ABSTRACT

Title of dissertation: ROLAX : LOCATION DETERMINATION TECHNIQUES IN 4G NETWORKS

Dongwoon Hahn, Doctor of Philosophy, 2012

Dissertation directed by: Professor Ashok K. Agrawala
Department of Electrical and
Computer Engineering

In this dissertation, ROLAX location determination system in 4G networks is presented. ROLAX provides two primary solutions for the location determination in the 4G networks. First, it provides techniques to detect the error-prone wireless conditions in geometric approaches of Time of Arrival (ToA) and Time Difference of Arrival (TDoA). ROLAX provides techniques for a Mobile Station (MS) to determine the Dominant Line-of-Sight Path (DLP) condition given the measurements of the downlink signals from the Base Station (BS). Second, robust RF fingerprinting techniques for the 4G networks are designed. The causes for the signal measurement variation are identified, and the system is designed taking those into account, leading to a significant improvement in accuracy.

ROLAX is organized in two phases: offline and online phases. During the offline phase, the radiomap is constructed by wardriving. In order to provide the portability of the techniques, standard radio measurements such as Received Signal Strength Indication (RSSI) and Carrier to Interference Noise Ratio(CINR) are used

in constructing the radiomap. During the online phase, a MS performs the DLP condition test for each BS it can observe. If the number of the BSs under DLP is small, the MS attempts to determine its location by using the RF fingerprinting.

In ROLAX, the DLP condition is determined from the RSSI, CINR, and RTD (Round Trip Delay) measurements. Features generated from the RSSI difference between two antennas of the MS were also used. The features, including the variance, the level crossing rate, the correlation between the RSSI and RTD, and Kullback-Leibler Divergence, were successfully used in detecting the DLP condition. We note that, compared to using a single feature, appropriately combined multiple features lead to a very accurate DLP condition detection. A number of pattern matching techniques are evaluated for the purpose of the DLP condition detection. Artificial neural networks, instance-based learning, and Rotation Forest are particularly used in the DLP detection. When the Rotation Forest is used, a detection accuracy of 94.8% was achieved in the live 4G networks. It has been noted that features designed in the DLP detection can be useful in the RF fingerprinting.

In ROLAX, in addition to the DLP detection features, mean of RSSI and mean of CINR are used to create unique RF fingerprints. ROLAX RF fingerprinting techniques include: (1) a number of gridding techniques, including overlapped gridding; (2) an automatic radiomap generation technique by the Delaunay triangulation-based interpolation; (3) the filtering of measurements based upon the power-capture relationship between BSs; and (4) algorithms dealing with the missing data.

In this work, software was developed using the interfaces provided by Beceem/Broadcom chip-set based software. Signals were collected from both the home

network (MAXWell 4G network) and the foreign network (Clear 4G network). By combining the techniques in ROLAX, a distance error in the order of 4 meters was achieved in the live 4G networks.

ROLAX: LOCATION DETERMINATION TECHNIQUES IN 4G NETWORKS

by

Dongwoon Hahn

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Ashok K. Agrawala, Chair/Advisor
Professor Mark A. Shayman
Professor Charles B. Silio Jr.
Professor Steven A. Tretter
Professor Atif M. Memon

Dedication

This dissertation is dedicated to my parents, Heungsoo Hahn and Mansoon Kim, my parents-in-law, Jaeak Kim and Yangbin Im, my sister, Sowon Hahn, my daughter, Seunghye Chloe Hahn, and my wife, Chong Mi Kim.

Acknowledgments

First, I am indebted to my advisor, Dr. Ashok Agrawala, for his continuous guidance and teaching throughout the years I spent at the University of Maryland. He has not only taught me how to do research but also taught me how to live. I will keep the lessons I learned from him throughout my life. He has been my greatest inspiration.

I am very grateful to my committee members, Dr. Mark Shayman, Dr. Charles Silio, Dr. Steven Tretter, and Dr. Atif Memon, for serving on my committee. Their teaching and feedback make this work stronger. I really appreciate Dr. Moustafa Youssef for his advises and guidance when I sought for approaches to solve problems in the area of wireless positioning.

Special thanks to those who encouraged me to start the journey to get the PhD degree. I am deeply grateful to Dr. Youngjoong Yoon, Dr. Donghyun Kim, Dr. Kyeongsoo Kim, and Mr. Donghoon Kwak for their support and encouragement, which enabled me to start this challenge after five years in industry.

I was lucky to work with Mr. Manoj Pansare, Dr. Padma Mundur, Mr. Brenton Walker, and researchers at Laboratory for Telecommunications Sciences during my graduate study. I really appreciate their support, and I am very grateful for giving me opportunities to pioneer a variety of technical fields.

Without the support from fantastic staff members at University of Maryland Institute for Advanced Computer Studies, Mr. Fritz McCall and Mr. Mike van Opstal, I would not be able to start my experimentation in 4G networks. I really

appreciate Mr. Raghu Narasimhan for his expertise and help which made the initial experimentation possible.

I really thank my dearest friends at the University of Maryland. It was fortunate to meet such wonderful people. I greatly appreciate support from fellow students of MIND lab, KGECE, and KGSYS.

I always thank my parents for their love and support. I need thousands of pages to describe everything I owed them. Without their support, I would not be able to start this challenge. I really appreciate my parents-in-law for their love and care. I believe I was able to go through hard times because of their support and pray. I sincerely thank my sister, my brothers-in-law, and my sister-in-law in Korea. It was a blessing for our family to live close to my sister when we had just come to Maryland. With her sincere help, my family could settle smoothly into new life in Maryland. I thank my aunt, my uncle, and my cousins, Jean, Susan, Terry, and Casey. They have been sincerely supportive, and our family always appreciates their invitations to their homes over holidays.

My daughter, Seunghye Chloe Hahn, is always my reason to live. She was born just before I started my PhD. Throughout my PhD years, she always makes me smile and cheers me up. I have no doubt that she will be the brightest and the most creative person in our family.

My deepest appreciation goes to my wife, Chong Mi Kim. I cannot express my gratitude enough for her love, care, sacrifice, consideration, and devotion to our family. I feel sorry that she sacrificed her career and her PhD study to support our family. I always relied upon her discretion and wisdom throughout my life. Thanks

to her, this work was possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADC	Analog-to-Digital Converter
AGC	Automatic Gain Control
A-GPS	Assistant GPS
ANN	Artificial Neural Networks
AP	Access Point
API	Application Programming Interface
BS	Base Station
BSID	Base Station ID
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CINR	Carrier to Interference Noise Ratio
CPICH	Common Pilot Channel
CQICH	Fast Channel Feedback
CSMA/CA	Carrier Sense Multiple Access / Collision Avoidance
DCD	Downlink Channel Descriptor
DFT	Discrete Fourier Transform
DL	Downlink
DLP	Dominant LOS Path
DL-MAP	Downlink Map
D-TDOA	Downlink TDOA
EIRP	Equivalent Isotropic Radiated Power
eNodeB	E-UTRAN Node B
ERP	Effective Radiated Power
E-UTRA	Evolved UMTS Terrestrial Radio Access
FBSS	Fast BS Switching
FCC	Federal Communications Commission
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPX	GPS eXchange Format
GSM	Global System for Mobile Communications
HSPA	High-Speed Packet Access
HSUPA	High-Speed Uplink Packet Access
IBSS	Independent Basic Service Set
IP	Internet Protocol
ISM	Industrial, Scientific and Medical radio frequencies
KML	Keyhole Markup Language
K-NN	K-Nearest Neighbor
LBS	Location-Based Service
LCR	Level Crossing Rate
LMU	Location Measurement Unit
LOS	Line of Sight
LS	Location Server
LS	Least Square

LTE	Long Term Evolution
MAC-SAP	Medium Access Control layer Service Access Point
MDHO	Macro Diversity Handoff
MDL	Minimum Discretization Length
MIMO	Multiple-Input and Multiple-Output
MS	Mobile Station
NAP	Network Access Provider
NDLP	Non Dominant Line-of-Sight Path
NLLS	Non-linear Least Square
NLOS	Non-line of Sight
NMEA	National Marine Electronics Association
NSP	Network Service Provider
OFDM	Orthogonal Frequency-Division Multiplexing
P-CCPCH	Primary Common Control Physical Channel
PCAP	Packet Capture
PDF	Probability Density Function
PDML	Packet Details Markup Language
PHY-SAP	Physical layer Service Access Point
PTP	Precision Timing Protocol
QoS	Quality of Service
RCPI	Received Channel Power Indicator
RLCR	RSSI Level Crossing Rate
RMS	Root Mean Square
RSNI	Received Signal to Noise Indicator
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indication
RTD	Round Trip Delay
RTT	Round Trip Time
SC	Spectral Centroid
SF	Spectral Flux
SR	Spectral Roll-off
SUI	Stanford University Interim model
SVM	Support Vector Machine
TDM	Time-division Multiplexing
TDOA	Time Difference of Arrival
TOA	Time of Arrival
TPC	Transmit Power Control
UCD	Uplink Channel Descriptor
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
USB	Universal Serial Bus
U-TDOA	Uplink Time Difference of Arrival
WCDMA	Wideband Code Division Multiple Access

WiMAX	Worldwide Interoperability for Microwave Access
XML	Extensible Markup Language

Chapter 1

Introduction

Wireless communication networks have been evolving, and the most advanced and widely deployed wireless communications networks, as of 2012, are 4G networks, which include IEEE 802.16e-based WiMAX and 3GPP Long Term Evolution (LTE). The number of mobile devices with 4G wireless access has been increasing, and, particularly, the number of WiMAX deployments reached 583 in 150 countries in 2011 [1].

As mobile devices become ubiquitous, it is feasible to offer context-aware services, by using location information to learn and approximate physical environment. Feasible location-based services (LBS) can vary, depending on the quality (e.g. accuracy) of the location information [2]. Location information is needed in services for both people and things. E911 is a representative location-based service that needs to determine location of people. Indeed, precise location determination is crucial in *Internet of Things*. When people process information, missing pieces of the information may be filled in by the human being's discretion. For instance, when a mobile cannot determine its location, the mobile user can manually enter the location information (e.g. zip code) in order to get a location-based service. Since things do not always have this level of intelligence, it is important to provide an accurate location for things.

There is an increasing need for *ubiquitous positioning* as more devices have computing capability and wireless connectivity. However, it has been very hard for a single positioning technology to provide a location determination solution for both indoor and outdoor. Global Positioning System (GPS) is widely used for outdoor positioning, and the GPS accuracy is typically between six to ten meters [3]. Receiver sensitivity of the GPS can be improved by getting aids from networks (e.g. Assisted GPS), or by getting corrected GPS signals from networks of ground towers (e.g. Differential GPS). However, GPS is typically not usable for indoor areas [4] due to attenuated signal power, multipath error, etc. Indoor location can be determined by using GPS pseudolites, but the installation of expensive equipment and the modification of the GPS receiver are required to use pseudolites. In order to provide the indoor location, Wi-Fi-based positioning has been widely used. However, the coverage of the Wi-Fi networks is typically limited to indoor areas or urban areas. Thus, in general, the Wi-Fi-based positioning cannot provide the outdoor location. Therefore, in order to provide the positioning for both indoor and outdoor areas, multiple positioning techniques such as GPS, Wi-Fi-based positioning, dead reckoning using inertial sensors, etc. have been combined.

Location determination, using the cellular-networks, has potential to be a solution for the ubiquitous positioning since cellular signals are essentially ubiquitous. Recent introduction of *femto cells* has increased indoor coverage of cellular networks. The Small Cell Forum and the WiMAX Forum have published the first standard to support the WiMAX femtocells in 2010, and IEEE 802.16m was developed with a consideration of supporting the femtocells. A formal femtocell specification has

been published for the LTE, too. There were 41 femtocell deployments as of June 2012 [5]. As femtocells are becoming widely deployed, it is expected that the 4G will provide better coverage in indoor areas.

In outdoor areas, the 4G-based location determination techniques can be used, either independently or complementarily to the GPS. In addition to frequency aiding from 4G networks, an initial user location, determined by 4G-based techniques, will allow mobile stations to choose satellites to initially seek acquisition, resulting in a shorter initial lock time. Doppler uncertainty can also be calculated if an initial approximate location can be provided [6]. The 4G-based positioning can perform better than the GPS under some circumstances. A comprehensive comparison is discussed in Section 2.3.1.

In 4G networks, sophisticated radio resource measurements such as Received Signal Strength Indication (RSSI) and Carrier to Interference Noise Ratio (CINR) are supported, and they have to meet certain accuracy requirements, while there are no such requirements in Wi-Fi which is widely used for indoor positioning(See Appendix A). Thus, the location determination in 4G networks can be vendor and device independent as long as those standard radio resource measurements are used in determining locations.

ROLAX is a set of robust location determination techniques in 4G networks, which can work in a variety of 4G wireless environment [1]. ROLAX provides a wireless positioning capability without any changes on the hardware of 4G mobile stations (MSs). It also does not require any changes on base stations (BSs), nor

[1]See Section 3.1 for a detailed description of the 4G wireless environment.

an installation of any special equipment. In ROLAX, the location is determined by radio resource measurements on a mobile station, for which the location has to be determined. ROLAX is designed, considering a wide applicability of the techniques across the devices from different vendors, and an extendibility to other 4G networks and future wireless networks. Thus, rather than using non-standard radio resource measurements, it uses a standard set of radio resource measurements, which is common in the most modern wireless networks such as WiMAX, LTE, IEEE 802.11-based Wireless Local Area Networks (WLAN), and IEEE 802.15.4-based ZigBee networks. These standard radio resource measurements include RSSI, CINR, and Round Trip Delay (RTD)[2].

This work targets achieving performance better than requirements set by the IEEE 802.16m [7]: 67 percentile of the Cumulative Distribution Function (CDF) of the position accuracy has to be less than 50 meters, and 95 percentile of the CDF of the position accuracy has to be less than 150 meters. This is the same as the wireless E911 location accuracy required by the Federal Communications Commission (FCC). According to experiments performed in this work, a mean distance error of 4 meters was achieved.

1.1 Contributions: Technical Components of ROLAX

ROLAX is composed of two major technical components, which enhance the positioning accuracy in the 4G networks. The first technical component is detection

[2]Specific terms, used in referring to each type of radio resource measurement, may differ between wireless network technologies.

of Dominant LOS Path (DLP) conditions in 4G networks and the second component is RF fingerprinting techniques designed for the 4G cellular networks.

1.1.1 Dominant LOS Path (DLP) Condition Detection in 4G Networks

To determine location in the cellular networks and other wireless networks, geometric techniques have been widely used. The most popular geometric techniques are trilateration and multilateration.

In the trilateration, to determine a location in a two-dimensional space, the ranges between a node, for which the location has to be determined, and at least three nodes with known locations have to be estimated. In order to estimate a range between a pair of nodes, time-based measurements such as Time of Arrival (ToA) or Round Trip Delay (RTD) can be used. While signal strength can be used to estimate the range, time-based measurements are typically preferred. When a radio signal travels through medium between a pair of nodes for time t , then a range, r, between them can be calculated by (1.1).

$$r = (c/n) \cdot t \tag{1.1}$$

where c is speed of light (299,792,458 meters/second), and n is a refractive index of the medium.

When the medium is air, n is 1.0003. The ToA is described in Figure 1.1. An example of system, using the ToA to determine location, is GPS.

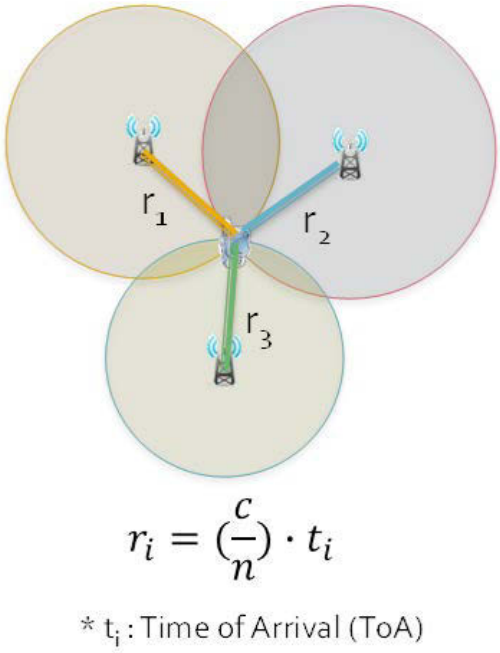


Figure 1.1: Geometric Techniques: Time of Arrival

In the multilateration, to determine a location in a two-dimensional space, a node measures differences in distances to at least two pairs of nodes with known locations. By finding an intersection of two hyperbolic curves, the location can be determined. The Time Difference of Arrival (TDoA) can be used to measure a difference in distance to a pair of nodes.

For the location determination of a MS in cellular networks by TDoA, either downlink or uplink signals can be used. In case of TDoA using the uplink signals (Uplink TDoA; U-TDoA), a reference signal transmitted by a MS is received by BSs or Location Measurement Units (LMUs). The TDoA is calculated between a pair of BSs/LMUs. The MS does not need a special hardware, and the location is determined on the network side. In case of the TDoA using the downlink signals, BSs transmit reference signals, and the TDoA is calculated by the MS. Enhanced

Observed Time Difference (E-OTD) is a downlink TDoA location determination technique used for Global System for Mobile Communications (GSM) and 3G. In E-OTD, Geometric Time Difference (GTD) is calculated by Observed Time Difference (OTD) and Real-Time Difference (RTD), which is the synchronization time difference between BSs. Since the timing difference is measured by the MS, modification on the MS is required to use E-OTD. In UMTS and LTE, Observed TDoA (OTDoA), an uplink TDoA technique, is supported.

In geometric techniques, major sources of error are the absence of dominant LOS path, multi-path, and multi-user interference [8]. If a signal, traveling over a Line-of-Sight (LOS) path, is not received, or if the LOS path is captured by the signals over other Non-Line-of-Sight (NLOS) paths, the estimated time of flight between two nodes has to contain a positive bias since the LOS path is always the shortest path between two nodes. The signals, affected by fast fading, change quickly, and it will make the range estimation difficult since the signal characteristics change over time very quickly. When the LOS path is not dominant, fast fading envelop of received signals is usually modeled by Rayleigh distribution (Rayleigh Fading) [9]. When the LOS path is dominant, fast fading envelop is usually modeled by Rician distribution (Rician Fading) [9]. The amount of amplitude variation in Rician fading is much less than that of the Rayleigh fading [10], thus it is expected to have more variance in signals at the absence of dominant LOS, which can harm the accuracy of the geometric techniques.

In ROLAX, techniques to determine the **Dominant LOS path (DLP)** condition, using standard radio resource measurements, are developed. Test statistics

and features that can be used to determine the DLP conditions are developed and presented in this work.

1.1.2 Robust RF Fingerprinting Techniques in 4G Networks

If there are not enough number of BSs under favorable conditions (at least three BSs are needed for ToA or TDoA), RF fingerprinting is used to determine the location. RF fingerprinting systems are described in Section 2.1.

The RF fingerprinting can be used initially without the DLP condition detection. Feature set for the RF fingerprinting is a super set of the feature set used for the DLP condition detection. Thus, the delay incurred by the RF fingerprinting is primarily introduced by the pattern matching.

The RF fingerprinting is widely used in Wi-Fi-based location determination for indoor areas. Since Wi-Fi is technically different from 4G, techniques developed for the Wi-Fi-based RF fingerprinting cannot directly applied to the 4G-based RF fingerprinting. The technical difference between Wi-Fi and 4G is described in Section 2.3.2. In addition, some techniques can be enabled in 4G while they are not readily applicable in Wi-Fi.

The primary disadvantage of the RF fingerprinting is that it requires a lot of labor and time to collect the measurements to build radiomaps. In ROLAX, automatic radiomap generation techniques, using Delaunay-triangulation-based interpolation and overlapped gridding, are presented. Features and pattern matching techniques are designed, and the missing measurement handling techniques are pre-

sented. A connection-based signal collection procedure is introduced to deal with long scanning time in 4G-WiMAX.

1.2 Organization

In Chapter 2, existing location determination techniques are described. Existing RF fingerprinting systems and previous research about the location determination in 4G networks are surveyed. Differences between the 4G-based location determination techniques and other ground-based and non-ground-based location determination techniques are discussed. In Chapter 3, the 4G wireless environment is described. In Chapter 4, offline signal collection and online location determination procedure of ROLAX are presented. The system and software architectures of ROLAX are also presented. In addition, 4G instrumentation, including hardware and software, used in ROLAX is presented. The 4G networks where the experimentation was performed are also described. Major technical components of ROLAX are presented in Chapter 5 and Chapter 6. Chapter 5 introduces ROLAX DLP condition detection techniques, and Chapter 6 introduces ROLAX RF fingerprinting techniques. Chapter 7 concludes this dissertation and discusses possible future work extending ROLAX.

Chapter 2

Existing Location Determination Techniques

2.1 RF Fingerprinting in Wireless Networks

In the geometric techniques such as ToA, TDoA, and Angle of Arrival (AoA), major sources of impairments are absence of LOS path, multipath, and shadowing effect. Thus, RF fingerprinting is usually used as an alternative to provide a positioning solution, particularly for indoor areas. In RF fingerprinting, a *radiomap* database, where location identifiers are associated with a set of features (RF fingerprints), is built and used for the location determination.

A location determination procedure, using RF fingerprinting techniques, is typically composed of two phases: *offline* and *online* phases. During the offline phase, the radiomap is built by collecting radio signals at locations with known location coordinates or identifiers. Each location, in the radiomap, is associated with a radio signal fingerprint, defined by a set of features extracted from the radio signal measurements such as RSSI. Examples of RF fingerprinting features include a mean value of the measurements over a certain time duration, a histogram of the measurements, and parameters of the distribution if the measurements are assumed to come from known distributions such as Gaussian or Exponential distributions. Measurements are collected per Access Point (AP) or Base Station (BS), so each location is associated with a vector of features. Radio measurements are typically

performed by issuing passive or active scanning in case of Wi-Fi. In 4G WiMAX, scanning is not very efficient for this purpose since wide scanning can take up to several minutes, and typically takes more than ten seconds.[1]

During the online phase, the mobile device, which needs to determine its location, measures radio signals, and features are extracted from them. The features generated during the online phase can be the same as the offline phase features. However, the online features can be different from the offline features. For instance, the offline feature can be the mean of the RSSI measurements while the online feature can be a single-shot RSSI measurement. Feature extraction can be done either on the mobile device or on the network side, depending on computing capability of mobile devices and wireless network bandwidth. In the former case, the extracted features are sent to the location server (LS), which can access the radiomap. In the latter case, the raw measurements are sent to the LS, and the features are extracted on the LS side. Many device drivers for wireless chip-sets report the processed value (e.g. average over a certain moving window) rather than the raw measurements.[2] In that case, features have to be extracted from the processed measurement values rather than from the raw measurements, or the processed values are used as the RF fingerprinting features as they are. The LS determines the location of the mobile device by using pattern matching between the feature vectors stored at the LS and the features generated during the online phase. By finding the best match and

[1]See Section 6.3.1 for further discussion about scanning time in 4G WiMAX.

[2]The RSSI can be also regarded as a processed value by itself since RSSI is usually calculated by averaging the samples from the ADC outputs. Appendix A.1 provides a detailed explanation.

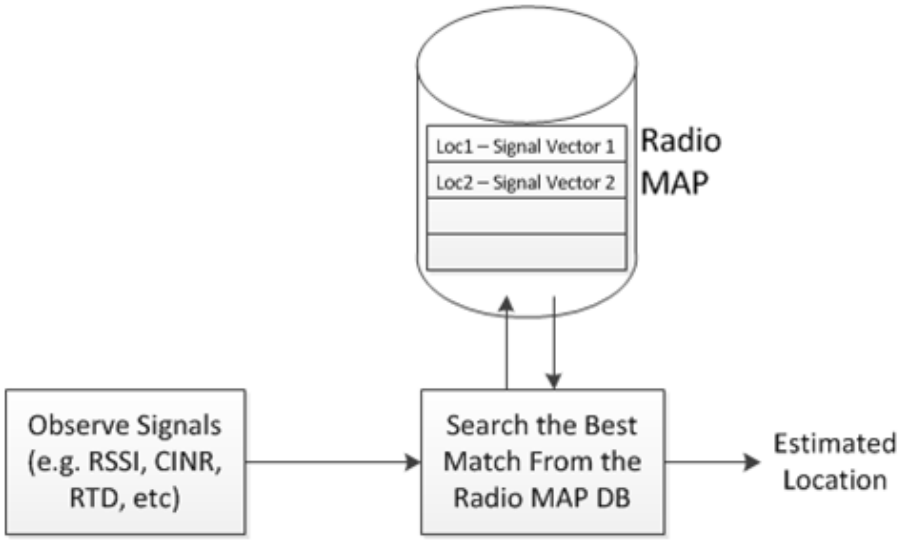


Figure 2.1: RF Fingerprinting System

reading the location identifier associated with it, the location is determined. The continuous space estimation may be needed, to get an accurate location on the continuous domain. A typical composition of a location determination system, using RF fingerprinting, is provided in Figure 2.1.

In order to design a location determination system using RF fingerprinting, the following has to be decided.

- **Underlying Measurements** - received power, round trip delay, etc.
- **Features Extracted from the Underlying Measurements** - average, median, variance, histogram, etc.
- **Pattern Matching** - K-nearest neighbor (K-NN), artificial neural networks (ANN), support vector machine (SVM), etc.
- **Distance Measure** - Euclidean distance, Manhattan distance, probability, etc.

- **Continuous Space Estimation** - center of mass, etc.

An example of the RF fingerprinting system is provided in Figure 2.2. In this example, there are three access points (APs). During the offline phase, radio survey is done at the chosen locations. In this example, RSSI is measured multiple times at each selected location, and mean of RSSI is calculated as a feature of RF fingerprint. Since there are three APs, the dimension of RF fingerprint feature vector is 3 by 1. During the online phase, RSSI is measured for each AP. The measurements can be done once or multiple times. In the former case, a single RSSI measurement is used as a feature by itself, and, in the latter case, features such as mean of RSSI can be extracted from multiple measurements. In either case, a feature vector of size 3 by 1 is generated during the online phase. By using a pattern matching technique, a location with the offline feature vector, which has the highest similarity with the online feature vector, is selected, and this location is reported as the location of the device being tracked. When the nearest neighbor is the used as the pattern matching, a distance measure (e.g. Euclidean distance) is chosen, and the distances are calculated between the online feature vector and the offline feature vectors. By finding an offline feature vector with the smallest distance, the location is determined. In this example, when the Euclidean distance is used, the location 1 provides the smallest distance between its associated offline feature vector and the online feature vector, so the location identifier for the location 1 is reported as the location of the device.

Previous works about RF fingerprinting are presented in Table 2.1. [11] also

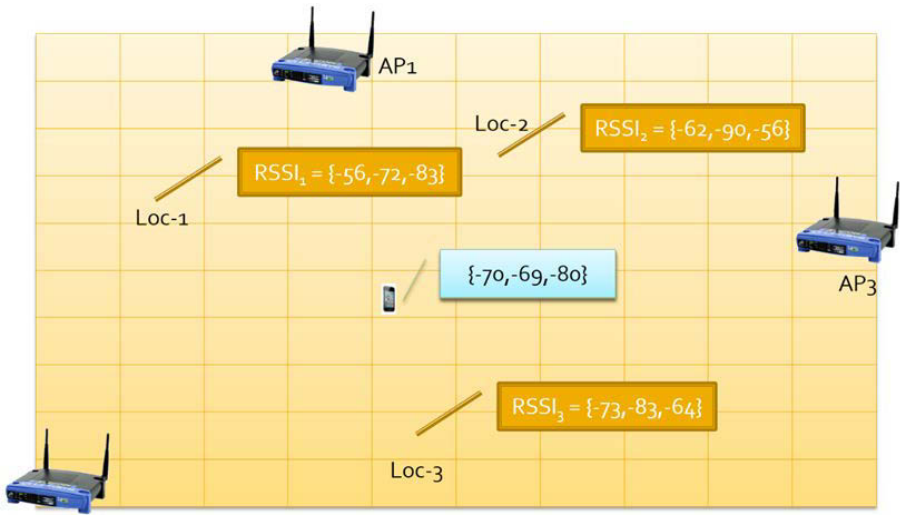


Figure 2.2: Example of Location Determination System Using RF Fingerprinting

provides an extensive list of previous RF fingerprinting research for indoor environment. In Chapter 6, RADAR and Horus are described in detail, and they are compared with ROLAX.

2.2 Location Determination in 4G WiMAX Networks

There is less research done about the location determination in 4G WiMAX networks. In [12], Bshara et al. proposed to use the fingerprinting depending on Received Signal Strength (RSS) observations for positioning and tracking, and presented the results on WiMAX networks. They proposed a static localization method called *BS-strict* where infinite penalty is given to non-matching values in the radiomap and the online phase measurement. For instance, if a BS is observed only in the online phase, then an infinite penalty is given to all candidate locations with measurements with regard to that BS. However, BS-strict has limited usability since there are frequent scanning misses during both offline and online phases, even though

Name / Institution	Network Type	Measurement	Pattern Matching
VUB (Bshara et al.) [12]	WiMAX	RSS, SCORE	Nearest Neighbor
Horus by UMD (Youssef et al.) [13]	Wi-Fi	RSSI	Probabilistic Method
RADAR by Microsoft [14, 15]	Wi-Fi	RSSI	Nearest Neighbor
Polaris Wireless and GATech [16]	GSM	RSSI	Nearest Neighbor
Ecole Polytechnique de Montreal [17]	Wi-Fi	RSSI	Artificial Neural Networks (ANN)
Tsinghua University [18]	Wi-Fi	RSSI	Support Vector Machine (SVM)

Table 2.1: Comparison of RF Fingerprinting Techniques

the signal quality is fair. In addition, BS-strict method is less reliable since signal quality can vary significantly over a small distance and over time. The underlying assumption of BS-strict method is that the scanning misses are primarily because of low signal quality, but scanning misses can be generated because of the capturing effect particularly when all sectors are operating on the same frequency band. They also proposed to use SCORE value, which has lower accuracy than the RSSI, but can be obtained from multiple BSs simultaneously. Another contribution made by this work is the utilization of particle filters to provide the dynamic location determination.

In [19], a cooperative localization, combining TDoA and RSS measurements from WiMAX/Wi-Fi hybrid networks, is presented. In the set-up of this work, a MS obtains a series of TDoA measurements with BSs. Following TDoA measurements, the MS exchanges packets with other MSs over the Wi-Fi connection, measuring RSS over Wi-Fi from other MSs. Upon receiving RSS measurements over Wi-Fi and TDoA measurements over WiMAX from the MS, the BSs fuse the data and determine the location of the MS. The TDoA measurements are used in estimating the positions by least square (LS) algorithm. These estimates are fed to the non-linear least square (NLLS) algorithm as the initial guesses.

Downlink TDoA (D-TDoA) and Uplink TDoA (U-TDoA) can be used in 4G WiMAX. D-TDoA can be used in 4G WiMAX only if Fast BS Switching (FBSS) / Macro Diversity Handoff (MDHO) is in progress. They are optional in mobile WiMAX, both for MS and BS. Because of the conditions that have to be met to use FBSS and MDHO, D-TDoA cannot be used under all circumstances. The required

conditions include that the BSs involving in FBSS / MDHO have to be synchronized based upon a common time resource and have the same frame structures. In addition, the BSs have to be on the same frequency assignment. The first condition can be met by using the GPS, which typically has 10 MHz clock, or by using IEEE 1588v2 Precision Timing Protocol (PTP) over Ethernet. In case of the PTP, the BSs are equipped with PTP slaves, which synchronize with the PTP grandmaster clock. PTP's accuracy is typically less than GPS's. The second condition can be usually met if the BSs are managed by the same Network Access Provider (NAP), and the NAP uses the same frame structures for its BSs. However, the third condition cannot be met under many deployments because each cell is likely to use a different frequency band for a better use of channel resource.

U-TDoA is more flexible than the D-TDoA because it can be used even though the BSs are not in the same frequency assignments. In both D-TDoA and U-TDoA, the MSs have to communicate with the serving BS and its neighboring BSs. Thus, it cannot fully utilize all the observable BSs since the MS is only able to communicate with BSs in its subscription service provider network.

2.3 Comparison of 4G-Based Location Determination Techniques with Other Techniques

2.3.1 Comparison with GPS (Global Positioning System)

GPS has the following disadvantages when it is compared with the 4G-based positioning:

- **Performance Degradation under Radio Shadow and Multipath** GPS-based positioning usually fails under radio shadow where a mobile device can only see a small number of satellites. Because initial lock needs higher power than tracking does, initial lock may fail, or take long time when the view of satellites is obstructed. GPS fails often in the urban areas with skyscrapers, due to shadowing and multipath. Other than ionospheric refraction, multipath is the major source of GPS error, which contributes the distance error in the order of ten meters [20]. When the RF fingerprinting is used in 4G networks, 4G-based positioning techniques work well even in the urban areas since (1) the base stations are deployed with a higher density in the urban areas, and (2) the RF fingerprints in the urban areas are more unique because of the multipath.
- **Susceptibility to Jamming** GPS is more exposed to jamming than ground-based networks since it is operating on a smaller power.[3] Relatively small powered jammer can overpower the legitimate GPS signals,
- **Performance Degradation in Indoor Areas** The accuracy of GPS deteriorates significantly (about 50 meters [21]) when the devices are indoor. Increasing deployments of 4G femto cells expect to provide a better 4G coverage for indoor areas.

[3]Typical received power from a GPS satellite is -127.5 dBm while the typical received power from cellular networks is between -70 dBm and -90 dBm.

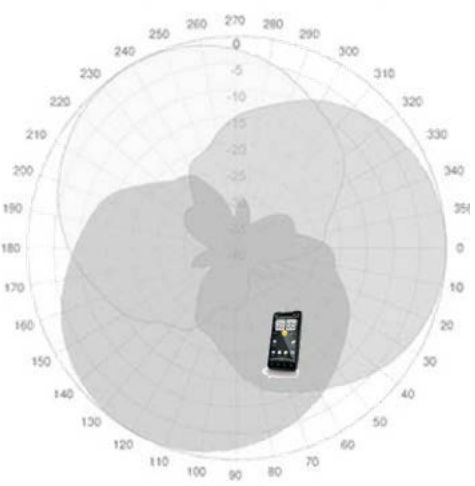
2.3.2 Comparison with Wi-Fi-Based Location Determination

The differences between Wi-Fi and 4G WiMAX include the following:

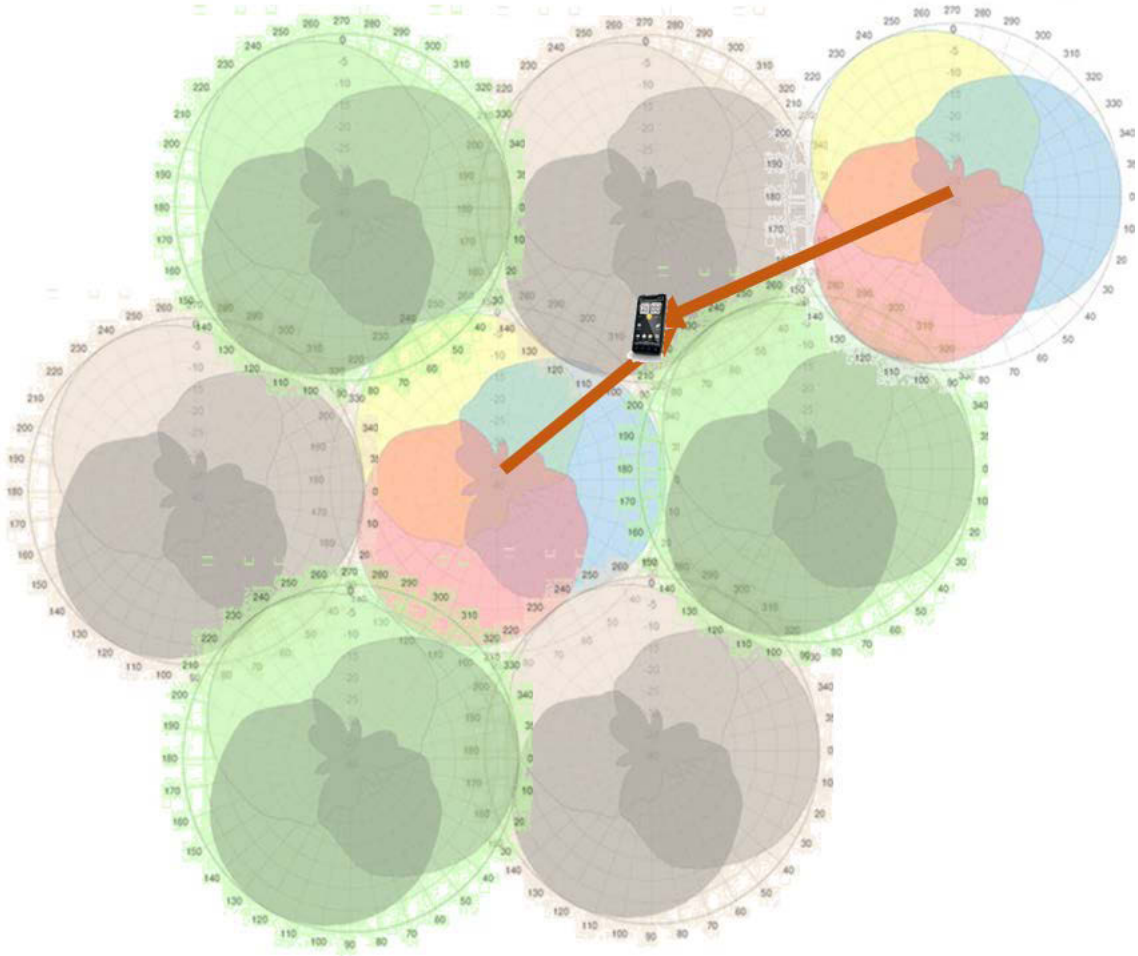
- **Scanning** In Wi-Fi, active or passive scanning is used to locate the access points (in infra structure mode) or peer Wi-Fi nodes (in Independent Basic Service Set mode). Scanning time differs by the wireless environment (e.g. number of the BSs, number of supported channels, etc.). A typical scanning time is in the order of seconds, while the wide scanning in 4G WiMAX can take up to several minutes. Section 6.3 provides detailed comparison between Wi-Fi and WiMAX scanning time.
- **Multiplexing and Capture Effect** Wi-Fi nodes are operating on the unlicensed Industrial, Scientific, and Medical (ISM) band with other Wi-Fi or non Wi-Fi devices. In Wi-Fi, nodes access medium based on Carrier Sense Multiple Access / Collision Avoidance (CSMA / CA). Although beacons are transmitted at a regular interval, an access point defers the beacon transmission when it detects that the channel is busy (delayed beacon) [22]. On contrary, the 4G WiMAX nodes operate on licensed bands. Thus, each base station can assume that the frequency band is solely assigned for its operation. In 4G WiMAX, the channel is accessed based upon Time Division Multiplexing (TDM), and there is no coordination in channel access between the sectors and base stations. Particularly when the frequency reuse factor of one is used, the nodes operating on the same frequency channel have to suffer from interference and capturing effect. CSMA potentially decreases the

impact of the capturing effect in Wi-Fi. Thus, the scanning in 4G WiMAX is likely to miss the existence of some base stations because of the capturing effect. For example, if two BSs are operating on the same frequency band, the mobile stations are likely to detect only the base station with a better signal quality whereas the signals from the non-detected base station can be fair enough. The frequency reuse pattern in 4G WiMAX can be denoted by (N_c, N_s, N_n) where N_c is the number of BS sites per cluster, N_s is the number of sectors per BS site, and N_n is the number of unique frequency channels required for reuse. N_c is the inter-cell frequency reuse factor, and N_n is the intra-cell frequency reuse factor [23]. Figure 2.3(a) shows an example of (1,3,1) reuse pattern and Figure 2.3(b) shows an example of (3,3,3) reuse pattern. In either case, a mobile station may see signals from multiple BS sectors on the same frequency channel, which are not coordinated each other. When a MS receives signals from multiple BSs at the same time, the one with a higher power is usually detected by the MS.

The scanning based measurements may result in very long location determination delay in 4G-WiMAX (Section 6.3 provides the scanning time in 4G-WiMAX from experimental results).



(a) (1,3,1) reuse pattern



(b) (3,3,3) reuse pattern

Figure 2.3: Frequency Reuse Patterns in 4G Cellular Networks

2.3.3 Comparison with 2G/3G Cellular Networks-Based Location Determination

2G and 3G cellular networks can be regarded as the predecessors of the 4G WiMAX network. The most widely deployed 2G networks are GSM and CDMA. The most popular 3G technology is WCDMA (UMTS). Each air interface operates on different frequency bands with different multiple access and modulation schemes. GSM is operating on 800–900MHz and 1.7–1.9 GHz band, CDMA IS-95 is on 800 MHz band, and WCDMA (UMTS) is on 1.9–2.1 GHz band.

4G WiMAX has a potential to provide the better positioning accuracy than 2G/3G. First, the channel bandwidth, typically used by the mobile WiMAX, is 10MHz (can be up to 20MHz), which is larger than 5MHz bandwidth used in UMTS and 200kHz bandwidth in GSM. The wider the bandwidth is the finer resolution of timing we have. In addition, the signal with the higher bandwidth helps in resolving multipath components.

Second, the frequencies of 4G WiMAX is higher than the frequencies used in 2G/3G networks. Although there is no global frequency band, WiMAX forum published three licensed spectrum profiles, which are 2.3 GHz, 2.5 GHz and 3.5 GHz. The density of the BS placement is higher in 2.5 GHz WiMAX than 2G/3G, and higher density of BSs helps in increasing the accuracy. A typical 2.5 GHz WiMAX cell radius in the rural area is 3.3 km, while a 1.9 GHz HSUPA cell radius is 7.5 km [24].

Chapter 3

Description of 4G Wireless Environment

3.1 Dynamic 4G Environment

In 4G, BSs are fixed, and they are typically installed on towers or on top of high-rise buildings. Between MSs and BSs, there are fixed obstacles, including buildings and natural surroundings such as hills, trees, etc. Mobility can be involved both MSs and the surroundings such as moving vehicles and crowds. The atmospheric environment can be dynamically changing over time of year and time of day. A MS typically can see multiple BSs, and multiple service providers may provide the service in the same area. For instance, in College Park, Maryland, where most of the experimentation for ROLAX was done, there are two NAPs: Clear and MAXWell. In 4G WiMAX, the geo-locations of the some BSs are known a priori, or can be obtained by receiving Location-based Services (LBS-ADV) messages on the downlink. The LBS-ADV messages provide absolute position of the transmitting BS, relative positions of the neighboring BSs, GPS time, GPS time accuracy, and frequency accuracy. However, the BS is not obliged to transmit the LBS-ADV messages. In general, the BS will not advertise the LBS information of the BSs operated by other NAPs. MSs have to be provisioned on the network side in order to achieve IP connectivity. This dynamic environment is depicted in Figure 3.1. ROLAX location determination techniques are designed considering this dynamic

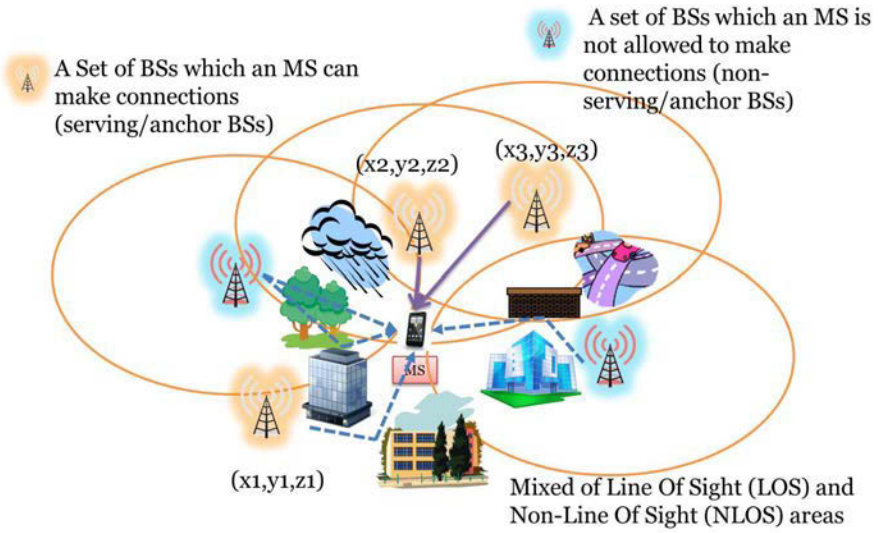


Figure 3.1: Dynamic 4G Environment

wireless channel environment.

3.2 4G Wireless Channel

The sources of the positioning errors differ by the channel environment and the types of measurements made for the positioning. For instance, the RTD value can be affected by the system delay. The Angle of Arrival techniques have significant errors if the reflected signal is stronger than the direct signal. The positioning errors, in geometric techniques, increase when the MS does not have a dominant LOS path with the BSs.

The fixed WiMAX, operating on 10–66 GHz band, does not support the NLOS operation, but the mobile WiMAX studied in this work supports the NLOS operation. In the urban area, the MS is highly likely to be located under NLOS and multipath from the BSs.

There are different scales of the fading in wireless networks: large-scale, medium-scale, and small-scale. The large-scale fading is primarily a function of the distance between a transmitter and a receiver, the transmitter gain, the receiver gain, and the frequency. The log-distance path loss is given in (3.1).

$$Path\ Loss(dB) = P_{tx} - P_{rx} = P_0 + 10\gamma \log_{10} \frac{d}{d_0} + N_g = 10\gamma \log_{10} d + C + N_g \quad (3.1)$$

where P_{tx} is the transmitted power in dBm, P_{rx} is the received power in dBm, P_0 is the path loss at the reference distance d_0, d is the distance between the transmitter and the receiver, γ is the path loss exponent, and N_g models the influence by the fading, which is usually modeled by the zero-mean Gaussian. C is the constant component in the path loss ($= P_0 - 10\gamma \log_{10} d_0$). C can be regarded as the system loss. The received power is primarily a function of the distance.

More sophisticated path loss model, such as *Okumura-Hata* model, was developed by empirical methods. Okumura-Hata model was later extended by *COST-Hata-Model*, and it is known as *COST 231* model [25]. COST-Hata-Model can be used for the frequency from 1.5 GHz to 2 GHz, MS antenna height from 1 to 10 meters, BS antenna height from 30 meters to 200 meters, and the distance between 1000 to 30000 meters. In [26], an optimized model based on COST-Hata-Model was developed to predict the path loss in the 2.3 GHz band. The COST-Hata-model is given in (3.2).

$$\begin{aligned} Path\ Loss(dB) =& 46.3 + 33.9log_{10}(f) - 13.82log_{10}(h_b) - ah_m \\ &+ (44.9 - 6.55log_{10}(h_b))log_{10}d + c_m \end{aligned} \tag{3.2}$$

where f is the frequency in MHz, d is the distance between the MS and the BS in kilometers, h_b is the BS antenna height above the ground level in meters, f is the frequency in MHz, ah_m is a correction parameter for urban and suburban environment each given as in (3.3) and (3.4), and c_m is a correction parameter for urban and suburban environment (3dB for urban area and 0dB for suburban area).

$$ah_m = 3.2(log_{10}(11.75H_r))^2 - 4.97 \tag{3.3}$$

$$ah_m = (1.1log_{10}f - 0.7)H_r - (1.56log_{10}f - 0.8) \tag{3.4}$$

where H_r is the height of the MS antenna in meters.

Stanford University Interim (SUI) Model was developed empirically by collecting measurements on 1.9 GHz band, and it supports three most common terrain types found in the United States [27]: hilly terrain with moderate-to-heavy tree densities (Category A), mostly flat terrain with moderate-to-heavy tree densities or hilly terrain with light tree densities (Category B), and flat terrain with light tree densities (Category C). The path loss under SUI model is given in (3.5).

$$Path\ Loss(dB) = A + 10\gamma log_{10}(\frac{d}{d_0}) + X_f + X_h + s \tag{3.5}$$

where d_0 is 100 meters, d is the distance between the BS and the MS in meters, s is a log-normally distributed factor used to account for the shadow fading (8.2dB–

10.6dB), X_f is the correction factor for the operating frequency, and X_h is the correction factor for the MS antenna height. A is defined by (3.6), and γ is defined by (3.7).

$$A = 20log_{10}(\frac{4\pi d_0}{\lambda}) \tag{3.6}$$

$$\gamma = a - bh_b + \frac{c}{h_b} \tag{3.7}$$

where λ is the wavelength in meters, h_b is the BS antenna height above the ground in meters (between 10 meters and 80 meters), and the constants a, b and c are given per terrain category type.

The LOS and NLOS conditions can be defined in terms of how much of the first *Fresnel zone* is free of obstacles. Fresnel zone is one of a number of concentric ellipsoids of revolution, which defines volumes in the radiation pattern of a circular aperture [28]. Some obstructions in the first Fresnel zone can be tolerated, and, as a rule of thumb, 60% of clearing is required in the first Fresnel zone to be regarded as LOS, but the recommended obstruction is 20% or less [28] (Figure 3.2).

The equation for calculating the Fresnel zone radius, at any point P in between the endpoints of the link, is provided in the (3.8).

$$F_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1 + d_2}} \tag{3.8}$$

where F_n is the nth Fresnel zone radius in meters, d_1 is the distance from one end in meters, d_2 is the distance from the other end in meters, and λ is the wavelength of the signal in meters.

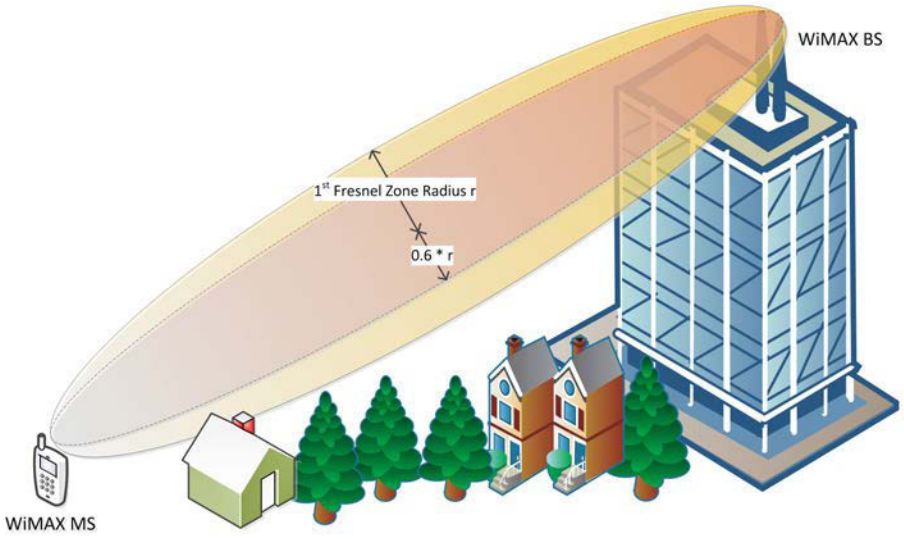


Figure 3.2: Fresnel Zone and NLOS

The maximum radius is obtained at the middle when d_1 is equal to d_2. Under this condition, the maximum radius of the first Fresnel zone in meters is provided in (3.9).

$$r = 8.657\sqrt{\frac{D}{f}} \tag{3.9}$$

where f is the frequency of the signal in gigahertz, and D is the distance between two ends in kilometers.

In case of 2.5GHz WiMAX, the radius and 60% radius of the first Fresnel zone are provided in Figure 3.3. At the distance of one kilometer, 4–5 meters of clearance is required from the ground at all locations between the MS and the BS so that the LOS is not significantly obstructed.

When the fast fading exists due to mobility of terminals or surroundings, the received signal envelope at the presence of a dominant LOS path is typically modeled

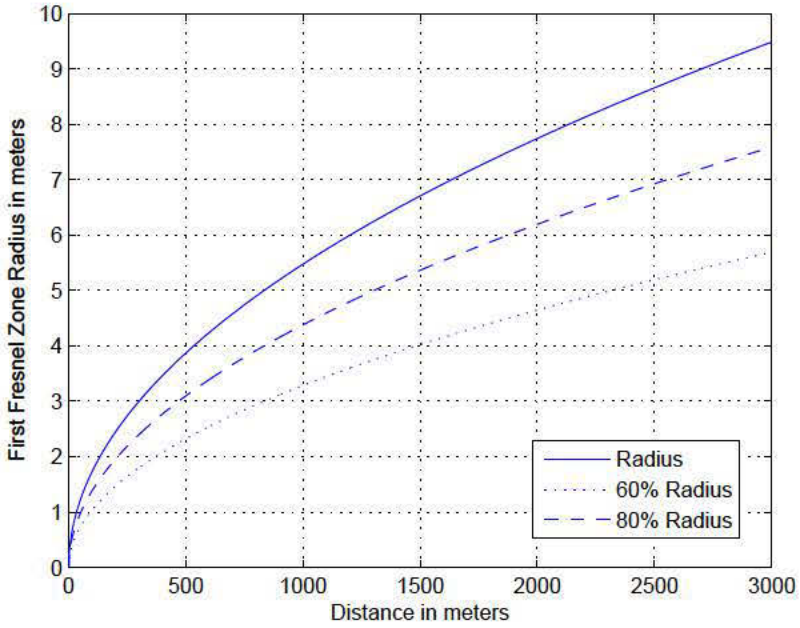


Figure 3.3: Radius of First Fresnel Zone (2.5 GHz WiMAX)

by the Rician Fading model. The Rician distribution [29] is given by (3.10).

$$f_{|r|}(x) = \frac{x}{\sigma^2} e^{-(x^2+\mu^2)/2\sigma^2} I_0\left(\frac{x\mu}{\sigma^2}\right), x \geq 0 \tag{3.10}$$

where μ^2 is the power of the LOS component, and I_0 is the 0th-order, modified Bessel function of the first kind.

The received signal envelope at the absence of a dominant LOS path is typically modeled by the Rayleigh Fading model. When μ is zero (dominant LOS path does not exist), we can observe that (3.10) becomes Rayleigh distribution. The K-factor is typically used to quantify how strong the LOS component is relative to the NLOS components [29]. K-factor is given by (3.11).

$$K = \frac{\mu^2}{2\sigma^2} \tag{3.11}$$

The spectral characteristics of the received signals vary by the significance of DLP condition. For instance, as K-factor decreases (the scattered paths are stronger than the LOS path), the level crossing rate increases [30].

Chapter 4

Location Determination Procedure and Architecture of ROLAX

The following is considered in designing ROLAX location determination system in the 4G networks.

1. **Adaptive to environment** ROLAX can be used in any environment (LOS, NLOS, multipath, shadowing, indoor, and outdoor).
2. **Small overhead** Location is determined by the everyday operations of the MS (e.g. preamble detection, scanning, periodic neighborhood listing, etc.). No special hardware or equipment is required to implement the techniques.
3. **Accuracy** ROLAX targets to provide an accuracy better than the existing ground-based solution on 4G WiMAX.
4. **Feasibility and Portability** Without expensive and sophisticated testing equipment, the location determination must be feasible and implementable. The ROLAX techniques can be used without using special equipment such as the network analyzer, which was used in some previous research such as [17]. The ROLAX techniques use the information available at the device driver level. The techniques use the readily available standard parameters and features of the 4G MS software so that the techniques can be widely applicable independent from a particular vendor's implementation.

The ROLAX location determination techniques are mainly composed of the DLP condition detection techniques and the robust RF fingerprinting techniques in 4G networks. The DLP condition detection techniques are described in Chapter 5, and the RF fingerprinting techniques are described in the Chapter 6.

4.1 Location Determination Procedure in ROLAX

ROLAX is organized in two phases: offline and online phases.

4.1.1 Offline Phase

During the offline phase, the radiomap is constructed. *Wardriving* is used when the measurements are collected in the outdoor area. The sequence chart for the offline phase is provided in Figure 4.1.

The MSs, used in offline signal collection, performs wide scanning multiples times, over the frequencies assigned for the 4G WiMAX operations, to find out the list of frequencies and bandwidths being used. The MS may have the list of frequencies and bandwidths assigned for the NAPs in its non-volatile memory. In this case, the MS may skip the wide scanning. For each frequency in the frequency list, the MS make a connection with a BS operating on that frequency. Typically, the BS with the best signal quality is connected with the MS. During the wardriving, we let the MS make the measurements continuously while performing handoff between the BSs on the same frequency band. In order to expedite the signal collection procedure, multiple MSs, each tuned to a frequency, can be used simultaneously. Once the

measurements are completed, features are extracted from the raw measurements, and the radiomap is created or is updated using them. Detailed explanation about the offline signal collection is provided in Section 6.3.

4.1.2 Online Phase

The MS's location is determined in the online phase. During the online phase, a MS performs a wide scanning to find out the list of BSs it can see. It may have the list of the BSs before it performs the wide scanning. In that case, the MS may skip the wide scanning. For each BS in the list, the MS performs the DLP test to see the channel condition between the MS and each BS. If the number of the BSs under the DLP condition is greater than three, MS attempts to estimate the range between itself and each BS under the DLP condition. The measurements collected during the scanning can be used in the range estimation. To obtain a better accuracy, the MS may attempt to make a temporary connection with each BS to collect additional measurements. Then the MS can apply geometric techniques such as TDoA and ToA to determine its location.

If there are not enough number of BSs under the DLP condition, or the level of precision does not meet the required quality of location, RF fingerprinting is applied to determine the location of the MS. For the BSs with which the MS is allowed to make a connection, the MS may make a temporary connection to obtain additional measurements if additional measurements are to the benefit of increasing accuracy. Next, the location is determined by using pattern matching between the observed

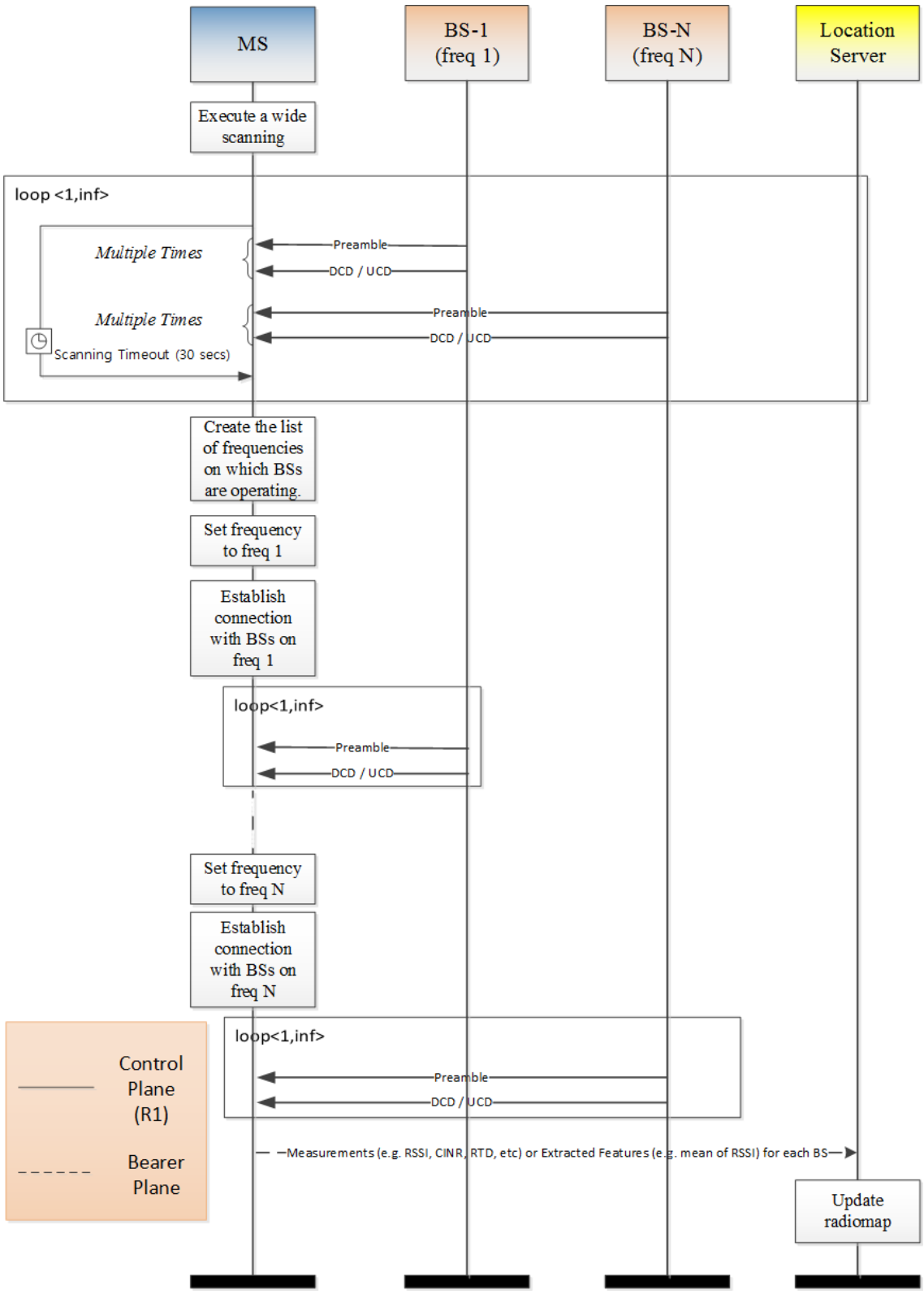


Figure 4.1: Sequence Chart of Offline Phase

signal fingerprint and the RF fingerprints in the location server (LS). The sequence chart for the online phase is provided in Figure 4.2.

4.2 System and Software Architecture of ROLAX

In this Section, the software and system architecture of the ROLAX location determination is described. The control plane of the system, for the offline signal collection, is provided in the Figure 4.3. The preambles are typically measured since it is modulated and coded in the same way while the modulation and coding for the data can by dynamically changed. On the WiMAX chain, the measurements are made on the 4G interfaces. On the GPS chain, the GPS location is logged so that the wireless measurement can be correlated with the location coordinates.[1]. In addition, it synchronizes the host clock with the GPS clock so that the timing drift in the host can be corrected.

On the 4G WiMAX chain, the *connection manager client* sets up the link by loading its configuration and making a connection command with regard to a particular BS. The *logging application* registers callback functions with the WiMAX library so that the WiMAX library can route the control packets registered with the callback functions upon receipt to the loopback interface. The *open-source packet*

[1]In this work, the GPS location is used as the true locations while it may contain some errors. For the deployment of ROLAX, more accurate GPS such as a survey grade GPS can be used to log the coordinates of the locations. Survey grade GPS can provide the accuracy in the order of centimeters [31] It is not expected to have a survey grade GPS in consumer electronics or mobiles due to its high cost.

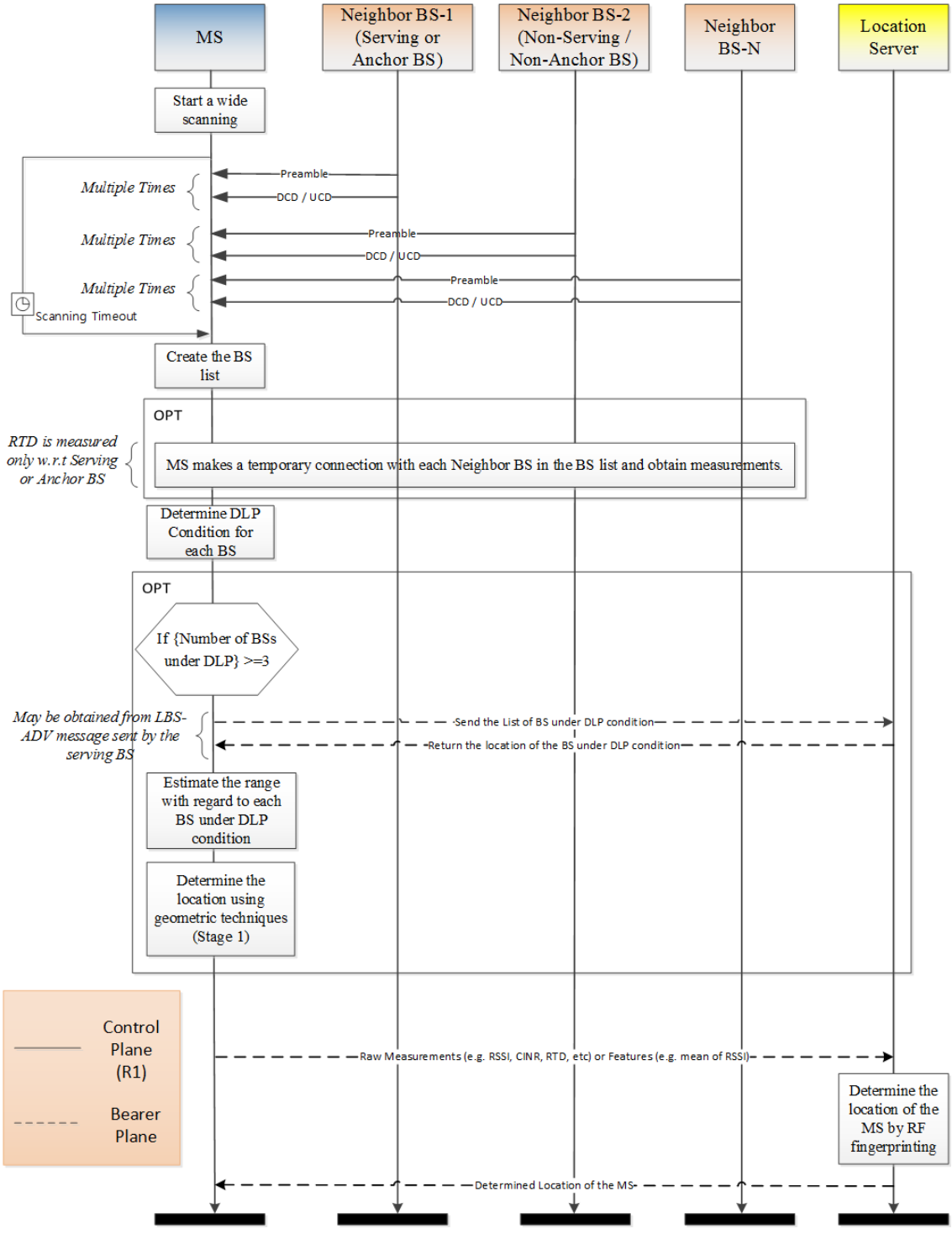


Figure 4.2: Sequence Chart of Online Phase

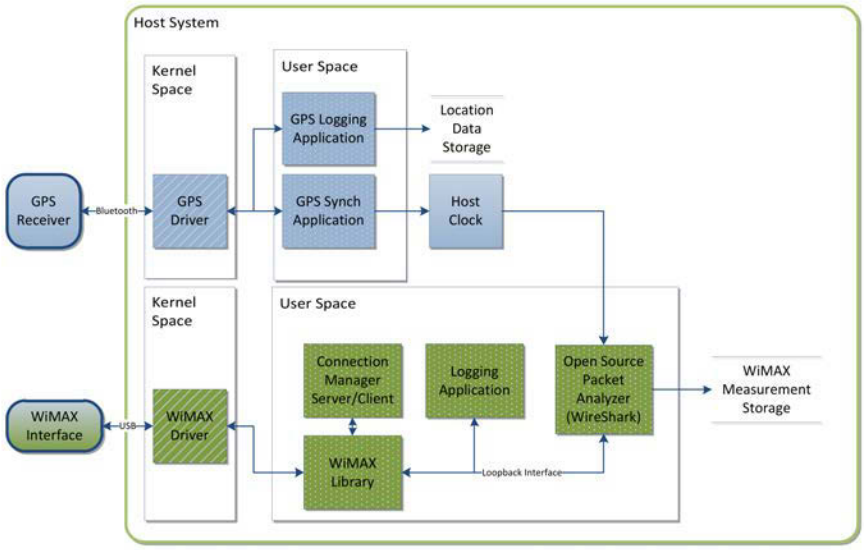


Figure 4.3: System Architecture for Offline Signal Collection

analyzer logs those control packets by monitoring the loopback interface. The log is stored in the *Packet Capture (PCAP)* format.

On the GPS chain, the *GPS synchronization application* periodically obtains the timing information from the GPS receiver and synchronizes the host clock according to the GPS clock.[2] In this way, the timing drift of the host can be corrected, and the radio measurement can be correlated with the location information. GPS logging application makes the location log in the NMEA 0183 format. The location log is updated every one second.

After collecting all the information needed to create the radiomap, they have to be merged and processed to produce the radiomap. The diagram of the software data flow for the radiomap creation is provided in the Figure 4.4.

The WiMAX log in PCAP format is translated to *Packet Details Markup*

[2]In this work, the time synchronization was performed every one minute

Language (PDML) format since Extensible Markup Language (XML) is easier to parse and handle. The NMEA log is converted to *GPS eXchange (GPX)* format for the same purpose. Both PDML and GPX are in XML format. The *PDML Parser* decodes the PDML log and generates the logs, such as neighbor report, BS parameter report, and frame statistics control packets, in a proprietary format. The frequency of the control packet differs by its type. For instance, the frame statistics control packet is generated about every 100 msec. Since about twenty frame statistics control packets are concatenated and encapsulated in a super-packet with a single time stamp, the arrival time, for each frame statistics control packet, has to be estimated. This procedure is performed within the PDML parser. The modem status such as "connected", "waiting for physical synchronization", etc. is correlated with each frame statistics control packet in this module. Both modem state and the frame statistics are keyed with the frame number. The information related to the BS (e.g. BSID) is retrieved from the BS parameter message. The frequency of BS parameter message is generated with 1Hz on average, and the frequency ranges from 1/6 Hz – 2 Hz. Each frame statistics control packet is tagged with the BS information, retrieved from the most recently received BS parameter message.

The neighbor report produced by the PDML parser can be fed to the *Keyhole Markup Language (KML) Creator*. By correlating the location data with the neighbor report, the KML creator produces a KML file, visualizing the measurements on the map.

By combining the BS parameter report, the neighbor report, the frame statis-

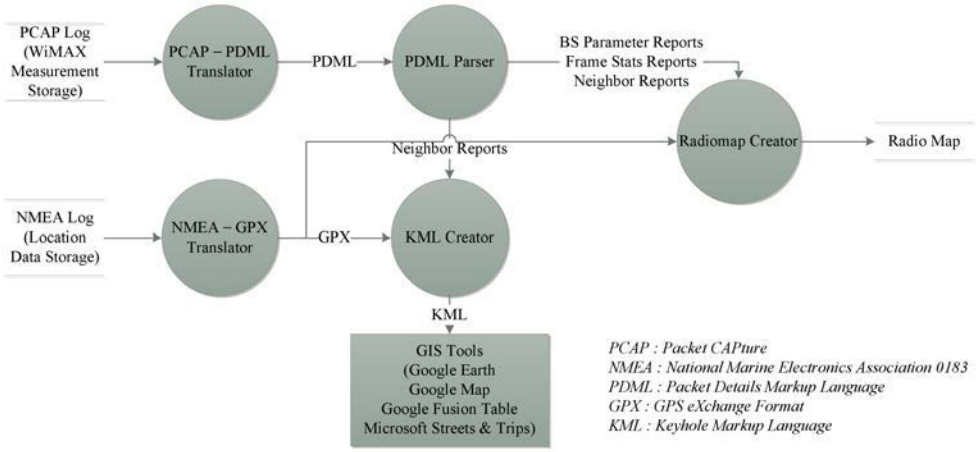


Figure 4.4: Software Data Flow Diagram for Offline Radiomap Generation

tics control packets, and the GPX, *Radiomap Creator* generates a radiomap. Radiomap Creator is composed of a number of blocks. The processing in each block is described in the Chapter 6. The diagram for the radiomap creator is provided in the Figure 4.5.

4.3 4G Instrumentation of ROLAX

The techniques developed in ROLAX were implemented and tested, using the data collected from live 4G networks in College Park, Maryland. For the DLP condition detection, the data is primarily collected from the MAXWell 4G network, while, for the RF fingerprinting, the data is collected from both Clear network and MAXWell 4G network. Software was implemented using Sprint 4G Development Kit, which supports the 4G device with Beceem (Broadcom) chip-sets. Motorola USBW100, which uses a Beceem chip-set, is used as the MS. In this chapter, the technical details of the 4G networks and hardware / software for both BS and MS

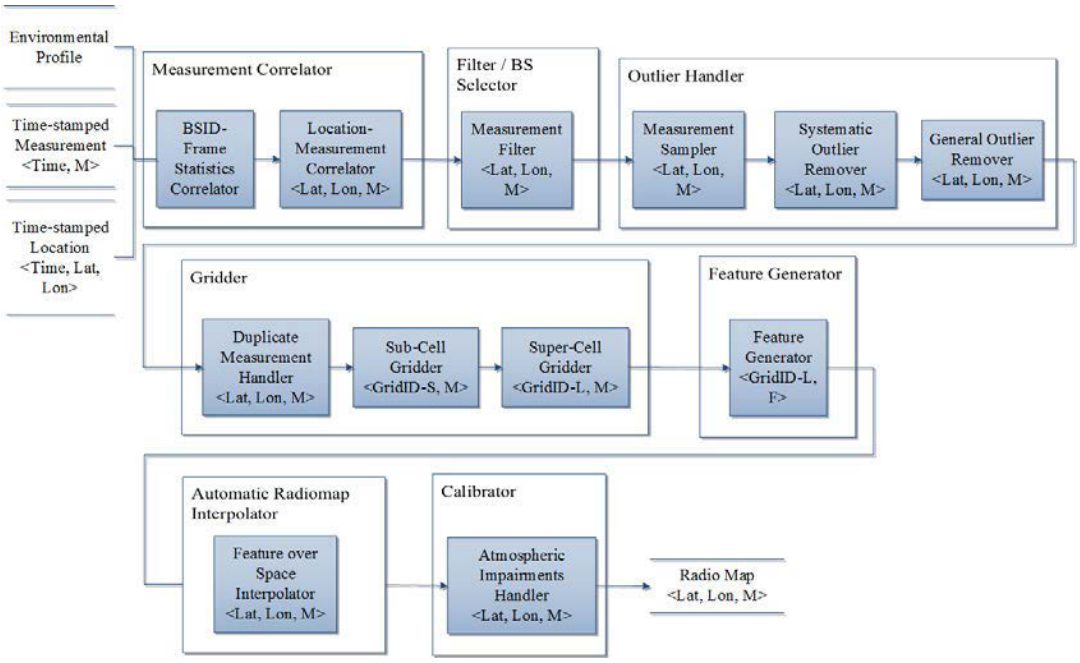


Figure 4.5: System Architecture for Radiomap Generator

are provided.

4.3.1 4G Networks

The data had been collected from the MAXWell 4G network operated by MAXWell Lab of University of Maryland as well as from the 4G WiMAX networks operated by Clear. At the time when this work was implemented (2011 and 2012), Clear was the 4G WiMAX Network Access Provider (NAP) in this area, and multiple Network Service Providers (NSPs) including Clear, Time Warner, Comcast, and Sprint provide the 4G service using the Clear 4G network.

4.3.1.1 MAXWell 4G Network

The MAXWell 4G WiMAX network is deployed in University of Maryland to promote research and development into applications for WiMAX mobile broadband networks. MAXWell lab obtained an experimental license from FCC to operate its network. It is composed of two base stations: one for outdoor service, and the other one for indoor experiments and tests.

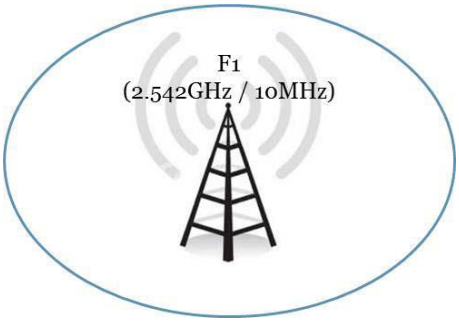
The summary of the MAXWell 4G network's FCC license is as follows:

- Experimental Radio Service (other than broadcast) under Part 5 of FCC rules.
- Frequency band: 2.4985–2.6875 GHz
- Output Power: 1 W (30 dBm)
- ERP (Effective Radiated Power): 50 W (46.99 dBm)

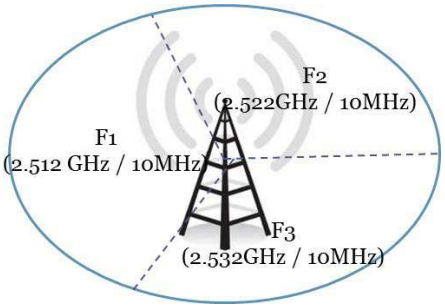
After a site survey performed on November 2010, the frequency and bandwidth assignments were updated to avoid interfering existing services, to better utilize the available channel by providing a higher bandwidth to users, and to provide a higher timing resolution for the timing-based location determination. The frequency assignment for the MAXWell 4G network, as of March 2012, is provided in Figure 4.6. As of March 2012, the equivalent isotropic radiated power (EIRP) of the BS, BS_EIRP, is 34dBm.[3]

The picture of MAXWell 4G network's outdoor RF heads is provided in Figure 4.7.

[3] BS_EIRP is given by $P_{Tx} + GANT\_BS\_Tx$, where P_{Tx} is the transmit power, and $GANT\_BS\_Tx$ is the transmit antenna gain.



(a) Frequency and Bandwidth Assignment for Indoor BS



(b) Frequency and Bandwidth Assignment for Outdoor BS

Figure 4.6: Frequency and Bandwidth Assignments in MAXWell 4G Network

Figure 4.7: MAXWell 4G Network's Outdoor RF Heads

4.3.1.2 Clear 4G Networks in College Park, Maryland

While we could configure and access the equipment of MAXWell 4G network, we could not access the equipment of Clear 4G network because it is operated by a commercial operator. Its configuration had to be inferred by capturing and decoding the management frames from the BSs such as Downlink Channel Descriptor (DCD) and Uplink Channel Descriptor (UCD).

However, the signals from the BSs operated by Clear can be observed by any MSs, and they are used for location determination in this work. Particularly, the RF fingerprinting in ROLAX uses the signals from the Clear 4G networks in addition to the signals from the MAXWell 4G network.

The locations of the Clear BSs can be also inferred from publicly accessible database such as AntennaSearch [32]. If the locations of the BSs are known, geometric techniques can be used for the mobile-based location determination even though the BSs are neither serving BSs nor anchor BSs. In this case, there could be some limitations because some measurements, such as timing measurements (e.g. RTD), can be only made with serving BSs or anchor BSs.

By using the scanning operation of the MS, the frequencies, bandwidth, and frequency reuse pattern could be found. The settings of Clear 4G network and MAXWell 4G network, in College Park area, are provided in Table 4.1. Clear BSs' EIRP is 42dBm, which is 8dB higher than MAXWell outdoor BS's EIRP.

According to a scanning test performed in College Park campus, 4.7 BSIDs were observed per scanning attempt on average.

NAP	Frequencies (MHz)	Bandwidth (MHz)	Frequency Reuse	BS_EIRP (dBm)	Number of Observed BSIDs
Clear	2630.5, 2647, 2657, 2667, 2673.5, 2683.5	10	3	42	25
MAXWell (outdoor)	2512, 2522, 2532	10	3	34	3
MAXWell (indoor)	2542	10	1	24	1

Table 4.1: Configurations of BSs Managed by Clear and MAXWell Lab in College Park, Maryland

4.3.2 4G Hardware

The specification of base station and mobile station hardware, operated by MAXWell 4G network and used in this work, is provided in Table 4.2.

4.3.3 4G Software

For the most of the experimentation in this work, Beceem (Broadcom) device driver, connection manager, and other logging facilities under Microsoft Windows were primarily used, while some of the experimentation used the Linux-WiMAX device driver for Intel WiMAX 6250 under the Linux operating system. The 4G WiMAX software used in this work is provided in Table 4.3.

Type	Model Number	Manufacturer	Comments
BS	WAP 400	Motorola	- Air-interface: IEEE 802.16e - Channel Bandwidth: 5MHz & 10MHz - Sectors: 3 - Duplex Mode: TDD - Frequency Reuse: 1 & 3
MS	Centrino Advanced-N + WiMAX 6250	Intel	- Air-interface: IEEE 802.16e
MS	USBw 100 (USBw 25100) [33]	Motorola	- Air-interface: IEEE 802.16e - Beceem Chipset (PHY D1000703, MAC 05-02-2976, BB Chip ID BECE0310, RF Chip ID 6600)

Table 4.2: Specification of WiMAX Base Station (BS) and Mobile Station (MSs) Hardware

Name	Developed By	Operating System	Supported Hardware	Software Components	Version
Linux WiMAX [34]	Intel & open source community	Linux (kernel $\geq$ 2.6.35)	Intel WiMAX / Wi-Fi Link 5x50 and 6250	firmware, driver, network service, and test configuration utility	Driver 1.5 i2400m firmware 1.5.0 WiMAX tools 1.4.4 WiMAX Network Service 1.5.1
Beceem	Beceem	Windows XP, Vista & 7	MSs using Beceem Chip-sets (e.g. Motorola w100)	firmware, driver, connection manager, and logging facilities	Driver 5.2.135.0 Firmware 5.2.2976 Library 05.02.0093

Table 4.3: Specifications of WiMAX Mobile Station (MSs) Software

Chapter 5

Dominant LOS Path (DLP) Condition Detection in 4G Networks

Major sources of errors in geometric techniques are the absence of the dominant LOS path (DLP) and the multipath. In this dissertation, when a MS has a dominant LOS path from a BS, it is said that the MS is under *DLP* (condition) with regard to that BS. When a MS does not have a dominant LOS path from a BS, the MS is said to be under *non-DLP/NDLP* (condition). When the MS can see multiple BSs, the MS can 1) discard the signals from the BSs under the non-DLP condition or 2) assign a weight to each BS according to the non-DLP condition in using the geometric techniques.

In ROLAX, a set of DLP condition detection techniques are designed by using the measurements available from the standard radio resource measurements such as RSSI, CINR, and RTD. We note that, compared to using a single feature, appropriately combined multiple features lead to a very accurate DLP condition detection. Machine learning techniques are used to detect the DLP condition from multiple features.

In Section 5.1, how the range between an MS and a BS under DLP condition can be estimated is demonstrated. In Section 5.2, a number of DLP detection techniques are presented. In Section 5.3, machine learning techniques used in combining multiple features to detect the DLP condition more accurately are described. In

each section, experimental results are provided.

5.1 Range Determination under DLP

In 4G WiMAX, RSSI can be passively measured by detecting preambles on the downlink, while measuring RTD involves dealing with the time advance, and the RTD measurement can be done only with regard to the serving or anchor BSs. RTD value is usually estimated through the ranging procedure, so the first value can be observed only after the initial ranging. Details about RTD and time advance are provided in Appendix A.3.

The range can be determined (1) by observing the change of power scale over distance, or (2) by observing the round trip time. In multipath, the received envelope is composed of the signal envelops over different paths, each with different path loss, phase shift, and Doppler spread. Since RSSI can be regarded as the power of the received envelope, the multipath can create either constructive or destructive effect on RSSI. In a non-DLP condition, it is very likely that the signals over the NLOS path are attenuated due to the presence of obstacles and the shadowing.

Since the LOS path is always shorter than NLOS paths, the RTD measured over the NLOS path is always larger than that over the LOS path. In order to estimate the range by RTD, it is desirable to detect the signal over the LOS path rather than signals over NLOS paths, but, in general, the receiver detects the strongest path rather than the shortest path. When the LOS path is not dominant, the detection of the LOS path is difficult due to the following reasons. First, there may not neces-

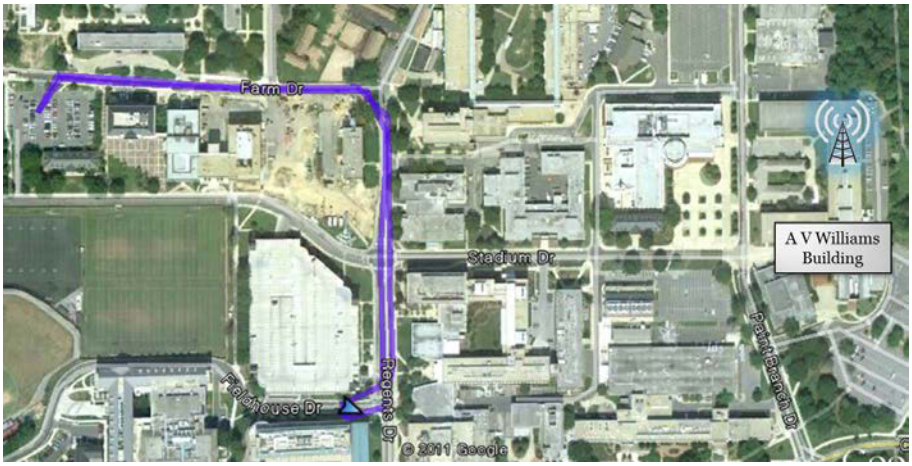
sarily be a LOS path when the attenuation by the obstruction is high. The signals over LOS may have worse signal quality than the signals over NLOS paths. In addition, the signals over the LOS path may have power below the receiver sensitivity of the device. Second, if the received time interval between the signal over LOS path and the signal of NLOS path is smaller than the symbol duration, the signals over the LOS and NLOS paths are overlapped so that they are not resolvable from each other. In this case, the RTD estimate is likely to increase. In addition, the correlation sidelobes appearing between the correlation peaks make the detection of the first path harder [35]. For the ranging purpose, it would be ideal to have a DLP condition.

A set of experiments were done under both DLP condition and non-DLP condition. In sum, both RSSI and RTD have the potential to be used in estimating the range under the DLP. Particularly with the RTD, a range estimation error in the order of 30 meters was achieved in the live 4G testbed. At the absence of DLP, the linear relation, between RSSI and logarithmic distance and between RTD and distance, is lost. RSSI is attenuated, and RTD contains an additional positive bias.

An area, with some light density of trees and houses, was chosen to see the signal characteristics under the DLP (Figure 5.1(a)). In addition, an area surrounded with buildings was chosen to see the signal characteristics under the non-DLP (Figure 5.1(b)).

The test was performed using wardriving, and the measurements were retrieved from the periodic neighbor report, which is generated once every three seconds. The wardriving vehicle was moving around 5–20 miles/hour, which leads to 6–27 meters

(a) DLP between the BS and the MS (Lakeland Rd, College Park, Maryland)



(b) Non-DLP between the BS and the MS (University of Maryland, College Park, Maryland)

Figure 5.1: Locations for RSSI/RTD Measurements for Range Estimation

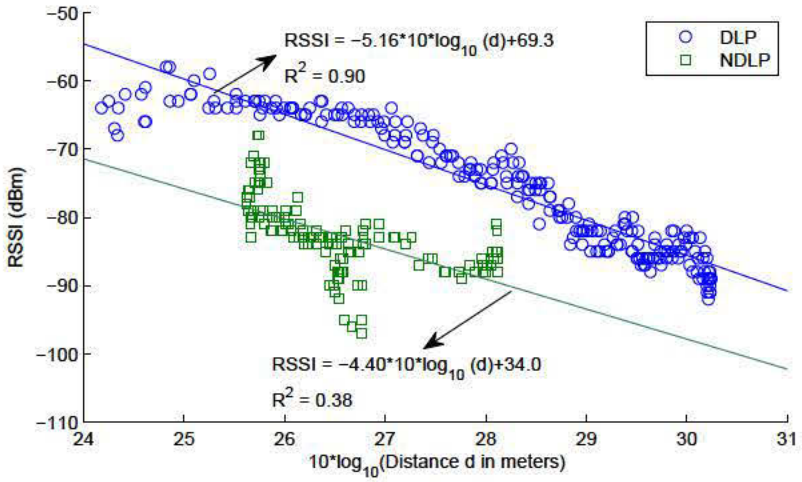
movement over three seconds duration. The measured RSSI and RTD values can be regarded as measurements over the last three seconds duration.[1] In addition, quantization of the RTD and the RSSI values limits the best resolution of the measurement and the estimated range (More details are provided in Appendix A.1 and Appendix A.3).

The RSSI measurement result is provided in Figure 5.2(a) and the RTD measurement result is in Figure 5.2(b). Note that the RTD value is a logical value rather than a physical round flight trip time.

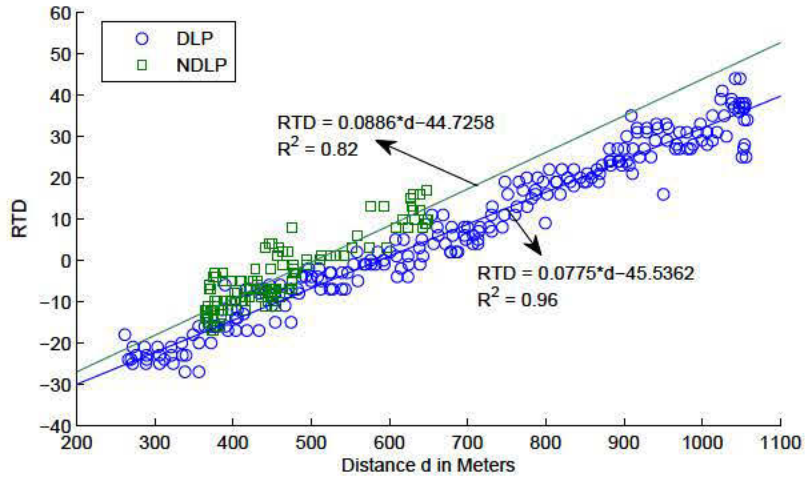
In Figure 5.2(a), the RSSI is not linear with the logarithmic distance under the non-DLP condition. In addition to its non-linearity, the value is attenuated by about 15dB because of the attenuation caused by the building structures. As seen in Figure 5.2(b), under the non-DLP condition, the linearity of RTD over the distance is weaker (R squared value of 0.82 as opposed to 0.96 under the DLP), and the RTD value is higher. It is consistent with the fact that the LOS path is always shorter than other paths, which result in a longer time of flight over the NLOS path. When the logical value of RTD is zero, the difference between estimated distances using the linear regression under the DLP and the non-DLP is about 80 meters. RSSI and RTD are correlated higher under the DLP than under non-DLP. The correlation coefficient ρ between RSSI and RTD is -0.95 under DLP while ρ is -0.58 under non-DLP (See Figure 5.9).

The range estimation errors, under DLP and non-DLP, are summarized in Table 5.1. The range estimation error by the linear regression significantly increases

[1]It is not known whether a one-shot value or an average value is reported.



(a) RSSI



(b) RTD

Figure 5.2: Measurements under DLP and Non-DLP and Their Linear Regression

Table 5.1: Range Estimation Errors under DLP and Non-DLP

Channel	Observed Parameter	Mean Error (meters)	Median Error (meters)	Is Gaussian?
DLP	RSSI	56.9	50.5	Yes (p=0.20)
	RTD	35.5	29.3	Yes (p=0.20)
NDLP	RSSI	410.8	407.2	No (p=0)
	RTD	81.5	70.8	No (p=0)

in non-DLP. The range estimation by RTD provides a better accuracy than the estimation by RSSI. The error distribution under DLP is Gaussian while it is not in non-DLP.

5.2 DLP Detection Techniques

Since the range estimation is affected by non-DLP condition, the MS has to determine whether it is under DLP or not by observing signals. In case of GSM, the NLOS error contributes positioning errors by 500–700 meters [36].

Many previous works assume the availability of signal envelope or other physical measurements, which is not usually available at the device driver interface level. In [37], Channel Impulse Response (CIR) from the received signals are used to extract features such as the kurtosis, the mean excess delay spread, and the root mean square delay spread to identify NLOS condition. In [38], the NLOS condition is determined by estimating the Rician K-factor from the CIR. Other NLOS condition

detection technique such as Wylie-Holtzman [39] relies on the range estimation to determine the NLOS condition.

Since most of the 4G WiMAX software provides limited access to the information at the physical layer, ROLAX considers how the information available at *MAC Service Access Point (MAC SAP)* or *PHY Service Access Point (PHY SAP)*, which is usually implemented at the firmware or device driver level, can be used to detect the DLP condition.

5.2.1 Wylie-Holtzman Technique Applied to RSSI

One of the famous NLOS detection techniques is Wylie-Holtzman technique [39]. In this technique, a priori knowledge of system noise is required. Under NLOS, NLOS error is assumed to be added to the range estimation, and it is assumed that the system noise is uncorrelated with the NLOS error. The range estimation $r_m(t_i)$ with regard to the BS m at time t_i is modeled by (5.1).

$$r_m(t_i) = L_m(t_i) + n_m(t_i) + NLOS_m(t_i) \tag{5.1}$$

where $L_m(t_i)$ is the true distance between the MS and the BS m at time t_i, $n_m(t_i)$ is the system measurement noise and $NLOS_m(t_i)$ is the NLOS error.

Because $n_m(t_i)$ and $NLOS_m(t_i)$ are assumed to be uncorrelated with each other, the variance at the presence of NLOS error is larger than the variance at the absence of NLOS error. In this technique, the distance estimations are smoothed and represented by $s_m(t_i) = \sum_{n=0}^{N-1} \hat{a}_m(n) t_i^n$ where $\{\hat{a}_m(n)\}_{n=0}^{N-1}$ is decided by the least

square technique. The variance is computed by $\hat{\sigma}_m = \sqrt{\frac{1}{K}\sum_{i=0}^{K-1}(s_m(t_i) - r_m(t_i))^2}$ where K is the number of measurements. Hypothesis testing is performed by calculating the difference between the known variance under LOS, σ_m, and $\hat{\sigma}_m$. The null hypothesis H_0 is $\hat{\sigma}_m = \sigma_m$, and the alternative hypothesis H_1 is $\hat{\sigma}_m > \sigma_m$. H_1 is accepted when $\hat{\sigma}_m > \kappa\sigma_m$.

In this work, the similar logic used in Wylie-Holtzman is used to determine the LOS condition by using the RSSI or RTD at a fixed location. When the RSSI is measured at the terminal, the RSSI with regard to signals from a BS under LOS condition ($RSSI_{LOS}$) can be given by (5.2).

$$RSSI_{LOS} = P_0 - 10\alpha \cdot log(d) - C + N_m \tag{5.2}$$

where P_0 is the transmit power of the BS, α is the path loss exponent, d is the distance between the terminal and the BS, C is the system loss, and N_m is the measurement noise, which is typically modeled by a Gaussian distribution.

Under the NLOS condition, the received signal strength, $RSSI_{NLOS}$ can be modeled by the addition of $RSSI_{LOS}$ and the NLOS noise N_{NLOS}. This is provided in (5.3).

$$RSSI_{NLOS} = RSSI_{LOS} + N_{NLOS} = P_0 - 10\alpha \cdot log(d) - C + N_m + N_{NLOS} \tag{5.3}$$

If the measurement noise, N_m, is uncorrelated with the NLOS Noise, N_{NLOS}, then the variance of $RSSI_{LOS}$ is always smaller than the variance of the $RSSI_{NLOS}$. It is shown in (5.4).

$$Var[RSSI_{LOS}] = Var[N_m] \leq Var[RSSI_{NLOS}] = Var[N_m] + Var[N_{NLOS}] \quad (5.4)$$

In ROLAX, the variance of the measurements is used to detect the channel condition with regard to a BS from a MS.

5.2.2 Level Crossing Rate

Level crossing rate is a measure of how the channel is rapidly changing due to the mobility of the terminals or surroundings. It is defined as $\xi_{LCR} = \frac{\eta_{NC}}{T}$, where η_{NC} is the number of crossings of the specified signal level, and T is the measurement duration. In [30], the observed level crossing rate is compared to the known value under a variety of Rician K factor to detect the Non-DLP condition. Since LCR increases as K-factor decreases, LCR can be used to detect the NLOS contribution to the fast fading.

In ROLAX, rather than calculating the LCR of the received signal envelop, standard radio resource measurements are used. For instance, when the RSSI is measure, the RSSI Level Crossing Rate (RLCR) is calculated to detect DLP condition. A priori knowledge of RLCR under DLP is needed. RLCR is expected to imply the Non-DLP condition because the coherence time is shorter under the stronger multipath condition (larger delay spread), and accordingly the RSSI values are likely to change more frequently. RLCR is defined in (5.5).

$$\xi_{RLCR} = max_{th} \frac{\eta_{th,RSSI}}{T} \quad (5.5)$$

where $\eta_{th,RSSI}$ is the level crossing rate at level th, and T is the measurement duration.

5.2.3 Correlation between RSSI and RTD

One of the observations made in Section 5.1 is that the correlation between RTD and RSSI is higher under DLP than under non-DLP. By calculating the correlation coefficient between the RSSI and the RTD, the DLP condition can be detected. It can be used to determine the DLP condition for a relatively large area since the variation of the RSSI and RTD has to be observed to identify the condition.

5.2.4 Kullback-Leibler Divergence

The distribution of the RSSI under non-DLP varies a lot over the locations. Therefore, the likelihood ratio test cannot be used because the conditional pdf of the measurement under non-DLP cannot be easily obtained while the conditional pdf under DLP may be available. Therefore, a nonparametric method using Kullback-Leibler divergence is used in discriminating DLP and NDLP. The Kullback-Leibler divergence is provided in (5.6).

$$D_{\mathrm{KL}}(P_{DLP}\|P) = \sum_i P_{DLP}(i) \log \frac{P_{DLP}(i)}{P(i)} \tag{5.6}$$

where $P(i)$ is the estimated probability to have a value i and $P_{DLP}(i)$ is the probability known a priori to have a value i under DLP condition.

Since Kullback-Leibler divergence can be infinity if $P(i) = 0$ for some i such

that $P_{DLP}(i) > 0$, the empirical pdf (Kaplan-Meier distribution) cannot be used in calculating the divergence value. Thus, the density is estimated by using kernel density estimation that is provided in (5.7) [40].

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \quad = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big) \tag{5.7}$$

where $K(\bullet)$ is the kernel - a symmetric but not necessarily positive function that integrates to one, and $h > 0$ is a smoothing parameter called the bandwidth.

The RSSIs are measured and quantized per antenna, and combined RSSI is calculated by using the formula like (5.8). Thus, the intervals between the possible combined RSSI values are not uniform.

$$RSSI(dBm) = 10log(\frac{P_1 + P_2}{1mW}) = 10log(10^{RSSI_1/10} + 10^{RSSI_2/10}) \tag{5.8}$$

where P_1 is the RSSI in Watt measured at the first antenna, P_2 is the RSSI in Watt measured at the second antenna, $RSSI_1$ is the RSSI in dBm measured at the first antenna, and $RSSI_2$ is the RSSI in dBm measured at the second antenna.

The bandwidth h was chosen so that the underlying probability density can be estimated while the artifacts can be smoothed out. If the bandwidth h is too small, peaks at the possible combined RSSI values are noted (Figure 5.3).

It has been noted a similar non-parametric test using Kullback-Leibler divergence was developed in [41], too. In [41], a general NLOS identification using the non-parametric test is provided. In this work, we specifically develop a non-parametric test to use the combined RSSI as the measurement in 4G networks.

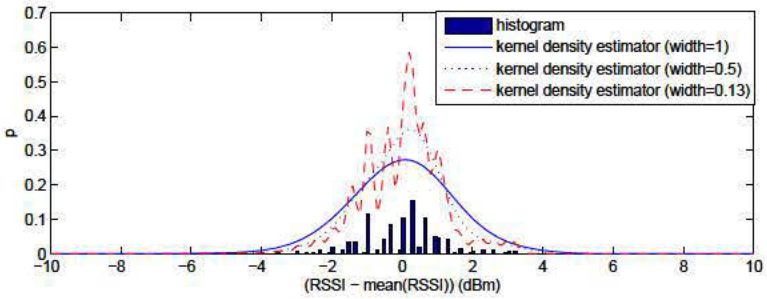
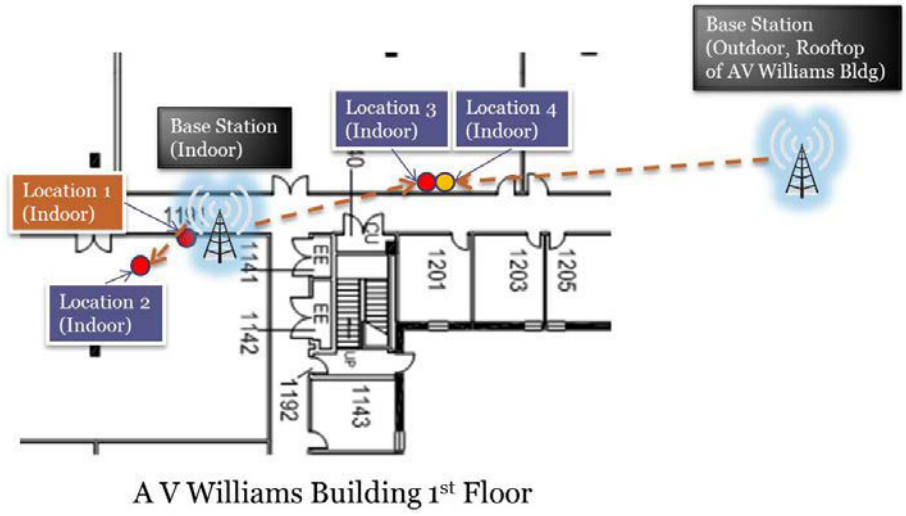


Figure 5.3: RSSI Histogram and Kernel Density Estimators with Different Width Values (Location 6 in Figure 5.4(b))

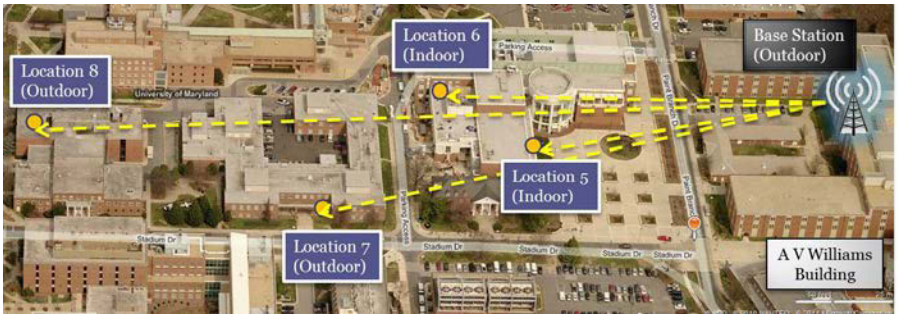
5.2.5 Experimental Result: DLP Detection Using Single Feature

The techniques designed in Section 5.2 are evaluated in the live 4G network. A number of indoor and outdoor locations, with various channel environments between the MS and the BS, were selected to gain a priori knowledge of RSSIs under DLP and NDLP. Four indoor locations (location 1–4, Figure 5.4(a)) and four outdoor locations (location 5–8, Figure 5.4(b)) were chosen. At location 1, a dominant LOS path existed between the BS and the MS. At location 2, some indoor obstacles (metal shelf and desks) were located between the MS and the BS within the same room, but there existed a dominant LOS path. At locations 1–3, the MS was connected to an indoor BS. At locations 4–8, the MS was connected to an outdoor BS.

The variance of the RSSI and RSSI Level Crossing Rate (RLCR) is provided in Figure 5.6. This experimental results show that both RSSI variance and RLCR can be used in detecting the DLP condition. At location 1 and 2, where a dominant LOS path exists, both RLCR and Variance[RSSI] are smaller than the values measured at other locations. However, at location 3, it was observed that the RLCR is much



(a) Indoor



(b) Outdoor

Figure 5.4: Locations for DLP Detection Experiment

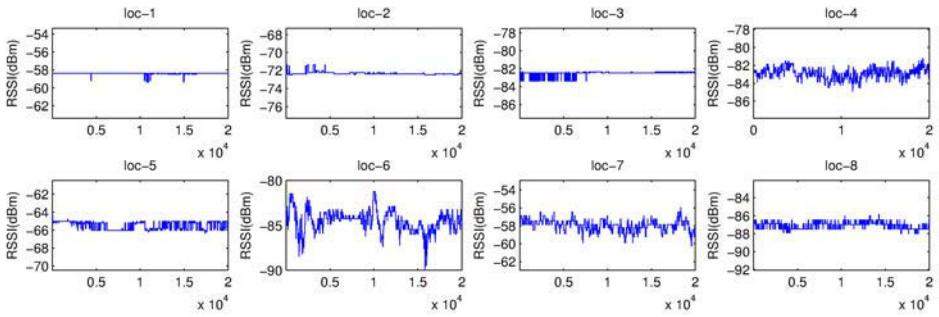


Figure 5.5: RSSI Values over Time under Various Locations

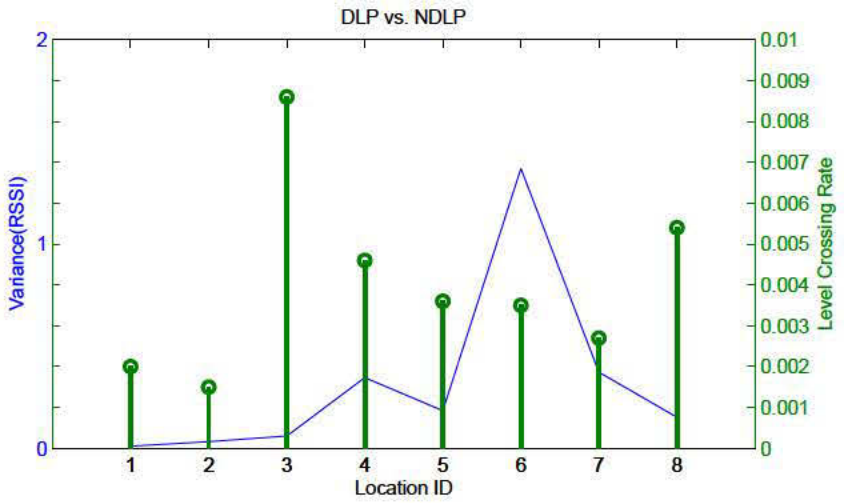


Figure 5.6: RSSI Variance and RSSI Level Crossing Rate under DLP and Non-DLP

higher than the values measured at the DLP locations while the Variance[RSSI] is slightly higher than the values at the DLP locations. This implies that combining the test statistics (features) would be beneficial to detect the DLP condition. In Section 5.3, it is demonstrated how features including RLCR and Variance[RSSI] can be combined and how they are effective in detecting the DLP condition.

The estimated RSSI kernel densities with width 0.5 at locations from 1 to 8 are provided in Figure 5.7. The Kullback-Leibler divergence values, calculated with regard to the estimated density at the location 1 (DLP), are provided in Figure 5.8. It shows that the divergence value at location 2 is notably smaller than the values at other locations. This shows the feasibility of using Kullback-Leibler divergence in detecting LOS/NLOS condition.

In order to evaluate the RSSI-RTD correlation technique described in Section 5.2.3, the data collected in Section 5.1 is used. As seen from the Figure 5.9, the correlation between RSSI and RTD is higher under DLP condition than under NDLP

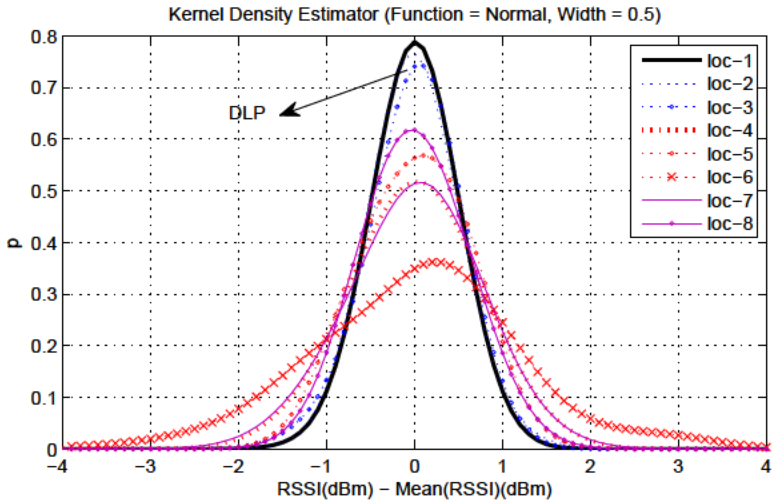


Figure 5.7: RSSI Kernel Density Estimator with Width 0.5

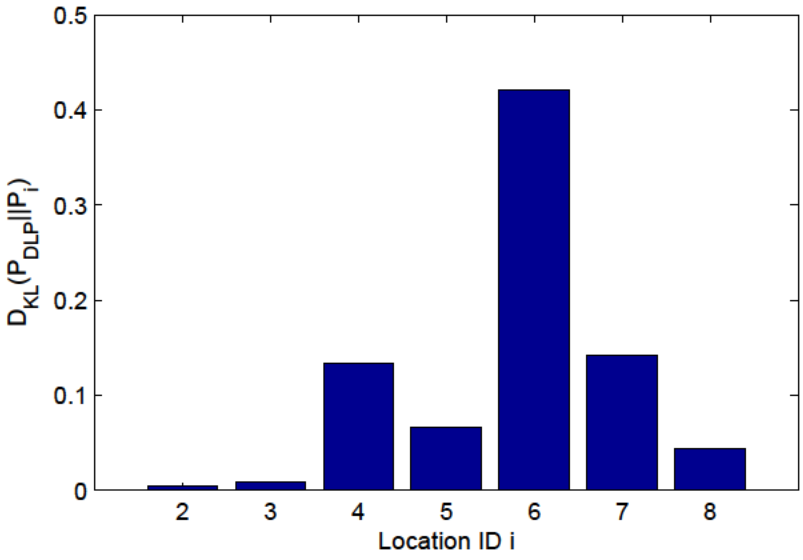


Figure 5.8: Kullback-Leibler Divergence between RSSI Density under DLP and Density at Other Locations

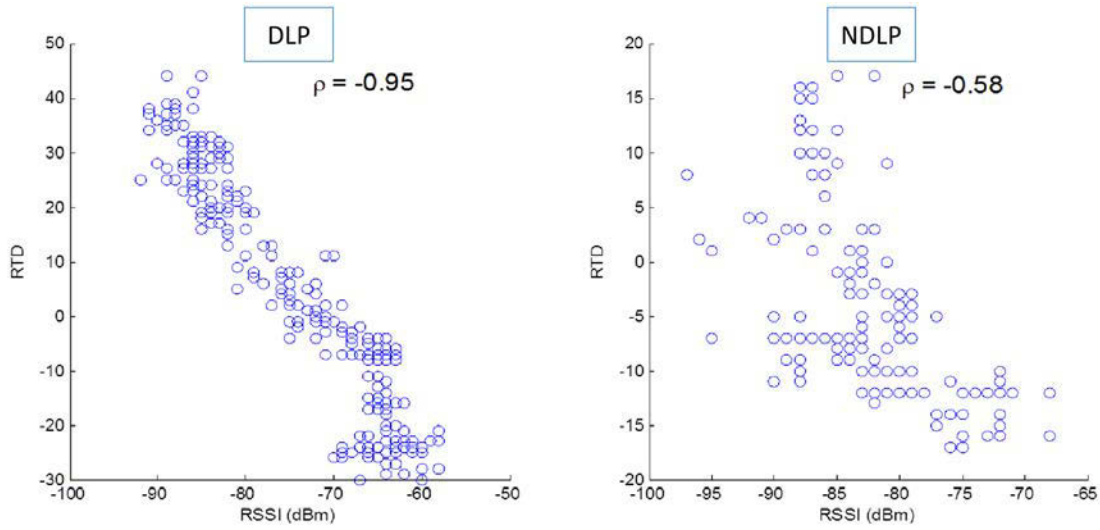
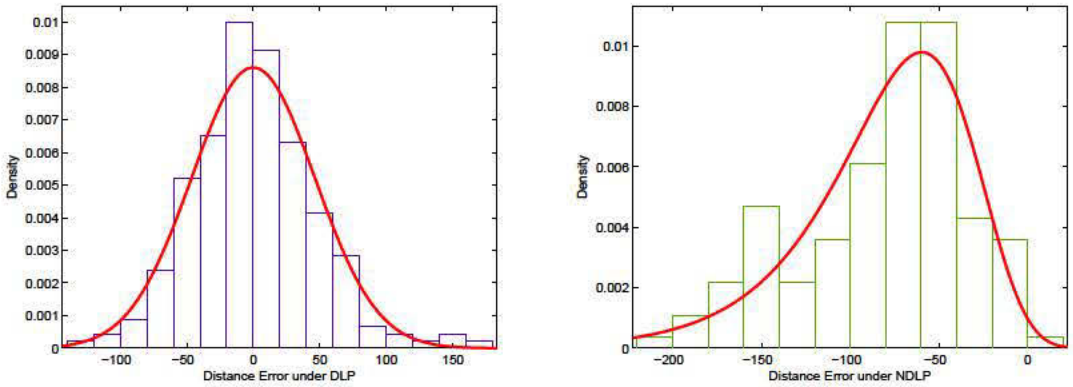


Figure 5.9: Correlation Between RSSI and RTD

condition.

5.2.6 Parametric Distribution Modeling for Errors under Non-DLP

The DLP detection experiment also showed another interesting result: the distribution of distance error was different under DLP and NDLP. The DLP error was normally distributed confirming the assumption made for the Gaussian measurement error (Figure 5.10(a)), but the error under NDLP is not normally distributed. In previous research, Gaussian, exponential, log-normal, and mixture of exponential and Gaussian have been used for modeling the ranging error in NLOS [42]. If the NLOS error is additive to measurement error, the error under NLOS, $r_{m,err}$, can be regarded as the summation of measurement errors n_m and NLOS errors $NLOS_m$. According to the data collected in Section 5.1, $r_{m,err}$ is observed to follow the *Extreme Value* distribution (Figure 5.10(b)) with our experimental data. The testing



(a) DLP (Gaussian Distribution) (b) NDLP (Extreme Value Distribution)

Figure 5.10: Distance Error under DLP and NDLP (Estimated by RTD)

was done by Lilliefors test. The pdf of extreme value distribution, with location parameter μ and scale parameter σ, is given in (5.9).

$$f(x) = \frac{1}{\sigma} e^{\left(\frac{x-\mu}{\sigma}\right)} e^{\left(-e^{\left(\frac{x-\mu}{\sigma}\right)}\right)} \tag{5.9}$$

Most of the errors were negative value confirming the fact that the NLOS path is always longer than the LOS path.

The *Weibull* distribution is a limited case of the Extreme Value distribution where x is bounded below or above [43]. If the measurement error can be assumed to be independent from the $NLOS_m$, then $r_{m,err}$ is a convolution between a Gaussian distribution and an unknown distribution corresponding to $NLOS_m$. There is a possibility to model $r_{m,err}$ as a convolution of the Weibull and the Gaussian distribution.

5.3 DLP Detection Techniques Using Multiple Features

As observed in the experiments in Section 5.2.5, the detection performance of each feature (test statistics) varies for each location. Thus, in this section, how to combine multiple features together to detect the DLP condition is presented. In addition, more features that can be used in detecting the DLP condition are introduced. Those features, in addition to some features discussed in the Section 5.2, are used as the input features for the pattern matching engine to decide a DLP condition. A number of pattern matching techniques including ANN are evaluated.

5.3.1 Underlying Measurements Used in Generating Multiple Features

Since the DLP detection has to be done within a reasonable time limit, the measurement duration for each BS to detect the DLP condition is assumed to be limited to one second. The following underlying measurements are made to generate features within one second. In 4G WiMAX, RSSI and CINR can be measured for each downlink (DL) preamble from the MS.

- Received Signal Strength Indication (RSSI) in dBm for each antenna ($RSSI_1$ and $RSSI_2$)
- Carrier to Interference Noise Ratio (CINR) in dB

Combined RSSI is calculated as in (5.8). An additional feature, RSSI difference (denoted by $RSSI_d$) is calculated by (5.10).

$$RSSI_d(dB) = |RSSI_1 - RSSI_2| \tag{5.10}$$

5.3.2 Feature Extraction from Underlying Measurements

Time features and spectral features are generated by processing each 200 consecutive underlying measurements ($RSSI$, $CINR$, and $RSSI_d$). The variance, the Level Crossing Rate(LCR), the spectral centroid, the spectral roll-off and the spectral flux of the underlying measurements are evaluated. The mean of $RSSI_d$ is also evaluated. The spectral centroid (SC), the spectral roll-off (SR) and the spectral flux (SF) are the features commonly used in speech recognition [44].

Let $x_i(n), n = 0, 1, ..., N-1$ be the underlying measurement (e.g. RSSI) samples of ith frame, where a frame is composed of N consecutive samples. A frame is not overlapped with other frames. N is 200 since a feature is generated for each 200 consecutive preambles received for one second duration. Let $X_i(m), m = 0, 1, ..., N-1$ be the corresponding DFT (Discrete Fourier Transform) coefficients.

The spectral centroid (SC) is the spectral center of the mass of the measurements. It is the weighted mean of the frequency components. The spectral centroid (SC) of ith frame is defined in the (5.11).

$$SC(i) = \frac{\sum_{m=0}^{N-1} m\,|X_i(m)|}{\sum_{m=0}^{N-1} |X_i(m)|} \tag{5.11}$$

The spectral roll-off (SR) is a measure of the skewness of the spectral distribution. It is the frequency sample, $m_c^R(i)$, below which more than $c\%$ of the magnitude

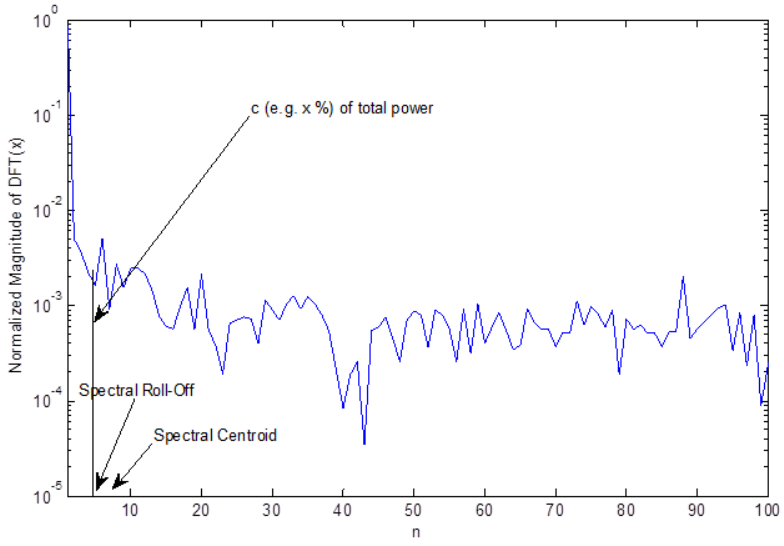


Figure 5.11: Example: Spectral Centroid and Spectral Roll-off of RSSI measurements

distribution of the DFT coefficients is concentrated [44]. For this frequency sample, the following is true:

$$\sum_{m=0}^{m_c^R(i)} |X_i(m)| >= \frac{c}{100} \sum_{m=0}^{N-1} |X_i(m)| \tag{5.12}$$

where c is chosen to be 85 in this study.

An example of spectral centroid and spectral roll-off calculated from RSSI measurements is provided in Figure 5.11.

Spectral flux (SF) is a measure of spectral change over time, and it is defined in (5.13). Each two hundreds consecutive measurements are divided into two sub-frames, each with one hundred measurement samples. The spectral flux of the ith frame is defined as in (5.13).

$$SF(i) = \sum_{m=0}^{N/2-1} (N_{i,1}(m) - N_{i,2}(m))^2 \tag{5.13}$$

where $N_{i,j}(m)$ is the normalized magnitude of the respective DFT coefficient of the jth subframe of the ith frame.

The reasons for choosing each feature are as follows:

- **Variance of RSSI / CINR**: In Section 5.2, it was discussed and demonstrated that the variance of the RSSI can be used in detecting DLP. Since CINR and RSSI are highly correlated as demonstrated in Section 6.4, the variance of CINR can used as the feature to detect the DLP condition.
- **Mean of RSSI difference**: When only the LOS path exists, the signal dissipation over the distance is a function of the distance from the transmitter, and it does not change significantly over a small distance. At the presence of the mobility on the channel and the stronger scattered paths, it is expected to see the signal magnitude variation over a small distance (smaller coherence distance). Thus, the RSSI difference between two antennas can be used as a feature to detect the DLP.
- **Level Crossing Rate of RSSI / CINR**: In Section 5.2.2, it is demonstrated that the Level Crossing Rate can be used in detecting DLP condition.
- **Spectral Centroid and Spectral Roll-off of RSSI / CINR**: The presence of the multi-path creates a changing environment that dissipates the signal energy in time [45]. It is expected to see more frequent change of RSSI and CINR under NDLP, which induces higher SC and SR.

- **Spectral Flux of RSSI / CINR**: It is expected that the underlying measurements have a higher spectral change over time as the channel condition changes over time.

Thirteen features were evaluated and ranked by three different criterion. The list of evaluated features is as follows:

- Standard deviation of the $RSSI$ ($Std[RSSI]$)
- Standard deviation of the $CINR$ ($Std[CINR]$)
- Standard deviation of the $RSSI_d$ ($Std[RSSI_d]$)
- Mean of $RSSI_d$ ($Mean[RSSI_d]$)
- Level Crossing Rate of the $RSSI$ (LCR_{RSSI})
- Level Crossing Rate of the $CINR$ (LCR_{CINR})
- Spectral Centroid of the $RSSI$ (SC_{RSSI})
- Spectral Centroid of the $CINR$ (SC_{CINR})
- Spectral Roll-off of the $RSSI$ (SR_{RSSI})
- Spectral Roll-off of the $CINR$ (SR_{CINR})
- Spectral Roll-off of the $RSSI_d$ (SR_{RSSI_d})
- Spectral Flux of the $RSSI$ (SF_{RSSI})
- Spectral Flux of the $CINR$ (SF_{CINR})

The processing between the underlying measurements and the extracted features is described in Figure 5.12.

The features are selected by three feature evaluators. The statistics used by each evaluators are Chi-square statistics, information gain, and symmetrical uncer-

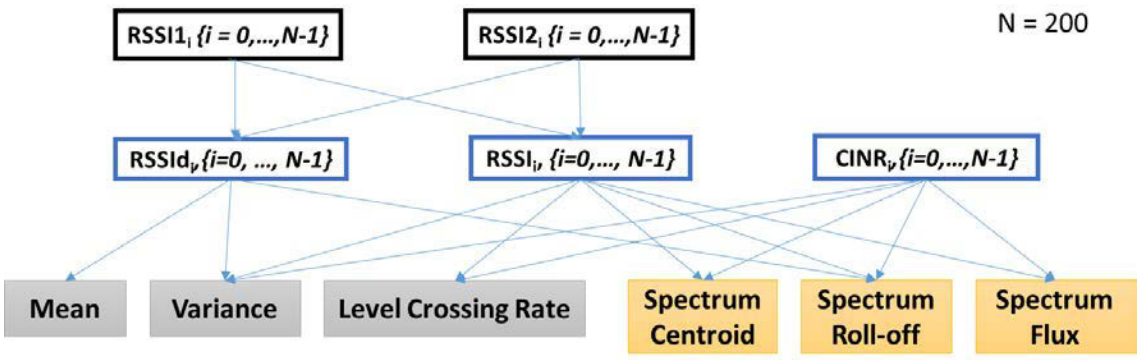


Figure 5.12: Processing between Underlying Measurements and Features

tainty. The statistics are calculated with respect to DLP/NDLP condition. All numeric attributes are grouped into bins by discretizing the numeric values using Fayyad and Irani's MDL (Minimum Discretization Length) method, which uses the information gain to find the best bins [46].

- **Chi-square Evaluator**: Chi-square statistic is calculated by comparing the observed count and the expected count under the assumption of the no association between the feature and the DLP/NDLP condition for each bin. The Chi-square statistics are defined as follows: $\chi^2 = \sum_{i=0}^{N_{bins}-1} \frac{(O_i - E_i)^2}{E_i}$ where N_{bins} is the number of bins, O_i is the observed count in the ith bin, and E_i is the expected count in the ith bin under the assumption of the no association.
- **Information Gain Evaluator**: Information gain can be regarded as the change in the entropy from a prior state without some information to a state that takes some information. It is also known as mutual information. Information gain is defined as follows: $IG(Class, Feature) = H(Class) - H(Class|Feature)$ where $Class$ is either DLP or NDLP, and H is the en-

tropy.

- **Symmetrical Uncertainty Evaluator**: Symmetrical uncertainty is a non-linear estimation of correlation between the feature and the DLP/NDLP condition. Symmetric uncertainty is defined as follows: $SU(Class, Feature) = 2\frac{H(Class)-H(Class|Feature)}{H(Class)+H(Feature)}$ where $Class$ is either DLP or NDLP, and H is the entropy.

The ranking of the features by each feature selector is provided in the Table 5.2. Since the top eight features are the same whichever evaluator is used, they are chosen in the DLP/NDLP detection. The selected features are as follows: SR_{CINR}, SC_{CINR}, $Std[CINR]$, LCR_{CINR}, SC_{RSSI}, SF_{RSSI}, $Mean[RSSI_d]$, and SF_{CINR}. The distributions of the selected features are provided in the Figure 5.13.

5.3.3 Experimental Results

The downlink signals from a BS in the MAXWell 4G network were measured. The experimental settings are as follows:

- Location: Rooftop of AV Williams Building, College Park, MD
- Number of DLP Locations: 12 (5 locations in the first test. 7 locations in the second test.)
- Number of NDLP Locations: 12 (5 locations in the first test. 7 locations in the second test.)
- 4G Network: MAXWell 4G Network (Motorola WAP 400 BS)
- Underlying Measurements: RSSI (per each antenna of the MS); CINR (Mea-

Table 5.2: Feature Ranking for DLP Detection

Ranking	Evaluator					
	Chi-Square Statistic		Information Gain		Symmetric Uncertainty	
	Score	Feature	Score	Feature	Score	Feature
1	3829	SR_{CINR}	0.5694	SR_{CINR}	0.5694	SR_{CINR}
2	3595	SC_{CINR}	0.5383	SC_{CINR}	0.5383	SC_{CINR}
3	3236	$Std[CINR]$	0.4847	$Std[CINR]$	0.4847	$Std[CINR]$
4	2285	LCR_{CINR}	0.345	LCR_{CINR}	0.345	LCR_{CINR}
5	1756	SC_{RSSI}	0.2446	SC_{RSSI}	0.2446	SC_{RSSI}
6	1417	SF_{RSSI}	0.1909	SF_{RSSI}	0.1909	SF_{RSSI}
7	1394	$Mean[RSSI_d]$	0.1907	$Mean[RSSI_d]$	0.1907	$Mean[RSSI_d]$
8	1179	SF_{CINR}	0.1753	SF_{CINR}	0.1753	SF_{CINR}
9	707.4	SR_{RSSI_d}	0.0871	SR_{RSSI_d}	0.0871	SR_{RSSI_d}
10	652.7	$Std[RSSI]$	0.083	$Std[RSSI_d]$	0.083	$Std[RSSI_d]$
11	652.1	$Std[RSSI_d]$	0.0816	$Std[RSSI]$	0.0816	$Std[RSSI]$
12	364.7	LCR_{RSSI}	0.045	LCR_{RSSI}	0.045	LCR_{RSSI}
13	63.77	SR_{RSSI}	0.01	SR_{RSSI}	0.01	SR_{RSSI}

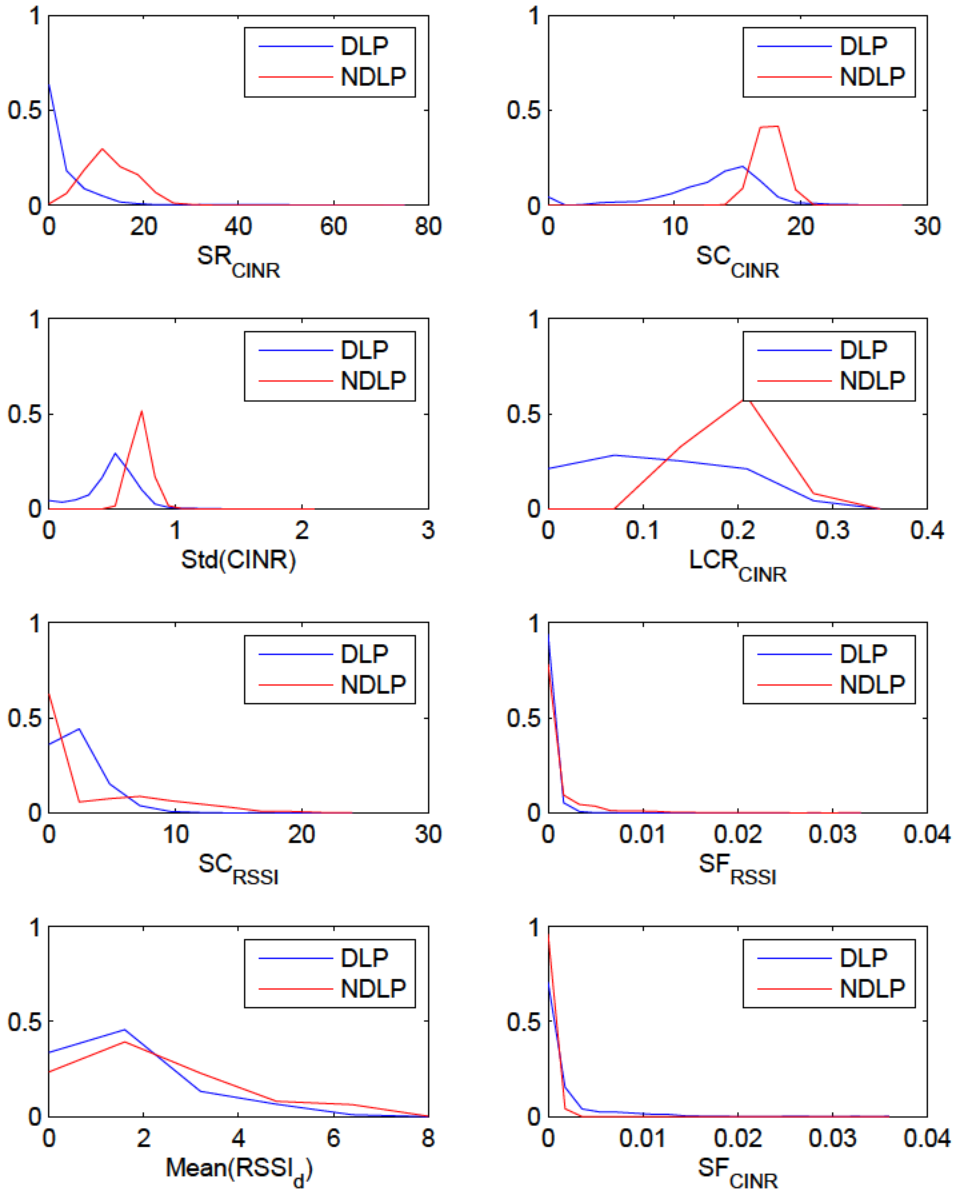


Figure 5.13: Distribution of Features Selected for DLP/NDLP Detection

(a) Test Locations in First DLP/NDLP Test (bird's eyes view)

(b) Test Locations in the First DLP/NDLP Test (zoomed view)

Figure 5.14: Test Locations in the First DLP/NDLP Test

sured on the preambles.)

- Number of Measurements per Each Location: 50,000 samples (250 seconds)
- Number of Components in a Feature Vector: 8 (See Section 5.3.2 for the list)
- Total Number of Feature Vectors Generated per Each Location: 250 (One feature vector generated per each second.)

The measurements were done at the same place for two days. The locations chosen as the DLP locations as well as the NDLP locations are provided in Figure 5.14 and Figure 5.15.

(a) Test Locations in Second DLP/NDLP Test (bird's eyes view)

(b) Test Locations in the Second DLP/NDLP Test (zoomed view)

Figure 5.15: Test Locations in the Second DLP/NDLP Test

5.3.3.1 DLP Detection by Using Neural Networks

A number of pattern matching techniques were evaluated, and one of the pattern matching techniques that provided a better detection performance was artificial neural networks. The configurations of the artificial neural networks used in DLP/NDLP detection are as follows:

- Nodes in the neural networks: Sigmoid.
- Algorithm: Back propagation algorithm.
- Number of Input Nodes: 8
- Number of Hidden Layers: 1
- Number of Neurons: 5

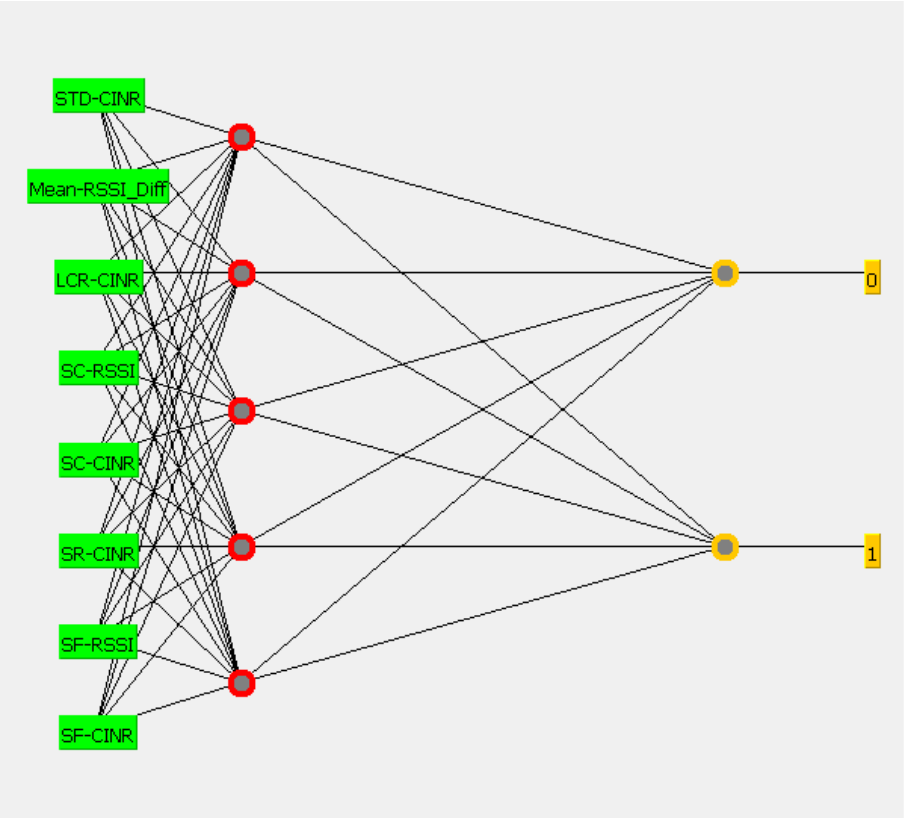


Figure 5.16: Diagram of Neural Networks for DLP/NDLP Detection

- Learning Rate (The amount the weights are updated): 0.3
- Momentum (Applied to the weights during updating): 0.2
- Attribute Normalization: Yes
- Training Time (The number of epochs to train through): 500

The diagram of this neural network is provided in Figure 5.16.

The performance of the artificial neural networks is evaluated with the stratified cross-validation with ten folds. In each iteration (out of ten iterations), nine folds are used for training while one fold is reserved for testing. In each fold, the number of DLP instances and NDLP instances are roughly about the same. The accuracy is calculated by averaging the values over iterations. The result is pro-

vided in the Table 5.3. With all eight features, the DLP condition can be accurately determined with 92%.

5.3.3.2 DLP Detection by Using K-Nearest Neighbor (K-NN)

An instance-based learning, K-NN, achieved a better performance than the neural networks. In K-NN, rather than creating a model from the training samples, the feature vectors of all training instances are stored with DLP/NDLP tag. For each test instance, the K nearest neighbors are found by calculating the distance between the feature vector of the test instance and the stored feature vectors. By using the voting, the DLP/NDLP condition is determined. Thus, odd values are chosen as the K value. When the K is seven, 94.5% of accuracy is achieved. The accuracy by the value of K is provided in Figure 5.17. Since the distance between each offline feature vector and online feature vector has to be calculated, the DLP/NDLP decision takes long time in K-NN if the number of stored instances is high.

5.3.3.3 DLP Detection by Rotation Forest

Rotation Forest is a meta classifier introduced by Rodriguez, J.J., Kuncheva, L.I., and Alonso, C.J. [47]. To create the training data for an underlying classifier, the feature set is randomly split into K subsets and principal component analysis (PCA) is applied to each set. A new feature set is reassembled while keeping all the components. An underlying classifier is trained for this data set.

Table 5.3: DLP/NDLP Detection Accuracy with Neural Networks

SR_{CINR}	SC_{CINR}	Std $[CINR]$	LCR_{CINR}	SC_{RSSI}	SF_{RSSI}	$Mean$ $[RSSI_d]$	SF_{CINR}	Accuracy
O	O	O	O	O	O	O	O	91.57%
O	O	O	O	O	O	O		90.93%
O	O	O	O	O	O			90.10%
O	O	O	O	O				89.87%
O	O	O	O					89.43%
O	O	O						88.87%
O	O							87.82%
O								87.08%
	O							85.17%
		O						83.67%
			O					77.08%
				O				58.90%
					O			55.22%
						O		53.30%
							O	61.77%

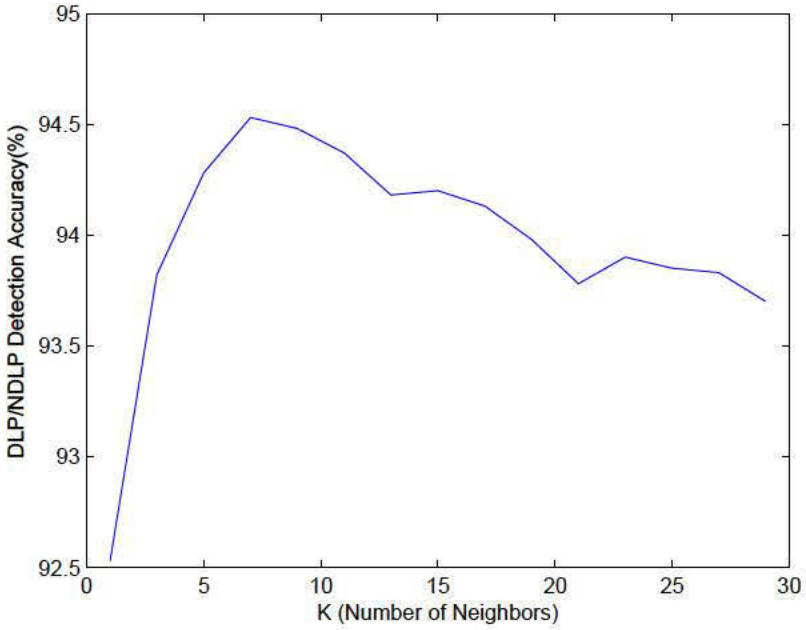


Figure 5.17: DLP/NDLP Detection Accuracy by K

With the data collected for the DLP/NDLP detection, Rotation Forest improves the performance, compared to the accuracy achieved by only using the underlying classifier. The results by the underlying classifiers are presented in Table 5.4. When the K-NN is chosen as the underlying classifier, 94.8% of accuracy is achieved. The Weka 3.6 [48] is used for the evaluation of the classifiers. In the table, the Weka class name is indicated for each classifier.

5.3.3.4 Evaluation of Other Pattern Matching Techniques for DLP Detection

Other classifiers were also evaluated, and their performance was not as good as that of the K-NN and Rotation Forest. The classifiers that attain better than

Table 5.4: DLP/NDLP Detection Accuracy Achieved by Rotation Forest

Classifier Name	Weka Class Name	Accuracy (%)	
		Without Rotation Forest	With Rotation Forest
K-NN (K=7)	IB7	94.5	94.82
Random Forest	Random Forest	93.72	94.32
Grafted C4.5 Decision Tree	J48graft	93.08	94.13
Logistic Model Trees	LMT	92.95	94.2
C4.5 Decision Tree	J48	92.85	94.17
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)	Jrip	92.78	94
K*	Kstar	92.72	93.95
Minimal Cost-complexity Pruning	Simple Cart	92.7	93.85
Best-first Decision Tree Classifier	BFTree	92.65	93.67
Functional trees	FT	92.65	93.88
Fast Decision Tree Learner	REP Tree	92.45	93.92
Artificial Neural Networks	Multilayer Perceptron	91.57	92.87
Support Vector Machine	LibSVM	91.38	92.07

90% of the accuracy are provided in Table 5.5. The classifiers included in the Table 5.4 are not included in this table. Weka 3.6 library is used for the evaluation of the classifiers.

5.3.4 Implication of DLP Detection Features on RF Fingerprinting

The mean of RSSI is primarily used in the previous research on RF fingerprinting (e.g. RADAR [14, 15]) in generating RF fingerprints. However, in this section (Section 5), it is demonstrated that two locations with the similar level of RSSI/CINR may have very different RSSI/CINR characteristics (different DLP condition).

Two locations are chosen to see if the DLP detection features can be used to improve the performance of the RF fingerprinting. In Figure 5.18, the mean of RSSI is about the same at the location 1 and 2. The mean values of RSSI and CINR at these locations are provided in Table 5.6.

Artificial neural networks are modeled to see how well two locations can be classified by using the features evaluated in the DLP detection. The configurations of the neural networks are as follows:

- Nodes in the neural networks: Sigmoid.
- Algorithm: Back propagation algorithm.
- Number of Input Nodes: 1, 13, 15 (depending on the choice of feature sets)
- Number of Hidden Layers: 3, 9, 10 (depending on the choice of feature sets)
- Number of Neurons: 5

Table 5.5: DLP/NDLP Detection Accuracy by Classifiers

Weka Class Name	Accuracy (%)
Bagging	93.45
Random Committee	93.32
Decorate	92.95
Nested Dichotomies	92.87
Ordinal Class Classifier	92.87
Classification Via Regression	92.75
IB1	92.53
PART	92.38
Random Subspace	91.87
Filtered Classifier	91.8
ADTree	91.5
DTNB	91.4
Random Tree	91.15
NBTree	91.07
Nnge	90.95
Ridor	90.87
Decision Table	90.23
LADTree	90.1
Bayes Net	90.02

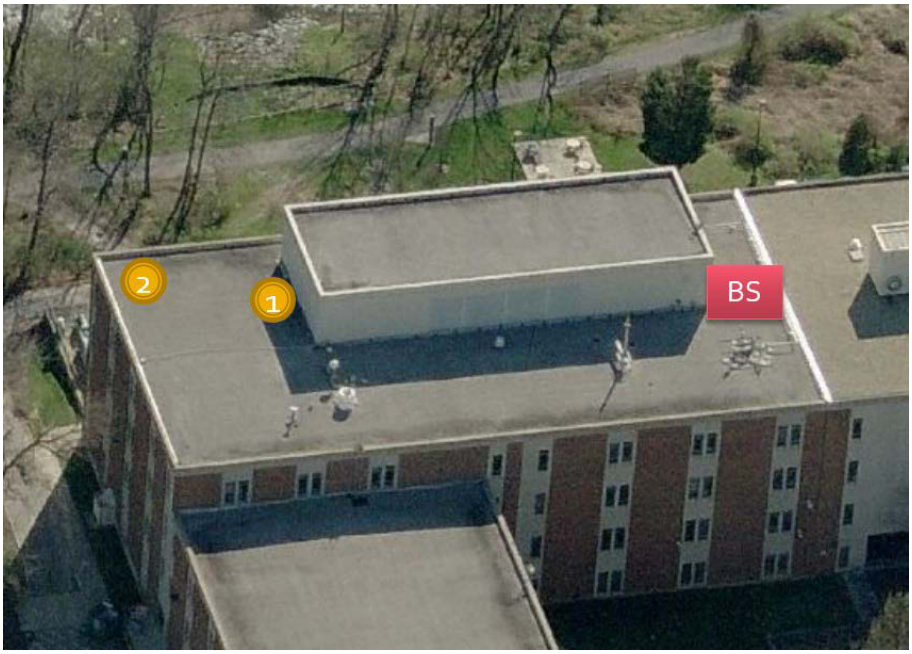


Figure 5.18: Locations for Evaluating DLP/NDLP Features for RF Fingerprinting

Table 5.6: Mean of RSSI and CINR at Locations in Figure 5.18

Location ID	Mean [RSSI]	Mean [CINR]
1	-63.4354	23.33866
2	-63.4862	26.32538

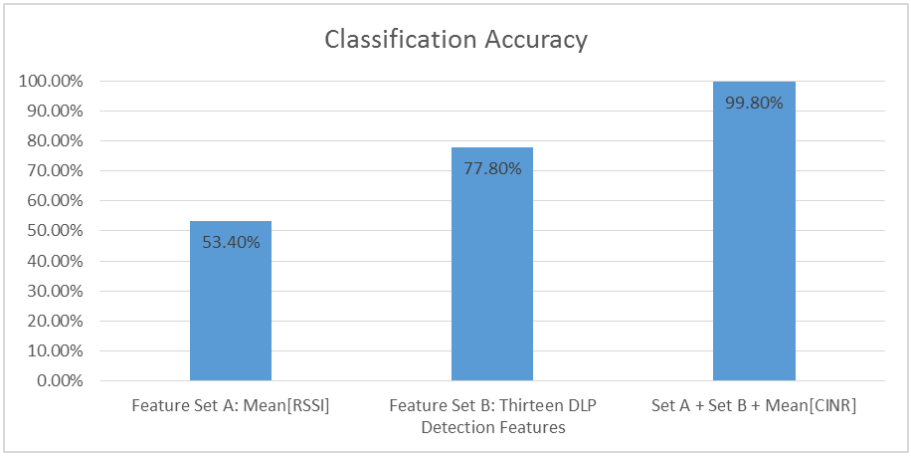


Figure 5.19: RF Fingerprinting Performance Improvement by Using DLP Detection Features

- Learning Rate (The amount the weights are updated): 0.3
- Momentum (Applied to the weights during updating): 0.2
- Attribute Normalization: Yes
- Training Time (The number of epochs to train through): 500

The classification accuracy was calculated after the stratified cross validation with ten folds (described in the Section 5.3.3.1). The result is provided in Figure 5.19. When only the mean of RSSI is used as the feature, two locations can be differentiated with the accuracy of 53%. When fifteen features including the mean of CINR and all DLP detection features evaluated in the Section 5.3.2 are used, the classification accuracy improves to 99.8%.

5.4 Summary

When the MS is under a NDLP condition, the range estimation error has significantly increased. In a set of experiments, the range was estimated by RSSI and RTD in live 4G networks. Under DLP, the range error was in the order of 30 meters, while, under NDLP, the range error was in the order of 70 meters. Thus, it was confirmed that the NDLP condition could be a crucial source of errors in geometric techniques such as ToA and TDoA.

In ROLAX, the DLP condition is determined from the RSSI, CINR, and RTD measurements. Since the RSSI is measured per each antenna of the MS, the RSSI difference ($RSSI_d$) between two antennas is also calculated and used as the measurement. The test statistics, including (1) the RSSI variance, (2) the RSSI level crossing rate, (3) the correlation between the RSSI and RTD, and (4) Kullback-Leibler Divergence between the RSSI distributions, have been successfully used in detecting the DLP condition.

It is noted that, compared to using a single feature, appropriately combined multiple features lead to a very accurate DLP condition detection. Multiple features are combined, and machine learning techniques are used to detect the DLP condition. Time features (variance, mean, level crossing rate) and spectral features (spectral centroid, spectral roll-off, spectral flux) were evaluated and ranked. Top eight features are chosen to detect the DLP condition: SR_{CINR}, SC_{CINR}, $Std[CINR]$, LCR_{CINR}, SC_{RSSI}, SF_{RSSI}, and $Mean[RSSI_d]$.

A number of pattern matching techniques are evaluated for the purpose of the

DLP condition detection. Artificial neural networks, lazy learning using K-nearest neighbor (K-NN), and a meta classifier called Rotation Forest are particularly used, while other pattern matching techniques were also evaluated. When the neural network is used, a detection accuracy of 92% is achieved, and, when the K-NN instance based learning is used, an accuracy of 94.5% was achieved. In case of Rotation Forest meta classifier, we can choose an underlying classifier. When K-NN is chosen as the underlying classifier for Rotation Forest, the best detection accuracy (94.8%) was achieved.

It has been noted that features designed in the DLP detection can be useful in the RF fingerprinting. Thus, the features developed for the DLP condition detection are used in ROLAX RF fingerprinting.

Chapter 6

ROLAX RF Fingerprinting in 4G Networks

In ROLAX, when the required accuracy cannot be achieved using signals from BSs under the DLP condition, RF fingerprinting is used in positioning the mobile station. If the target area is large, then the location estimation by the RF fingerprinting may take a lot of time. Therefore, a location determined by the geometric techniques can be used to choose the search domain of the radiomap in order to decrease the computation of the RF fingerprinting. In order to design a location determination system using RF fingerprinting, the following has to be considered.

- How to design features?
- How to deal with the variation of measurement data?
- How to deal with the missing data?

In ROLAX, fourteen features, including the features developed for the DLP detection, are designed to construct the radiomap. The sources of the measurement variation over a distance and time are identified, and techniques are developed to deal with the variation. A major drawback of the RF fingerprinting techniques is that it takes a tremendous amount of time to build the radiomap. Time and labor needed to build the radiomap increase as the resolution of the radiomap increases. To ease the signal collection, automatic radiomap generation techniques are developed and used in ROLAX. In ROLAX, a novel measurement technique for building

a radiomap is designed since the existing techniques developed for other wireless networks such as Wi-Fi cannot be applied to 4G due to the technical discrepancies.

6.1 Comparison with Existing RF Fingerprinting System

The most widely known RF fingerprinting systems include RADAR [14, 15] and Horus [13]. In this section, ROLAX is compared with RADAR and Horus. RADAR and Horus provide the location determination solution in Wi-Fi networks, particularly for indoor areas. In order to determine the location, the online measurement and features stored in the radiomap are compared using pattern matching techniques. In both RADAR and Horus, the signal strength (RSSI) is measured on the client device with regard to the beacons transmitted from the Access Points (APs) during the offline and online phases.

During the offline phase, the radiomap is created by collecting measurements at known locations. A series of RSSI measurements are collected to create feature vectors in both RADAR and Horus. In RADAR, sample mean of RSSI with regard to each AP is calculated. In Horus, a distribution of the RSSI with regard to each AP is estimated. The distribution can be a parametric distribution (e.g. Gaussian with an estimated mean and an estimated variance) or a non-parametric distribution (e.g. Histogram of RSSI).

In ROLAX, in addition to RSSI, CINR is used as the underlying measurement. RSSI can be measured per each antenna of the terminal, so RSSI difference ($RSSI_d$) can be retrieved from the RSSI measurements to create additional features.

RTD is also considered, but due to its low measurement frequency, RTD is not currently used in ROLAX. In ROLAX, fourteen features, including time features (e.g. mean, variance) and spectral features (e.g. spectral centroid, spectral roll-off), are generated from the underlying measurements. In Wi-Fi, the measurements are performed upon receiving the beacons from the AP while, in 4G, the measurements are performed on the downlink preambles from the BS. While a typical Wi-Fi beacon interval is 100 msec, a typical 4G WiMAX preamble interval is 5 msec. Because of the higher measurement frequency, the spectral feature generation is more feasible in 4G WiMAX than in Wi-Fi. When only a single feature is used in ROLAX, the mean of the RSSI is used as the feature.

During the online phase, RSSI measurements are collected for each AP in RADAR and Horus, and these measurements can be used as the online phase feature by itself. In RADAR, distances are calculated between the online phase feature vector and the offline feature vectors, and Euclidean distance is primarily used as a distance measure, while other distance measure can be also used. In Horus, the probability to get the online feature vector is calculated given an offline feature (parametric or non-parametric RSSI distribution) in the radiomap. Probability can be regarded as a distance measure to calculate a distance between an offline feature and an online feature. In both RADAR and Horus, a nearest neighbor search is performed given the calculated distances between the online feature vector and the offline feature vectors in the radiomap.

In ROLAX, the online and offline feature vectors have the same set of features. The pattern matching techniques, such as artificial neural networks (ANN) and

Rotation Forest, have been used in ROLAX to determine a location given an online feature vector. However, other pattern matching techniques can be used as long as they can deal with the missing values because the measurement miss is a frequent event in 4G WiMAX. When a single feature is used, nearest neighbor search is performed in ROLAX while other pattern matching technique can be also used.

The comparison between RADAR, Horus, and ROLAX is summarized in Table 6.1.

The three systems are compared in an example. In this example, measurements are performed at three locations during the offline phase with regard to three APs. In RADAR, the mean of RSSI is calculated to create the offline features. The offline features associated with a location i is denoted by $RSSI_i$. The created offline features are $RSSI_1 = [-56, -72, -83]$, $RSSI_2 = [-62, -90, -56]$, and $RSSI_3 = [-73, -83, -64]$. During the online phase, RSSI is measure for each AP, and it is denoted by $RSSI_{online}$. $RSSI_{online}$ is $[-62, -81, -75]$. The Euclidean distance, D_i is calculated between $RSSI_{online}$ and $RSSI_i (i = 1, 2, 3)$. (Figure 6.1)

In Horus, the distribution of the RSSI is estimated for each location in the radiomap. In this example, parametric distribution using the Gaussian is used. The offline features associated with a location i is denoted by $Hist_i$. The created offline features are $Hist_1 = [N(-56, 3), N(-72, 4), N(-83, 2)]$, $Hist_2 = [N(-62, 4), N(-90, 4), N(-56, 1)]$, and $Hist_3 = [N(-73, 4), N(-83, 4), N(-56, 1)]$ where $(N(\mu, \sigma^2))$ is the Gaussian distribution with a mean μ and a variance σ^2. The probability, P_i to get $RSSI_{online}$ of $[-62, -81, -75]$ is calculated for each $Hist_i$. (Figure 6.2)

In ROLAX, the offline features at location i is denoted by $Feature_i$. Since

Table 6.1: Comparison between RADAR, Horus, and ROLAX

	Underlying Measurements	Offline Feature	Online Feature	Distance Measure	Pattern Matching
RADAR	$RSSI$	Sample Mean	$RSSI$	Euclidean Distance	Nearest Neighbor
Horus	$RSSI$	Parametric Distribution, Non-parametric Distribution (Histogram)	$RSSI$	Probability (Maximum Likelihood)	Nearest Neighbor
ROLAX (single feature)	$RSSI$	Sample Mean	$RSSI$	Euclidean Distance / Manhattan Distance	Nearest Neighbor
ROLAX (multiple features)	$RSSI$ and $CINR$ (RTD in future)	14 Features generated from $RSSI$, $CINR$, and $RSSI_d$	Same as offline features	N/A	ANN, Rotation Forest, etc.

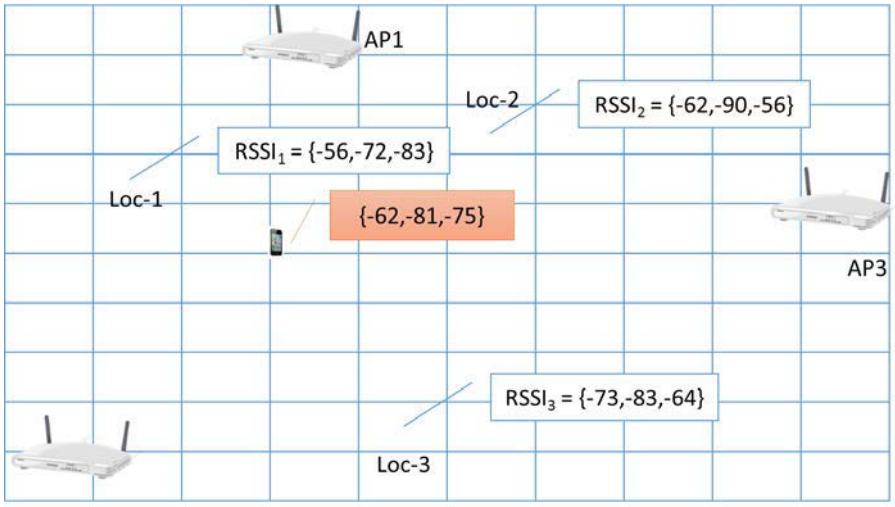


Figure 6.1: RADAR Example

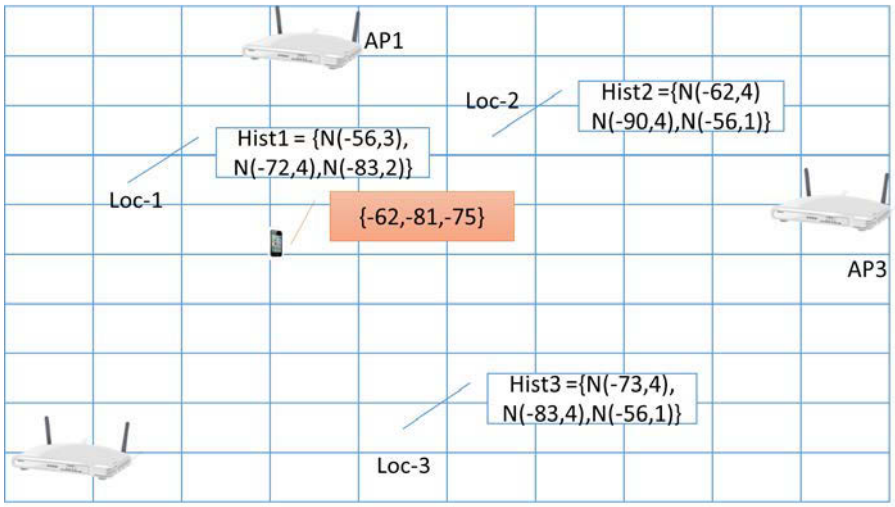


Figure 6.2: Horus Example

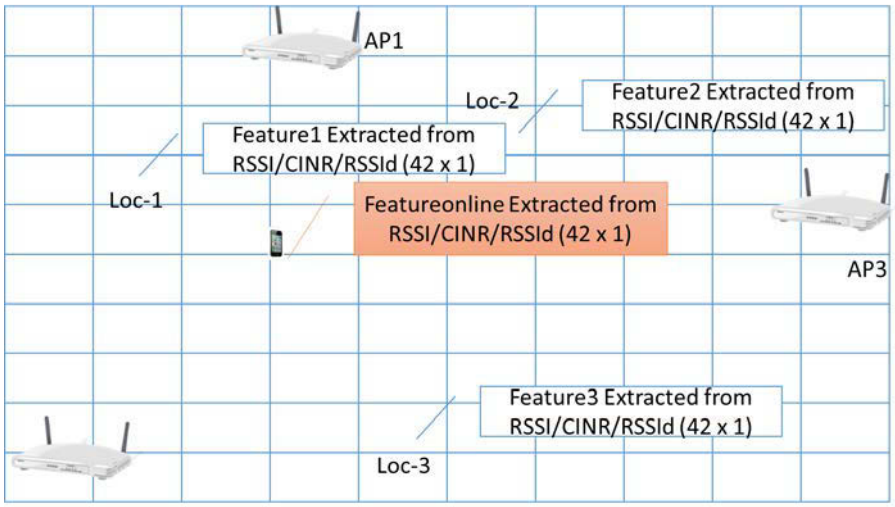


Figure 6.3: ROLAX Example

there are three APs, 14 features are generated for each AP. Thus, the size of $Feature_i$ is 42 (14*3) by 1. During the online phase, an online feature vector with the same size is generated from the measurements. By using a pattern matching such as ANN, the location is determined.(Figure 6.3)

The client may miss the presence of the AP, but how to manage the missing value is not well addressed in both RADAR and Horus. In ROLAX, the handling of missing values is strongly needed because there are very frequent measurement misses. ROLAX is designed to deal with missing values by using techniques including overlapped Gridding, interpolation, and distance measures that can deal with missing values.

6.2 Quality of Measurement Data

The measurements vary over time and over (small) distance. The quality of measurement data matters in both the offline and the online phases. In ROLAX, to provide an accurate geo-location, the sources of the signal variation and outliers are identified, and each of them is handled.

The outliers can be generated either by errors (e.g. systematic errors, measurement errors), or by having legitimate data with very different values from its surrounding values. For the power and time related measurements, the variation of the radio channel condition known as fading can be a significant source of the variation. The causes of measurement variation are summarized as follows:

- **Systematic error** - implementation error of the software, etc.
- **Measurement error** - combination of unknown measurement error sources
- **Channel fading** - fading over frequency/time/location
- **Atmospheric propagation impairments and thermal noise** - impairments by rain, atmospheric absorption, fog, snow, and atmospheric multipath

6.2.1 Systematic Error

One systematic cause of outliers is the *systematic implementation errors* of the 4G hardware or the software. For instance, the device driver of Beceem chip-set reports a RTD value of zero under two different conditions.[1] Zero RTD value can be a legitimate value, but it can be also reported when the link is almost lost. It is

[1]The RTD value reported by the software of Beceem chip-set MS is in the range of [-39,104].

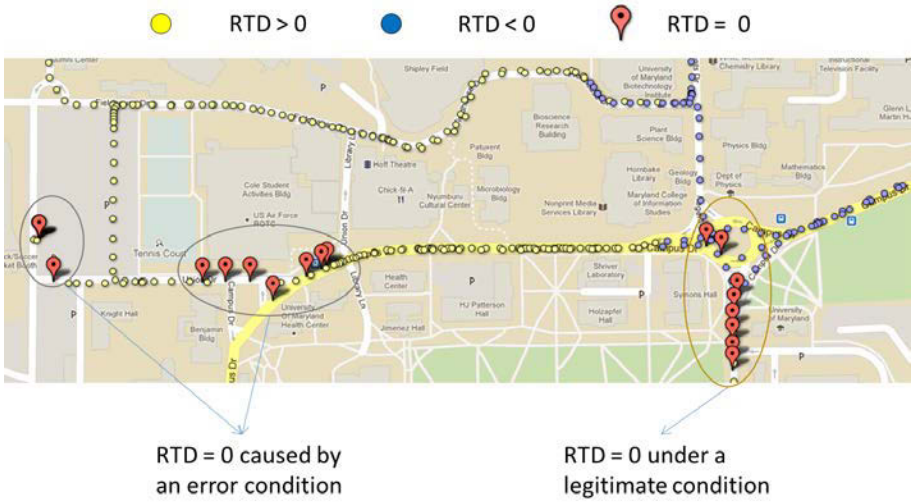


Figure 6.4: Example of Systematic Errors in Reporting RTD

conjectured that zero value is reported under a certain error condition. Thus, when the RTD value is zero, it would be hard to tell whether the measured RTD value is zero or an indication of a certain error condition (Figure 6.4).

Similarly, CINR of -10 dB and RSSI of -123 dBm are observed when the physical synchronization is lost in the Beceem chip-set software. They are the minimum values set each for RSSI and CINR in the Beceem device driver. The outliers, generated by systematic errors, have to be handled before processing the data further.

6.2.2 Measurement Errors

The measurements come with the measurement errors, which are usually modeled by the Gaussian distribution. This assumption can be justified by the Lindeberg Central Limit Theorem (CLT). Let X_i be a sequence of random variable with finite mean m_i and variance σ_i^2. s_n is defined by $\sqrt{\sum_{j=1}^{n} \sigma_j^2}$. If, for every $\varepsilon > 0$, the condition given in (6.1) is met, then the distribution of the $\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i)$ converges

in distribution to a standard Gaussian distribution [49].

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E\Big[(X_i - \mu_i)^2 \cdot \mathbf{1}_{\{|X_i-\mu_i|>\varepsilon s_n\}}\Big] = 0 \tag{6.1}$$

where $\mathbf{1}_{\{\dots\}}$ is the indication function.

Thus, as long as the sources of errors are independent and meet a certain conditions, the combined errors can be modeled by the Gaussian distribution.

6.2.3 Channel Fading

The fading of the radio channel causes the variation in the power and the time measurement such as RSSI, CINR, and RTD. At a fixed location, the amplitude can vary over frequency, time, and location because of multipath, Doppler spread, etc.

6.2.3.1 Shadow Fading

The surrounding obstacles between the BS and the MS can cause shadow fading, and the measured signals under the shadow fading can be very different from the average value and between two locations. The shadowing over a number of locations is usually modeled by log-normal distribution.

The implications of the shadowing on the generation of the radiomap are as follows:

- The measurement value in the unvisited area can be predicted precisely as long as the affect by obstacles on the radio signals can be identified and known a priori. The adjusted path loss model can be used to construct the power-

related measurement values in the area that could not be visited during the offline signal collection.

- At a particular location, the characteristic of the radio signals does not vary vastly due to the shadowing as long as the surroundings between the MS and the BS are static. Since the surroundings can change over the time (e.g. the density of the tree leaves is different between summer and winter), the measurement values can change over time. Thus, the *profiling* of the radiomap over the time (e.g. months, seasons) is needed.

6.2.3.2 Frequency-selective Fading

The *coherence bandwidth*, B_c, is the bandwidth over which the channel is considered to be flat. It is inversely proportional to the delay spread of the channel, and the relaxed ballpark definition is $B_c = \frac{1}{5\sigma_\tau}$, where the σ_τ is the rms delay spread. When the coherence bandwidth is smaller than the bandwidth of the channel, the frequency-selective fading can be observed. The change of amplitude and the phase is different per radio frequency channel in the frequency-selective channel. The rms delay spread in 4G WiMAX is dependent on the terrain type, and, under the Stanford University Interim (SUI) model, it ranges between 0.111 μsec (SUI-1 channel) to 5.240 μsec (SUI-6 channel) [27] when the omni-directional antenna is used. These rms delay spread values result in the coherence bandwidth from 0.038 MHz to 1.8 MHz, which is much larger than the channel bandwidth used by an OFDM subcarrier. When the 10MHz bandwidth is used, the subcarrier spacing in

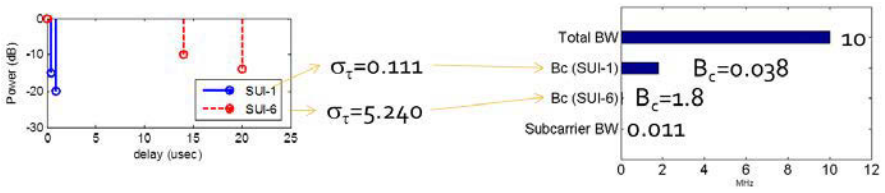


Figure 6.5: Coherence Bandwidth in 4G WiMAX

4G WiMAX is 0.011 MHz. If the 30° directional antenna is used, the coherence bandwidth increases more than twice the coherence bandwidth under the use of the omni-directional antenna. Therefore, we can assume that the fading is not frequency-selective in general. The channel is not considered flat over a 10MHz band, but it is considered flat over an OFDM subcarrier.

The RSSI value is calculated typically at the baseband, as a function of ADC outputs with the Gaussian distribution assumption for the ADC inputs and outputs,[2] and the sample power is averaged over the samples, each on a different subcarrier in a preamble.[3] Even though the channel is frequency-selective, the variation over the frequencies is averaged out because of the way the RSSI value is calculated.

6.2.3.3 Fast Fading

In the fast-fading channel, primarily due to the mobility, the received signal strength varies over time relatively faster (the coherence time T_c of the channel is small relative to the delay constraint of the channel). In this case, the radio receiver typically uses the time diversity since the signal may be at a deep fade at time t

[2]See the equation (A.1) in Appendix A.1.

[3]In 4G WiMAX OFDMA PHY, a preamble is composed of a single OFDM symbol. In the case of 10MHz bandwidth, an OFDM symbol is composed of 1024 subcarriers.

while it may not be at time $t + \delta t$.

If the channel is completely static, then the coherence time is merely a function of the coherence distance (D_c/v where D_c is the coherence distance, and v is the speed of the MS). If the Doppler spread can be found, then $T_c \approx \frac{1}{f_m}$. If the mobility of the MS dominates over the mobility of the surroundings, then $T_c \approx \frac{1}{vf_c/c}$, where v is the speed of the MS, f_c is the center frequency, and c is the speed of light.

In 2.5 GHz 4G WiMAX networks such as MAXWell 4G WiMAX network, the coherence time is 200 msec at the speed of 1.2 miles/hour, 10 msec at the speed of 27 miles/hour, and 4 msec at the speed of 60 miles/hour [29]. Therefore, the amplitude and the phase of the signal can be regarded as flat over the duration of a preamble, which is composed of a single OFDM symbol spanning over 0.1 msec. During the offline signal collection in this work(Section 6.6.2), the vehicle was driven at a speed below 10 miles/hour.

The autocorrelation functions over 100 lags (500 msec) for RSSIs measured at eight different locations are provided in Figure 6.6. This figure shows that the RSSI values are highly correlated within, at least, 100 msec window.

The implications of the coherence time on the radiomap generation are as follows:

- The measurements have to be done multiple times to absorb the variability of the signals over the time.
- The interval between the measurements can be larger than the coherence time since the channel does not change significantly over the coherence time.

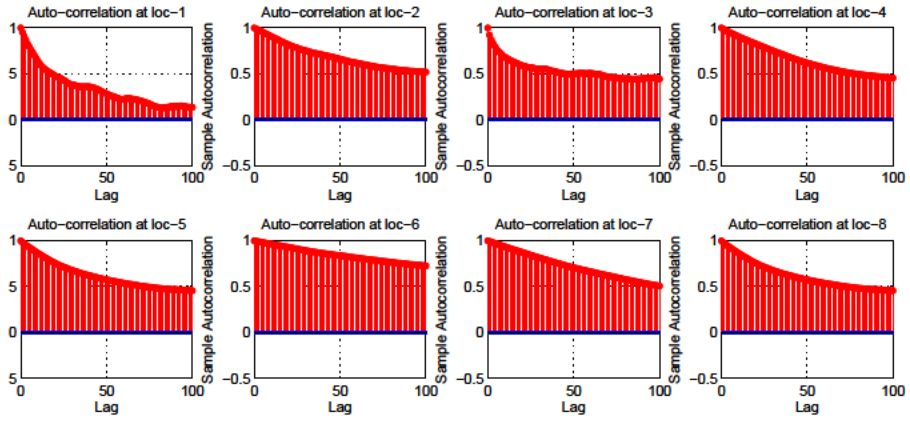


Figure 6.6: Autocorrelation of RSSI Measurements at Eight Locations (1 Lag = 5 msec)

In building our radiomap, the measurement was made per preamble, so the measurement interval was 5 msec. Taking the coherence time into account, the sampling of the measurement can be done to decrease the size of the data and accordingly ease the computation complexity.

6.2.3.4 Small Scale Fading(Fading over Short Distances)

Under the small-scale fading, the rapid fluctuation of the signals can be found not only over the time but also over the sub-wavelength distance.[4]

Coherence distance D_c is the distance over which a fading channel appears to be uncorrelated. Thus, a MS, receiving signals with fair quality at the location l, may encounter a deep null by traveling a short distance larger than D_c. Coherence

[4]The wavelength when the 2.5GHz frequency is used is in the order of 11 cm.

distance is the dual of *angular spread*[5], which refers to the statistical distribution of the angle of the arriving energy. Large angular spread implies that the channel energy is coming from many different directions, and the coherence distance is small [29]. When there is no dominant path, and the angles of the arrivals are uncorrelated, the in-phase component and quadrature-component follow the Gaussian distribution at a fixed time because of the central limit theorem. Thus, the signal envelope follows the Rayleigh distribution. *Rayleigh fading* is the channel model typically assumed when there is no apparently strong LOS. In addition, it eases in computing the coherence time because a uniform angular spread is assumed [50]. The coherence distance in an omnidirectional Rayleigh channel is given by (6.2) [45].

$$D_c = \frac{9\lambda}{16\pi} \tag{6.2}$$

In case of the 2.5GHz mobile WiMAX, $D_c = \frac{9 \cdot 3\times10^8/2.5\times10^9}{16\pi} = 0.021(meter) = 2.1(cm)$. If the directional antenna is used, D_c increases because fading contributions are minimized by the spatial filtering effect of the antenna pattern [51]. In general, the coherence distance is usually greater than 0.5λ (6 cm in case of 2.5GHz), and the antenna spacing for the BS with low-medium and high antenna heights is in the order of 10 to 20 λ (1.2–2.4 meters) if 120° antenna is used [27].

In this work, Motorola USBw 25100 was primarily used as the 4G MS for the experimentation. It has dual omni-directional antennas, and the distance between them is 2.5 cm [33]. The antennas are located on the left and right ends of the USB

[5]Angular spread is defined by $\Lambda = \sqrt{1 - \frac{|F_1{}^2|}{F_0^2}}$ where $F_n = \int_0^{2\pi} p(\theta) exp(-jn\theta) d\theta$ and $p(\theta)$ is the angular distribution of multipath power [45].

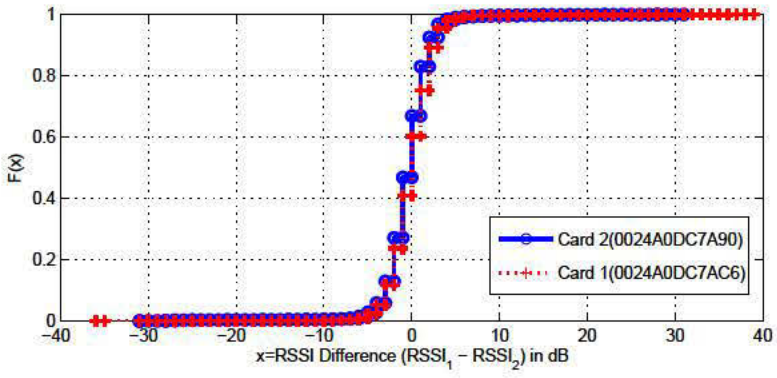


Figure 6.7: Empirical CDF of RSSI Difference per Card

adapter. During the offline phase, the RSSI values were measured at each antenna.

The difference between RSSI values measured from both antennas were recorded and analyzed. Let us denote the RSSI difference by $RSSI_d = RSSI_1 - RSSI_2$, where $RSSI_1$ is the RSSI measured at the first antenna, and $RSSI_2$ is the RSSI measured at the second antenna. To see if the choice of the MS adapters can affect the distribution of the $RSSI_d$, the Kaplan-Meier estimate of the CDF for the $RSSI_d$ per card was generated (Figure 6.7). Both cards made measurements for five frequency bands with 10MHz bandwidth, but they scanned a different set of BSs.

The mean, median, maximum, and minimum of $RSSI_d$ values are provided in Table 6.2. More than 95% of the $RSSI_d$ values were between -4dB to 4dB, but the $RSSI_d$ could be over 30dB. The median and average values of the $RSSI_d$, measured by both cards, are close to zero as expected.

At some locations, large $RSSI_d$ value is observed. 5529 RSSI measurement samples were collected, by making measurements for the preambles from a Clear BS on 2673.5 MHz over 27 seconds, at a fixed location, and the result is provided in the

Card (MAC Address)	Mean $[RSSI_d]$	Median $[RSSI_d]$	Std $[RSSI_d]$	Min $[RSSI_d]$	Max $[RSSI_d]$	Corrcoef ($RSSI_1$, $RSSI_2$)
0024A0DC7A90	-0.45	0	5.09	-31	31	0.91
0024A0DC7AC6	0.02	0	2.56	-36	39	0.93

Table 6.2: Statistics of RSSI Difference R_d

Figure 6.8. At this location, large $RSSI_d$ is observed over the whole measurement duration, with a mean of 8.7(dB) and a standard deviation of 8.5(dB). $RSSI_1$ was consistently higher than $RSSI_2$. At this location, the correlation coefficient between the $RSSI_1$ and the $RSSI_2$ is 0.26, which is much smaller than correlation coefficient of 0.91 calculated over all samples measured by the same card. Thus, it has been confirmed, by the observation, that the coherence distance can be smaller than 2.5 cm in some environment. Indeed, the coherence distance varies according to the multipath environment (angular spread).

The RSSI values measured per antenna are combined to calculate the combined RSSI as in (5.8). The locations where one of the antennas is located at deep radio null can have outliers in the radiomap. If one of the antennas is located at a deep radio null, the combined RSSI can be decreased as much as 3dB.

These observations about the coherence distance have implications on the filtering and the processing of the measurements for the radiomap as follows:

1. To capture the radio signal pattern completely in the radiomap, the resolution

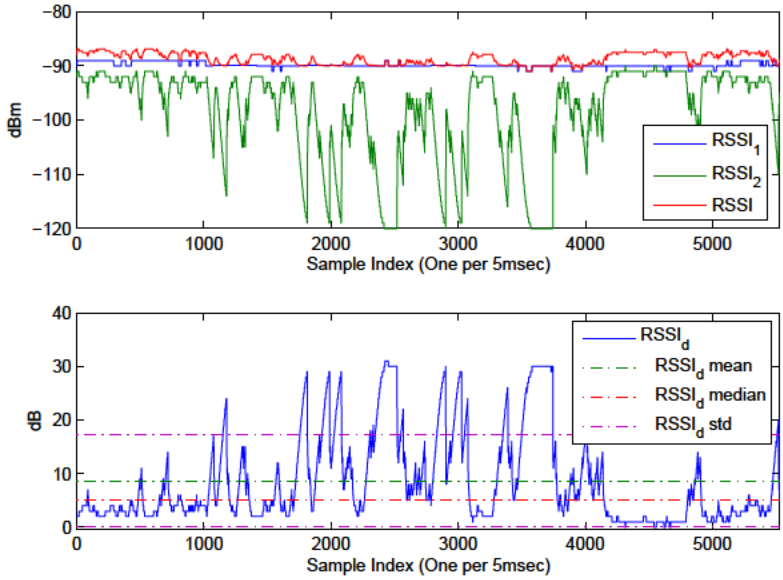


Figure 6.8: RSSI Per Antenna and RSSI Difference at Location with One Antenna under Radio Null

of the radiomap has to be as small as the coherence distance. It is difficult because it requires very dense measurements since the coherence distance can be in the order of centimeters in the 2.5GHz 4G network. Since the channel is dynamic, due to the movements of the obstacles in the surroundings, the small-scale signal pattern is not consistent over time. Since it is difficult to capture the small-scale variation of the radio signal pattern, and the small-scale variation can change over time, capturing the medium-scale variation would be the best.

2. By combining the measurements in a small area (grid), the effect by the measurements at the deep null can be diminished.

3. When the RSSI difference is higher than a certain threshold value, the mea-

surement can be declared as an outlier and should not be included in generating the radiomap.

6.2.4 Atmospheric Propagation Impairments

The signals are attenuated at the presence of atmospheric propagation impairments such as rain, atmospheric absorption, fog, snow, atmospheric multipath, etc. There are few previous researches studying the atmospheric impairments on 4G networks. [52] evaluated the environmental effects on the signals by using ITU-R and other models on the 2–66 GHz band.

Rain attenuation can be a significant issue particularly in fixed WiMAX networks using the frequency over 10GHz. At this frequency band, the rain attenuation can be a dominating factor as the attenuation by multipath is lower [53]. However, the rain attenuation at 2.3–2.5 GHz is negligible [52]. At the rain rate of 102 mm/hour (4 inches/hour), 2.4GHz signal and 5.8GHz signal is attenuated each by 0.05dB/km and 0.5dB/km [54].

The attenuation by snow is primarily due to the moisture content of the particles, and it is usually less than the rain attenuation [52]. Thus, snow attenuation can be regarded as negligible at 2–3GHz band.

The attenuation by atmospheric gases, due to dry air and water vapor for 1–350GHz band, is studied in [55]. The attenuation at 2–3GHz at sea level is 0.007–0.008 dB/km, so this effect can be regarded as negligible for our purpose.

The attenuation by vapor components in the nature such as **fog** is also

negligible unless the frequency is above 100GHz. Even at the 10GHz, the attenuation by fog is less than 0.01 dB, and the attenuation at 2–3GHz band is much less [52].

The **thermal noise** can cause a variation of the power measurement over the time. It is calculated by (6.3).

$$P_{dBm} = 10log_{10}\left(k_B T \times 1000\right) + 10log_{10}(B) \tag{6.3}$$

The temperature change, from -20° Celsius to 40° Celsius, results in 0.9 dB difference in thermal noise. For example, in Washington DC, the average low temperature in January is -4° Celsius, and average high temperature in July is 33° Celsius. The thermal noise has a range of 0.7 dB in Washington DC region over a year. The effect by the temperature change is higher than that by the rain attenuation.

4G WiMAX can operate on 2–66GHz, and the most popular band in US is 2.5GHz. The US license free spectrum, at 5.8 GHz and 3.65 GHz band, has a few 4G vendors building products. The 4.9 GHz will be used for public safety (Homeland security band).

The summary of the atmospheric propagation impairments is as follows:

- In case of 2.3–2.5GHz, most of the atmospheric impairments can be ignored since their contribution is more than 10 times smaller than the RSSI quantization size.
- For the 3–5GHz operation, the rain attenuation needs to be considered in calibrating the readings particularly if the precipitation rate is high.

- Since the thermal noise can cause about 1dB variation over the time of the year, it would be better if the power measurement can be calibrated according to the temperature when the measurements are made.

The location determination system and the context aware system such as Rover [56] can be beneficial to each other. The technologies developed in this work can provide location information to the location context server. At the same time, to increase the location accuracy provided by the location context server, it can benefit from the other context servers in the context-aware ecosystem. For instance, the rain rate and temperature information from weather context server can enhance the accuracy of the location determination system.

6.2.5 Effect of Outliers

The outliers have to be removed before data processing. The filtering of the outliers can be done before the gridding and interpolation in ROLAX. Particularly, when the Euclidean distance is used as the measure in a pattern matching algorithm such as K-NN, outliers is the dominant component affecting the distance between two RF fingerprints because of the squaring operation. Thus, choice of distance (similarity) measure is dependent on the frequency of the outliers if they are not filtered out. If the outliers occur frequently, the distance measure has to be chosen or designed so that it is not sensitive to the presence of outliers. For instance, Manhattan distance can be used in this case since it is less subjective to outliers than Euclidean distance is. It is discussed further in Section 6.5.1.1.

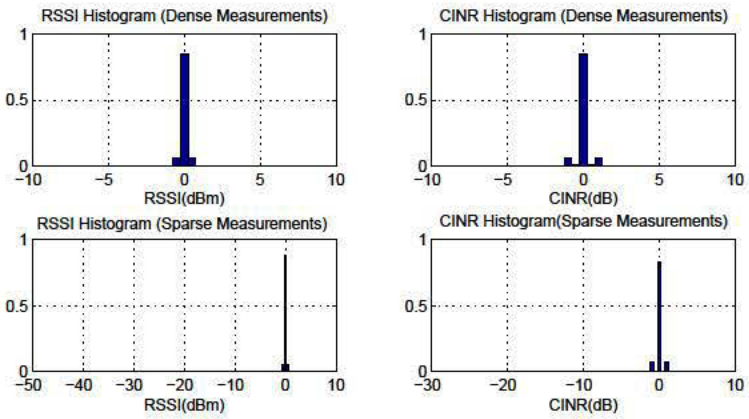


Figure 6.9: Histograms of Measurement Difference from Nearest Location

The difference M_d, between the measurement at a location l_i and the measurement at the nearest location from l_i, is calculated. Histograms of M_d were constructed to see the frequency of the outliers (Figure 6.9). The histograms were constructed for both the dense measurement case and the sparse measurement case (each with 130,000 scattered locations and 7695 scattered locations in roughly 18,000 m^2 area). It was observed that more than 98% of the M_d is with $[\mu - 3 \cdot \sigma, \mu + 3 \cdot \sigma]$ window (Table 6.3).

Detection of outliers can be regarded as an anomaly detection problem, which is also known as exception mining. Popular anomaly detection techniques include model-based techniques, proximity-based techniques, and density-based techniques [57].

Unless the data points in a grid are very close to each other, the physical distance has to be considered in calculating outlier score. When the data points are RSSI or RTD values, the density can be defined by (6.4) by extending the definition given in [57]. This density can be used in a density-based anomaly detection to filter

Parameter	Measurement Density	Range $[M_d]$	Mean $[M_d]$	Std $[M_d]$	Prob $(M_d \in [\mu - \sigma, \mu + \sigma])$	Prob $(M_d \in [\mu-2\cdot\sigma, \mu+2\cdot\sigma])$	Prob $(M_d \in [\mu-3\cdot\sigma, \mu+3\cdot\sigma])$
RSSI	dense (0.7 samples/m^2)	[-5.5, 5.5]	0.00	0.39	0.87	0.97	0.98
	sparse (0.04 samples/m^2)	[-42, 3.5]	-0.01	0.53	0.85	0.86	0.99
CINR	dense (0.7 samples/m^2)	[-5.5, 5.5]	0.00	0.44	0.96	0.99	1.00
	sparse (0.04 samples/m^2)	[-20, 6]	0.00	0.49	0.83	0.85	0.99

Table 6.3: Frequency and Range of Measurements Difference from Nearest Location M_d

the outliers in the radiomap.

Definition 1 *Outlier Density*

$$density(\mathbf{x}, k) = \left(\frac{\sum_{\mathbf{y} \in (\mathbf{x},k)} \frac{distance(\mathbf{x},\mathbf{y})}{D(\mathbf{x},\mathbf{y})}}{|N(\mathbf{x}, k)|} \right)^{-1} \tag{6.4}$$

where $N(\mathbf{x}, k)$ is the set containing k-nearest neighbors of $\mathbf{x}$, $|N(\mathbf{x}, k)|$ is the size of that set, $\mathbf{y}$ is the nearest neighbor, $distance(\mathbf{x}, \mathbf{y})$ is the distance between measurements at $\mathbf{x}$ and $\mathbf{y}$, and $D(\mathbf{x}, \mathbf{y})$ is defined as in (6.5).

$$D(\mathbf{x}, \mathbf{y}) = \begin{cases} log_{10} \|\mathbf{x} - \mathbf{y}\| & \text{if the measurement is RSSI} \\ \|\mathbf{x} - \mathbf{y}\| & \text{if the measurement is RTD} \end{cases} \tag{6.5}$$

The rationale behind this definition is that the RSSI linearly changes over the logarithmic distance while RTD linearly changes over the distance. The data points with low density are removed as outliers.

6.3 Generation of 4G Radiomap

6.3.1 Signal Collections for Radiomap Generation

Scanning operations are typically used in building a radiomap for wireless networks such as Wi-Fi. In building radiomaps for 4G networks, scanning can be used, but it has a huge limitation because the scanning typically takes longer in 4G than in Wi-Fi. Thus, in 4G, it is hard to use scanning operations in building a radiomap for large areas while meeting high accuracy requirements.

In 4G, during the initial entry scanning, the MSs passively receive preambles and decode Downlink Maps (DL-MAPs) from BSs. The presence of a BS is reported when the MS can achieve a downlink synchronization (i.e. when the DL-MAPS are decoded) and obtain uplink parameters from the UCD. The MSs listen to the frequencies listed in its configuration on the non-volatile memory. This takes a minimum of two frames at each channel [58]. The maximum time between DL MAPs can be as high as 11 seconds [59]. Once the MS decodes the DL-MAP/DCD and UCD, it obtains all channel information and link measurements it needs to perform the next stage of the network entry (ranging process). Then it switches the frequency channel to look for another BS.

The number of frequency bands being used differs by the region and the regulatory domain. In a standard 4G WiMAX configuration for the United States, two bandwidths (5MHz and 10MHz) in 22 frequency bands are needed to be scanned. In the area under study (College Park, Maryland), as of March 2012, six frequency bands were being used by Clear, and three frequency bands were being used by the MAXWell 4G network. Both of the networks use 10MHz bandwidth.[6] It was observed that decreasing the number of frequencies to be scanned, from twenty-two to nine, does not help so much in decreasing the total scanning time. It is believed that most of the scanning time is spent on achieving synchronization rather than making a decision whether the channel is being used or not. Scanning time also depends on the scanning mode, which differs by the association level (0, 1 or 2) of

[6]MAXWell 4G network used to use one frequency for all three outdoor sectors and 5MHz bandwidth in 2010. Some experiments were performed under this setting.

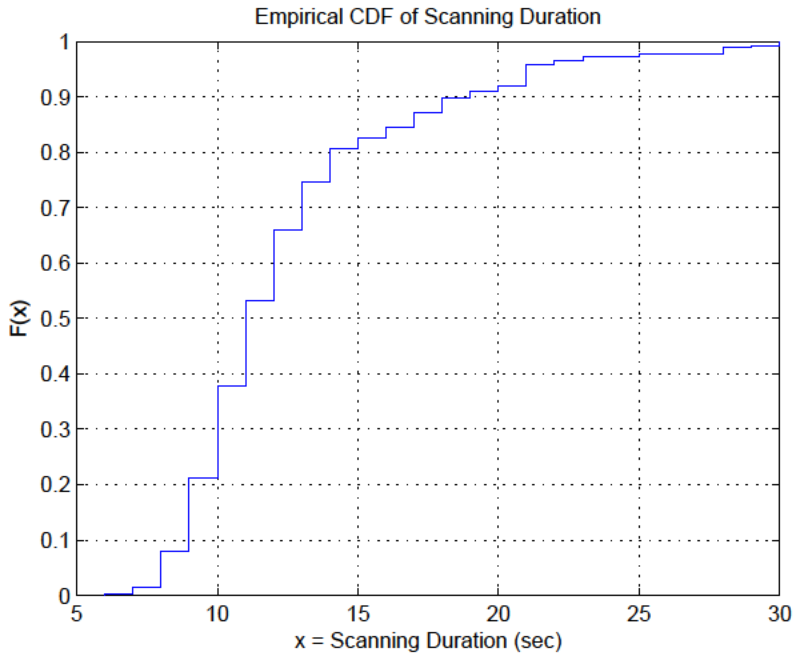


Figure 6.10: Empirical CDF of Scanning Duration

the MSs.

The mean scanning duration for nine frequency bands, measured during the experiments in Section 6.6.2, was 12.5 seconds, and the standard deviation was 4.3 seconds. The scanning duration could not be over 30 seconds since the scanning is configured to terminate if it takes more than 30 seconds. The Linux-WiMAX [34] connection manager warns that the wide scanning may take several minutes. The empirical cumulative distribution function (CDF) of WiMAX scanning time is provided in the Figure 6.10.

This scanning duration value is about 40 times higher than the scanning duration in IEEE 802.11 Wi-Fi. The average scanning duration in Wi-Fi is in the order of 250msec on 2.4GHz band [60].

Therefore, the classical radiomap generation techniques based upon scanning

cannot be used in 4G unless the accuracy requirement is very low. In this work, a novel radiomap generation technique is introduced. The radiomap is generated in the following procedure. If the target area is large, then the target area has to be segmented into smaller areas. For each segmented area, perform the following procedure.

1. Perform continuous scanning of mode 1 (scanning with no association) at the target area. Create a hashtable T_{BS} with the frequency as the key K_{freq} and the array list of the BSIDs $BSID_{K_{freq}}[]$ that were seen on the frequency K_{freq} as the value. Hash function is denoted by $H_{BS}(K_{freq})$.
2. For each key (frequency) K_{freq} of H_{BS},
   (a) Set the frequency of the connection manager to K_{freq}.
   (b) Attempt to connect the BSs ($BSID_{K_{freq}}[]$) on the frequency K_{freq}.
   (c) Once the connection is made with a BS, start wardriving over the target area.

Instead of scanning in the step 1), neighbor advertisement messages (MOB_NBR-ADV) can be used. Neighbor advertisement messages provide the information about the neighbor BSs of the serving BS, which is typically managed by the same service provider. In order to receive these messages, the connection has to be made with a BS (serving BS). This information also can augment the information collected in the step 1.

Due to the limitation of the software, it is hard to build a radiomap per a BSID. The software used in this work only lets the user to choose the frequency

band rather than BSID. MSs scan their neighbor BSs during the scanning interval assigned by the BS for the handoff. Unless the handoff is disabled, the MS attempts to handoff to other BSs on the same or other frequency channels. Thus, the firmware of the MSs was configured not to perform handoff during the radiomap generation. However, even though the handoff is disabled, the MSs attempt and execute handoff to the other BSs on the *same* frequency channel with the software used in this work.

When the scanning is used in building the radiomap, 4.8 signal quality readings per minute (RSSI and CINR) are obtained for each BS on average. Since one reading is generated per 12.5 seconds on average, hardly can wardriving be used. Only one reading is obtained while driving 55 meters when the vehicle is driven at 10 miles/hour.

If signal collection procedure of ROLAX is used to build the radiomap, 12,000 signal quality readings per minute can be obtained. In other word, one reading can be obtained while driving 2.2 cm distance.

The wardriving vehicle used in this work is equipped with the following hardware and software components.

- Two Windows XP netbook computers installed with 4G software (firmware, device driver, connection manager, logging facilities, Wireshark, etc.) and PC clock synchronization software, which synchronizes the clock of the PC with the GPS clock.
- Two 4G WiMAX MS receivers (USB dongle type).
- One GPS receiver with Bluetooth interface: It provides the location log-

Figure 6.11: Wardriving Vehicle with Two 4G Receivers

ging and the clock synchronization (once every one second) so that the time stamped measurement data can be correlated with the time stamped location.

The vehicle has to go over the area at least N_{freq}={the number of frequencies being used in the target area} times. In other words, the vehicle needs N_{freq} receivers each tracking the signals on a frequency band. The measurement locations are scattered along the wardriving path.

6.3.2 Gridding

In [61], an approach called *gridding* was used (Figure 6.12). The RSSI histogram was constructed using all measurements inside each grid cell. The resolution has to be reduced to the size of the cell, but this approach can construct scalable radiomap by selecting an appropriate resolution that suits the needs of the selected location-based services. If the cell size is too small, we cannot obtain enough samples to create features for some pattern matching techniques. For instance, in order to apply the probability distribution-based techniques (e.g. Horus), enough number of measurements is needed to estimate the empirical distribution function. By the

Figure 6.12: Gridding

strong law of large numbers, the empirical distribution function $\hat{F}_n(t)$ converges to the true distribution $F(t)$ almost surely as $n \to \infty$. Gridding helps to obtain enough number of samples at the expense of the resolution.

In ROLAX, the gridding approach is extended to provide a better resolution while maintaining the size of the cell. It is called *overlapped gridding* (Figure 6.13). In this approach, the measurements in a small cell can be a part of multiple large cells. While the resolution remained as the size of the small cell, the measurements in the large cell are combined to create a RF fingerprint. The sizes of the small cell and the large cell have to be chosen carefully by considering the measurement

Figure 6.13: Overlapped Gridding

density and the target resolution required by the location-based service.

6.3.3 Interpolation of Measurements by Delaunay Triangulation

The locations of the measurement samples are scattered and non-uniformly distributed. Because of the data misses, caused by a variety of the factors including the capturing effect and the uncontrollable handoff between the BSs on the same frequency band, the data samples are scattered along the wardriving path. The gridding and the overlapped gridding can be extended by the *scattered data interpolation* fitting a smooth surface through the scattered samples.

In ROLAX, interpolation of the measurements is done by using *Delaunay triangulation* over two-dimensional space. A Delaunay triangulation for a set P of points in the plane is a triangulation $DT(P)$ such that no point in P is inside the circumcircle of any triangle in $DT(P)$ [62]. The Delaunay triangulation is used in scattered data interpolation because of its favorable properties such as the rejection of sliver-shaped triangles and the empty circumcircle property [63].

The interpolation, based on Delaunay triangulation, produces a surface of the form $V = F(X)$, where the location matrix is given by $X = [LAT\ LON]$ (LAT is a column vector of latitudes, and LON is a column vector of longitudes of the measurement locations), and the measurement (e.g. RSSI, CINR, etc.) associated with the locations X is given by V. The surface can be evaluated at any query location QX, using $QV = F(QX)$ where QX lies within the convex hull of X, and QV is the interpolated measurement value. When there is more than one measurement at a location, average value of the measurements is used for that location.

In ROLAX, linear interpolation was used. In linear interpolation, the interpolated value of a query point is given by the weighted summation of the measurement value on the vertices of the enclosing triangle.

Since the interpolated points cannot be outside the convex hull in the most of interpolation methods, the boundary of the convex hull polygon defines the coverage of the radiomap.[7] If the density of the measurement locations is sparse, then it would

[7]There are some interpolation techniques, such as nearest neighbor interpolation method, which can interpolate the points outside the convex hull.

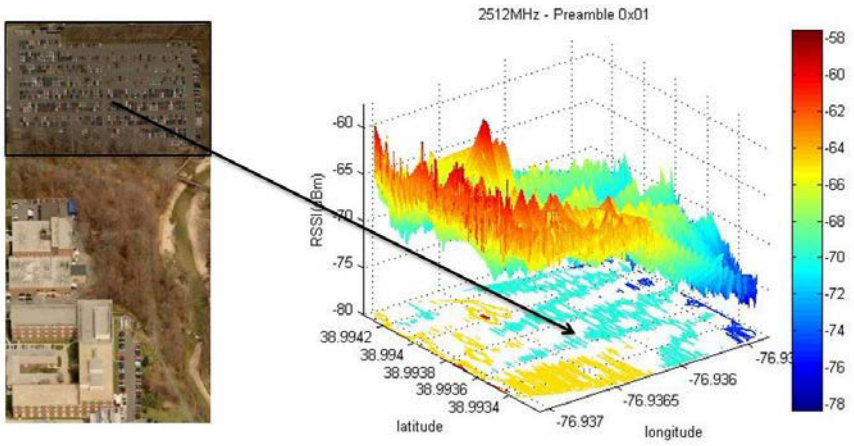


Figure 6.14: Linearly Interpolated RSSIs using Delaunay Triangulation

be hard to capture the small-scale variation over the space by the interpolation. An example of the interpolated RSSI radiomap is provided in Figure 6.14.

The interpolation method has to be chosen or designed so that the method can reflect the physical phenomenon and the systematic variation of the observed parameter over the space. When the RSSI is measured, it is better to design the interpolation method considering the fading of the RF signals. Particularly, the large-scale fading known as path loss (described in Chapter 3) needs to be considered. The path loss prediction is mostly based upon the statistical (empirical) method. If the shadow fading effect on the path loss in (3.1) is ignored, the path loss exponent γ and the system loss C in the equation have to be found.

The logarithm of the distance can be approximated by a linear function when the location is far from the BS (between 200 meters and 1000 meters). In the College Park, Maryland area, the average radius of the 4G cell is about 700 meters. In the longer distance, RSSI interpolations using the linear method do not deviate from

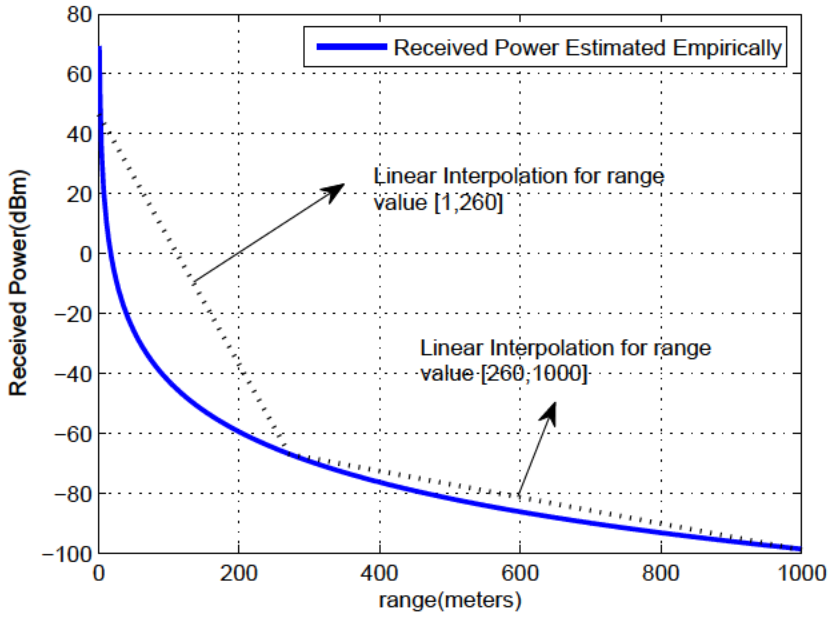


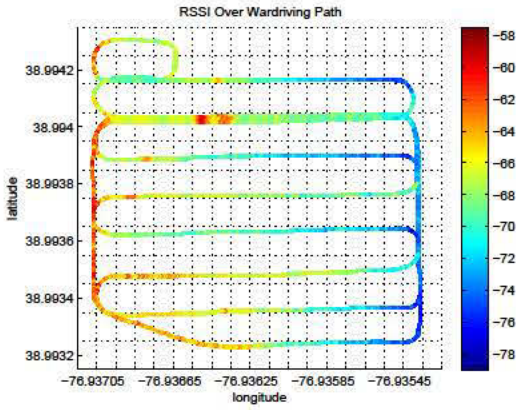
Figure 6.15: Linear Interpolation Applied to Interpolate RSSI Values

the theoretical path loss model (a linear function of the logarithm of the distance) a lot since they can be linearly approximated quite well. However, RSSI interpolation values deviate a lot from the underlying path loss model in the shorter distance. It is demonstrated in the Figure 6.15.

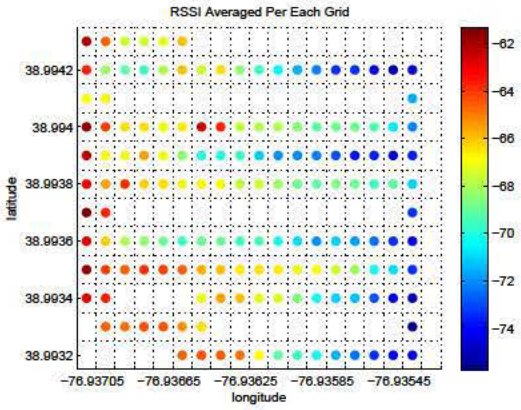
When the RTD is measured, the linear method works because the time of flight is proportional to the distance under the LOS (Figure 5.2(b)).

6.3.4 Illustration of Gridding and Interpolation Procedure

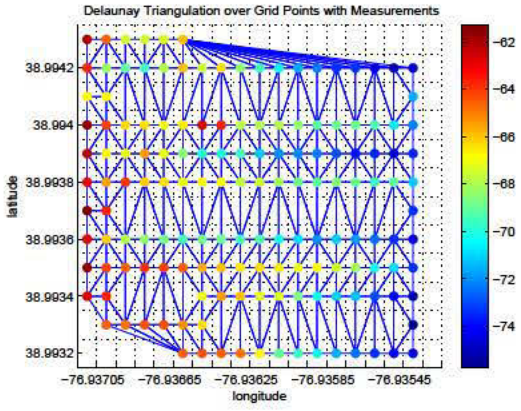
The procedure of gridding and interpolation is illustrated in Figure 6.16 for the grid size of 8 meters × 10 meters. The interpolated values, overlaid with the wardriving path, for a smaller grid size (0.8 meter × 1 meter) is provided in Figure 6.17.



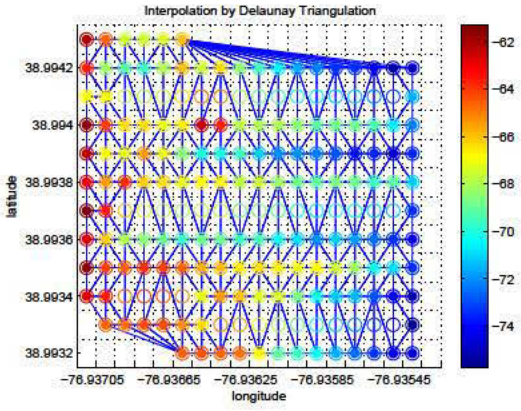
(a) RSSI is Measured over Wardriving Path.



(b) Mean of RSSI Values Is Calculated for Each Grid.



(c) Delaunay Triangulation Is Done over Centroids of Grids.



(d) Linear Interpolation Is Done by Using the RSSI Values on the Vertices of Triangles.

Figure 6.16: Illustration of Gridding and Interpolation Procedure (RSSI, Grid Size = 10 meters × 8 meters)

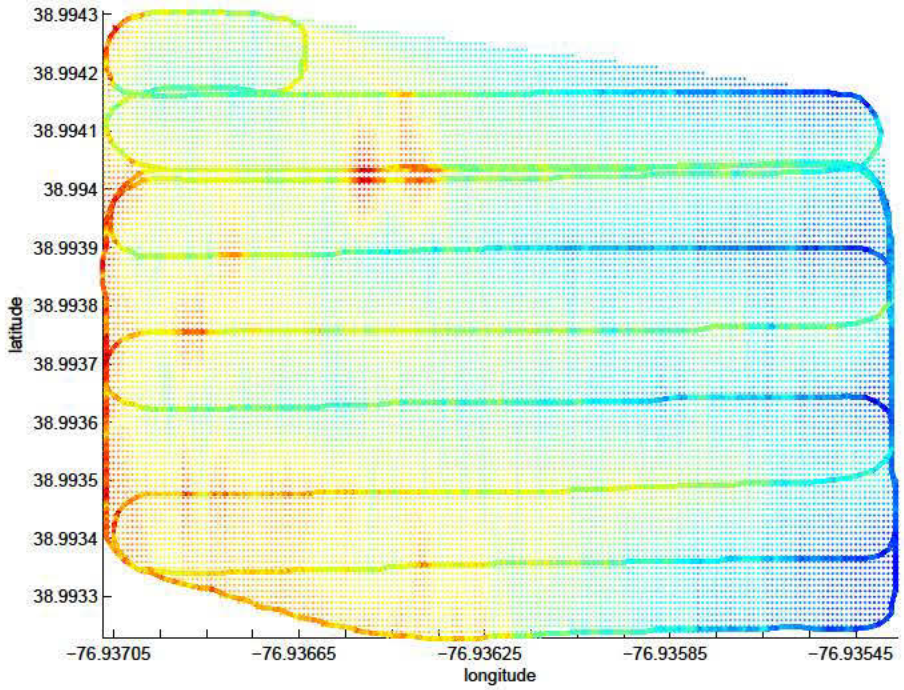


Figure 6.17: Interpolations with Smaller Grid (RSSI, BS on 2512kHz, Grid Size = 1 meter × 0.8 meter)

6.4 Fingerprint Feature Selection

6.4.1 Comparison of Underlying Measurements

Features are extracted from the underlying measurements such as RSSI, CINR, and RTD. Detailed description of each measurement type is provided in the Appendix A. Each measurement type has different limitations, and they are summarized below.

- **Types of BSs** : While RSSI and CINR can be measured with regard to any BSs by passively monitoring the signals from them, RTD can be measured only with regard to the serving or the anchor BSs. Since many service providers provide web-based authentication, the MS may be able to measure the RTD with a BS, which belongs to a foreign network.
- **Measurement Frequency** : RSSI and CINR measurements are made for each downlink preamble of the frames. In OFDMA PHY of IEEE 802.16e, preambles are typically sent once every 5 milliseconds, but it can be between 2 milliseconds and 20 milliseconds (2 msec, 2.5 msec, 4 msec, 5 msec, 8 msec, 10 msec, 12.5 msec, and 20 msec) [64]. The measurement frequency of the RTD is much less than that of RSSI and CINR. Since RTD measurement is typically involved with the calculation of the time advance, RTD measurement can be updated during the ranging, in which the correct timing offset between the MS and the BS is acquired. Thus, RTD measurement can be made only after the initial ranging between the MS and the BS. In addition, the RTD measurement value is typically updated during the periodic ranging.

In OFDMA PHY of IEEE 802.16e, the periodic ranging interval is defined by T4, which can be as large as 35 seconds [64]. In MAXWell 4G network, the periodic ranging interval was set to five seconds at the time this work was done (2012). The device driver used in this work updates the RTD value once every three seconds. Thus, even though the periodic ranging interval can be reduced, the RTD sampling frequency could not be increased over 1/3 Hz. This low frequency makes it hard to use the wardriving because the low measurement frequency results in coarse resolution.[8] The sample RTD measurements confirmed that the RTD measurement frequency is fairly low to be used to create the radiomap with a fine resolution (Figure 6.18).

- **Requirement for the demodulation lock** : While RSSI can be measured without the receiver demodulation lock, CINR and RTD can be measured only after the receiver demodulation lock is acquired. Particularly, the RTD can be measured only after the downlink synchronization and the acquisition of the uplink parameters are made.

The summary of the comparison between the underlying measurement types is provided in the Table 6.4.

In 4G WiMAX, the MS performs transmit power control (TPC) during its operation to combat the near-far effect while the BS does not perform the transmit power control. It makes harder to obtain consistent RSSI and CINR readings on uplink since the MS adjusts its transmit power. In addition, the power on the UL

[8] 22 meters of resolution if the speed of the wardriving vehicle is 10 miles/hour and the RTD measurement is performed every five seconds (resolution = $\{speed\} / \{measurement\ frequency\}$).

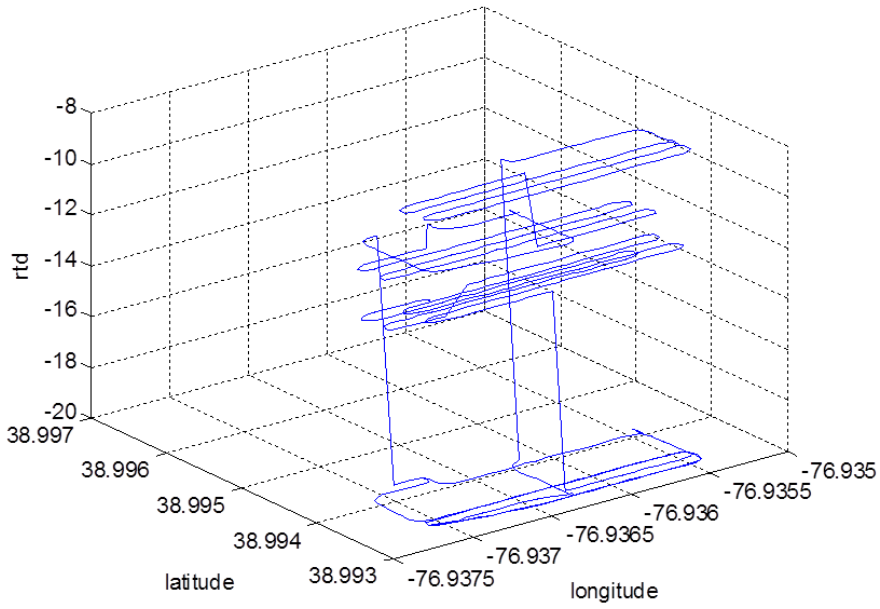


Figure 6.18: Example of RTD Measurements by Wardriving (MAXWell BS on 2512GHz)

Table 6.4: Comparison of Measurement Types

	Types of BSs	Measurement Frequency	Demodulation Lock Required	UL Parameter Acquisition Required
RSSI	Any BSs	50 Hz – 500 Hz	No	No
CINR	Any BSs	50 Hz – 500 Hz	Yes	No
RTD	Serving BSs Anchor BSs	maximum 1/35 Hz	Yes	Yes

channel is dependent upon the number of active carriers, so it is hard to obtain consistent reading when the UL traffic is dynamically changing. Thus, it would be best to measure the preamble of the DL channel rather than UL channel for the measurements in the RF fingerprinting. The BS can obtain the measurement at the MS side (DL channel) from the scanning result report (MOB_SCN-REP) message.

6.4.2 Relationship between Underlying Measurements

Throughout the University of Maryland, College Park campus, the RSSI, RTD, and CINR values were measured with regard to each RF sector in MAXWell 4G network using the periodic neighbor report. A high positive correlation is observed between RSSI and CINR, but the correlation is low and negative between RSSI/CINR and RTD. This value is negative because, on the large scale, RSSI decreases and RTD increases as the range between the MS and the BS increases (Table 6.5 and Figure 6.19).

Since the correlation between RSSI/CINR and RTD is low, combining RSSI and RTD has the potential to increase the uniqueness of the RF fingerprints. RSSI and RTD are less correlated under the non-DLP as discussed in the Section 5.1.

6.4.3 Feature Extraction from Underlying Measurements

In the RF fingerprinting, the features are extracted from the signal measurements (e.g. RSSI, CINR, RTD) during both the offline phase and online phase. If a shorter positioning delay is preferred, a single measurement can be used as a

BSID (last 8 octets)	$\rho_{RSSI,CINR}$	$\rho_{RSSI,RTD}$	$\rho_{CINR,RTD}$
01:00:00:11	0.95	-0.43	-0.44
01:00:00:12	0.98	-0.64	-0.61
01:00:00:13	0.97	-0.69	-0.64

Table 6.5: Correlation Coefficients between Measurement Types

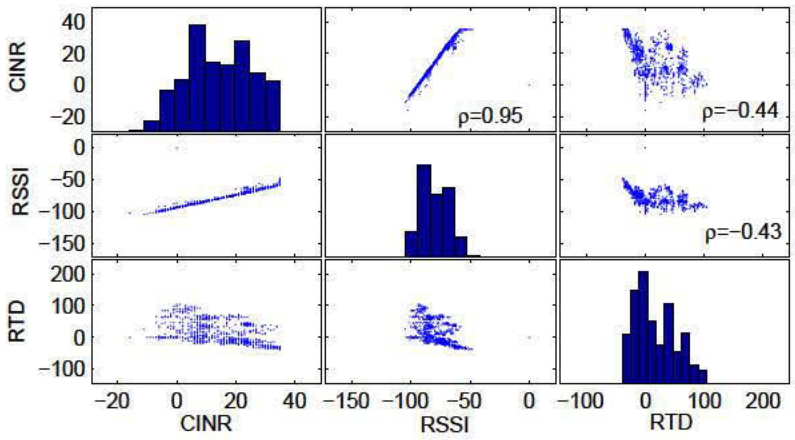


Figure 6.19: Scatter Plots between Measurement Types

feature by itself. In ROLAX, features are extracted from RSSI and CINR. Since the RTD measurement frequency is fairly low, features are not extracted from RTD. The same set of underlying measurements, used in the DLP detection, are used for the RF fingerprinting: RSSI per each antenna and the CINR. Combined RSSI and RSSI difference ($RSSI_d$) are calculated from the underlying measurements.

In ROLAX, the mean of the RSSI ($Mean[RSSI]$) and the mean of the CINR ($Mean[CINR]$) are used as features in addition to DLP detection features. The mean of RSSI is chosen since the received signal strength, on a large scale, is affected by the distance and the presence of the obstacles between the MS and the BS. The mean of the RSSI is used primarily in the previous research such as RADAR. An important observation is made during the DLP detection study: while two locations have the similar average level of the signal strength (e.g. RSSI) or the signal quality (e.g. CINR), the underlying radio characteristics can be fairly different. For example, a location under the DLP condition and another location under the non-DLP condition, from a BS, may have the similar level of the RSSI and/or CINR, but they may have different spectral signal characteristics (e.g. level crossing rate).

Therefore, the time and spectral features evaluated in DLP detection are also included in the feature set. The SR_{RSSI} is not selected due to its poor performance in discriminating the DLP condition. The list of the features used in constructing the RF fingerprints is as follows:

- Mean of the $RSSI$ ($Mean[RSSI]$)
- Mean of the $CINR$ ($Mean[CINR]$)

- Standard deviation of the $RSSI$ ($Std[RSSI]$)
- Standard deviation of the $CINR$ ($Std[CINR]$)
- Standard deviation of the $RSSI_d$ ($Std[RSSI_d]$)
- Mean of $RSSI_d$ ($Mean[RSSI_d]$)
- Level Crossing Rate of the $RSSI$ (LCR_{RSSI})
- Level Crossing Rate of the $CINR$ (LCR_{CINR})
- Spectral Centroid of the $RSSI$ (SC_{RSSI})
- Spectral Centroid of the $CINR$ (SC_{CINR})
- Spectral Roll-off of the $CINR$ (SR_{CINR})
- Spectral Roll-off of the $RSSI_d$ (SR_{RSSI_d})
- Spectral Flux of the $RSSI$ (SF_{RSSI})
- Spectral Flux of the $CINR$ (SF_{CINR})

6.4.3.1 Offline Feature Extraction

During the offline phase, for each grid, the signals are combined, and features are extracted. When a single feature (e.g. mean of RSSI) is generated per each grid, all measurements are combined to create a single feature. When multiple feature sets are generated per each grid, in order to have the same number of feature vectors for each grid, artificially generated feature sets are added.

Let us denote the number of measurements in series needed to generate a feature, by N_{mf}, the number of measurements made in a grid i with regard to BS j by n_{ij}, the number of feature sets generated for each grid by N_{fs}, and the required

minimum number of measurements to generate the features by N_{min}. Thus, in order to produce N_{fs} feature sets in a grid, $N_{fs} \cdot N_{mf}$ measurements are required. N_{min} is much smaller than $N_{fs} \cdot N_{mf}$ ($N_{min} << N_{fs} \cdot N_{mf}$).

When multiple feature sets are generated per each grid, the offline features are generated as follows:

For each grid i and for each BS j,

- **Case 1** if $n_{ij} >= N_{fs} \cdot N_{mf}$, choose $N_{fs} \cdot N_{mf}$ measurements in series and extract features from them.
- **Case 2** if $n_{ij} < N_{fs} \cdot N_{mf}$ and $n_{ij} >= N_{min}$, choose $\left\lfloor \frac{n_{ij}}{N_{mf}} \right\rfloor \cdot N_{mf}$ measurements in series and extract $\left\lfloor \frac{n_{ij}}{N_{mf}} \right\rfloor$ feature sets from them. Create a parametric or non-parametric probability distribution from the extracted features. Generate $N_{fs} - \left\lfloor \frac{n_{ij}}{N_{mf}} \right\rfloor$ feature sets according to the created distribution.
- **Case 3** if $n_{ij} < N_{min}$, assign "NaN" value for the features with regard to BS j in the grid i.

6.4.3.2 Online Feature Extraction

In ROLAX, the online phase measurements can be made by 1) the scanning operation or 2) by making a physical connection with each BS. With the scanning, the measurement frequency is fairly limited - in the order of one second for each BS. If a connection can be made to the BS, the measurement frequency is the same as the BS's preamble transmission frequency.

When the scanning is used for the signal collection, the extraction of the

spectral features is not very feasible (200 measurements take in the order of 200 seconds). However, when the physical connection is actually made for the signal collection, the features can be extracted in the same way as when the offline feature extraction is made. If multiple features are used in ROLAX, the MS makes a connection to each BS to collect enough number of measurement samples.

6.5 Pattern Matching for RF Fingerprinting

In the previous research about RF fingerprinting, a number of techniques have been used for matching between online and offline RF fingerprints. They include K-nearest neighbor (K-NN), probabilistic method, support vector machine (SVM), artificial neural networks (ANN), etc. In [11], pattern matching techniques for the RF fingerprinting were compared to each other. It showed that the K-NN performs better than neural networks in accuracy and precision at the expense of complexity during the online phase. K-NN performs better than parametric probabilistic techniques with normality assumptions, but other previous research has demonstrated that non-parametric probabilistic technique performs better than K-NN in terms of accuracy [13].

6.5.1 K-NN (K-Nearest Neighbor)

When a single feature is used in ROLAX, K-NN pattern matching is used.

6.5.1.1 Distance Measures

To use the K-nearest neighbor (K-NN), similarity (distance) measure has to be defined.

The grid points in the radiomap is given as $\{g^j\}_{j=1}^{L}$, where L is the number of locations in the radiomap. Let an offline feature vector recorded at the location g^j in the radiomap (the RF fingerprint at location g^j) denoted by $\hat{M}_{off}^j = [\hat{M}_1^j\ \hat{M}_2^j \dots \hat{M}_N^j]$, where N is the number of the base stations in the radio map, and M_i^j is the offline feature vector or scalar with regard to BS i at the location g^j. Let the online feature vector denoted by $\hat{Y}_{on} = [\hat{Y}_1\ \hat{Y}_2 \dots \hat{Y}_N]$ where $\hat{Y}_i$ is the online feature vector or scalar with regard to BS i. The distance measures have to be defined between the online feature and the offline feature.

If the measurement cannot be made with regard to the BS i, then the corresponding feature is set to NaN (scalar) or $[NaN\ NaN \dots NaN]^T$ (vector). It may be because the BS is out of range from the MS, or the MS fails to detect this BS due to some reasons. The difficulty comes in dealing with this NaN value caused by data misses.

Let the distance between a RF fingerprint at location g^j and the online phase feature with regard to a BS i denoted by d_i^j. d_i^j is provided in (6.6) where f is the distance function to be chosen.

$$d_i^j = f(\hat{M}_i^j, \hat{Y}_i) \tag{6.6}$$

Distance measure, between a RF fingerprint $\hat{M}_{off}^j$ at location g^j and the $\hat{Y}_{on}$,

can be defined as in (6.7).

$$D^j = w\left(\hat{d^j}\right) \tag{6.7}$$

where $\hat{d^j} = [d_1^j\ d_2^j \dots d_N^j]$, and w is a function calculating the overall distance given $\hat{d^j}$.

$f(\hat{M_i^j}, \hat{Y_i}) = (\hat{M_i^j} - \hat{Y_i})^2$ and $w(\hat{d^j}) = \sqrt{\sum_{i=1}^{N} d_i^j}$ result in the classical *Euclidean distance*. *Manhattan distance* is less sensitive against outliers than the Euclidean distance is. In case of Manhattan distance, $f(\hat{M_i^j}, \hat{Y_i}) = \left|\hat{M_i^j} - \hat{Y_i}\right|$ and $w(\hat{d^j}) = \sum_{i=1}^{N} d_i^j$.

6.5.1.2 Algorithms Dealing with Missing Values

Since the online and offline feature may contain NaN values for some non-detected BSs, the distance has to be defined considering the non-detected BSs. In [12], two algorithms dealing with the non-detected BSs were introduced. In *Classical Fingerprinting* algorithm, NaN values are ignored. In *BS-strict* algorithm, infinite penalty is given to the non-matching NaN values. For instance, if a BS is observed only in the online phase, then an infinite penalty is given to all candidate locations with measurements with regard to that BS. While [12] showed the accuracy of BS-strict is better than that of classical fingerprinting, our data show a different result. Because the data misses are frequent events, and the data misses are not always generated by having the MS far from the BSs, the BS-strict algorithm tends to result in the worst performance with the 4G data collected in the experiments. One

of the causes for the data misses is due to the capturing effect. It is probable that [12] collected data from a WiMAX network with different configuration. For instance, the WiMAX networks, from which they collected the data, may be operating over non-overlapping frequency bands.

In this work, in addition to the algorithms in the [12], new algorithms have been designed and used.

Classical Fingerprinting (Alg-CF) In classical fingerprinting, the offline and online measurement pair with non-matching NaN values is ignored, and the distance is set to zero. When both offline and online measurements are NaN, the distance is also set to zero. When both of them are not NaN, the distance is calculated. The distance function in classical fingerprinting is provided in (6.8).

$$f(\hat{M}_i^j, \hat{Y}_i) = \begin{cases} g(\hat{M}_i^j, \hat{Y}_i) & \text{if } \hat{M}_i^j \neq NaN \text{ and } \hat{Y}_i \neq NaN \\ 0 & \text{otherwise} \end{cases} \tag{6.8}$$

where g is a distance function defined for two non-NaN inputs.

As an exception case, when $\hat{M}^j$ and $\hat{Y}$ are all NaNs, D^j is set to ∞.

BS-Strict (Alg-BSS) In the BS-Strict algorithm, infinite penalty is given to the online and offline feature pair if either of them is NaN. The distance function in BS strict fingerprinting is provided in (6.9).

$$f(\hat{M}_i^j, \hat{Y}_i) = \begin{cases} 0 & \text{if } \hat{M}_i^j = NaN \text{ and } \hat{Y}_i = NaN \\ g(\hat{M}_i^j, \hat{Y}_i) & \text{if } \hat{M}_i^j \neq NaN \text{ and } \hat{Y}_i \neq NaN \\ \infty & \text{otherwise} \end{cases} \tag{6.9}$$

where g is a distance function defined for two non-NaN inputs.

Modified Classical Fingerprinting (Alg-MCF) Modified Classical Fingerprinting is designed by combining ideas from the classical fingerprinting and BS Strict fingerprinting. The classical fingerprinting may result in higher errors when the number of matching non-NaN pairs is small. When the K-nearest neighbor pattern matching is used, the performance of classical fingerprinting is not good, particularly when the number of neighbors K used in pattern matching is small. It is demonstrated in Section 6.6.2.4.

Two modifications are done to improve the performance. First, only when the number of non-NaN pairs is larger than a certain value, the distance is calculated.[9] It is intended to emphasize the identity of BSs in calculating the distance. Otherwise, infinite penalty is given. Secondly, the distance measure is normalized by using the number of matching non-NaN pairs.

Fingerprinting with an Assumed Threshold Value (Alg-TH) In this algorithm, an assumed threshold value replaces the NaN value. It works well if the data misses are primarily because the signal from the BSs is not fair enough. For instance, the minimum RSSI value of -123 dBm can replace NaN value if the RSSI feature is

[9]Threshold value of three is used in the experiment in Section 6.6.2

used. It may not perform well if the data misses are due to the capturing effect. The captured signal may be much larger than the assumed threshold value. The distance function in classical fingerprinting is provided in (6.10). In ROLAX, the similar approach is used to deal with missing values in ANN when multiple features are used.

$$f(\hat{M}_i^j, \hat{Y}_i) = \begin{cases} 0 & \text{if } \hat{M}_i^j = NaN \text{ and } \hat{Y}_i = NaN \\ g(\hat{M}_i^j, \hat{Y}_i) & \text{if } \hat{M}_i^j \neq NaN \text{ and } \hat{Y}_i \neq NaN \\ g(\hat{M}_i^j, TH_{type}) & \text{if } \hat{M}_i^j \neq NaN \text{ and } \hat{Y}_i = NaN \\ g(TH_{type}, \hat{Y}_i) & \text{if } \hat{M}_i^j = NaN \text{ and } \hat{Y}_i \neq NaN \end{cases} \tag{6.10}$$

where TH_{type} is the threshold value set for a feature type (e.g. mean RSSI) *type*, and g is a distance function defined for two non-NaN inputs.

6.5.1.3 BS Filtering

The measurement vector created during offline phase contains missing data frequently. The sources of missing data can be the marginal signal quality as well as the capturing effect.

The frequency reuse pattern in 4G WiMAX can be denoted by (N_c, N_s, N_n) where N_c is the number of BS sites per cluster, N_s is the number of sectors per BS site, and N_n is the number of unique frequency channels required for reuse (See Section 2.3.2). In MAXWell 4G network, (1, 3, 1) pattern had been used by the end of 2010. Since 2011, it has used (1, 3, 3) pattern.

As discussed in Section 2.3.2, a MS may be in a communication range from multiple BS sectors operating on the same frequency. In the frequency reuse pattern with $N_n = 1$, all sectors are operating on the same frequency band. The coverage by each sector is overlapped with the coverage by other sectors (See Figure 2.3). Particularly, at the intersection of the sectors, a MS observes multiple BSs. Due to fading, the signal equality from each sector may vary over time. Even though $N_c > 1$ and $N_n > 1$, multiple BSs on the same frequency band may be observed at a location because the signals from other network cluster can travel to a network cluster, which shares the same frequency channels, with a weaker power.

Since the 4G network may not be fully loaded, the weaker signal can be detected just because of no transmission from other BSs on the same frequency band. When more than two sector antennas are transmitting over the overlapped time duration on the same set of carriers, the frames with weaker power are captured by the stronger ones.

In addition, when the signal is around the marginal quality, it can be easily affected by the fading so that the signal quality frequently goes below the receiver sensitivity of the MS.

In an experiment, scannings were attempted 30 times at nine locations where eight 4G BSs were visible (Figure 6.20). Out of 270 attempts, only 7% of the scannings can detect all eight BSs. In this experiment, the BS-A and BS-B (with square marker) were operating on the same 2.617 GHz. Even though the signals from a BS-A were with fair RSSI (around -50 dBm), the scanning missed the BS-A quite often because the signals from BS-B captured the signals from BS-A. The

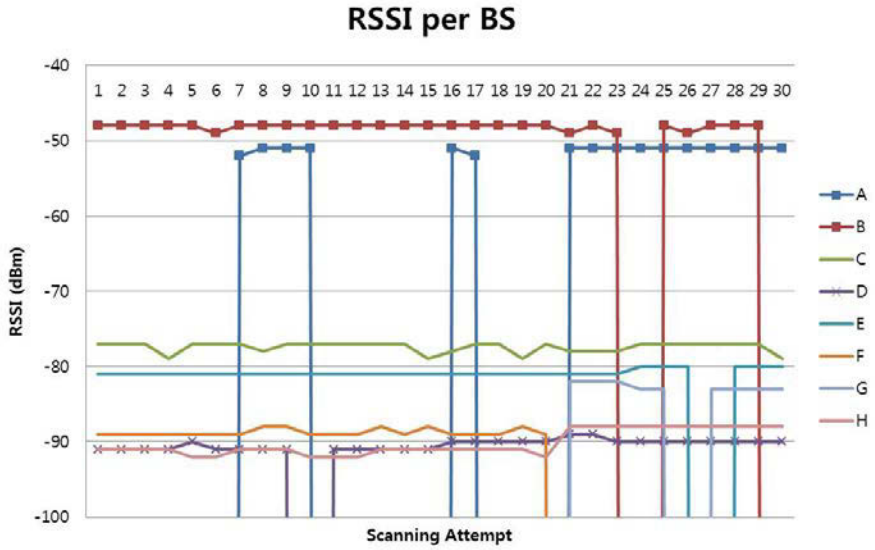


Figure 6.20: WiMAX Scanning Misses

signals from BS-D were missed some times because it had marginal signal quality.

These observations help in creating the following rules to deal with the scanning misses in 4G:

- **Scanning misses by power capture**
  - Only one BS per frequency band is considered in matching the RF fingerprints. Or,
  - The frequency with multiple BSs is discarded.
- **Scanning misses by marginal signal quality** If the signal quality is below a certain threshold, the measurement is not considered in matching the RF fingerprints.

6.5.2 Artificial Neural Networks

Artificial neural networks (ANN) are primarily used as the pattern matching technique in ROLAX when multiple features are used because (1) ANN can prioritize the feature by adjusting the weights during the modeling, and (2) ANN can deal with missing value (feature with "NaN" value) during the offline modeling and the online classification.

Let us denote the number of BSs in the target area by N_{BS}, the number of features generated with regard to each BS by N_{fb}, and the number of grids by N_g.

The configurations of the artificial neural networks used for RF fingerprinting are as follows:

- Nodes in the neural networks: Sigmoid.
- Algorithm: Back propagation algorithm.
- Number of Input Nodes: $N_{fb} \cdot N_{BS}$
- Number of Hidden Layers: 1
- Number of Neurons: $\left\lfloor \frac{N_{fb} \cdot N_{BS} + N_g}{2} \right\rfloor$
- Number of Output Nodes: N_g
- Learning Rate (The amount the weights are updated): 0.3
- Momentum (Applied to the weights during updating): 0.2
- Attribute Normalization: Yes
- Training Time (The number of epochs to train through): 500

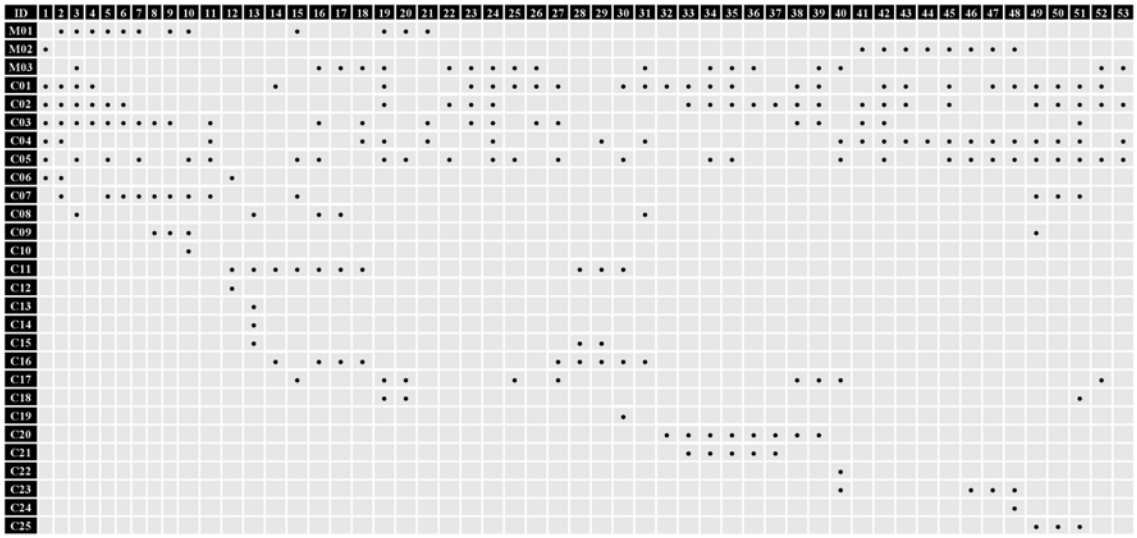


Figure 6.21: BS Detection Bitmap at 53 locations at University of Maryland, College Park (Mxx:BSs in Maxwell 4G Network, Cxx: BSs in Clear 4G Network)

6.6 Experimental Results

6.6.1 4G Site Survey in College Park, Maryland

To see if the RF fingerprinting in 4G is feasible, 53 locations in University of Maryland, College Park were chosen, and 4G scanning was performed. The list of the BSs observable at each location is provided in the Figure 6.21. 28 BSIDs including 3 BSIDs in the MAXWell 4G network and 25 BSs in Clear 4G network were observed. Out of 53 locations, only two locations share the same set of BSs.

6.6.2 RF Fingerprinting with Mean of RSSI Feature

During the offline phase in this experiment, wardriving was performed to collect the signals for generating the radiomap. During the online phase, 24 locations

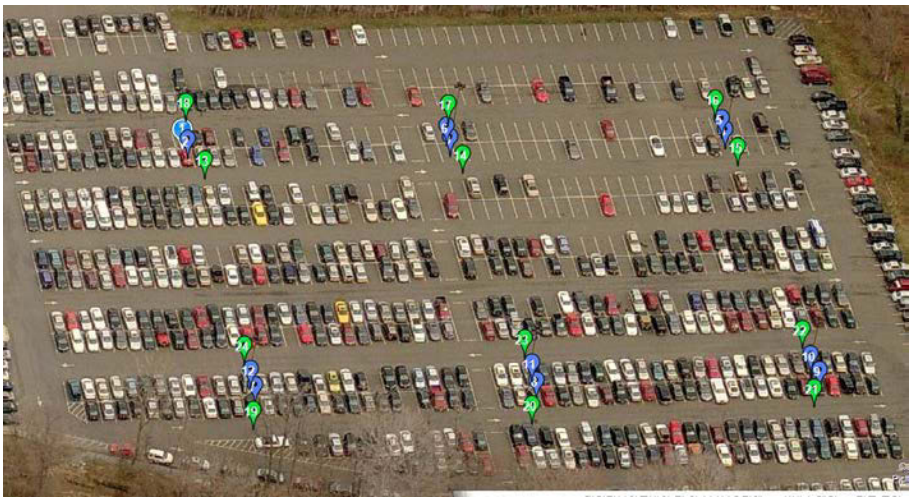


Figure 6.22: Locations in Online Phase for Experiment in Section 6.6.2

were selected, and signals were collected (Figure 6.22). The experiment was done at a campus parking lot in College Park, Maryland. The size of parking lot is 123 meters by 150 meters. Median distance errors were calculated over all locations in the online phase. The achieved accuracy obtained during this experiment showed a high potential in achieving better accuracy than those achieved in the previous research.

The number of BSIDs observed in the test area was twelve. The following set-ups were used.

- Measurement technique: wardriving and techniques described in Section 6.3
- Observed measurements: RSSI
- Pattern matching techniques: K-nearest neighbor (equal weight for each neighbor)
- Distance measure: Euclidean distance and Manhattan distance
- Feature: mean

- Missing value handling algorithm: classical fingerprinting (Alg-CF), modified classical fingerprinting (Alg-MCF), BS-strict fingerprinting (Alg-BSS), and fingerprinting with an assumed threshold value (Alg-TH)
- Continuous domain estimation: center of mass
- Interpolation of the measurements over space: by Delaunay triangulation with linear interpolation method. Duplicate measurements are replaced by the mean value.
- Gridding: Grid-L, Grid-S, Grid-O, and Grid-LS (See Table 6.6)
- BS Selection: All BSs vs. BSs in MAXWell 4G network.
- BS and Frequency Filtering: All frequencies with detected BSs vs. frequencies with a single BS.
- Outlier Remover: Outliers by systematic errors were removed.[10]
- Distance calculation by the latitude and longitude values: Great-circle distance in radian multiplied by the mean radius of the Earth in meters (6371000 meters)

The number of measurements with regard to each BS was mostly between 150,000 and 200,000. The locations where the measurement was made for each BS are demonstrated in Figure 6.23. This figure shows that the measurements were not made completely for those BSs captured by other BS(s) on the same frequency band.

After the interpolation, about 85–90% of the area in the radiomap was covered

[10] All RSSI measurements with -123 dBm with the modem state of lost physical synchronization or unknown state are removed.

	Grid Size (meters by meters)	Overlapped Grid	Interpolation Resolution (meters by meters)	Number of Points
Grid-L	8 by 10	No	8 by 10	228
Grid-S	0.8 by 1	No	0.8 by 1	19314
Grid-O	8 by 10 (supercell) 0.8 by 1 (subcell)	Yes	0.8 by 1	22800
Grid-LS	8by 10	No	0.8 by 1	20091

Table 6.6: Gridding and Interpolation Options

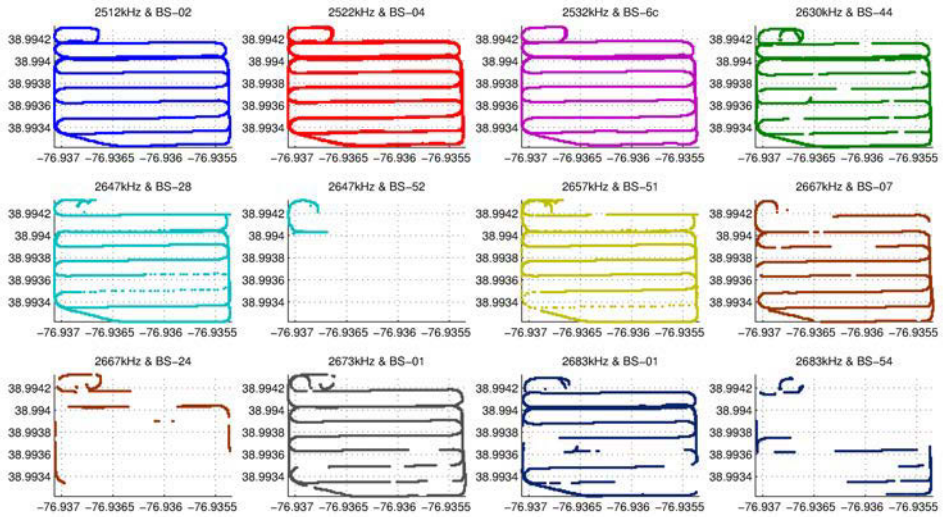


Figure 6.23: Locations of the Measurements for Each BS

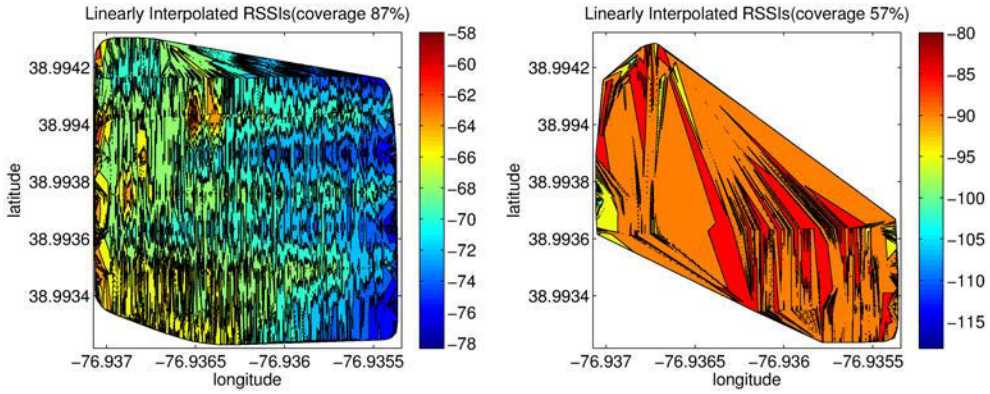


Figure 6.24: Heatmap of Linearly Interpolated RSSIs by Coverage

with regard to the most of the BSs. For those BSs with captured signals and marginal signal quality, the number of measurements and the portion of the covered area were limited (5% in the worst case). The comparison between BS radiomaps with a large coverage and a small coverage is provided in Figure 6.24. The observed BSs' configurations, the number of measurements, and the percentage of the area covered by interpolation (area within the convex hull) compared to the whole area is provided in the Table 6.7. Two 4G cards were used as MSs, and each MS collected signals from different set of BSs.

The distance error by randomly choosing the positions ranges between 60–80 meters. This value can be regarded as the worst-case error and can be compared to the achieved distance error.

6.6.2.1 Improvement by Interpolation

The distribution of the interpolated measurement values was calculated for each BS in order to see the range of the measurement in the test area as opposed to

Table 6.7: BS Configurations and Coverage

Card (last 2 octets)	Freq (kHz)	BSID (last 8 octets)	Preamble Index	Number of Measure-ments	Coverage
c6	2512000	01:00:00:11	2	168484	0.87
90	2522000	01:00:00:12	4	152838	0.87
c6	2522000	01:00:00:12	4	149630	0.87
90	2532000	01:00:00:13	6c	143484	0.87
90	2630500	02:02:0c:74	44	198478	0.87
c6	2647000	02:02:0a:54	28	182662	0.91
c6	2647000	02:02:14:95	52	10841	0.05
90	2657000	02:02:1a:01	51	150760	0.91
c6	2667000	02:02:29:a6	24	108637	0.89
c6	2667000	02:02:27:d6	7	50175 0	0.63
90	2673500	02:02:24:be	1	163655	0.9
c6	2683500	02:02:1c:c1	54	139329	0.87
c6	2683500	02:02:54:be	1	78612 0	0.57

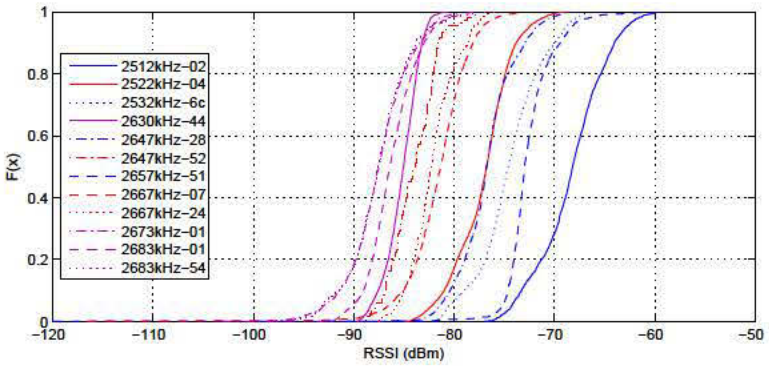


Figure 6.25: Distribution of the Interpolated RSSIs under Test Area

the variability of the measurements at a given location. This distribution was calculated before outliers were removed. As there were no BSs located in the proximity of the test area, the small-scale variation might dominate over the large-scale variation. Hence, the location determination is a challenging task under this circumstance, and a fair accuracy is very hard to be achieved by a single online measurement.

With the grid size of 0.08 meter $\times$ 0.1 meter, the difference between the third and the first quantile interpolated RSSI values was between 1.9–3.8. With the grid size of 8 meters $\times$ 10 meters, the range was between 2.2–4.4. In case of CINR, the range was between 2.1–3.7 with the smaller grid size. Thus, about 50% of the RSSI and CINR measurements values were within 2dB range. The CDF of the interpolated RSSI, with the grid size of 0.8 meter $\times$ 1 meter, is provided in Figure 6.25.

Distance error was calculated in order to see the gain by using the interpolation on the measurement data. For the comparison, the following settings were used.

- Gridding: Grid-L and Grid-S.

- Distance measure: Euclidean distance and Manhattan distance.
- Distance error: calculated and averaged over the results from two MSs.
- BS selection: all BSs
- Missing data handling algorithm: Alg-MCF
- Outliers by systematic errors are disregarded.

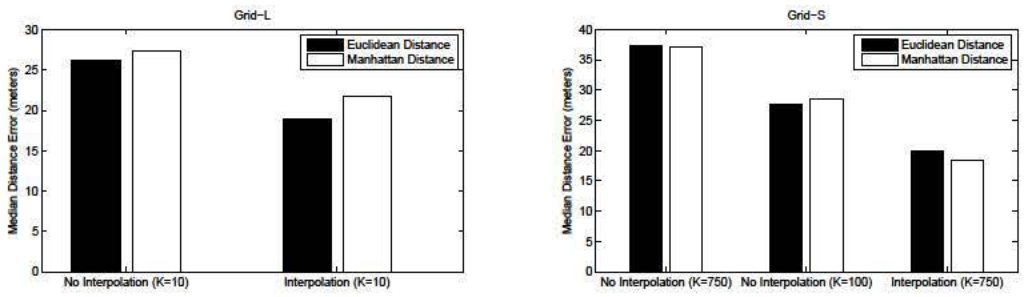
When the interpolation is not used, many points in radiomap tend to contain larger number of NaNs corresponding to invisible BSs simply because there are points, which were not visited during the wardriving. Thus, there are a smaller number of points with a rather complete measurement vector. Much smaller K value for K-nearest neighbor has to be chosen when the interpolation is not used. Otherwise, the distance error gets worse.

This gain is demonstrated in Figure 6.26. The gain obtained by applying the interpolation was different by the gridding options and the distance measures. The gain by the interpolation ranged between 20–35% when appropriate K value was selected for each interpolation and gridding option pair. The figure shows that smaller K value is preferred when the interpolation is not used.

6.6.2.2 Improvement by Removal of Outliers

For the comparison, the following settings were used.

- Gridding: Grid-L (K=10), Grid-S (K=750) and Grid-O (K=1000).
- Distance measure: Euclidean distance and Manhattan distance.
- Distance error: calculated and averaged over the results from two MSs.



(a) Grid-L (b) Grid-S

Figure 6.26: Improved Distance Error by Interpolation

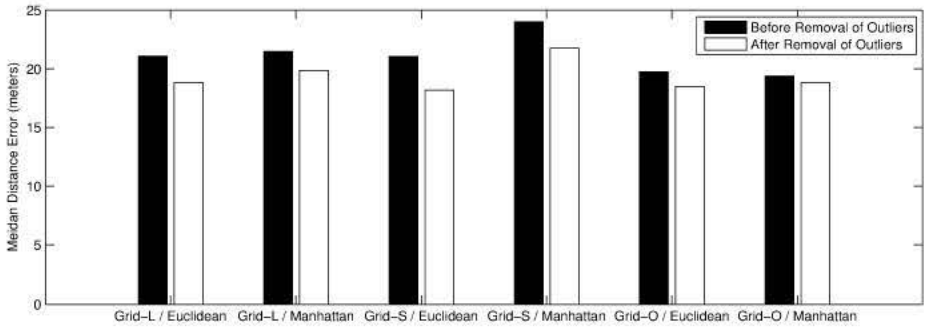


Figure 6.27: Improved Distance Error by Removal of Outliers

- BS selection: all BSs
- Missing data handling algorithm: Alg-MCF

The portion of outliers generated by the lost physical synchronization was very small, which was 0.027% of total measurements (458 occurrences out of 1697585 measurements), but removal of them increased the distance error performance by 3–13%. This gain is demonstrated in Figure 6.27.

6.6.2.3 Improvement by Gridding

By decreasing the resolution of the gridding and the interpolation, a radiomap with a higher resolution was generated. If the grid size is too small, then the radiomap tends to have overfitted interpolation values. In addition, higher resolution results in a higher computation time in both offline and online phases. Four different gridding settings were used and compared. It is described in Table 6.6. The following settings were used.

- Number of neighbors in K-NN: K = 10 for Grid-L, K = 750 for Grid-S, and K= 1000 for Grid-O and Grid-LS.
- Distance measure: Euclidean distance
- Missing data handling algorithm: Alg-MCF
- Distance error: Calculated and averaged over the results from two MSs
- BS Selection: All BSs
- Outliers by systematic errors are disregarded.

The RSSI radiomap with regard to a BS in MAXWell 4G network is provided in Figure 6.28 over the four gridding options.

The results show that Grid-O and Grid-LS have a small gain over Grid-S in achieving a smaller distance error. Compared to Grid-S, Grid-O and Grid-LS improved the distance error by 8–15% when Alg-MCF and Euclidean distance were used. The result is provided in Figure 6.29. Grid-LS with Euclidean distance measure achieved a median distance error of 16.9 meters. In this experiment, high density of samples were obtained. It is expected that the distinction between gridding

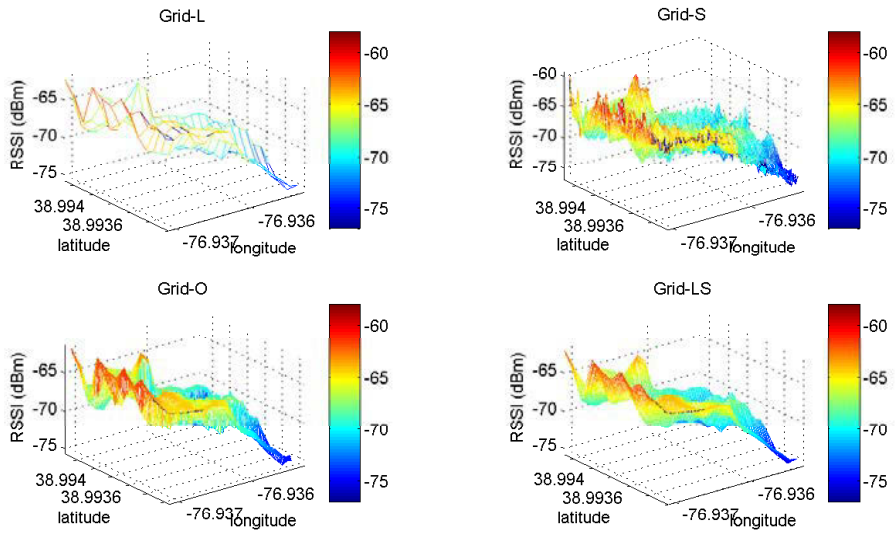


Figure 6.28: RSSI Radiomap for BS in MAXWell 4G Network over Gridding Options

options may be clearer when the measurements are sparse.

6.6.2.4 Improvement by Missing Value Handling Algorithms

The distance error was calculated over different data algorithms. For the evaluation, the following settings were used.

- Gridding: Grid-L with K=10.
- Distance Measure: Euclidean distance and Manhattan distance.
- Distance Error: calculated and averaged over the results from two MSs.
- BS Selection: all BSs, MAXWell BS.
- Outliers by systematic errors are disregarded.

The result is provided in Figure 6.30.

Alg-BSS resulted in the worst accuracy over all cases, and sometimes it resulted in worse result than random selection. Thus, this algorithm is regarded non-usable

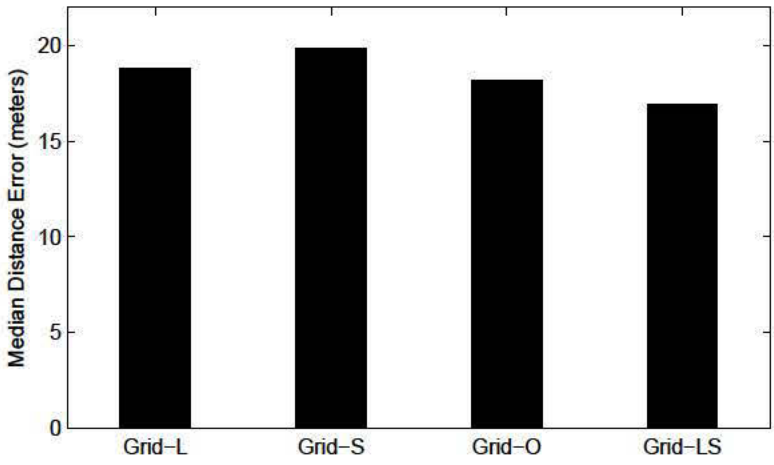


Figure 6.29: Comparison of Gridding Options

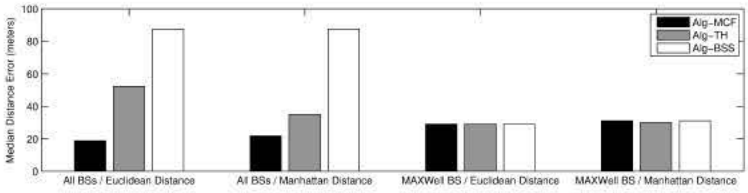


Figure 6.30: Comparison between Missing Data Algorithms, BS Selection Schemes, and Distance Measures

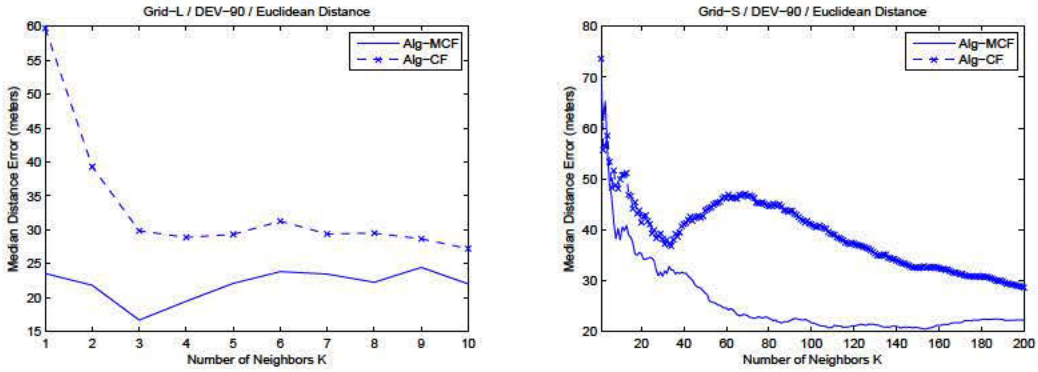
at the presence of frequent missing data.

When all BSs were taken into account, Alg-MCF performed better than Alg-TH. When only the MAXWell BS was taken into account, Alg-TH and Alg-BSS performed similar to Alg-MCF. Signals from the MAXWell BS were not frequently missed during the experiment because each MAXWell sector was operating on a channel, which is not shared with other BSs, and the signal quality was fair in the experiment area since MAXWell BS is located relatively in proximity.

Alg-TH assumes the missing data are primarily due to the marginal signal quality. When there are many signal misses due to other causes such as the capturing effect, its assumption fails. Thus, it is believed Alg-TH can perform reasonably well in the area where each BS is assigned a unique frequency band such as in MAXWell 4G network.

Between Alg-CF and Alg-MCF, Alg-MCF performs better than Alg-CF for small K values. Alg-CF tends to pick up the locations with small number of non-NaN matching pairs, so it does not perform well particularly when the K is small. The distance error over the number of K in two gridding settings is provided in Figure 6.31. Apparently, Alg-MCF performed better than Alg-CF for small K values. In order to decrease the computation complexity, small K value is preferred. Thus, Alg-MCF is preferred to Alg-CF.

In addition, it is interesting to see distance error of about 30 meters was achieved by only using the RF fingerprint with regard to a single BS. In MAXWell 4G network, a single BS controller unit manages three RF sectors.



(a) Grid-L (b) Grid-S

Figure 6.31: Comparison between Alg-CF and Alg-MCF over Number of Neighbors K

Under the circumstances where the BSs capture each other on the same frequency band, artificial outliers are generated by the assumption made in Alg-TH. Thus, with Alg-TH, Manhattan distance measure performs better than Euclidean distance measure since Manhattan distance is less susceptible to outliers. With Alg-MCF, both distance measures perform about the same.

6.6.2.5 Improved Distance Error by Filtering of BSs

The presence of multiple BSs on the same frequency band causes missing data due to capturing effect as well as the uncontrollable handoff between the BSs on the same frequency band. It was attempted to filter the signals from all the BSs sharing the frequency bands with other BSs. For the evaluation, the following settings were used.

- Gridding: Grid-L with K=10, Grid-S with K=750, Grid-O and Grid-LS with K=1000

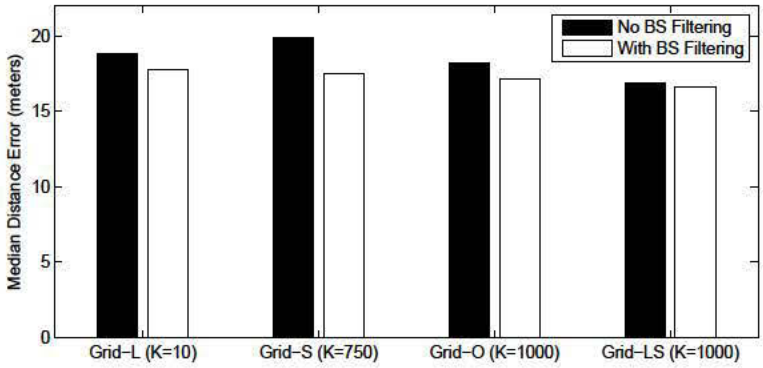


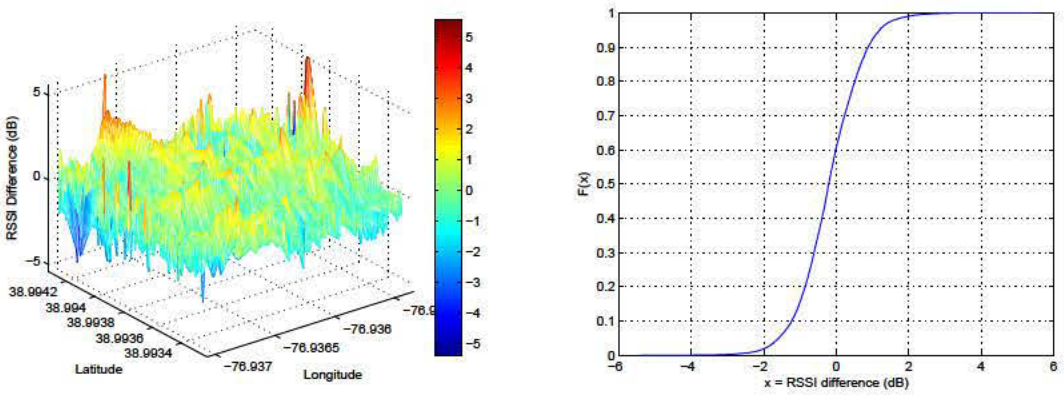
Figure 6.32: Improved Distance Error by Filtering

- Distance Measure: Euclidean distance.
- Distance Error: calculated and averaged over the results from two MSs.
- BS Filtering: all BSs vs. all BSs except those share the frequency with other BSs
- Outliers by systematic errors are disregarded.

This filtering combined with techniques described so far resulted in the smallest distance error, which is 16.6 meters. The result is provided in Figure 6.32. The gain in distance errors differs by gridding options and distance measure. The gain by filtering ranged between 2–12%.

6.6.2.6 Measurement Variation over WiMAX Cards

The measurement difference between two WiMAX cards was calculated to see the variance between WiMAX MS devices. With regard to a sector of MAXWell 4G network, the RSSI values were measured by two MS devices. The calculation was done for the interpolated values with Grid-L.



(a) Heatmap of the RSSI Difference

(b) Empirical CDF

Figure 6.33: Distribution of RSSI Difference between Hardware Devices

The result shows that more than 90% of the RSSI measurement was within ± 2dB. The result is provided in the Figure 6.33. The mean of absolute value of the RSSI difference was 0.5(dB). The mean of absolute value of the CINR difference was also 0.5 (dB).

6.6.2.7 Sampling of Measurements

Since the measurements were made for each preamble, the original measurements were made at a 200 (1/sec) sampling rate. Downsamplings were done to these samples, decreasing the sampling rate to 40 (1/sec) and 10 (1/sec). The radiomaps generated with the sampling rates 40 and 10 were compared with the original radiomap. The Pearson correlation coefficient, ρ, is calculated as the similarity measure. The original radiomap and the radiomaps with reduced sampling rates are very similar: ρ values are 0.997 and 0.996. Therefore, the sampling rate of the measurements can be reduced to the coherence time, decreasing the processing

overhead. The sampling rate has to be chosen depending on the wardriving vehicle speed since the coherence time is dependent on it.

6.6.3 RF Fingerprinting with Multiple Features

The experiment was done at the same location where the experiments in Section 6.6.2 were performed. The setting of the experiment is as follows:

- Grid Size: 8 meters by 11 meters
- Number of BSs: 11
- Number of Instances per Grid: 20
- Number of Measurements Used to Generate a Feature: 200
- Number of Measurements per Grid: 44000 (200 * 20 * 11)
- Number of Grids: 200
- Number of Features: 14

The summary of the Neural Networks modeled for the RF fingerprinting is as follows:

- Nodes in the neural networks: Sigmoid.
- Algorithm: Back propagation algorithm.
- Number of Input Nodes: 154
- Number of Hidden Layers: 1
- Number of Neurons: 177
- Number of Output Nodes: 200
- Learning Rate (The amount the weights are updated): 0.3

Table 6.8: RF Fingerprinting Performance with Multiple Features and Neural Networks

Number of Features	Features	Accuracy (%)	Distance Error (meters)
14	All features	97.94%	4.02
12	All features but mean(RSSI)/mean(CINR)	97.57%	4.14
3	mean(RSSI), mean(CINR), and mean(RSSId)	89.26%	6.07

- Momentum (Applied to the weights during updating): 0.2
- Attribute Normalization: Yes
- Training Time (The number of epochs to train through): 500

Since the estimated location is represented by the centroid of the grid, and the measurements are combined per grid during the online phase in this evaluation of the RF fingerprinting (i.e. measurements are made while the mobile station is moving around during the online phase), the distance error has to be calculated taking these into account. The distance error is calculated numerically, and the average distance error and the accuracy are presented in the Table 6.8.

When the rotation forest is used for the pattern matching with the neural networks as its underlying classifier, 98.5% of accuracy (4.00 meters of distance

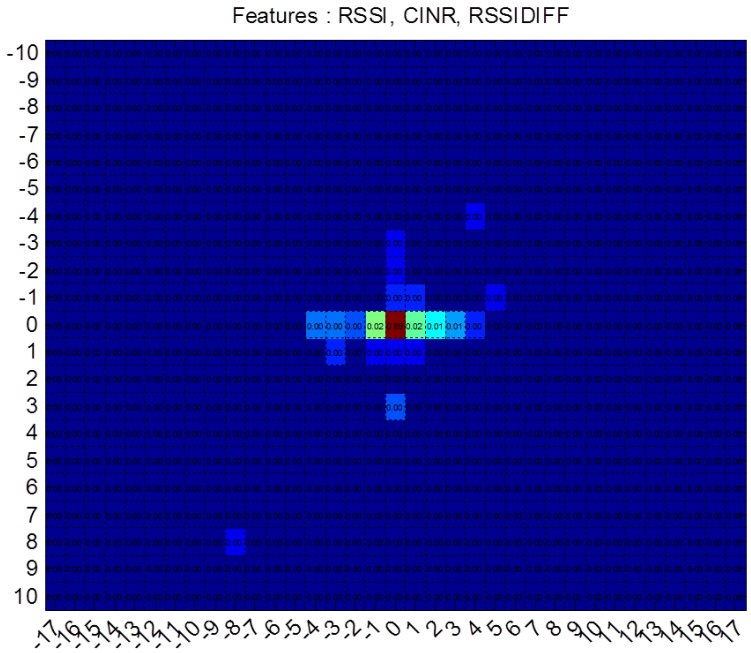
error) was achieved.

The distribution of the estimated locations (grid), when only three features (See Table 6.8) are used, is provided in the Figure 6.34(a). The distribution of the estimated locations (grid), when all fourteen features in Table 6.8 are used, is provided in the Figure 6.34(b). As seen from those figures, the estimated locations are distributed around the true location. When all features are used, it can be seen that the estimated locations are within the true grid with very high probability (98%) and all estimated grids are very close to the true locations.

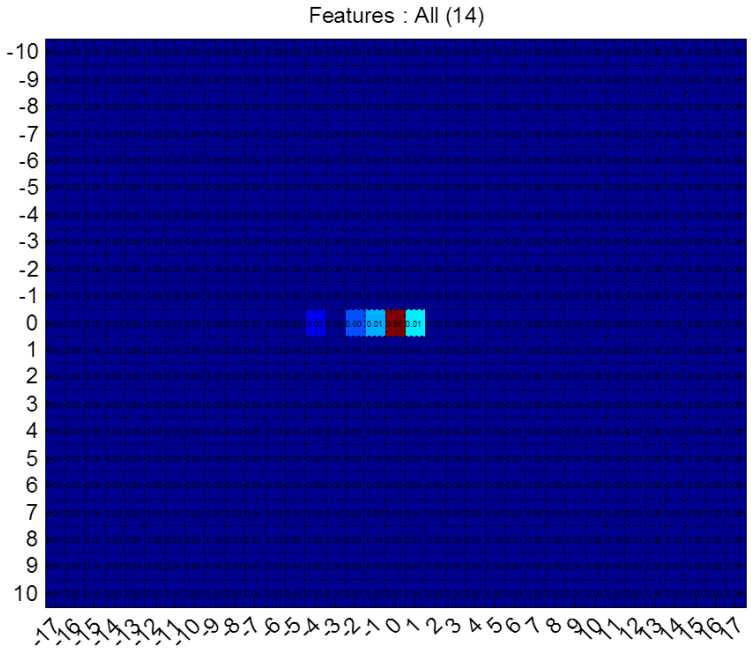
6.7 Summary

The causes for the signal measurement variation are identified, and the system is designed considering those, leading to a significant improvement in accuracy. Systematic errors, measurement errors, channel fading (shadow fading, fast fading, small scale fading), atmospheric propagation impairments in 2.5GHz 4G WiMAX are studied. It has been also noted that the measurement misses are frequently caused by the power capture between BSs. Techniques are developed to deal with each source of measurement variation and measurement misses.

In ROLAX, in addition to the DLP detection features (top twelve features evaluated in Section 5.3.2), mean of RSSI and mean of CINR are used to create unique RF fingerprints. The same set of features are used during the online phase. In total, fourteen features are used to create the RF fingerprint in ROLAX. Neural networks are trained by the radiomap data and used in determining the location.



(a) Three Features



(b) Fourteen Features

Figure 6.34: Distributions of Estimated Locations (Grid)

When a single feature is used for a smaller positioning latency, mean of RSSI is used as the feature, and K-NN pattern matching is used.

ROLAX RF fingerprinting techniques include: (1) a number of gridding techniques, including overlapped gridding; (2) an automatic radiomap generation technique by the Delaunay triangulation-based interpolation; (3) the filtering of measurements based upon the power-capture relationship between BSs; and (4) algorithms dealing with the missing data.

In order to deal with the measurements misses due to power capturing, handoff, etc., the overlapped gridding with the interpolation based upon Delaunay triangulation is designed. It also helps to deal with the small scale variation and to get enough number of samples to construct features for the radiomap.

Algorithms to deal with the missing values, in both offline and offline phase measurements, are developed for K-NN pattern matching. In addition to two algorithms introduced from previous work, two addition algorithms (Alg-MCF and Alg-TH) are designed. Particularly, modified classical fingerprinting (Alg-MCF) performs better than other algorithms.

With the combination of the techniques, an average distance error of 4 meters was achieved on the live 4G networks composed of MAXWell 4G Network and Clear 4G Network.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this dissertation, ROLAX location determination system in 4G networks is presented. There is an increasing need for ubiquitous positioning, but it has been very hard for a single location technology such as GPS to provide the ubiquitous positioning capability. Since the cellular signals are virtually ubiquitous and the indoor coverage is improved by increasing deployment of femto cells, 4G-based location determination has the potential to provide the location solution ubiquitously.

Geometric techniques such as Time of Arrival (ToA) and Time Difference of Arrival (TDoA) have been widely used to pinpoint the mobile terminal's location in wireless networks. The major sources of the impediments in geometric techniques are NLOS error, multipath error, and co-channel or inter-channel interference. Particularly, the signals over NLOS can cause positive bias on the timing-based measurements. The signal variation due to multipath and interference also make precise location difficult.

In RF fingerprinting systems, we can provide a location determination solution without using geometric techniques. In that way, the NLOS and multipath errors can be dealt with. For indoor positioning, RF fingerprinting has been successfully used by using the Wi-Fi signals. Due to the technical difference between 4G and

Wi-Fi, the techniques developed for the Wi-Fi networks cannot be directly applied to the 4G networks. Since the signal characteristics vary over time and over a small distance, it is challenging to develop RF fingerprinting techniques that are robust against the signal variation and the signal misses. Another challenge in the RF fingerprinting is that it takes very long time and lots of labor to collect the signals in order to build the radiomap. Thus, automatic radiomap generation techniques are highly needed.

ROLAX provides two primary solutions for the location determination in 4G networks. First, it provides techniques to detect the error-prone wireless conditions in geometric approaches of Time of Arrival (ToA) and Time Difference of Arrival (TDoA). Dominant Line-of-Sight Path (DLP) is a term coined in this work to refer to the channel condition where Line-of-Sight path is dominant over the Non-Line-of-Sight path. ROLAX provides techniques for a MS to determine the DLP condition given the measurements of the downlink signals from the BS. Second, the robust RF fingerprinting techniques for 4G networks are designed and used in ROLAX. The causes for the signal measurement variation are identified, and the system is designed taking those into account, leading to a significant improvement in accuracy.

The problems are approached by using Explanatory Data Analysis (EDA). By collecting data from the live 4G networks, the sources of signal variations are identified, and the techniques are developed to deal with each source of signal variation. Machine learning is used in both the DLP detection and the RF fingerprinting. By collecting signals from the offline phase, a priori knowledge about the signals under the DLP and the non-DLP condition is gained. In the RF fingerprinting, the signals

collected during the offline phase are used to create the radiomap that contains the signal pattern at each location. The techniques developed in this work were tested in live 4G networks.

ROLAX provides a location determination solution that can be used independently from the vendor and the device. In order to provide the portability of the solution, ROLAX uses a set of standard radio resource measurement that is supported in the most of the wireless networks—RSSI, CINR, and RTD. Typically, a MS has two antennas for the MIMO and the antenna diversity, and the measurement can be made per each antenna. Since these standard radio resource measurements have to meet a certain accuracy required by the standard, the signal pattern gathered during the offline phase can be compared with the online phase measurement whatever device is used. Due to this characteristic, it may be feasible to apply the ROLAX techniques to other wireless networks that support the similar set of radio resource measurements (e.g. LTE).

In ROLAX, the system architecture and the software architecture, for the offline signal collection and the online location determination, are designed. Software for the MS is developed for both offline and online phase operations. In this work, the software was developed using the interfaces provided by Beceem/Broadcom chipset based software(device driver, connection manager, logging facilities, etc.). Signals were collected from both the home networks and the foreign networks by the wardriving. Since scanning takes a long time in 4G, the MS makes a connection with each BS to gather enough number of measurements.

ROLAX is organized in two phases: offline and online phases. During the

offline phase, the radiomap is constructed by the wardriving. During the online phase, a Mobile Station (MS) performs the DLP condition test for each Base Station (BS) it can observe. If the number of the BSs under DLP is small, the MS attempts to determine its location by using the RF fingerprinting.

In ROLAX, the DLP condition is determined from the RSSI, CINR, and RTD (Round Trip Delay) measurements. A number of signal features that can be used to detect the DLP condition were designed. Features generated from the RSSI difference between two antennas of the MS were also used. The features, including the variance, the level crossing rate, the correlation between the RSSI and RTD, and Kullback-Leibler Divergence, were successfully used in detecting the DLP condition. We note that, compared to using a single feature, appropriately combined multiple features lead to a very accurate DLP condition detection. In addition to time features, spectral features, such as spectral centroid, spectral roll-off, and spectral flux, are used in the DLP detection using multiple features. A number of pattern matching techniques are evaluated for the purpose of the DLP/NDLP condition detection. Artificial neural networks, lazy learning using K-nearest neighbor (K-NN), and a meta classifier called Rotation Forest are particularly used in DLP detection, while other pattern matching techniques were also evaluated in this work. When the Rotation Forest is used with the K-NN, a detection accuracy of 94.8% was achieved in the live 4G networks. It has been noted that features designed in the DLP detection can be useful in the RF fingerprinting. Thus, the features developed for the DLP condition detection are used in ROLAX RF fingerprinting.

In ROLAX, in addition to the DLP detection features, mean of RSSI and mean

of CINR are used to create unique RF fingerprints. ROLAX RF fingerprinting techniques include: (1) a number of gridding techniques, including overlapped gridding; (2) an automatic radiomap generation technique by the Delaunay triangulation-based interpolation; (3) the filtering of measurements based upon the power-capture relationship between BSs; and (4) algorithms dealing with the missing data. In order to deal with the measurements misses due to the power capturing, handoff, etc., the overlapped gridding with the interpolation based upon Delaunay triangulation is designed. Interpolation and gridding each improves the performance by 20–35% and 8–15%. Algorithms to deal with the missing values, in both offline and offline phase measurements, are developed. Particularly, modified classical fingerprinting (Alg-MCF) performs better than other algorithms in K-nearest neighbor pattern matching.

By combining the techniques in ROLAX, a distance error in the order of 4 meters was achieved in the live 4G networks.

7.2 Future Work

ROLAX has been demonstrated on live 4G mobile WiMAX (IEEE 802.16e) networks. Since it uses the standard radio resource measurements, it would be applicable to other 4G networks such as WiMAX2, LTE, and LTE-advanced. In LTE, the radio resource measurements similar to those available in WiMAX (IEEE 802.16e) are provided. The power measurements in the LTE include Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ). LTE also

supports other power measurements for the mobility between LTE (E-UTRAN) and 2G/3G networks. In the LTE, the round trip time between an UE and an eNodeB can be measured on the eNodeB side by Time Advance Type 1 and Type 2. The detailed description of the measurements available in LTE is provided in Appendix B. Since LTE supports similar radio resource measurements to those in the mobile WiMAX, the techniques developed in ROLAX are expected to be applicable in the LTE with a limited modification.

In this work, it has been demonstrated that there is a strong possibility to create a more unique RF fingerprint by combining the features from the timing measurements (e.g. RTD) and the features from the power measurements (e.g. RSSI, CINR). Due to the low measurement frequency of the RTD, it was hard to create a RF fingerprint using the timing measurements. If the RTD measurements with a higher measurement frequency can be supported, the timing measurements can be considered in the future ROLAX RF fingerprinting. The RTD with a limited measurement frequency can be still used, for instance, in clustering the radiomap before the RF fingerprinting to reduce the positioning delay.

Another possible future extension of ROLAX is to make it be adaptive to a given LBS Quality of Service. The LBS quality of service is defined by horizontal accuracy, vertical accuracy, latency, QoS class (assured and best), LBS priority, and age in the 4G. In this work, it has been demonstrated that each techniques will result in the different positioning delay and the accuracy. For instance, using a single feature in the RF fingerprinting results in the worse accuracy than using multiple features, but it has a smaller positioning delay. Each pattern matching evaluated

in this work has a different accuracy and a computation time complexity. Thus, by choosing the technical components given the LBS Quality of Service, ROLAX can be evolved to perform suiting the needs of the Location-based Service.

Appendix A

Measurements Available in Mobile WiMAX (IEEE 802.16e)

Received Signal Strength Indication (RSSI), Carrier to Interference Noise Ratio (CINR), and Round Trip Delay (RTD) are widely used in wireless networking operation to evaluate the signal quality. Not only mobile WiMAX but also other wireless networks such as IEEE 802.11 and IEEE 802.15.4 supports the measurement of similar metrics. Each network has its own definition and requirements for those metrics. They are typically used in link adaptation and the selection of the peer (e.g. Access Point, Base Station). In this work, they are used in composing RF fingerprint vector.

A.1 Received Signal Strength Indication (RSSI)

RSSI is more clearly defined in WiMAX than in Wi-Fi. In Wi-Fi, the definition of RSSI is not consistent between vendors. In general, it is relative level of signal power measured at the RF front end of the Wi-Fi receiver. Each chip vendor implements the RSSI measurement in a different way, and there is no accuracy, resolution, and reporting range (e.g. maximum and minimum value) requirements. Therefore, this situation has made the RF fingerprinting in Wi-Fi hard to be deployed widely because the user's Wi-Fi device during online phase is highly likely to use different vendor's Wi-Fi chipset from the one used during the generation of

the radiomap (offline phase). In addition, the user chooses the Wi-Fi device from a huge set of selections, so the level of performance can differ from user to user a lot. In order to fix this situation, Received Channel Power Indicator (RCPI) has been defined in IEEE 802.11k but it has not widely implemented yet. Also since the RCPI measures the entire frame rather than only the preamble, the value may be changing as the link rate is being adapted to wireless channel particularly because the different transmit power level is used for each channel rate.

In WiMAX (IEEE 802.16e), measuring RSSI does not necessarily require receiver demodulation lock. Therefore, RSSI measurements offer reasonably reliable channel strength assessments even at low signal levels. Statistics shall be quantized in 1 dB increments, ranging from -40 dBm to -123 dBm. The relative accuracy of a single signal strength measurement shall be ± 2 dB with an absolute accuracy of ± 4 dB. The standard deviation shall be quantized in 0.5 dB increments [64].

Measurements by MS can be reported to its serving BS over the scanning result report message (MOB_SCN-REP). Not only the mean RSSI but also mean CINR, relative delay, and BS RTD for the requested BSs specified in the scanning result report message are reported. The RSSI measurement shall be performed on the frame DL burst preamble, and values are averaged over the measurement period [64].

The method of RSSI measurements is up to the vendor as long as the measurements meet the requirements set in [64]. One possible method to calculate the RSSI is proposed in the (A.1) [64] [65].

$$RSSI \;=\; 10^{-(G_{rf}/10)} \frac{1.2567 \times 10^4 V_c^2}{(2^{2B})R} \left(\frac{1}{N} \sum_{n=0}^{N-1} |Y_{I \text{ or } Q}[k,n]| \right)^2 \; mW \qquad \text{(A.1)}$$

where B is ADC precision, number of bits of ADC, R is ADC input resistance (Ω), V_c is ADC input clip level (V), G_{rf} is analog gain from antenna connector to ADC input, $Y_{I \text{ or } Q}[k,n]$ is n-th sample at the ADC output of I or Q-branch within signal k and N is number of samples.

In this method, the RSSI is estimated at the output of Analog Digital Converter (ADC), and it is assumed that the inputs and outputs of the ADC are Gaussian with zero mean. RSSI is calculated as a function of ADC parameters (precision, input resistance, input clip level), analog gain from antenna connector to ADC, and the sample values at the ADC output of I and Q-branch.

A.2 Carrier to Interference Noise Ratio (CINR)

It is demonstrated that CINR can be used in RF fingerprinting in this work. CINR is typically used to judge the signal quality at the receiver end. One of its usages is for the computation of the handoff trigger.

In the IEEE 802.11, the measurement of CINR is not supported by the specification. IEEE 802.11k supports the measurement of the Received Signal to Noise Indicator (RSNI).

In mobile WiMAX, measuring CINR requires receiver lock while RSSI measurement does not require receiver lock. Thus, CINR is measurable only when the signal quality is fair enough to obtain the receiver lock. CINR provides information

on the actual operating condition of the receiver, including interference and noise levels and signal strength.

There are two types of CINR: physical CINR and effective CINR. The effective CINR is a function of physical CINR, varying channel conditions, and implementation margin. One possible method to estimate the CINR is by normalizing the mean-squared residual error of detected data symbols by the average signal power. It is given in equation (A.2) [64] [66].

$$CINR[k] = \frac{\sum_{n=0}^{N-1} |s[k,n]|^2}{\sum_{n=0}^{N-1} |r[k,n] - s[k,n]|^2} \tag{A.2}$$

where $r[k,n]$ is the received sample n within message measured at time index k in frame units, $s[k,n]$ is the corresponding detected or pilot sample (with channel state weighting), and N is the total number of samples within a message.

In (A.2), the average signal power $\sum_{n=0}^{N-1} |s[k,n]|^2$ is normally kept constant by the operation of the Automatic Gain Control (AGC).

Depending on the frequency reuse configuration (one or three), the reported CINR shall be the estimate over different sets of the subcarriers. If the frequency reuse configuration is one, then the CINR shall be the estimate of the average CINR over all subcarriers of the preamble except the guard and the DC subcarriers.

In this work, physical CINR with the frequency reuse configuration of one was obtained and used as RF fingerprints.

MS can transmit CINR information to the BSs by using the REP-RSP MAC message or fast-feedback channel (CQICH).

It is quantized in 1 dB increments, ranging from a minimum of -20 dB to a maximum of 40 dB. Relative and absolute accuracy of a CINR measurement derived from a single message shall be ±1dB and ±2dB.

A.3 Round Trip Delay (RTD)

RTD can be measured only with regard to the serving BSs or anchor BSs in WiMAX. How RTD can be calculated is implementation-dependent, but the latest time advance taken by MS can be used as RTD according to the IEEE 802.16 standard release in 2009 [64]. Time advance is calculated to adjust the timing of the MSs during the ranging procedure. Ranging adjusts each MS's timing offset such that it appears to be co-located with the BS. The MS shall set its initial timing offset to the amount of internal fixed delay, implementation-specific delays, etc. It is conjectured that the history of time advances has to be accumulated and used in calculating the RTD. How the software used in this work calculates the time advance is not known. The RTD values observed with the Beceem software are logical values, and they can be mapped to the physical time value, which can be converted to the physical distance by backward engineering.

The RTD value's unit is $1/F_s$ where F_s is the sampling frequency. Since the F_s is a function of channel bandwidth in WiMAX, increasing bandwidth results in higher resolution in RTD measurement. The F_s is given as

$$F_s = floor(n \cdot BW/8000) \times 8000 \tag{A.3}$$

where BW is the channel bandwidth and n is the sampling factor. Sampling factor is set as follows: for channel bandwidths that are a multiple of 1.75 MHz, then n = 8/7; else, for channel bandwidths that are a multiple of any of 1.25, 1.5, 2, or 2.75 MHz, then n = 28/25; else, for channel bandwidths not otherwise specified, then n = 8/7.

Thus, in case of 5MHz channel bandwidth, the F_s is 5.6MHz, and accordingly the distance resolution given by $(speedoflight)/F_s$ is 53.6 meters. In case of 10MHz channel bandwidth, F_s is 15MHz, and the corresponding distance resolution is 20 meters.

Appendix B

Measurements Available in LTE

LTE is another 4G standard developed by 3GPP. LTE is specified in release 8 and release 9 documents. LTE is evolved to LTE-Advanced in the following 3GPP releases. In LTE, the mobile terminal is referred to User Equipment (UE), and the base station is referred to E-UTRAN Node B (eNodeB or eNB). UE measures two power parameters on reference signals, which are Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ).

RSRP is a type of signal strength measurement, and it is defined in 3GPP TS 36.214 (Evolved Universal Terrestrial Radio Access; Physical layer; Measurements) [67]. RSRP is the linear average over the power contributions (in [W]) of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth. For RSRP determination, the cell-specific reference signals R0 according TS 36.211 [68], shall be used. If the UE can reliably detect that R1 is available, it may use R1 in addition to R0 to determine RSRP.

RSRQ is a type of signal quality measurement, and it is also defined in 3GPP TS 36.214 [67]. Evolved UMTS Terrestrial Radio Access (E-UTRA) Carrier Received Signal Strength Indicator (RSSI), comprises the linear average of the total received power (in [W]) observed only in OFDM symbols containing reference symbols for antenna port 0, in the measurement bandwidth, over N num-

ber of resource blocks by the UE from all sources, including co-channel serving and non-serving cells, adjacent channel interference, thermal noise etc. RSRQ is the ratio $N \cdot RSRP/E-UTRAcarrierRSSI$, where N is the number of Resource Block's(RB's) of the E-UTRA carrier RSSI measurement bandwidth. The measurements in the numerator and denominator shall be made over the same set of resource blocks.

For the mobility between LTE and 3G UMTS, other set of measurements can be supported. For the mobility between LTE(E-UTRAN) and 3G(UTRAN), FDD CPICH RSCP, FDD carrier RSSI and FDD CPICH Ec/No, TDD carrier RSSI, and TDD P-CCPCH RSCP are supported. For the mobility with GSM, carrier RSSI is supported.

LTE also supports the measurement of round trip delay between the UE and eNodeB. eNodeB can report a type of timing measurement referred by the Timing Advance, and it is also defined in in 3GPP TS 36.214 [67]. There are two types of Timing Advance measurements: the Timing Advance Type 1 and Timing Advance Type 2. UE also may have the capability to report the RTT. UE can make Rx-Tx time difference which defined by $T_{UE-RX} - T_{UE-TX}$, where T_{UE-RX} is the UE received timing of downlink radio frame i from the serving cell, defined by the first detected path in time, and T_{UE-TX} is the UE transmit timing of uplink radio frame i. In type 1, the UE reports this Rx-Tx time difference to its serving eNodeB, and the eNodeB calculates its own receive transmit time difference. The Type 2 measurement is the receive transmit time difference at the eNodeB. Potentially, Type 1 provides a better timing measurement accuracy [69]. It is not clear if the

UE can measure the RTT by itself in LTE.

Bibliography

[1] WiMAX Forum. Industry research report May 2011. Technical report, WiMAX Forum, 2011.

[2] Kai Siwiak and Jerry Gabig. 802.15.4IGa Informal call for application response-real time location service (RTLS) applications, range and accuracy requirements in P802.15.4. Technical report, IEEE 802.15.4IGa, 2003.

[3] Guenther Retscher and Allison Kealy. Ubiquitous positioning technologies for intelligent navigation systems. In *Proceedings of the 2nd Workshop on Positioning, Navigation and Communication (WPNC05) & 1st Ultra-wideband Expert Talk (UET'05)*, pages 99–108, 2005.

[4] Garmin. What is GPS? `http://www8.garmin.com/aboutGPS/`. [Online; accessed 15-September-2012].

[5] Informa Telecoms and Media. Small cell market status. Technical report, Small Cell Forum, 2012.

[6] M. Venkatachalam, K. Etemad, W. Ballantyne, and B. Chen. Location services in WiMax networks. *Communications Magazine, IEEE*, 47(10):92 –98, october 2009.

[7] Mark Cudak et al. IEEE 802.16m system requirements. Technical Report IEEE802.16m-07/002r9, IEEE 802.16m, 2009.

[8] Alan Bensky. *Wireless Positioning Technologies and Applications*. Artech House, Inc., Norwood, MA, USA, 2007.

[9] Jason W. p. Ng. *Space-Time Array Communications: Vector Channel Estimation and Reception*. Imperial College Press, London, UK, UK, 2007.

[10] Charles Chien. *Digital radio systems on a chip: a systems approach*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[11] Tsung-Nan Lin and Po-Chiang Lin. Performance comparison of indoor positioning techniques based on location fingerprinting in wireless networks. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 2, pages 1569 – 1574 vol.2, june 2005.

[12] M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen. Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on WiMAX networks. *Vehicular Technology, IEEE Transactions on*, 59(1):283 –294, jan 2010.

[13] Moustafa Youssef and Ashok Agrawala. The Horus location determination system. *Wirel. Netw.*, 14:357–374, June 2008.

[14] P. Bahl and V.N. 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*, volume 2, pages 775 –784 vol.2, 2000.

[15] P. Bahl, A. Balachandran, and V. Padmanabhan. Enhancements to the RADAR user location and tracking system, 2000.

[16] Jian Zhu, S. Spain, T. Bhattacharya, and G.D. Durgin. Performance of an indoor/outdoor RSS signature cellular handset location method in Manhattan. In *Antennas and Propagation Society International Symposium 2006, IEEE*, pages 3069 –3072, july 2006.

[17] C. Nerguizian, C. Despins, and S. Affes. Indoor geolocation with received signal strength fingerprinting technique and neural networks. In Jose de Souza, Petre Dini, and Pascal Lorenz, editors, *Telecommunications and Networking - ICT 2004*, volume 3124 of *Lecture Notes in Computer Science*, pages 866–875. Springer Berlin / Heidelberg, 2004. 10.1007/978-3-540-27824-5_114.

[18] Zi-Ning Zhen, Qing-Shan Jia, Chen Song, and Xiaohong Guan. An indoor localization algorithm for lighting control using RFID. In *Energy 2030 Conference, 2008. ENERGY 2008. IEEE*, pages 1 –6, nov. 2008.

[19] Carlos Leonel Flores Mayorga, Francescantonio della Rosa, Satya Ardhy Wardana, Gianluca Simone, Marie Claire Naima Raynal, Joao Figueiras, and Simone Frattasi. Cooperative positioning techniques for mobile localization in 4g cellular networks. *International Conference on Pervasive Services*, 0:39–44, 2007.

[20] Anis Drira. GPS navigation for outdoor and indoor environments. Master's thesis, The University of Tennessee, Knoxville, 2006.

[21] G.M. Djuknic and R.E. Richton. Geolocation and assisted GPS. *Computer*, 34(2):123 –125, feb 2001.

[22] 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. *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999)*.

[23] M. Maqbool, M. Coupechoux, and P. Godlewski. Comparison of various frequency reuse patterns for WiMAX networks with adaptive beamforming. In *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pages 2582 –2586, may 2008.

[24] Qualcomm Incorporated. 3G provides mobile broadband today:An overview of 3G, its evolution, and some perspectives on mobile WiMAX. Technical report, Qualcomm Incorporated, January 2008.

[25] COST Action 231. Digital mobile radio towards future generation systems, final report. Technical report, European Communities, 1999.

[26] T. S Priya. Optimised COST-231 hata models for WiMAX path loss prediction in suburban and open urban environments. *Modern Applied Science*, 4(9):P75, 2010.

[27] V. Erceg et al. Channel models for fixed wireless applications. Technical Report IEEE 802.16a-03/01, IEEE 802.16a, 2003.

[28] Wikipedia. Fresnel zone — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/wiki/Fresnel_zone`, 2012. [Online; accessed 19-October-2011].

[29] Jeffrey G. Andrews, Arunabha Ghosh, and Rias Muhamed. *Fundamentals of WiMAX : understanding broadband wireless networking*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall, Upper Saddle River, NJ, 2007.

[30] Kegen Yu, Ian Sharp, and Y. Jay Guo. *Ground-based wireless positioning*. IEEE Press series on digital & mobile communication. Wiley, Chichester, West Sussex, U.K. ; Hoboken, NJ, 2009.

[31] U.S. geological survey - global positioning application and practice. `http://water.usgs.gov/osw/gps/`, September 2012.

[32] Anteannsearch. `http://www.antennasearch.com/`, July 2011.

[33] Motorola USBw 100 USB adaptor for mobile WiMAX. `http://www.motorola.com/web/Business/ProductLines/Motowi4/USBw100/Documents/_Staticfiles/USBw_100_DataSheet.pdf`.

[34] Linux WiMAX. `http://www.linuxwimax.org/`, May 2011.

[35] Zafer Sahinoglu, Sinan Gezici, and Ismail Guvenc. *Ultra-wideband positioning systems : theoretical limits, ranging algorithms, and protocols*. Cambridge University Press, Cambridge, UK ; New York, 2008.

[36] M.I. Silventoinen and T. Rantalainen. Mobile station emergency locating in gsm. In *Personal Wireless Communications, 1996., IEEE International Conference on*, pages 232 –238, feb 1996.

[37] I. Guvenc, Chia-Chin Chong, and F. Watanabe. NLOS identification and mitigation for UWB localization systems. In *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, pages 1571 –1576, march 2007.

[38] F. Benedetto, G. Giunta, A. Toscano, and L. Vegni. Dynamic LOS/NLOS statistical discrimination of wireless mobile channels. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 3071 –3075, april 2007.

[39] M.P. Wylie and J. Holtzman. The non-line of sight problem in mobile location estimation. In *Universal Personal Communications, 1996. Record., 1996 5th IEEE International Conference on*, volume 2, pages 827 –831 vol.2, sep-2 oct 1996.

[40] Wikipedia. Kernel density estimation — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/wiki/Kernel_density_estimation`, 2012. [Online; accessed 19-October-2012].

[41] S. Gezici, H. Kobayashi, and H.V. Poor. Nonparametric nonline-of-sight identification. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, volume 4, pages 2544 – 2548 Vol.4, oct. 2003.

[42] Mohammad Heidari and Kaveh Pahlavan. A Markov model for dynamic behavior of ToA-based ranging in indoor localization. *EURASIP J. Adv. Signal Process*, 2008:82:1–82:14, January 2008.

[43] Mary Natrella. *NIST/SEMATECH e-Handbook of Statistical Methods*. NIST/SEMATECH, July 2010.

[44] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, 4th edition, 2008.

[45] Theodore S. Rappaport. *Wireless communications : principles and practice*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, Upper Saddle River, N.J., 2nd edition, 2002.

[46] Usama M. Fayyad and Keki B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *IJCAI*, pages 1022–1029, 1993.

[47] J.J. Rodriguez, L.I. Kuncheva, and C.J. Alonso. Rotation forest: A new classifier ensemble method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1619 –1630, oct. 2006.

[48] Weka Machine Learning Project. Weka. `http://www.cs.waikato.ac.nz/~weka`.

[49] Robert B. Ash and Catherine Doléans-Dade. *Probability and measure theory*. Harcourt/Academic Press, San Diego, 2nd edition, 2000.

[50] Gregory D. Durgin. *Space-time wireless channels*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, Upper Saddle River, NJ, 2003.

[51] P. Smulders. Statistical characterization of 60-GHz indoor radio channels. *Antennas and Propagation, IEEE Transactions on*, 57(10):2820 –2829, oct. 2009.

[52] S. Enoch and I. Otung. Propagation effects in WiMAX systems. In *Next Generation Mobile Applications, Services and Technologies, 2008. NGMAST '08. The Second International Conference on*, pages 425 –430, sept. 2008.

[53] Zerihun Abate. *WiMax RF systems engineering*. Artech house mobile communications series. Artech House, Boston, 2009. Zerihun Abate. ill. ; 24 cm. Includes bibliographical references and index.

[54] Frank Ohrtman, Konrad Roeder, and Books24x7 Inc. Wi-fi handbook building 802.11b wireless networks, 2003.

[55] ITU-R. Attenuation by atmospheric gases. Technical Report RECOMMENDATION ITU-R P.676-5, ITU-R, 2001.

[56] C.B. Almazan, M. Youssef, and A.K. Agrawala. Rover: An integration and fusion platform to enhance situational awareness. In *Performance, Computing, and Communications Conference, 2007. IPCCC 2007. IEEE Internationa*, pages 582 –587, april 2007.

[57] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Addison Wesley, Boston, 2006.

[58] R. Rouil and N. Golmie. Adaptive channel scanning for IEEE 802.16e. In *Military Communications Conference, 2006. MILCOM 2006. IEEE*, pages 1 –6, oct. 2006.

[59] P. Boone, M. Barbeau, and E. Kranakis. Strategies for fast scanning and handovers in WiMAX/802.16. In *Access Networks Workshops, 2007. AccessNets '07. Second International Conference on*, pages 1 –7, aug. 2007.

[60] D. Murray, M. Dixon, and T. Koziniec. Scanning delays in 802.11 networks. In *Next Generation Mobile Applications, Services and Technologies, 2007. NGMAST '07. The 2007 International Conference on*, pages 255 –260, sept. 2007.

[61] M. Ibrahim and M. Youssef. CellSense: A probabilistic RSSI-based GSM positioning system. In *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, pages 1 –5, dec. 2010.

[62] Wikipedia. Delaunay triangulation — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/wiki/Delaunay_triangulation`, 2012. [Online; accessed 14-October-2011].

[63] Mathworks. Interpolating Scattered Data. `http://www.mathworks.com/help/matlab/math/interpolating-scattered-data.html`, 2012. [Online; accessed 25-October-2012].

[64] IEEE standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems. *IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004)*, pages C1 –2004, 29 2009.

[65] Choongill Yeh, Hyoungsoo Lim, and Dongseung Kwon. RSSI measurements. Technical Report IEEE C802.16d-03/92, IEEE 802.16d, 2003.

[66] Brian Eidson and Anader Benyamin-Seeyar. Proposed addition of section 8.3.3.1.7 [channel quality measurements at the subscriber station] to document 80216ab_01/01r2. Technical Report IEEE 802.16abc-01/51r1, IEEE 802.16ab, 2001.

[67] 3rd Generation Partnership Project. Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical layer; measurements (release 9) TS 36.214 V9.2.0, 2010.

[68] 3rd Generation Partnership Project. Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 9) TS 36.211 V9.1.0, 2009.

[69] Stefania Sesia, Issam Toufik, Matthew Baker, and ebrary Inc. *LTE–the UMTS long term evolution from theory to practice*. Wiley, 2011.