1376665

[14] M. Durr, F. Gschwandtner, C. Schindhelm, and M. Duchon, “Secure and privacy-preserving cross-layer advertising of location-based social network services,” in *Mobile Computing, Applications, and Services*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, J. Zhang, J. Wilkiewicz, and A. Nahapetian, Eds. Springer Berlin Heidelberg, 2012, vol. 95, pp. 331–337. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-32320-1_21

[15] S.-H. Fang, W.-C. Lai, and C.-M. Lee, “Privacy considerations for cloud-based positioning,” in *ITS Telecommunications (ITST), 2012 12th International Conference on*, Nov 2012, pp. 527–531.

[16] M. Fredrikson and B. Livshits, “Repriv: Re-envisioning in-browser privacy,” *Microsoft Research*, 2010.

[17] D. Goldschlag, M. Reed, and P. Syverson, “Onion routing,” *Commun. ACM*, vol. 42, no. 2, pp. 39–41, Feb. 1999. [Online]. Available: http://doi.acm.org/10.1145/293411.293443

[18] M. Gruteser and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*. ACM, 2003. [Online]. Available: http://doi.acm.org/10.1145/1066116.1189037

[19] F. Gschwandtner and C. Schindhelm, “Spontaneous privacy-friendly indoor positioning using enhanced wlan beacons,” in *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, Sept 2011, pp. 1–8.

[20] S. Guha, A. Reznichenko, K. Tang, H. Haddadi, and P. Francis, “Serving ads from localhost for performance, privacy, and profit,” in *Proceedings of Hot Topics in Networking (HotNets)*, New York, NY, October 2009. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=120737

[21] M. Hardt and S. Nath, “Privacy-aware personalization for mobile advertising,” in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS ’12. New York, NY, USA: ACM, 2012, pp. 662–673. [Online]. Available: http://doi.acm.org/10.1145/2382196.2382266

[22] C. Hasan, S. Ahamed, and M. Tanviruzzaman, “A privacy enhancing approach for identity inference protection in location-based services,” in *Computer Software and Applications Conference, 2009. COMPSAC ’09. 33rd Annual IEEE International*, vol. 1, July 2009, pp. 1–10.

[23] A. Henry, “How retail stores track you using your smartphone (and how to stop it),” 2015. [Online]. Available: http://lifehacker.com/how-retail-stores-track-you-using-your-smartphone-and-827512308

[24] H. Jeremy, “Android phone’s battery use can reveal user location,” 2015. [Online]. Available: http://spectrum.ieee.org/tech-talk/telecom/wireless/android-phones-battery-use-can-reveal-user-location/

[25] J. Jiang, G. Han, C. Zhu, Y. Dong, and N. Zhang, “Secure localization in wireless sensor networks: A survey (invited paper),” in *Journal of Communications, vol. 6, no.6*, 2011, pp. 460–470.

[26] T. Jiang, H. J. Wang, and Y.-C. Hu, “Preserving location privacy in wireless lans,” in *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*, ser. MobiSys ’07. New York, NY, USA: ACM, 2007, pp. 246–257. [Online]. Available: http://doi.acm.org/10.1145/1247660.1247689

[27] H. Kido, Y. Yanagisawa, and T. Satoh, “An anonymous communication technique using dummies for location-based services,” in *Pervasive Services, 2005. ICPS ’05. Proceedings. International Conference on*, July 2005, pp. 88–97.

[28] A. Kupper, *Location-based Services: Fundamentals and Operation.* John Wiley & Sons, 2005.

[29] A. Kushki, K. Plataniotis, and A. Venetsanopoulos, “Indoor positioning with wireless local area networks (wlan),” in *Encyclopedia of GIS*, S. Shekhar and H. Xiong, Eds. Springer US, 2008, pp. 566–571. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-35973-1_629

[30] L. Lazos and R. Poovendran, “Serloc: Secure range-independent localization for wireless sensor networks,” in *Proceedings of the 3rd ACM Workshop on Wireless Security*, ser. WiSe ’04. New York, NY, USA: ACM, 2004, pp. 21–30. [Online]. Available: http://doi.acm.org/10.1145/1023646.1023650

[31] J. Machaj, P. Brida, and R. Piche, “Rank based fingerprinting algorithm for indoor positioning,” in *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, Sept 2011, pp. 1–6.

[32] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, “L-diversity: Privacy beyond k-anonymity,” *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, Mar. 2007. [Online]. Available: http://doi.acm.org/10.1145/1217299.1217302

[33] A. Malm, “Mobile location-based services,” 2012. [Online]. Available: http://www.berginsight.com/ReportPDF/ProductSheet/bi-lbs4-ps.pdf

[34] N. Ozer, C. Conley, H. O’Connell, E. Ginsburg, and T. Gubins, *Location-Based Services: Time for a Privacy Check-In*, ACLU of Northern California, 2010.

[35] W. Qiang, X. Zhiwei, and Q. Shengzhi, “An enhanced k-anonymity model against homogeneity attack,” in *Journal of software,2011*, October 2011.

[36] K. Shin, X. Ju, Z. Chen, and X. Hu, “Privacy protection for users of location-based services,” *Wireless Communications, IEEE*, vol. 19, no. 1, pp. 30–39, February 2012.

[37] R. Shokri, G. Theodorakopoulos, G. Bindschaedler, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec, “Lpm: Location-privacy and mobility meter,” 2015. [Online]. Available: http://icapeople.epfl.ch/rshokri/lpm/doc/

[38] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, “Quantifying location privacy,” in *Security and Privacy (SP), 2011 IEEE Symposium on*, May 2011, pp. 247–262.

[39] S. Shrestha, J. Talvitie, and E. Lohan, “On the fingerprints dynamics in wlan indoor localization,” in *ITS Telecommunications (ITST), 2013 13th International Conference on*, Nov 2013, pp. 122–126.

[40] J. Tsai, P. Kelley, and N. Sadeh, "Location-sharing technologies: privacy risks and controls," *Research conference on communication, information and internet policy (TPRC)*, 2009.

[41] F. van Diggelen, "Indoor gps theory implementation," in *Position Location and Navigation Symposium, IEEE*, 2002, pp. 240–247.

[42] M. Youssef, A. Agrawala, and A. Udaya Shankar, "Wlan location determination via clustering and pro**b**ability distributions," in *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, March 2003, pp. 143–150.

[43] H. Yun, D. Han, and C. C. Lee, "Understanding the use of location-based service applications: Do privacy concerns matter?" in *Journal of Electronic Commerce Research*, vol. 14, 2013.

[44] X. Zhang, X. Gui, F. Tian, S. Yu, and J. An, "Privacy quantification model based on the bayes conditional risk in location-based services," *Tsinghua Science and Technology*, vol. 19, no. 5, pp. 452–462, Oct 2014.

[45] G. Zhong and U. Hengartner, "A distributed k-anonymity protocol for location privacy," in *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, March 2009, pp. 1–10.

[46] Y. Zhu, D. Ma, D. Huang, and C. Hu, "Enabling secure location-based services in mobile cloud computing," in *Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing*, ser. MCC '13. New York, NY, USA: ACM, 2013, pp. 27–32. [Online]. Available: http://doi.acm.org/10.1145/2491266.2491272