Syed Ibrahim Khandker

RF fingerprinting for user locationing in LTE/WLAN networks

Master's thesis in Information Technology

August 22, 2016

University of Jyväskylä

Department of Mathematical Information Technology

Author: Syed Ibrahim Khandker

Contact information: bd.syed@gmail.com

Supervisor: Tapani Ristaniemi

Title: RF fingerprinting for user locationing in LTE/WLAN networks

Työn nimi: Käyttäjien paikantaminen LTE/WLAN verkossa radiotaajuus sormenjälkien avulla.

Project: Master's thesis

Study line: Web Intelligence and Service Engineering

Page count: 53+7

Abstract: Location based applications and services have become popular in many wireless communication devices. This thesis study presents a performance evaluation of Radio Frequency (RF) fingerprinting framework in heterogeneous Long Term Evolution (LTE) and Wireless Local Area Networks (WLAN) using Minimization of Drive Testing (MDT) measurements which allow automated construction of extensive RF fingerprint training databases. Utilization of MDT data could create additional opportunities for service provider and application developers. Typical RF fingerprint consist of radio measurements from multiple LTE transceiver stations and WLAN access points. Regardless environmental conditions, received signal strength indicator pattern for difference reference points, set a fingerprint of radio conditions for the specific location. Based on RF fingerprinting framework for positioning by using MDT measurements specified in LTE release 10, performance of locating user equipment was studied by signal strength mean value algorithm using grid based method at outdoor environment. To find out the optimal grid size, three different grid size were used. Source of positioning error and effects have been extensively investigated. Cumulative distribution function has been used to express the result. This study suggests that, MDT based RF fingerprint locationing system can provide a good basis for the network based proximity detection.

Keywords: LTE, WLAN, Radio Frequency, Fingerprint, Positioning

Suomenkielinen tiivistelmä: Käyttäjien sijaintiin perustuvat sovellukset ja palvelut ovat nousseet suosioon monissa langattomissa laitteissa. Tämä tutkimus esittelee suoritus arvion radiotaajuus (RF) "sormenjälkien " käytön heterogeenisissä Long Term Evolution (LTE) verkoissa, Wireless Local Area Networks (WLAN) verkoissa käyttäen Minimization of Drive Testing (MDT) mittauksia joiden avulla voidaan rakentaa laajoja RF sormenjälki datavarastoja. MTD datan hyödyntäminen voisi luoda uusia mahdollisuuksia palveluiden tarjoajille sekä sovellusten luojille. Tyypillinen RF sormenjälki koostuu monista lähetin asemista ja WLAN tukiasemista saaduista mittauksista. Maantieteellisistä syistä riippumatta vastaanotetun signaalin vahvuus eri referenssi pisteiltä luo radio sormenjäljen jonka voi tarkentaa tietylle alueelle. RF sormenjälkien perusteella sijain paikannus käyttämällä MTD mittauksia määritettyna LTE julkaisussa 10, Käyttäjien laittaiden paikantamisen suoritusta tutkittiin signaalin vanvuuden arvon algoritmilla käyttmällä ruudukkoon perustuvaa metodia ulkoilma ympäristössä. Sijainti virheiden lähdettä ja vaikutuksia on tutkittu laajasti. RF sormenjälkiin pohjautuvien metodien suorityskykyä on arvioitu Vastaan otetun signaalin vahvuuden (Received Signal Strenght) (RSS) Euclideaanin etäisyyden ilmaisemalla kumulatiivisella jakelu funktiolla. Tämä tutkimus ehdottaa, että MTD:hen pohjautuvat RF sormenjälki paikannus systeemi voi tarjota hyvän pohjan verkkoon pohjautuvan etäisyyden paikantamiseen.

Avainsanat: Radiotaajuus, Sormenjälki, Paikannus

Glossary

AP	Access Point
BS	Base Station
BTS	Base Transceiver Station
CDF	Cumulative Distribution Function
CTS	Common Training Signature
DB	Database
DCM	Database Correlation Method
EDGE	Enhanced Data Rates for GSM Evolution
EGSM	Extended Global System of Mobile Communications
GCU	Grid Cell Unit
GLONASS	Globalnaya Navigazionnaya Sputnikovaya Sistema
GNSS	Global Navigation Satellite System
GPRS	General Packet Radio Service
GPS	Global Positioning system
GSM	Global System for Mobile Communications
HSDPA	High Speed Downlink Packet Access
HSUPA	High Speed Uplink Packet Access
KNN	K Nearest Neighbor
LTE	Long Term Evolution
MDT	Minimization of Drive Test
MTS	Multiple Training Signature
NIC	Network Interface Card
RAM	Random Access Memory
RF	Radio Frequency
RFID	Radio Frequency Identification
RSRP	Reference Signal Received Power
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RX	Receiver

SON	Self Organizing Network
TDOA	Time Difference Of Arrival
TETRA	Terrestrial Trunked Radio
TOA	Time Of Arrival
TX	Transmitter
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
WCDMA	Wide band Code Division Multiple Access
WLAN	Wireless Local Area Network

List of Figures

List of Tables

Contents

1 Introduction

The proliferation of wireless networks and mobile devices have indicated a growing interest in the location based systems and services, the main feature of such services is that, the service is provided to the user as a function of user's location (Bahl and Padmanabhan 2000). Therefore, accurate position of User Equipments(UE) is very important. Positioning in wireless networks depends on many factors, for example the mobility of users, dynamic nature of the indoor or outdoor environment and nature of radio signals in that environment (R. Mondal et al. 2013). Users expect same level of positioning accuracy regardless whether they are in indoor or outdoor environment in a rural or urban area. So far no single positioning method, including Global Positioning System (GPS), works well in all environments (Olsson et al. 2013). Receivers in Global Navigation Satellite Systems (GNSS) such as GPS or Globalnaya Navigazionnaya Sputnikovaya Sistema (GLONASS) often receives incorrect location estimation while operating in urban area due to high raised building which often block the straight path between receivers and navigation satellites (Ben-Moshe et al. 2011).

In a LTE/WLAN network, the UE receives signal from multiple Base Transceiver Stations (BTS) and Access Points (AP), these Received Signals Strength Indication (RSSI) value or path lost measurements along with the known location can create a radio pattern or fingerprint of radio conditions for that particular geographical area. The location of the reference fingerprint is normally measured by accurate reference position measurement, e.g., GNSS. Accumulated fingerprints of a geographical area can be used to create a database (DB) or a Radio Environment Map (REM) so that UE with no position information can estimate it's location by comparing current radio condition with REM or DB, what can be defined as Database Correlation Method (DCM) which consists of two phases: training or database creation phase, testing or positioning phase. Among all the non-standard positioning methods which are included into the LTE Release 9, RF fingerprinting is the most cost-efficient solution for indoor WLAN positioning (Milioris et al. 2014), (Chan, Baciu, and Mak 2008) , as well as for outdoor mobile cellular positioning in densely built urban environments (Zhu and Durgin 2005), (Laitinen, Lahteenmaki, and Nordstrom 2001).

Expected aspects of an ideal DCM dependent positioning system are self-learning capability, environmentally adaptive, capability of building up knowledge set that store actual observation, and employ intelligent data analysis mechanism (Olsson et al. 2013). To achieve such type of positioning system it needs huge amount of dataset from all geographical positioning of a network at different environmental conditions. Minimization of Drive Tests (MDT) has been proposed in LTE Release 10 which reduces the huge cost and efforts associated with the conventional drive test measurement procedure to collect user experiences of a network from different coverage geographical area. MDT provides a framework for gathering user reported location-aware radio measurements from commercial UE that can be used for creating and maintaining RF fingerprint databases (Hapsari et al. 2012; Johansson et al. 2012). By MDT data, operators can autonomously build and update large database for RF fingerprinting from various locations of users , containing user experience along with available location information, from UEs without extra hardware installation. RF fingerprint from an UE compared with RF fingerprint training database where possible location information for nearly same pattern fingerprint is recorded, it would be possible to provide user's location information by using own resource of the network.

The motivations for this thesis are to find an efficient user locationing system for the heterogeneous network, exploiting already existing radio infrastructures like IEEE 802.11 or LTE, and meet the challenge of radio condition variation. Real data regarding received signal strength and respective base transitive station ID / access point ID was collected from about 500 meter square urban area of Tampere region. Grid-based RF fingerprinting technique (Mondal, Ristaniemi, and Turkka 2015) was used to determine user's position.

The results show that, the performance of RF fingerprint depends on resemblance of training and testing data. For 95% of testing data, positioning error was around 70 meters while there was high radio condition variation. This suggests that, in a small cell networks, MDT based RF fingerprint positioning can provide a good basis for location assisted Radio Resource Management (RRM) algorithms such as network based proximity detection.

The second chapter contains comparative study of different positioning systems, followed by RF fingerprinting positioning in details in chapter three. Different aspects, radio signal behaviours and effects are discussed in chapter number four. RSS based positioning algorithms are explained in the next chapter. This thesis's research method and experimental procedures are stated in chapter six and seven respectively. Experiment related discussion and some analysis were made in chapter eight, followed by conclusion at the last. All experiments related results are attached in the appendices section.

2 General Positioning Techniques

2.1 Cell of origin

This is the simplest positioning system to find out an UE's location in a network. From a mobile device it is possible to determine to which cell it is connected and the pre-recorded location of that cell will be counted as the position of the mobile device. In this case, positioning accuracy depends on the size of the cell, few hundred meters to few kilometres depending on the coverage.

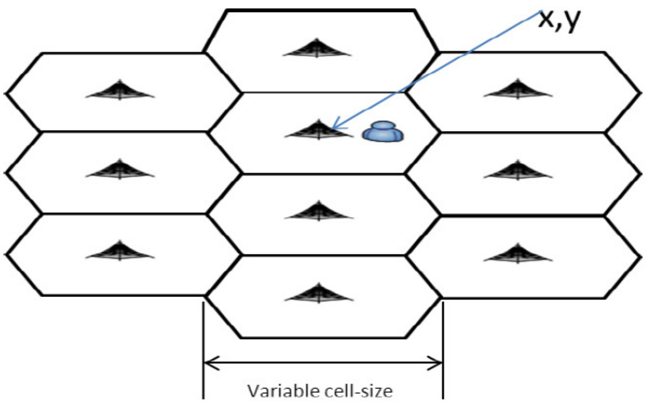

Figure 1. Cell of origin positioning technique.

It is not a precise locationing technique but provides positioning information at a low cost with very fast response and usable for almost all existing equipments.

2.2 Signal level triangulation

The path loss increases when the distance between the transmitter and the receiver increases and vice versa. By calculating the path loss or the signal level from different access points it is possible to calculate the distance between access points and mobile device. As we know

$$P_r = P_t G_t G_r \frac{\lambda^2}{(4\pi d)^2} \qquad (2.1)$$

Here

G_t and G_r are the transmitter and receiver antennas' gains

d is the distance between transmitter and receiver

P_t transmitted power

P_r received power

λ is wavelength

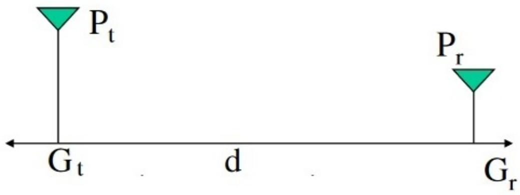

Figure 2. Relation between path loss and distance.

By calculating the distance from three BTSs/APs we can use the triangulation method to determine the coordinate of the UE. The coordinate of the BTSs/APs should be pre recorded.

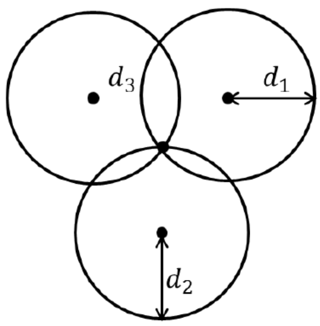

Figure 3. Signal level triangulation technique.

But only in an idle condition when transmitter transmits omni directionally signal level drops at a same rate to all directions, but in real condition presence of the obstacles vary. Therefore, this method might be effected by the obstacles and by the weather conditions.

2.3 Time of arrival

Determining the signal's travel time from the Tx point to the Rx point we can calculate the distance between them, by counting the Tx point's coordinate as a reference point it is possible to get the position of the receiver.

$$d = t \times c \tag{2.2}$$

d is the distance from transmitter to receiver , speed c is constant 3.00×10^8 m/s. By calculating the distance from three different access points we can apply the triangulation method to get the coordinate of the receiver.

Time Of Arrival(TOA) needs very precisely synchronised clock between the transmitter and the receiver, only 1 micro second of error can cause 300 meters of error ("Overview of 2G LCS Technnologies and Standards" 2001). Sophisticated positioning system e.g., GPS uses atomic clock to avoid any kind of clock synchronization related problem.

2.4 Time difference of arrival

The Time Difference Of Arrival (TDOA) method measures the receiver's position based on signal arrival time difference between two or more pair of access points.

Access point = AP, receiver = R, Time = t, Time Difference = d

t_1 = Required time from AP_1 to R

t_2 = Required time from AP_2 to R

$d_1 = t_1 - t_2$

with help of AP_3 it is possible to calculate d_2 and d_3, each d place the R in a hyperbolic curve

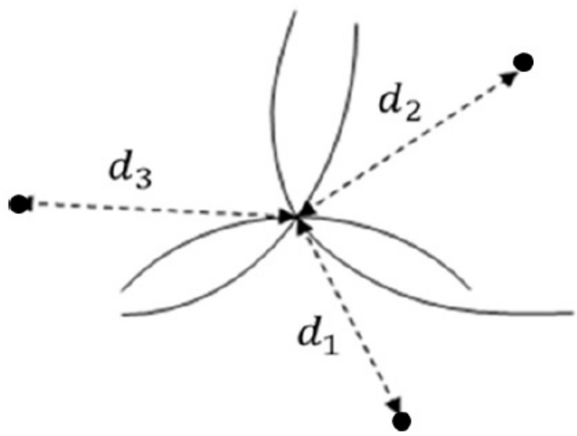

Figure 4. Time difference of arrival technique.

now by trilateration technique it is possible to get the location information of R. It is not so popular technique due to high cost, additional antenna, large infrastructure and need of the location equipment at each access point.

2.5 Angle of arrival

In the angle of arrival technique, UE's signal need to be received by at least two Base Stations (BS), for implementing this particular technique, BS should have additional equipment to determine the compass direction of the received signal. By using two reference position of the BSs and the two measured angles it is possible to determine the UE's position.

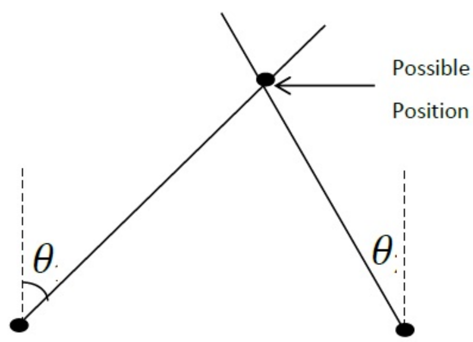

Figure 5. Angle of arrival technique.

The advantage of this technique is that, it supports mobile devices regardless the age or the model as there is nothing to do at receiver side. Disadvantage is BS need to have, receive signal's angle detection facility which is complex enough. This method helps law enforcement department to trace criminals.

2.6 GPS

Global Positioning System (GPS) so far the most popular and reliable positioning system. It is a satellite based positioning system developed by US department of defence for military use, in the year 2001 it was opened for civilian use. Currently consist of 31 satellites in service where at least 3 of them require to get 2D positioning of a receiver. Trilateration method is used to determine the receiver's position by measuring the distance from 3 satellites to the receiver, presence of 4th satellite opens the option to measure the elevation. As it works based on TDOA method, synchronized time accuracy with other satellites is must, therefore atomic clock is used.

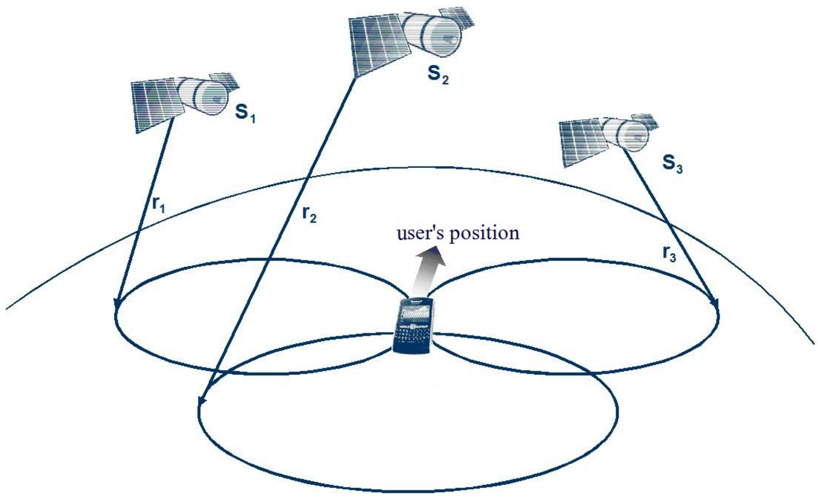

Figure 6. Global positioning system.

GPS requires direct line of sight from the satellites therefore does not provide good performance at indoor environment, at outdoor environment GPS has on average 5 to 10 meters positioning error due to ionosphere, multipath and noise problem (Matosevic, Salcic, and Berber 2006). There are some other unpopular satellite based positioning systems e.g., The Globalnaya Navigazionnaya Sputnikovaya Sistema (GLONASS), India's Indian Regional

Navigation Satellite System, China's BeiDou Navigation Satellite System etc.

2.7 Infra-red

As the infra-red ray can not go through an obstacle, this system is mostly used in small and free environment e.g., home, office, warehouse etc. It needs direct line of sight between the transmitter and the receiver. Transmitter / tag periodically transmit infra-red beacon containing unique ID of that transmitter so receivers can understand nearby which receiver the tag is, it's functionality is same as Radio Frequency Identification (RFID). Task like, to know the location of nearby printer or patient's position in a hospital this type of positioning method is used.

2.8 Radio frequency identification

In RFID system a unique ID or serial number is being transmitted by using radio wave, depending on that ID or serial number receiver can recognize the tag and does further action. Most countries have assigned the 125 or 134 kHz areas of the spectrum for low-frequency RFID systems, and 13.56 MHz is generally used around the world for high-frequency RFID systems. There are two type of RFID tags active and passive. Active tag can transmit radio signal while passive tag can only be read by tag reader. RFID system consist of antenna and transponder.

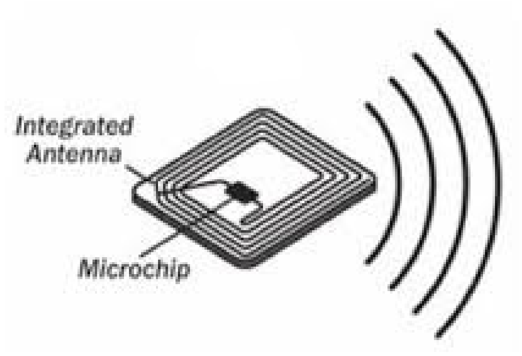

Figure 7. Radio frequency identification system.

Depending upon output power generally active RFID tag's range is upto 20 meters. A reader can read the tag and decode the data while the tag comes in the range. Nowadays RFID system is used in the identification of products moving through harsh assembly process, as a door key or as a reference point.

2.9 Ultrasonic

In ultrasonic based positioning system high frequency sound wave (above 18 kHz) is used to determine the distance between sensor and object by evaluating the echo of the sound. Time interval between Tx of sound wave and Rx of echo, reveals the distance. Radar, sonar, and ultrasonography are the example of this type of positioning.

3 Radio Frequency Fingerprinting

3.1 RF fingerprinting concept

Nowadays RF fingerprinting based positioning system is drawing the attention of researcher for its cost effective and efficient features. A RF signal attenuates over a distance that travels, the mean of RSS can be predicted by several path loss models (Pahlavan and Krishnamurthy 2002), or can be measured locally. This distance dependency property can be transformed into location dependency of RSS if there are multiple signals from different reference points. In this method base stations or access points' ID with corresponding RSS value and location information of the geographical area (called RF fingerprint) is recorded in a database. It is assumed that, each location of the area of interest has the unique fingerprint (Pahlavan, Li, and Makela 2002). At positioning state, real time fingerprint is compared with nearest matched fingerprint from the database through the algorithm to determine UE's current location. Apart of the database creation, no additional equipment is needed. In this thesis work various experiments were done based on different techniques and got acceptable level of accuracy.

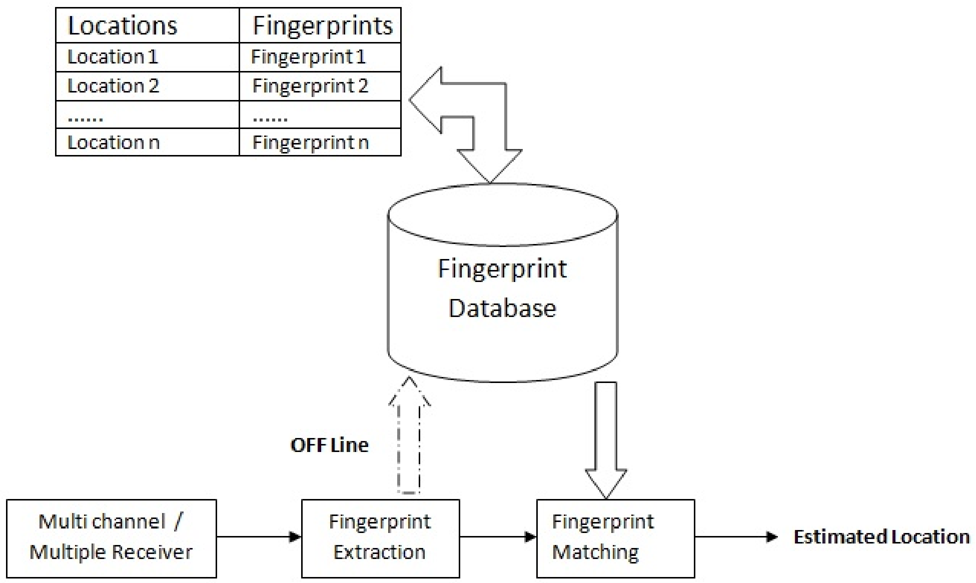

Figure 8. Principles of RF Fingerprinting

A general concept of the RF fingerprinting techniques is demonstrated in the figure 8 . It is consist of two phase, off-line or training phase and online or testing phase. At off line phase RF fingerprints of the desired area are collected through multiple devices, but to collected this huge amount of data from every possible point of the coverage is a time consuming and costly process. 3GPP release 10 ("3GPP TR 32.827" 2010), opens the opportunity to collect data autonomously from user devices called Minimization of Drive Test (MDT), what can be used to create fingerprint database. During the locationing or testing phase, the real time MDT measurements will be compared with existing DB elements. All UE may not have the positioning features therefore, some signatures of that location may not contains the coordinates. To get rid of this Brunato suggested that, the location information can be saved as RSS-real coordinate tuple or RSS-custom indicator value tuple (Mauro Brunato 2002) . The later one gives option to add custom features (e.g., elevation, orientation) while the first one solve the regression problem. The following figure shows the use of MDT as RF fingerprint DB, where by applying DCM, the estimated position could be retrieved.

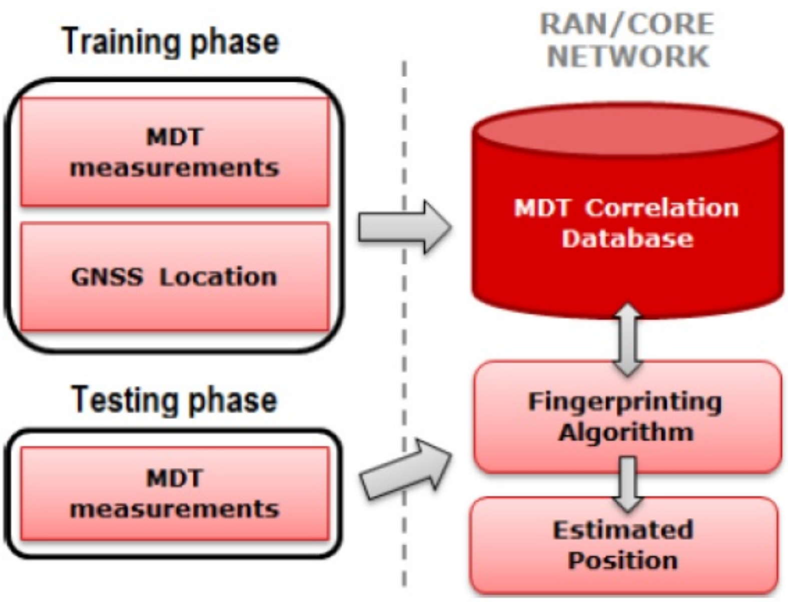

Figure 9. MDT based RF fingerprinting

3.2 Minimization of drive test data

For providing a good quality of service, mobile networks like GSM, UMTS,LTE and TETRA must be monitored regularly. Network providers need to know the radio measurements from different locations of the network with some parameters (e.g., time, load and angle of antenna etc.). These measurements are carried out by cars with the measurement equipments. In order to reduce the operational expenditure and efforts, and to increase network performance and quality, 3GPP release 10 ("3GPP TR 32.827" 2010) addresses the automation of the measurements and configurations called minimizaion of drive test. The idea was to use the logged devices of the network to collect measurement data.

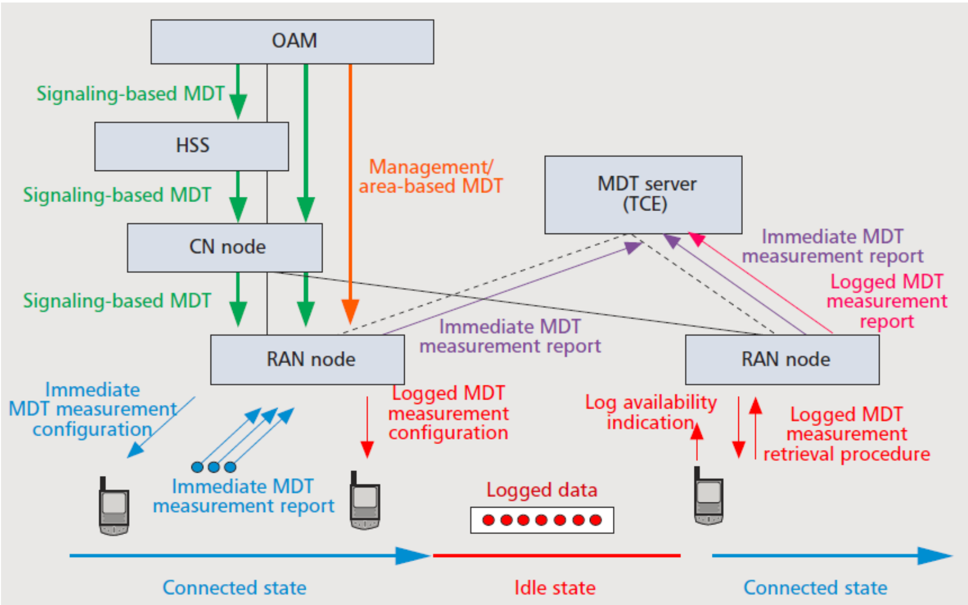

Figure 10. Use of user's equipments for MDT data

Figure 10 illustrates field measurement data collection procedure in a mobile network. In the 3GPP TR 36.805 ("3GPP TR 36.805" 2009) the 3GPP TSG RAN group defined the constraints for MDT, it says that

- The operator shall be able to configure the UE measurements independently from the network configuration.
- The operator shall have the possibility to configure the logging in geographical areas.

MDT helps network providers to know about service status and adopt a Self Organizing Network (SON) on the other hand for this thesis's experimental purpose we can use the MDT data to create a large and extensive correlation databases for RF fingerprint positioning.

3.3 Sample and signature

From a geographical location, experienced MDT measurements generally reference signal received power value and the location information are called sample.

RSRP of BS1	RSRP of BS2	RSRP of BS3	....	RSRP of BS N-1	RSRP of BS N	Location Information

Figure 11. RF fingerprint Sample

The properties and size of the sample depends on the purpose of use of it. Generally in the densely populated area more than 30 BTS/AP signal can be received from any geographical point, but for locational fingerprinting purpose we need not all of those, extensive amount of data make the system loaded. If user's orientation and elevation are expected to be a feature of positioning, at the sample collocation phase those information also need to be recorded, but it would be nearly impossible to enable elevation feature since we are not capable to record data for the third dimension. Generally after collecting the raw data, samples are reshaped to make it ready for further use.

This reasearch (Bahl and Padmanabhan 2000) shows that, due to change in user's orientation, RSS value varies up to 5 dB, therefore if each sample is considered as unique fingerprint than there will be the possibility of ambiguity. Beside this, fraction of a millimetre may create a new fingerprint what is unnecessary. To retain the error as low as possible and to make the system smarter, the average of nearly common samples are counted as a signature of that location. This signature concept reduces the number of fingerprints and the error of positioning. Figure 12 shows the formation of the signature from the samples, the amount of samples require for a signature is depends on the design of the DB. In this thesis study, minimum two sample is required to be a signature, if any unique sample is found that it has no similarity with others than was counted as a noise and was eliminated.

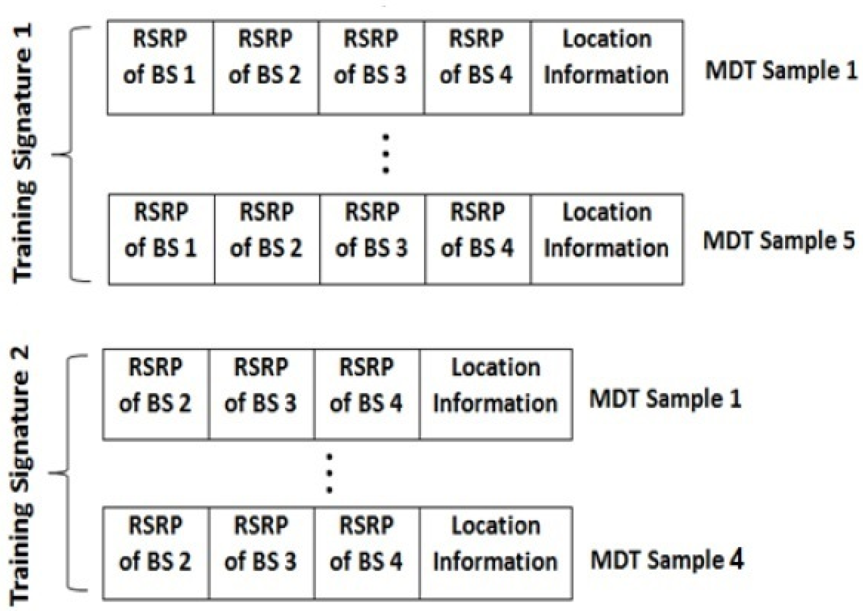

Figure 12. RF Signature

Signature 1 is formed by the average of 5 closely related samples, the mean value of the samples location information is used as signature's location. At data collection phase we recorded latitude and longitude as location information, apart of those other features e.g., orientation and altitude could not be answered.

3.4 Training and testing data

The data what will be use for radio map creation is called as training data. At online phase user may send his radio measurements to the system and ask his position or service provider may detect user's radio measurement from MDT data to offer a service, this instant radio measurements are called testing data. In some experiments, few chunks of data were pulled off from the DB to make some testing signatures while remaining part were used for training data purpose, they will be called as homogeneous data. In some other experiments two different DB were used for training and testing purpose, in this case they are named as heterogeneous data.

4 Received Signal Strength Factors

4.1 Distribution of RSSI

Many researchers have relied on RSSI, a sensor information to determine object's location, but we should realize that the RSSI is not intended to use for determining location, therefore its distribution is expected to be different from other studies of radio wave propagation. According to Sklar's large scale fading model (Sklar 1997) the average RSSI is distributed lognormally, therefore it has the predictability and follows one of the standard path loss model (Pahlavan and Krishnamurthy 2002), while (Xiang et al. 2004) utilized shape filtered on empirical distribution to estimate the RSSI distribution. However, Kaemarungsi found a significance discovery in his research (K. Kaemarungsi 2006) where he observed the RSSI become more symmetric and can be closely approximated by lognormal distribution when the received signal is weak (approximately after -75 dBm), on the other hand the distribution of RSSI becomes significantly left-skewed when the received signal is strong (approximately above -55 dBm). The following figure shows the left-skewed behaviour of RSSI [1]

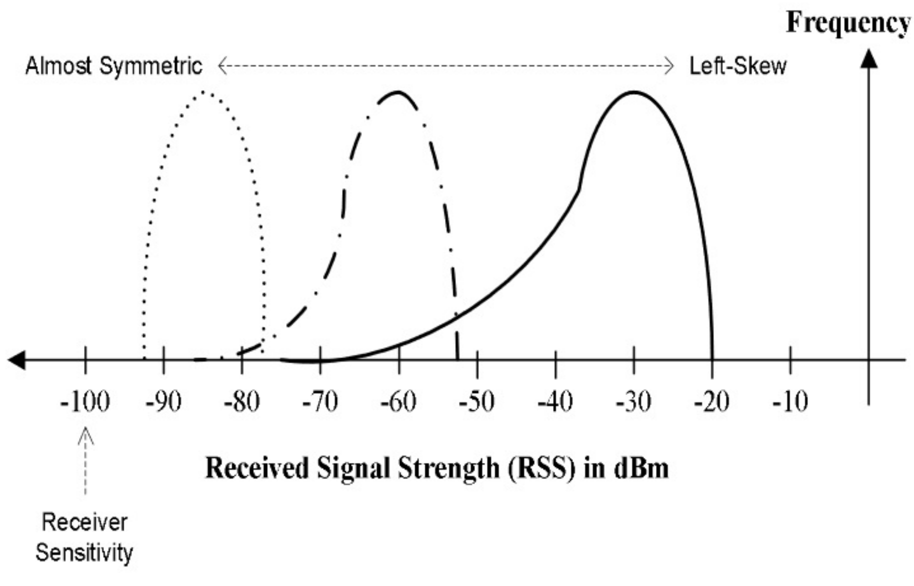

Figure 13. Distribution of RSSI

1. Figure 13 is taken from the Phd thesis (Kamol Kaemarungsi 2005)

He suggested that, the possible reason of the change in distribution is the non-linear mapping between the actual RF energy and the reported RSS values of wireless cards, since small RF power causes small variations. He also concluded that, the narrower range between the lowest limit of RSSI and the highest receivable RSSI at a far away location also contributes for symmetric distribution at the edge of the range. This phenomenon indicates the likelihood of error for weaker signals. In perfect condition, if it is possible to measure small RF energy, long left-skewed distribution is expected across all RSS level. Radio mapping based on correct distribution of RSS would make the location detection more accurate.

4.2 Path loss

Path loss is the reduction in power density of an radio signal as it propagates through channel. This path loss makes the change of RSRP what we are trying to use as an indicator of the change of location.

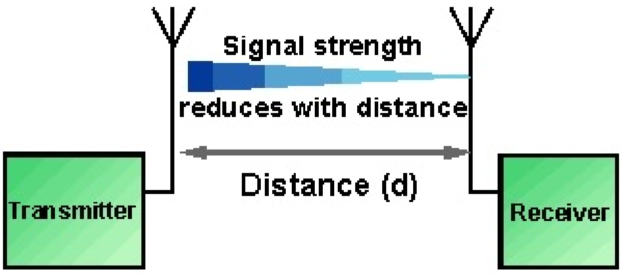

Figure 14. Path loss of radio signal

The linear path loss of the channel is the ratio of transmit power to receive power.

$$L = \frac{P_t}{P_r} \tag{4.1}$$

here L is the path loss, P_t is the transmitted power and P_r is the received power. This formula can be defined in dB

$$P_L dB = 10 log_{10} \frac{P_t}{P_r} \tag{4.2}$$

When signal is transmitted though the air channel it is affected by environmental factors 4.3.4. According Friis transmission equation free space path loss is

$$P_r = P_t G_t G_r \frac{\lambda^2}{(4\pi d)^2} \tag{4.3}$$

Empirically, RSS after the loss can be defined by

$$P_r \propto d^{-\gamma} \tag{4.4}$$

Here γ is called the path loss exponent, typical value of γ is at

Free space = 2

Urban area = 2.7 to 3.5

Suburban area = 3 to 5

Indoor = 1.6 to 1.8

Total path loss formula can be simplified in dB as

$$L = 10\gamma log_{10}(d) + c \tag{4.5}$$

where c expresses the constant value of system losses. Hence the path loss exponent depends on environment and system losses can vary one to another (e.g. training system and testing system) radio measurement may get affected.

4.3 Effect

4.3.1 User's effect

According to the user's comfortability the UE what shall be used to collect MDT data, will be placed nearby different parts of the user's body or in the pocket of vehicle. Placement, movement and presence of user plays a vital role on the mean and the spread of RSSI value.

The user's body influences the RSS distribution by spreading the range of RSS values by a significant amount. Kaemarungsi observed that, the standard deviation is reduced from approximately 3.00 dBm to 0.68 dBm when the user is absent, the reason behind it is the human body reflects or scatterer of the signal and causes the received signal to fluctuate more than vice versa. User movement also makes the signal stronger or weaker (toward or backward of the source), this study (Zhang et al. 2015) it was observed that, the human body movement can cause instability on RSSI

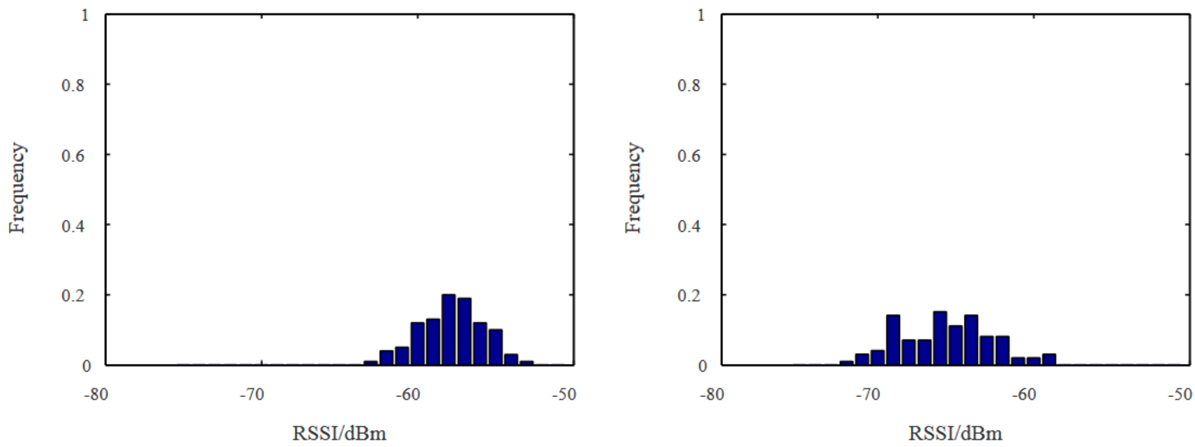

Figure 15. Effect of user body on histogram of the same RSS.

The figure's left part shows distribution of RSSI while user is moving towards the the source of signal, right part shows while moving away from the signal source. In the movement user's body may obstruct the direct path of signal propagation, we know the resonance frequency of water is 2.4 GHz and the human body is consist of 70% water. Therefore, the signal is absorbed when the user obstructs the signal and RSSI value comes down, Kaemarungsi's research shows that the mean RSSI was attenuated by 9.32 dB due to the obstruction from the human body. We can see, presence of human body spreads the range of RSSI, makes the signal weaker. Movement towards the source makes signal stronger, these type of uncertain circumstances may cause inaccurate radio measurement.

4.3.2 Multipath fading effect

Radio waves flow from the transmitter's antenna to the receiver's antenna but not always point to point, it can be affected by various propagation effects such as reflection, diffraction, and scattering. The multipath fading effect is generated by the constructive or destructive combination of RF signal at receiver side.

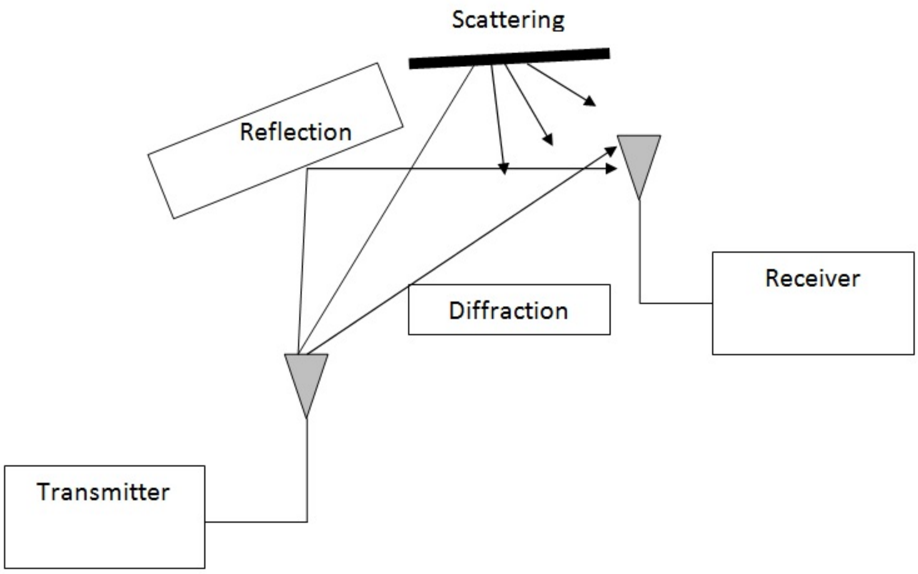

Figure 16. Radio wave propagation.

Therefore there would be multipath interference, it may causes rapid changes in signal strength over a small travel distance or time interval (Fang, Lin, and Lee 2008). This fading incurs a characteristic mismatch between the off-line recorded data and the online measurement, some research utilize a sensor network to adapt the environmental dynamics (Chen et al. 2005).

4.3.3 Device diversity

Difference in network interface card can make difference in RF fingerprint (Youssef, Agrawala, and Shankar 2003), quality and performance of all IEEE 802.11b cards from different vendors are not same. Moreover, different vendors choose to measure RF energy differently. For example, Cisco's 802.11 card measures RSS value based on 100 levels, while the Atheros chipset does by 60 levels, later converted into signal strength by device driver software (Bardwell 2002). As the mapping between the actual RF energy and the RSS range are not same for all manufacturer, the choice of NICs can affect the radio mapping. Wider range of RSSI with good granularity is expected for better location fingerprinting performance.

4.3.4 Other effects

The following physical factors influence small-scale fading in the radio propagation channel:

1. **Motion of the stations** - The relative motion between base station and mobile station creates a "Doppler effect" on each of the multipath components that results in random frequency modulation.

2. **Motion of surrounding objects** - Due to the relative motion of the objects in a radio channel, a time vary Doppler shifts occurs. If surrounding objects motion dominates the speed of the mobile station than this effect outruns the fading.

3. **Transmission bandwidth** - The received signal will be distorted if transmitted signal's bandwidth exceeds the bandwidth of the channel.

4. **Effects of weather** - Different weather factors effects are discussed below

- The hot air of summer and the cold air of winter changes the air density, radio condition of a geographical area is affected by this phenomenon.
- The clouds contain large amounts of water which have a different effect on radio signal propagation compared with air. In some cases while two signal travels opposite direction of each other through same point of clouds, those ionize the area and act as a reflector.
- Lightning is the differentiated charges among clouds, ground, and conducting objects. It can cause significant electrical noise in the channel.
- Large amount of rain within the radio wave path greatly reduce the signal level. On the other hand, as it is consist of dipole property of water, it can behave as a reflector thus enhancing the signal.
- Fog is water vapor, its properties are much like clouds. Nearby fog can weaken VHF,UHF range signal.
- Snow can create a tremendous amount of static electricity. As it made of water it also exhibits some of the characteristics of rain and fog.

5 Signal Strength Based Positioning Algorithms

5.1 Signal strength mean value

The algorithms estimate the location or position from signature of RSS vectors by learning from previous examples of location dependent RF fingerprint. Signal strength mean value algorithm indicates the distance between a transmitter and receiver as a function of the RSS difference. The computed RSS differences are expressed in a signal strength difference vector in which the number of elements represents the amount of matched APs / BTSs between real time and database values. This RSS distance can be express as

$$d = \sqrt{\sum_{i=1}^{n} (RSS_{training} - RSS_{testing})^2} \tag{5.1}$$

where $RSS_{training}$ is RSS value of AP_i or BTS_i at data collection mode, $RSS_{testing}$ is that of positioning mode. n is the amount of matched APs / BTSs. It is very important to take into account the number of n, because lower n value will lead a comparatively minimal distance value with less confidence level of positioning, because of presence of less reference points. However, this research (Bahl and Padmanabhan 2000) suggests that the increasing in the number of n above 3 yields a little benefit due to the inherent noise in the radio measurements imposing a limitation on the localization accuracy.

This algorithm is expected to give better result for homogeneous data, that means there should not be huge variety between training and testing signature. But the layout of the environment is always changing which are putting effects on radio condition of that area e.g., obstacle's position and weather. Moreover presence of the new electronic equipments generate noise and may create coverage hole. If signal level fluctuates highly between training and testing phase it can add error in positioning while calculation is done through this algorithm. But it is obvious that environment will be changing always, by putting some weighted value in this algorithm e.g., counting the number of common access points or base transceiver stations it is expected to get better result.

5.2 kNN algorithm

The k-Nearest Neighbor (kNN) algorithm is a non-parametric method used for classification and regression. The kNN algorithm measures the distance between a query scenario and a set of scenarios in the data set. In RF fingerprinting positioning system to calculate the UE's coordinates the kNN algorithm uses coordinate of k signatures in the database to find out the k-th signature having the smallest RSS distance $d(x||x_i)$. To avoid large scale error, it is the common practice to use an average of these k location coordinates. If ω denotes the coordinate of nearest neighbors than location of the testing signature ($\widehat{\omega}$) is

$$\widehat{\omega} = \frac{1}{k}\sum_{i=1}^{k}\omega_i \tag{5.2}$$

For large scale database where the nearest neighbor cluster size is bigger, this algorithm gives better performance, it is also suitable for noisy training data. For too much fluctuation in training data, additional feature added with kNN algorithm is called weighted kNN algorithm. In this research (Mauro Brunato 2002) weighted kNN algorithm is expressed with weights proportional to the inverse of the distance between signal strengths

$$\widehat{\omega} = \frac{\sum_{i=1}^{k}\frac{1}{d(x||x_i)+d_0}\times\omega_i}{\sum_{i=1}^{k}\frac{1}{d(x||x_i)+d_0}} \tag{5.3}$$

here d_0 is a small real constant used to avoid division by zero. $d(x||x_i)$ is the radio condition between signal strength vector x and x_i measured at training and testing phase respectively. Numerous distance measures are available to find out similarity between vectors e.g., Kullback-Leibler divergence, Euclidean distance, Minkowski distance and Hausdor Distance.

6 Research Method

6.1 Approach

The research approach of this thesis is explaining characteristics of radio signals, finding out sources of positioning error, explaining the possible reason for the error with evidence. This thesis shows some comparative study of different techniques to narrow down the positioning error. Radio mapping was done twice for the same place for a long time interval to observe the change of the radio condition. One of the targets of this research is to use MDT data to build fingerprint database, at Self Organizing Network (SON) implementation it would be an effective feature. Beside this, the another significant difference between this particular study and traditional Wifi signal based positioning study is that, it has been done for outdoor environment, both LTE and WLAN signals was recorded as the reference signal at training phase. As we have seen in the chapter 4 for many reasons signal level fluctuates, therefore instead of single sample, radio signature (average of common samples) have been calculated for the geographical points. Earlier in this department (Mathematical Information System, University of Jyvaäskylä) experiments was done on simulation data (R. Mondal et al. 2013), for this thesis work real data was used. The purpose of the experiments were in a manner so no real-world effects are missed, for comparing smooth and noisy radio mapping, most of the experiments were done in contrast of homogeneous and heterogeneous data. The computer program Matlab was used to process data and experiments purpose. For the simplicity of the language, the station ID of LTE BTS and the WLAN AP will be expressed together as ID. The DB of 14 September 2014 will be called as 1st DB, the 15 May 2015 dated collected data will be called as 2nd DB.

6.2 Grid based method

In grid based positioning method, the total area from where data was recorded is divided into small Grid Cell Units (GCU), those are associated with training signatures. It is also a matter of thorough research that, what would be the size and shape of the grid. For example if a NIC report N level of RSS than the unique locations for the πr^2 area would be N. Ideally, the presence of another source of signal (AP2) should increase the number of unique location exponentially but in real it does not happen due to common locations.

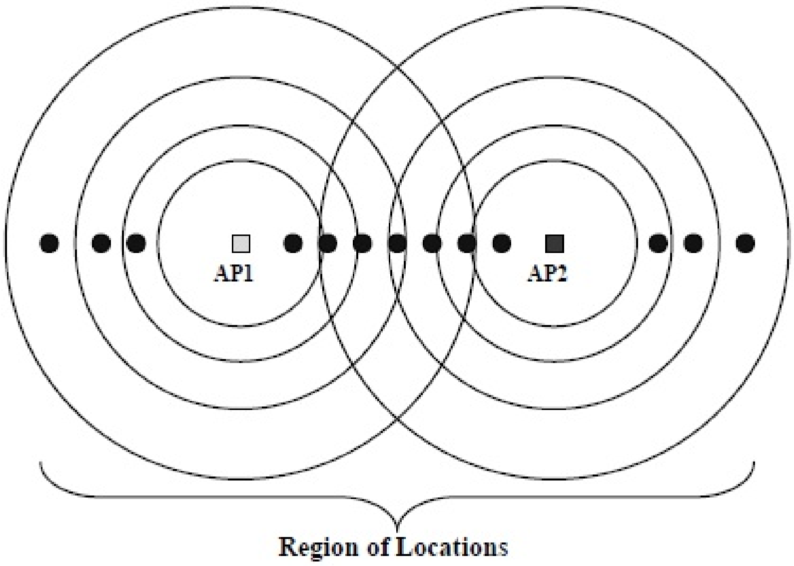

Figure 17. Location distribution in the coverage

For the coverage of $2\pi r^2$ area by two access points the maximum unique location would be N^2 the area per unique location is of $2\pi r^2$ / N^2. The effective grid spacing would be

$$Grid_s = \sqrt{2\pi r^2 / N^2} \tag{6.1}$$

So the grid size depends on the coverage and on the amount of the RSS level. In the sub-chapter 4.3.3 we have seen RSS level varies vendor to vendor. Moreover, to find out coverage radius of each APs/BTSs is a time consuming task. For the simplicity of this thesis work all calculation is done for square shape grids of three different sizes $20m^2$, $10m^2$, and $5m^2$. In every GCU there could be one or several training signatures, for the larger grid size multiple signatures in a grid would increase the probability of being accurate, on the other hand only one signature (the common signature) per GCU contains all the available IDs of that grid so system gains the efficiency to analyse all the testing signatures with minimal similarity. It may happen that, in a grid there is no representation of the signatures because of the area was not accessible during training phase or that place was in the coverage hole or due to huge presence of the obstacles, fewer amount of available station IDs failed to form a signature.

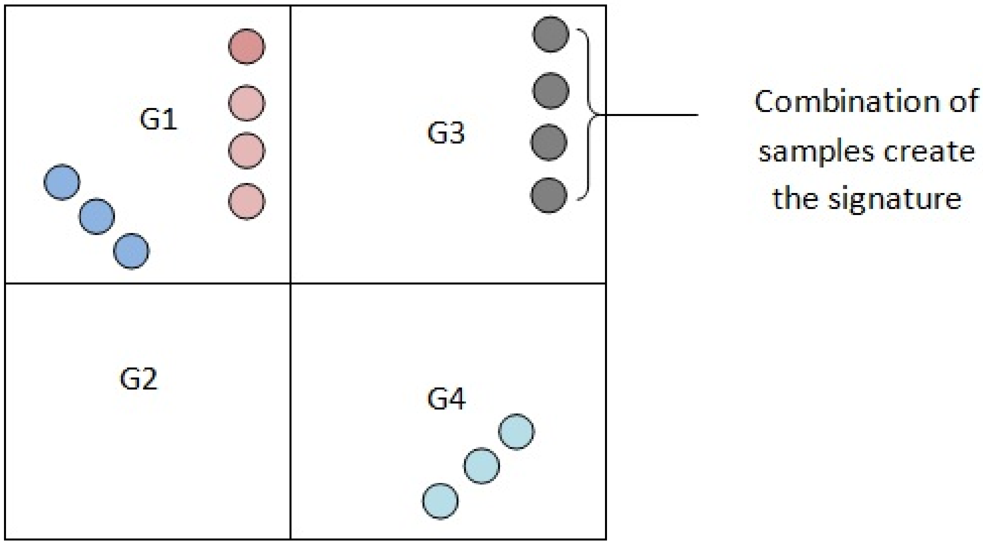

Figure 18. Grid cell unit

The figure 18 shows that, the area of interest is divided into 4 grids. A set of samples are counted as a signature, i.e., G1 contains 2 signatures. Estimated location information of the signature depends on which method is being used, normally the standard candidates are

- The best match training signature's location.
- Average of the closely matched training signatures' location.
- The center position of the GCU.

6.3 Data collection

Data was collected twice from an about 500 meter square area of Tampere, Finland (61.4978°N, 23.7610°E). On the 14th September 2014, the first time data was collected by MIT department's doctoral student Riaz Uddin Mondal while the weather condition was a sunny day of the Autumn. Second time I collected data on the 15th May 2015, it was a partially rainy day of the summer time. The environmental conditions of that location is, a city area with numbers of high rise buildings. There were many WLAN access points and LTE base station available there. By keeping the device in hand data were collected by walking and by bicycling to get the user movement effect. From all accessible location data were recorded.

Figure 19. Data measurement area, Tampere

The figure shows the environmental conditions of the area from where data were collected. Signal would be less interfered at free park area but high rise building and moving objects could contribute on multipath fading and signal interference. An android operating system based Samsung Galaxy S5 mobile phone and an network measurement application called "Nemo Handy" were used for recording the field data. This application can thoroughly measures the on air interface of EGSM, GPRS, EDGE, WCDMA, HSDPA, HSUPA, and WiFi 802.11b/g wireless network ("Nemo Handy" 2014). 1st and 2nd databases contain 33000 and 87930 samples respectively. Data was reshaped through MatLab and finally every signature was arranged as

1 LTE BTS ID + 6 WLAN AP ID + 1 LTE RSRP + 6 WLANs RSRP + Latitude, Longitude

The reason behind 1 LTE signal is that for some other research 1st time (14 September 2014, before starting of this thesis work) data was collected as that. To make comparative study on same type of data second time also same format was maintained. Before creating radio mapping both database were thoroughly checked to eliminate AP/BTS without ID or

RSS value, and any sample that missing the locational information. The device's sampling frequency was 2 samples per second for WLAN and 1 sample per 5 seconds for LTE. The technical specification of the device is

- Android OS, v4.2.2 (Jelly Bean)
- Quad-core 1.6 GHz Cortex-A15 and quad-core 1.2 GHz Cortex-A7
- Network: 2.5G (GSM/ GPRS/ EDGE): 850 / 900 / 1800 / 1900 MHz; 3G (HSPA+ 42Mbps): 850 / 900 / 1900 / 2100 MHz; 4G (LTE Cat 3 100/50Mbps)
- Connectivity: WiFi 802.11 a/b/g/n/ac (HT80), GPS / GLONASS, NFC, Bluetooth 4.0 (LE), IR LED (Remote Control), MHL 2.0 2 GB RAM
- Sensors: Accelerometer, gyro, proximity, compass, barometer, temperature, humidity, gesture.

The following figure shows the device used for raw data collection:

Figure 20. Data collection device

7 Experiments

Based on the collected data, several experiments were done by changing some parameters (e.g. grid size, amount of matched ID numbers, signature per grid etc.) to observe their effects. Results are expressed by Cumulative Distribution Function(CDF) to see the distribution of positioning error. Statistically 68% of the observations fall within 1st standard deviation of the mean, 2nd standard deviation of the mean covers 95% of the observations.

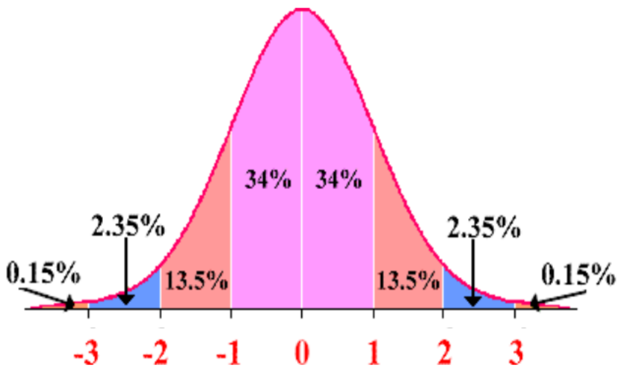

Figure 21. 68-95 rule

Therefore 68-95 rule has been applied through out the experiments. Data from both DB was used, appendixes contain the full experimental results.

7.1 Shortest RSS distance

In this experiment, each GCU contains multiple numbers of signatures, the lowest RSS value distance was considered as the key parameter for choosing the best matched training signature for the testing signature. The RSS distance was calculated by Euclidean formula 5.1 .

7.1.1 Multiple training signatures per GCU

2nd DB was used to make the training DB, to test how the homogeneous data behave some random chunks of data was pulled off from training database (and those were not used for the training DB) to make the testing signatures. As the both data (training and testing) from same source (of same day), it would give us idea that how the data would behave in ideal condition. 16246 training signatures were located in the 500m^2 area. We tested approximately 390 testing signatures (amount of training and testing signatures slightly vary depending on grid size). Following figure expresses the result.

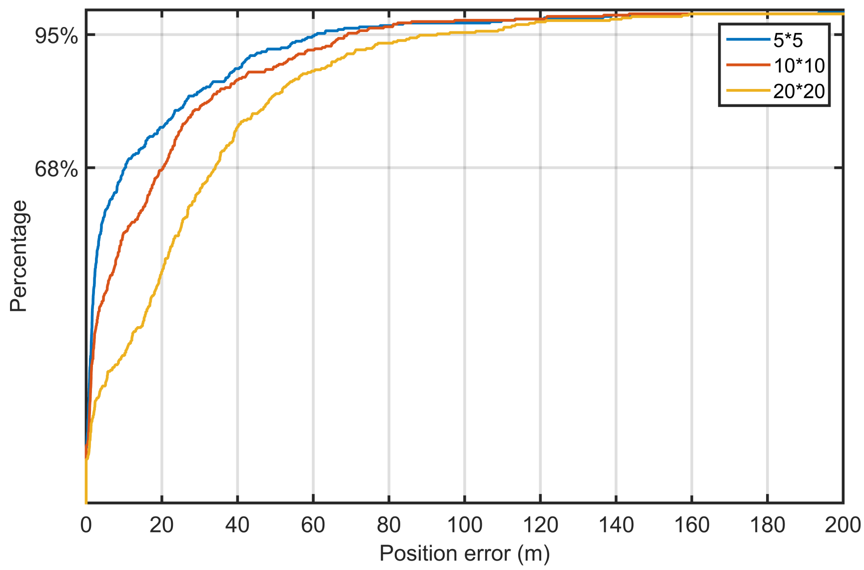

Figure 22. MTS per GCU in three grid size

The figure 22 shows CDF for three different size grid cell (5m^2, 10m^2 and 20m^2). It indicates that the smaller grid size provides the better result (detailed result in the appendix RSS distance method). But smaller grid size needs more calculating resource and time. Moreover, there is always a threshold value of grid size after the result would fall again, in this study threshold were observed at 4.15m^2. Researcher group of this department studied grid size optimization (Mondal, Ristaniemi, and Turkka 2015). So size of grid size need to be optimized with calculation efficiency.

For many reasons RSS can be affected or may not behave according theory4, therefore only one (shortest RSS distance) reference value may contain erroneous value. For the same set up another experiment was done; instead of the shortest RSS distant training signature's location, average of 5 minimal RSS distant training signatures' average location was considered as the testing signature's location. Against of a testing signature if less than 5 close training signatures were found, than the average of available numbers of testing signature's location were counted as the position of the testing signature. Experimental result shows

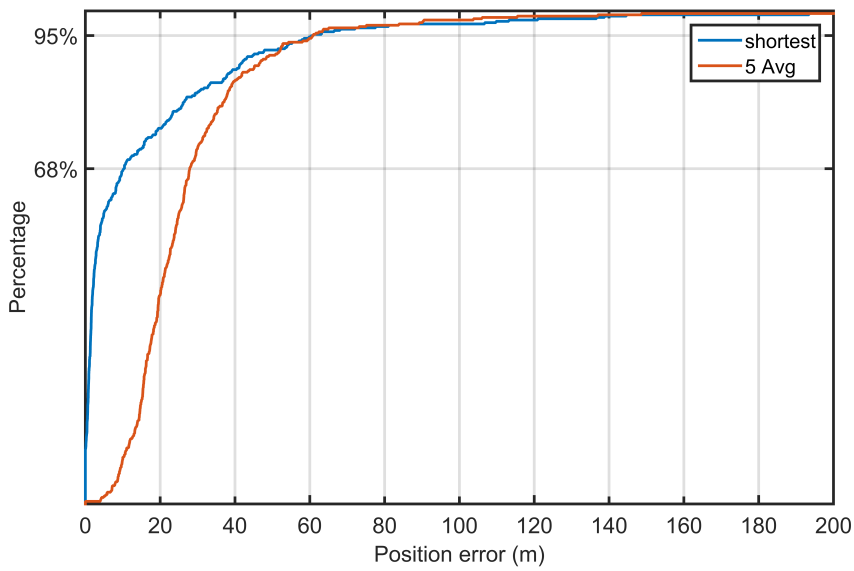

Figure 23. Shortest and average RSS distance scenario

The shortest RSS distance formula gives better result till approximately 93% . From this observation we can come to two possible conclusion:

1. The shortest RSS distance provides the best result. (But it does not as we can see cumulative error is very high.)
2. As the experiment was done in outdoor environment, often the RSS got randomly effected by many contributors. These randomness aggravated the error. Therefore, more reference points (station ID matching) may increase the probability of accuracy.

Bahl and Padmanabhan did there RADAR experiment in indoor environment where signal fluctuation was comparatively less than that of outdoor environment. Moreover, their experiment was done on WLAN signal only, this experiment contains LTE signal too, so their concept of maximum 3 ID matching is not reflecting better result here. It is obvious that the lower number of ID matching condition would make the system more efficient and flexible. The following figure shows the relation among the amount of matched station ID, positioning error and system's analysing efficiency for homogeneous data.

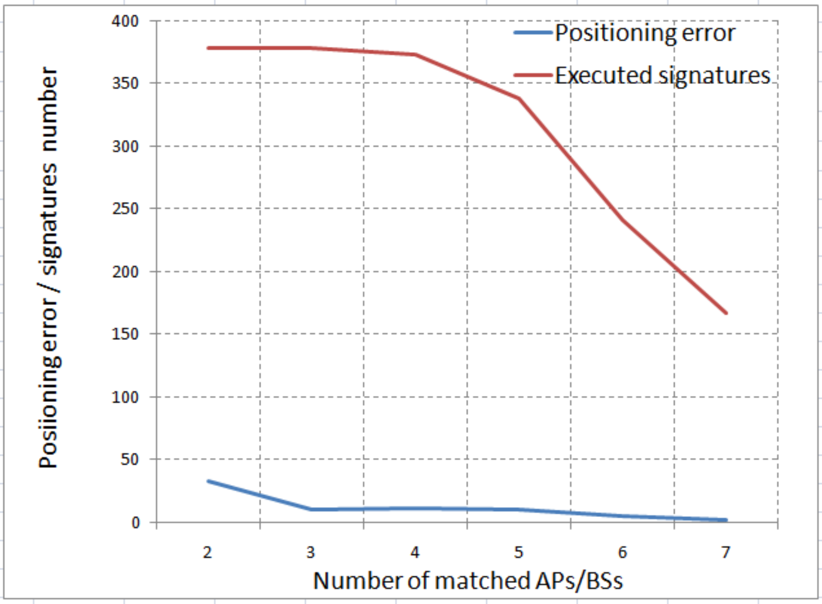

Figure 24. Amount of matched ID vs positioning error

For the $5m^2$ grid cell size, the error come down while the number of matched ID are increased, beside that the system efficiency get decreased if more ID match condition is applied. Out of 378 location retrieval requests only 167 were possible to answer for 100% ID match. Moreover, in the noisy training and testing database it would be hard to find out such high scale ID matching. To observe this phenomenon in cross DB, another experiment was carried out. The 2nd DB was set as the training DB and 378 testing signatures were collected from the 1st DB. 273 requests were possible to answer for 3 ID match with the error of 39.95 meters for 95% , while only 10 requests were answered with 24.10 meters of error when the condition was raised till 6 ID match. Total result sheet for this experiment can be found in the appendix RSS distance method on cross DB.

7.1.2 Common training signature per GCU

Increasing the number of matched ID, reduce the analysing efficiency of the system, but if we can combine all the signatures of a grid cell into a single signature, that training signature would contain all the IDs of that cell, therefore the portability of ID matching with testing signatures would be increased, experiment was done on 2nd DB. As the location of the common signature, we used two methods; 1st method is, the grid center is the location of the testing signature. 2nd method is, the average of the locations of the testing signatures in a GCU is the location of the testing signature. Compare with Multiple Training Signature (MTS) per GCU this Common Training Signature (CTS) per GCU gives following result

Table 1. Comparison of MTS and CTS

Method	Grid size					
	20m2		10m2		5m2	
	68%	95%	68%	95%	68%	95%
MTS per GCU	34.08	92.45	20.30	69.00	10.31	60.70
CTS per GCU, mean location	58.22	139.78	53.46	120.38	38.87	110.47
CTS Per GCU, grid center coordinate	59.95	141.21	54.22	120.10	38.36	110.19

From the appendices RSS distance method for CTS, mean location and RSS distance method for CTS, Grid center we can see now the system can handle all testing signatures for a greater range of common AP/BTS number but location accuracy falls down for both 68% and 95%. For the grid center location and average of training signatures location concepts give almost similar result.

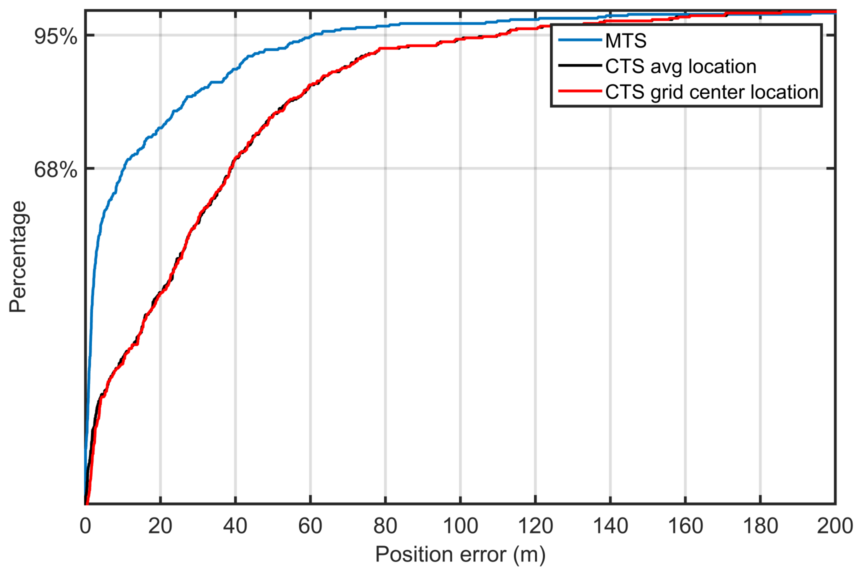

Figure 25. MTS vs CTS

The result shows that the both method of CTS (violet and red line) are laying down than

MTS method (blue line). The possible reason for this is we do not know the physical location of training signatures so when we making the average of that sometimes it may put effect constructively sometimes destructively.

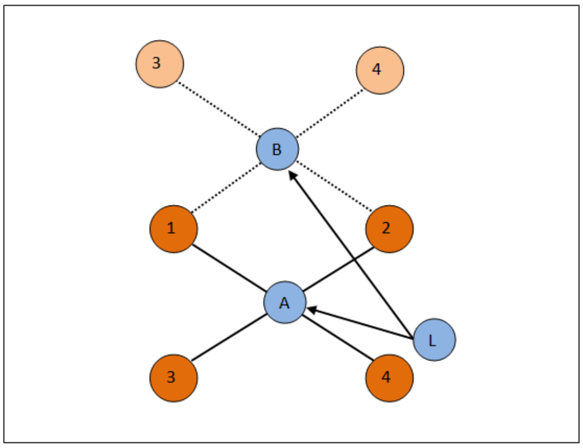

Figure 26. Average of MTS

Let assume, a testing signature's real location is L and it is surrounded by four training signatures 1,2,3 and 4. As a mean value we get point A is the positioning of the testing signature, the positioning error is LA. If location of 3 and 4 would have been at upside (expressed by dotted line) than average location would have been B, the positioning error is LB. But LB > LA therefore sometimes making the average is enhancing the error.

7.1.3 Observations

- Smaller size grid provides better result.
- The more IDs match between training and testing signature the better result. But we need to set optimal amount of IDs matching because more IDs matching decrease the amount of analysed signatures.
- Average of 5 minimal RSS value distance strategy provides better result at 98% whereas the shortest RSS distance signature gives better result at 65% level
- CTS per GCU does not give better result for smaller amount of ID matching, but we need to find out better result with minimum number of matched station ID.

7.2 Weighted RSS distance

From the first experiment we observed that, in the grid based positioning method, the shortest RSS distance is not the sole key parameter specially for outdoor positioning. Formula 5.1 indicates, for less amount of matched ID (the less n value) the RSS distance d would be less but figure 24 shows that for the smaller n value the positioning error is higher. So in this experiment a combine strategy of "shortest RSS distance and amount of matched station ID" is used. The positioning criteria is, against a testing signature first need to check which training signatures contain most amount common IDs. While candidates are more than one, than check the shortest RSS distance (same as experiment 7.1.1). If t is the expected training signature out of a GCU X than

$$t \in X_{GCU} \tag{7.1}$$

$$|\, t \,| = max\ matched\ ID$$

$$|\, t \,| = min\ RSS$$

Experiment was done on homogeneous data (training data and testing data were from same DB of 15th May, 2015). For at least 3 ID matching condition it gives following result

Table 2. The RSS distance method and the weighted RSS distance method on same DB

Error based on shortest RSS distance		Error based on weighted RSS distance		Grid size	Number of analysed testing signature
68 %	95%	68%	95%	m^2	out of approximately 388
34.08	92.45	28.91	76.02	20	371
20.30	69.00	12.68	51.13	10	381
10.31	60.70	4.81	31.01	5	378

It shows a significant improvement of taking more ID matching effect into account. $5m^2$ grid gives overwhelming improvement, about 50% of positioning error reduction at both 68% and 95%. Experiment's detailed result on appendix Weighted RSS distance method. I think one of the cause for this exceedingly better result is that, there is not huge fluctuation between training and testing DB as those from same source. To see the same effect on noisy DB, the same experiment was done on cross DB, data from two different day. Experiment's result reveals that, an acceptable level of testing signatures were been able to analyse with more accurate positioning compared with only shortest RSS distance experiment on cross cross data.

Table 3. Comparison of the RSS distance method and the weighted RSS distance method on cross DB

Error based on shortest RSS distance		Error based on weighted RSS distance		Grid size	Number of analyzed testing signature
68%	95%	68%	95%	m^2	out of approximately 388
45.62	87.07	42.17	83.48	20	190
46.08	90.94	37.02	81.61	10	243
39.95	108.42	32.06	67.11	5	273

At 95% error comes down till 67.11 meters, it shows also stability as error are decreasing gradually along with the size of grid where we observed a haphazard trend without ID matching effect taking into account. This effect filters the result for being mixed with calculation error. Detailed result in the appendix Weighted RSS distance method is applied on cross DB.

7.3 Signature cluster

Attenuation hampers radio signal to travel to a long distance but sometimes due to scattering or reflection it can be received from far away. But the probability of these unusual behaviour are less than the probability of normal radio signal behaviour. From the true location, this possibility would certainly create a densely cluster of matched signature to the closer point of the reference signature (i.e. testing signature) rather than at the far distance (false position). For the remote area, it may result a scattered cluster. From the real data, a testing signature and surrounded training signatures (at least 3 common ID) has been plotted

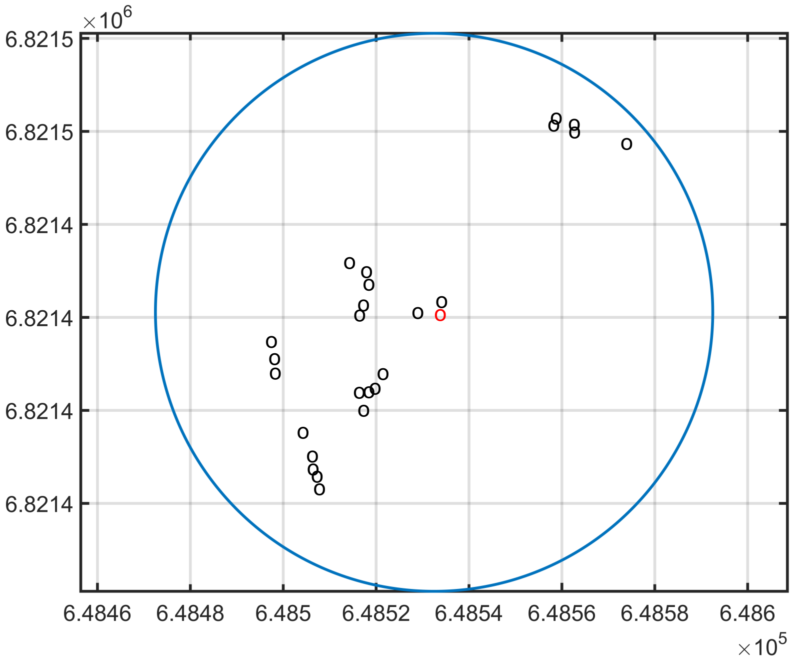

Figure 27. Signatures plotting

The red dot and the black dots are the physical location of the testing signature the training signatures respectively. These were chosen according there RSS proximity. It can be seen that, the amount of remote signatures from red dot are fewer than that neighbor area. This phenomenon could be use as a filtering element to rectify the positioning error. An experiment was done based on cluster size effect. The strategy was

1. Finding out from which grid cell maximum ID are matching. Because the probability of more ID matching from far distance due to randomness is less.
2. Among those grid cells, which one is the most densely clustered.
3. Finally from that cell, the shortest RSS distance training signature's location is the testing signatures location.

The experiment was done on cross data. 2nd DB was used for radio mapping and 1st DB was used as source of testing signatures, the reason for that is the 2nd DB had more amount of sample compare with the 1st DB, therefore 1st DB would give a smoother radio mapping for the area of interest. The comparative result shows that

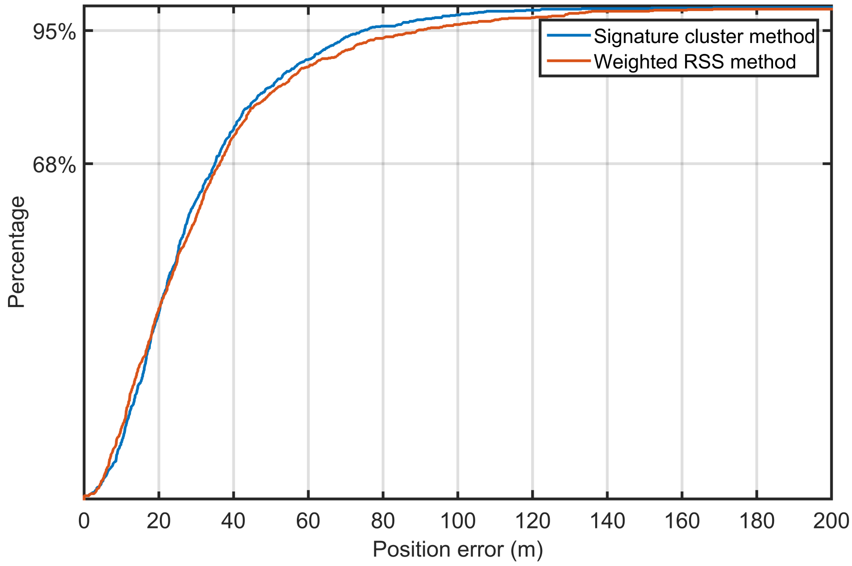

Figure 28. Signature cluster method VS weighted RSS distance method

Till thirty percentile both result are nearly same however, from that point till end we see little bit of improvement. For at least three ID matching for the $5m^2$ grid the experiment's result below:

Before this experiment

For 68% positioning error is 35.57 meters

For 95% positioning error is 79.52 meters

90% testing signature can be analysed (822 out of 940)

After this experiment

For 68% positioning error is 35.60 meters

For 95% positioning error is 73.00 meters

90% testing signature can be analysed (822 out of 940)

From the appendix of the Signature cluster method we can see, this method gives better result even with minimum number of matched ID because of finding the true cluster.

8 Result Analysis

In the previous chapter, we have seen some experiments to find out user's location in the LTE/WLAN network. The minimum positioning error for a cross DB experiment were 32.06 meters and 67.11 meters error for 68% and 95% respectively which remain well below the E911 emergency positioning requirements ("FCC wireless 911 requirements" 2001). By signature cluster method we managed to rise the percentage of analysed testing signature from 80% to 98% for at least two station ID matching on cross database operation. We observed even better result in homogeneous data even though that does not reflect the real world situation but indicates that, in dense small cell networks, MDT training databases can provide a good basis for network based positioning system.

At data collection phase, available WLAN and LTE signals were recorded. Time difference between two DB was very long, therefore probability of change in NIC or BTS is pretty high, what would have created noise in DB. Due to large inequality between DBs we had to emphasis on amount of matched ID number between online and off-line data but for harmonious DBs three common ID should give acceptable level of accuracy. From figure 24 we can see 2, 3 or 4 common ID strategy is giving almost same result beyond that, the accuracy is exceedingly well but analysing capability falls down dramatically.

Smaller grid size were giving better performance linearly on homogeneous data, for dissimilar DB this trend was not found. So it is hard to guess which grid size may provide the best result, in this thesis we found an odd 4.15 meter square grid provided the best result. However, figure 17 tells, for the best grid size selection we need to know coverage radius and physical distance among the APs/BTSs which is literally impossible, while radio mapping would be done through MDT data. This research (Mondal, Ristaniemi, and Turkka 2015) used multi-objective genetic algorithm for selecting the best possible GCU layout over the area of interest so that RF fingerprinting can give the optimal output, but it takes long processing time. May be through some very smart and complex models, positioning accuracy can be improved in a large scale but we need to keep the system as simple as possible, simplicity would save the resource and time, what would make the idea worth to implement in commercial use. Here is the list for possible cause of error for our set up:

1. Signal interference, there were lot of stationary and moving obstacles.
2. Weather difference, sunny and rainy day of two different seasons.
3. Scattered position of LTE base transceiver stations and WLAN access points.
4. Very noisy database, 9 months time difference between training and testing DB.
5. Improper RSS distribution concept, there was no tool available for considering real RSS distribution scenario.
6. Another possible source of error is expecting same behaviour and distribution for LTE and WLAN signal.

Some discussions regarding improvement:

Interference is a fundamental property of radio signal we can not get rid off that. Moreover, weather, multi path, heat and cold these environment related effects are also unavoidable. There are many models available for handling environmental dynamics, so it is possible to implement one of them, comparatively better result is expected.

While collecting data, we recorded all available APs, BTSs and their corresponding signal level. It is normal that, density of APs and BTSs will not be even everywhere. In some place more than enough, in some place only few. We can not put stations where fewer of them but if we would have eliminated excess APs and BTSs from densely clustered area, we could avoid those excess station's weak signal's high fluctuation at far distance.

We collected data only for 2 times, but in real MDT implementation data can be collected round the clock automatically so it would be possible to get updated database always. But that time device diversity 4.3.3 related problem may occur. LTE-TDD signal frequency range is 1850 MHz to 3800 MHz while WLAN signal are transmitted in a range of 2.4 GHz to 5.9 GHz. As WLAN works on comparatively higher frequency range it would be attenuated faster than LTE. For same distance difference WLAN signal level change would be greater than that of LTE. We could not measure the ratio so we assumed same level of change. It could be an area of improvement.

At higher signal level signal behave log-normally but at lower signal level we see symmetric distribution. There are only few papers available on RSSI distributional variation. Proper constant for correction this variation would improve the result.

9 Conclusion

In the thesis, some theories and RF fingerprinting based locationing techniques have been discussed. Several experiments were done to identify the source of error. Step by step, errors were filtered and finally positioning error came down till 32.06 meters and 67.11 meters positioning error for 68% and 95% of testing signatures respectively. It is not the conclusive result, for more or less amount of testing signature the error may vary but it is just an overall idea of RF fingerprint based user locationing. The possible reason for this error is, the time difference between two database was 9 months, therefore there were less similarity between them. There were no available tools to handle environmental effects so radio mapping were not perfectly accurate. Selection of optimal grid size is also a crucial matter, as the variation in result were seen due to change in grid size. We observed better result for homogeneous data, it indicates that updated MDT data based radio mapping would provide better result. Some further research can make RF fingerprinting based locationing technique more effective as we felt necessity of those researches in this thesis study.

1. Consideration of correct RSSI Distribution and environmental dynamics eliminated radio mapping. So far few studies have been done to create a radio map with taking care of proper RSSI distribution effect. Some good papers are available on environmental factors, we need to implement these effects while creating radio mapping and while handling user's position request at online phase.
2. RF fingerprint based positioning for only LTE network. There have been huge amount of location fingerprinting based research for WLAN network specially for indoor positioning. As coverage of WLAN network are smaller than LTE network, positioning accuracy would be better at WLAN network. It would be interesting to see the positioning system only in the LTE network, it would be more challenging because of high signal fluctuation and more interference by the obstacles in a bigger coverage. MIT department of University of Jyvaskyla has done a research on MDT assisted LTE RF fingerprint framework on simulated data (R.U. Mondal et al. 2014), it would be interesting to investigate this issue on real data.

Bibliography

"3GPP TR 32.827". 2010. Visited on March 15, 2016. http://www.qtc.jp/3GPP/Specs/32827-a10.pdf.

"3GPP TR 36.805". 2009. Visited on May 11, 2016. http://www.qtc.jp/3GPP/Specs/36805-900.pdf.

Bahl, P., and V. N. Padmanabhan. 2000. "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,* volume 2, 775–784 vol.2. doi:10.1109/INFCOM.2000.832252.

Bardwell, J. 2002. *Converting Signal Strength Percentage to dBm Values.* Visited on April 29, 2016. https://www.researchgate.net/file.PostFileLoader.html?id=55df0dd85cd9e3ee388b45d7&assetKey=AS:273840230338570@1442300006396.

Ben-Moshe, B., E. Elkin, H. Levi, and A. Weissman. 2011. *Improving Accuracy of GNSS Devices in Urban Canyons.* Visited on October 24, 2015. http://2011.cccg.ca/PDFschedule/papers/paper46.pdf.

Chan, E. C. L., G. Baciu, and S. C. Mak. 2008. "Wireless Tracking Analysis in Location Fingerprinting". In *2008 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications,* 214–220. doi:10.1109/WiMob.2008.30.

Chen, Yi-Chao, Ji-Rung Chiang, Hao-hua Chu, Polly Huang, and Arvin Wen Tsui. 2005. "Sensor-assisted wi-fi indoor location system for adapting to environmental dynamics". In *Proceedings of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems,* 118–125. ACM.

Fang, S. h., T. n. Lin, and K. c. Lee. 2008. "A Novel Algorithm for Multipath Fingerprinting in Indoor WLAN Environments". *IEEE Transactions on Wireless Communications* 7, number 9 (): 3579–3588. ISSN: 1536-1276. doi:10.1109/TWC.2008.070373.

"FCC wireless 911 requirements". 2001. Visited on July 11, 2016. https://transition.fcc.gov/pshs/services/911-services/enhanced911/archives/factsheet_requirements_012001.pdf.

Hapsari, W. A., A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire. 2012. "Minimization of drive tests solution in 3GPP". *IEEE Communications Magazine* 50, number 6 (): 28–36. ISSN: 0163-6804. doi:10.1109/MCOM.2012.6211483.

Johansson, J., W. A. Hapsari, S. Kelley, and G. Bodog. 2012. "Minimization of drive tests in 3GPP release 11". *IEEE Communications Magazine* 50, number 11 (): 36–43. ISSN: 0163-6804. doi:10.1109/MCOM.2012.6353680.

Kaemarungsi, K. 2006. "Distribution of WLAN received signal strength indication for indoor location determination" (): 1–6. doi:10.1109/ISWPC.2006.1613601.

Kaemarungsi, Kamol. 2005. "Design of Indoor Positioning Systems Based on Location Fingerprinting Technique". AAI3188962. PhD thesis.

Laitinen, H., J. Lahteenmaki, and T. Nordstrom. 2001. "Database correlation method for GSM location". In *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd,* volume 4, 2504–2508 vol.4. doi:10.1109/VETECS.2001.944052.

Matosevic, M., Z. Salcic, and S. Berber. 2006. "A Comparison of Accuracy Using a GPS and a Low-Cost DGPS". *IEEE Transactions on Instrumentation and Measurement* 55, number 5 (): 1677–1683. ISSN: 0018-9456. doi:10.1109/TIM.2006.880918.

Mauro Brunato, Roberto Battiti. 2002. *Statistical Learning Theory for Location Fingerprintingin Wireless LANs.* Visited on April 29, 2016. http://disi.unitn.it/~brunato/pubblicazioni/ComNet.pdf.

Milioris, Dimitris, George Tzagkarakis, Artemis Papakonstantinou, Maria Papadopouli, and Panagiotis Tsakalides. 2014. "Low-dimensional signal-strength fingerprint-based positioning in wireless LANs". *Ad Hoc Networks* 12 (1–3): 100–114. doi:10.1016/j.adhoc.2011.12.006..

Mondal, R. U., T. Ristaniemi, and J. Turkka. 2015. “Genetic algorithm optimized grid-based RF fingerprint positioning in heterogeneous small cell networks”. In *2015 International Conference on Location and GNSS (ICL-GNSS),* 1–7. doi:10.1109/ICL-GNSS.2015.7217160.

Mondal, R., J. Turkka, T. Ristaniemi, and T. Henttonen. 2013. “Positioning in heterogeneous small cell networks using MDT RF fingerprints”. In *Communications and Networking (BlackSeaCom), 2013 First International Black Sea Conference on,* 127–131. doi:10.1109/BlackSeaCom.2013.6623395.

Mondal, R.U., J. Turkka, T. Ristaniemi, and T. Henttonen. 2014. “Performance evaluation of MDT assisted LTE RF fingerprint framework”. In *Mobile Computing and Ubiquitous Networking (ICMU), 2014 Seventh International Conference on,* 33–37. doi:10.1109/ICMU.2014.6799054.

“Nemo Handy”. 2014. Visited on March 15, 2016. http://www.elsinco.com/images/doku/Anite-Nemo_Handy_2-70.pdf.

Olsson, Magnus, Shabnam Sultana, Stefan Rommer, Lars Frid, and Catherine Mulligan. 2013. *EPC and 4G Paket Network: Driving the Mobile Broadband Revolution.* 2nd edition. 43. Miami: Academic Press.

“Overview of 2G LCS Technnologies and Standards”. 2001. Visited on March 12, 2016. ftp://www.3gpp.org/workshop/Archive/0101LCS/Docs/PDF/LCS-010019.pdf.

Pahlavan, K., and P. Krishnamurthy. 2002. *Principles of Wireless Networks: A Unified Approach.* Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR. ISBN: 9780130930033. https://books.google.fi/books?id=qRYfAQAAIAAJ.

Pahlavan, K., Xinrong Li, and J. P. Makela. 2002. “Indoor geolocation science and technology”. *IEEE Communications Magazine* 40, number 2 (): 112–118. ISSN: 0163-6804. doi:10.1109/35.983917.

Sklar, B. 1997. “Rayleigh fading channels in mobile digital communication systems .I. Characterization”. *IEEE Communications Magazine* 35, number 7 (): 90–100. ISSN: 0163-6804. doi:10.1109/35.601747.

Xiang, Z., S. Song, J. Chen, H. Wang, J. Huang, and X. Gao. 2004. “A wireless LAN-based indoor positioning technology”. *IBM Journal of Research and Development* 48, number 5.6 (): 617–626. ISSN: 0018-8646. doi:10.1147/rd.485.0617.

Youssef, M. A., A. Agrawala, and A. Udaya Shankar. 2003. “WLAN location determination via clustering and probability distributions”. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on,* 143–150. doi:10.1109/PERCOM.2003.1192736.

Zhang, Kaiqing, Hong Hu, Wenhan Dai, Yuan Shen, and Moe Z Win. 2015. “Indoor localization algorithm for smartphones”. *arXiv preprint arXiv:1503.07628.*

Zhu, Jian, and G. D. Durgin. 2005. “Indoor/outdoor location of cellular handsets based on received signal strength”. In *2005 IEEE 61st Vehicular Technology Conference,* volume 1, 92–96 Vol. 1. doi:10.1109/VETECS.2005.1543256.

Appendices

Experiment : Signature cluster

Database : Cross DB. 1st and 2nd database

Details : For comparing weighted RSS distance method and signature cluster method.

Amount of matched ID	Positioning error for weighted RSS distance(in meter)		Positioning error for Signature cluster (in meter)		Analysed signatures	Grid Size
out of 7	68%	95%	68%	95%	out of 940	m^2
2	39.55	93.74	38.43	80.63	862	$20m^2$
3	41.52	92.05	40.36	80.38	566	
4	40.75	85.92	38.29	79.49	207	
2	37.23	89.47	36.96	76.97	911	$10m^2$
3	37.92	85.32	37.59	80.43	737	
4	37.22	80.87	36.62	75.85	390	
2	36.27	89.15	34.97	74.79	926	$5m^2$
3	35.57	79.52	35.60	73.00	822	
4	34.38	72.85	33.35	68.04	525	

Experiment : Weighted RSS distance method

Database : Single DB, 2nd database

Details : To observe the result of wighted RSS distance method for MTS per GCU on homogeneous data.

Amount of matched ID	Positioning error based on weighted RSS distance (in meter)		Number of analysed Signatures	Grid size
out of 7	68%	95%	out of 388	m^2
2	15.28	59.12	386	$20m^2$
3	28.91	76.02	371	
4	34.24	77.63	306	
5	29.43	66.80	147	
6	9.28	40.03	46	
2	4.56	39.41	383	$10m^2$
3	12.68	51.13	381	
4	18.07	52.87	365	
5	17.63	49.87	260	
6	7.58	31.01	134	
2	2.19	24.46	378	$5m^2$
3	4.81	31.01	378	
4	8.37	40.15	373	
5	9.62	40.07	338	
6	4.97	31.37	241	

Experiment : Weighted RSS distance method is applied on cross DB

Database : Cross DB. 1st and 2nd database

Details : To observe the result of weighted RSS distance method for MTS per GCU on heterogeneous data.

Amount of matched ID	Positioning error based on weighted RSS distance (in meter)		Number of analysed Signatures	Grid size
out of 7	68%	95%	out of 388	m^2
2	37.64	91.17	301	$20m^2$
3	42.17	83.48	190	
4	43.24	67.87	68	
5	34.18	77.20	11	
6	29.43	34.22	2	
2	35.04	88.63	310	$10m^2$
3	37.02	81.61	243	
4	36.82	75.59	130	
5	32.48	110.84	29	
6	32.16	34.22	4	
2	34.19	73.94	314	$5m^2$
3	32.06	67.11	273	
4	31.59	65.35	178	
5	29.53	57.11	69	
6	24.64	45.55	10	

Experiment : RSS distance method on cross DB

Database : Cross DB. 1st and 2nd database

Details : To observe the result of only RSS distance method for MTS per GCU on heterogeneous data.

Amount of matched ID	Positioning error based on lowest RSS distance (in meter)		Positioning error based on average of 5 minimal RSS distance (in meter)		Analysed Signatures	Grid size
out of 7	68%	95%	68%	95%	out of 388	m^2
2	53.57	103.31	49.49	90.40	301	$20m^2$
3	45.62	87.07	42.88	83.48	190	
4	43.24	67.87	36.98	69.72	68	
5	35.11	77.20	34.18	77.20	11	
6	29.43	34.22	29.43	34.22	2	
2	57.24	110.44	52.00	89.16	310	$10m^2$
3	46.08	90.94	43.76	84.84	243	
4	36.82	75.48	36.73	69.48	130	
5	34.42	110.84	33.45	82.25	129	
6	32.16	34.22	28.31	34.22	4	
2	55.61	124.29	56.29	100.83	314	$5m^2$
3	39.95	108.42	40.96	81.22	273	
4	31.73	70.12	33.02	64.88	178	
5	31.54	57.11	35.38	57.95	69	
6	24.64	45.55	23.57	45.55	10	

Experiment : RSS distance method

Database : Single DB, 2nd database

Details : To observe the result of the RSS distance method for MTS per GCU on homogeneous data.

Amount of matched ID	Positioning error based on shortest RSS distance (in meter)		Positioning error based on average of 5 minimal RSS distance (in meter)		Analysed Signatures	Grid size
out of 7	68%	95%	68%	95%	out of 388	m^2
2	32.84	87.97	37.15	87.75	386	$20m^2$
3	34.08	92.45	37.79	76.77	371	
4	34.35	81.74	35.27	68.27	306	
5	28.60	66.80	31.96	61.54	147	
6	9.28	40.03	9.87	44.66	46	
7	1.96	9.77	2.04	9.77	27	
2	32.18	141.64	39.01	94.51	383	$10m^2$
3	20.30	69.00	31.47	62.98	381	
4	22.62	63.58	29.92	63.41	365	
5	18.87	52.94	24.77	50.70	260	
6	7.58	31.01	14.48	32.96	134	
7	2.56	7.55	3.42	8.66	77	
2	33.09	106.27	41.71	90.23	378	$5m^2$
3	10.31	60.70	27.89	60.82	378	
4	11.03	51.40	25.84	50.55	373	
5	10.16	44.54	21.63	42.58	338	
6	5.07	31.87	12.81	32.58	241	
7	2.18	6.62	3.17	10.63	167	

Experiment : RSS distance method for CTS, mean location

Database : Single DB, 2nd database

Details : To observe the result of the RSS distance method for CTS per GCU on homogeneous data. Mean of the training signatures locations as testing signature's location.

Amount of matched ID	Positioning error based on lowest RSS distance (in meter)		Number of analysed Signatures	Grid size
out of 7	68%	95%	out of 388	m^2
2	92.88	207.45	all	$20m^2$
3	58.22	139.78	all	
4	44.32	97.26	all	
5	33.30	79.65	all	
6	21.09	53.98	all	
7	13.00	42.50	all	
2	79.25	190.46	all	$10m^2$
3	53.46	120.38	all	
4	37.54	89.21	all	
5	27.19	66.68	all	
6	15.25	45.16	all	
7	7.04	27.13	all	
2	69.95	187.53	all	$5m^2$
3	38.87	110.17	all	
4	25.37	68.51	all	
5	13.87	45.94	all	
6	5.95	31.95	all	
7	3.48	18.88	all	

Experiment : RSS distance method for CTS, Grid center

Database : Single DB, 2nd database

Details : To observe the result of the RSS distance method for CTS per GCU on homogeneous data. Grid center was counted as the testing signature's location.

Amount of matched ID	Positioning error based on lowest RSS distance (in meter)		Number of analysed Signatures	Grid size
out of 7	68%	95%	out of 388	m^2
2	93.45	209.79	all	$20m^2$
3	59.95	141.21	all	
4	45.90	95.39	all	
5	33.62	79.56	all	
6	21.87	54.58	all	
7	13.72	41.56	all	
2	79.20	189.32	all	$10m^2$
3	54.22	120.10	all	
4	37.97	89.99	all	
5	27.40	66.00	all	
6	15.54	44.99	all	
7	7.42	27.36	all	
2	69.98	187.98	all	$5m^2$
3	38.56	110.19	all	
4	25.68	69.63	all	
5	13.84	45.80	all	
6	5.93	31.02	all	
7	3.84	19.08	all	