TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

AURYN PINK SODERINI

OUTDOOR POSITIONING ALGORITHMS BASED ON LTE AND WIFI MEASUREMENTS

Master of Science Thesis

Examiners: Assoc. Prof. Elena-Simona Lohan
Prof. Robert Piché
Examiners and topic approved by the Faculty of Computing and Electrical Engineering on May 6, 2015

ABSTRACT

AURYN PINK SODERINI

Outdoor Positioning Algorithms Based on LTE and WiFi Measurements
Tampere University of Technology, 76 pages
December 2016
Master's Degree Programme in Electrical Engineering
Major: Wireless Communications Circuits and Systems
Examiners: Assoc. Prof. Elena-Simona Lohan and Prof. Robert Piché

Keywords: Outdoor positioning algorithms, LTE WIFI RSS-based positioning, Statistical models, Coverage area, Path loss

This scientific work focuses on outdoor positioning in WLAN and 4G wireless cellular networks based on extensive collection of radio measurements from WiFi and LTE signals, supplied with Global Positioning System location information and time stamp.

The objective of work is to explore the performance of different RSS-based outdoor positioning algorithms in terms of distance error and database. The objective includes, through simulating experiments, to verify if accuracy is in compliance with the E911 requirements. The scope of the research is to establish an energy-efficient and low-latency solution for accurate and reliable outdoor positioning based on cellular networks.

Two probabilistic models based on coverage-area and path-loss are studied and implemented whereas the more common deterministic model based on classical-fingerprinting is used as benchmark for assessing the performance. The advantage of using probabilistic model over deterministic is that only few parameters per transmitter identification need to be stored and hence there is a significant reduction of the database size. Results show that statistical models suffer accuracy loss to some extent but nevertheless, the decrease in accuracy is not significant with respect to the requirements imposed by FCC.

PREFACE

Based on industrial-provided measurements, this research work was carried out at the Department of Automation Science and Engineering of Tampere University of Technology. Here Prof. Robert Piché, leader of the Personal Positioning Algorithm research group, closely supervised the work supplying what was necessary to perform the required activities and leading me towards a completion of postgraduate studies. In this department I received the support and expertise from the personnel in the research group that hosted me. From the department of Electronics and Communication Engineering, Assoc. Prof. Elena-Simona Lohan, group leader for the mobile-based and satellite-based positioning activities, supervised the work. She guided my research activities by a countless number of meetings, discussions and contributions.

I was appointed the thesis position on date 30.03.2015. The thesis was commissioned by Dr. Jussi Turkka who, on behalf of the Finnish company Magister Solution Ltd., provided an extensive collection of radio measurements from live 4G and WiFi communication networks for the development of outdoor positioning algorithms. He co-supervised the work, supporting what was needed to overcome the theoretical radio network-related issues and in understanding and interpreting of dataset analysis outcomes.
I would like to thank Associate Professor Elena-Simona Lohan, Professor Robert Piché, and Doctor Jussi Turkka for offering the opportunity to research in this topic and for providing their invaluable support and guidance throughout the time. By demonstrating competence and dedication to the research, they inspired me confidence in pursuing the object.

The research group I worked with at the time of doing the thesis were cohesive and amiable, very welcoming to me when arrived as a new member. I appreciate them for being friendly and caring in a way that made me feel comfortable. I am very grateful to them for sharing their time and supporting me along this path of maturity. I would like to express my gratitude in particular to, Dr. Simo Ali-Löytty, Dr. Pavel Davidson, Dr. Matti Raitoharju, Dr. Helena Leppäkoski, Henri Nurminen and Philipp Müller.

In Finland, since I came in 2013, I have met many new people. In particular, I feel lucky for meeting Akshay Kankariya and Sergi Roca who have had a very positive impact in my life. During my student career at TUT I had the opportunity to learn more about Wenbo Wang and to share with him several activities related to our common interests of study. I appreciate him for being consistent and punctual and for his work and help.

Finally I would like to thank my family, my lifelong army, for their unconditional support and love. I dedicate this thesis to my nephew Manuel who was born in July 2015, a few months earlier starting to write.

Tampere, 30.04.2016

Auryn Pink Soderini

CONTENTS

Abstract i
Preface ii
Abbreviations vi
Symbols x

1. Introduction 1
1.1 Author's contribution 2
1.2 Thesis outline 4
2. LTE-based positioning 6
2.1 LTE signal structure 6
2.2 LTE positioning signal 9
2.3 Non RSS-based LTE positioning 10
2.3.1 Positioning signal parameters overview 10
2.3.2 LTE positioning concept model 13
2.3.3 LTE positioning methods 14
2.4 RSS-based LTE positioning 17
2.4.1 LTE measurement: computation and mapping 18
2.4.2 RSRP-based positioning 19
2.4.3 Fingerprinting approach 20
2.4.4 Classical fingerprinting: a deterministic approach 21
3. WIFI-based positioning 24
3.1 WLAN standard 24
3.2 Non-RSS-based WiFi positioning 27
3.3 RSS-based WiFi positioning 28
4. Statistical channel models for positioning purpose 31
4.1 Path-loss models 31
4.2 Parameters estimation: path-loss and shadowing 33
4.3 Power map 40

5. RSS-based position estimation with statistical information 42
5.1 Coverage area based positioning 42
5.2 Path-loss based positioning 45
6. Measurement-based analysis and results 48
6.1 Measurement data description 48
6.1.1 Introduction 48
6.1.2 Data import and processing 50
6.1.3 Empirical RSS statistic 51
6.1.4 Estimation data 54
6.2 Transmitter localization 58
6.2.1 Introduction 58
6.2.2 Estimation algorithm 58
6.2.3 Setup algorithm 60
6.2.4 Histogram: mutual distance 61
6.2.5 Summary 63
6.3 Positioning results 63
6.4 Database size and reduction factor 66
7. Conclusion and future work 67

Bibliography 68

ABBREVIATIONS

3GPP	Third Generation Partnership Project
4G	Fourth Generation
A-GNSS	Assisted Global Navigation Satellite System
AOA	Angle of Arrival
AP	Access Point
BCC	Binary Convolutional Coding
BS	Base Station
BSSID	Basic Service Set Identifier
CA	Coverage Area
CCA	Clear Channel Assessment
CCF	Cross Correlation Function
CID	Cell Identifier
CKK DQPSK	Complementary Code Keying DQPSK
CP	Cycle Prefix
CRS	Cell-specific Reference Signal
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
DBPSK	Differential Binary PSK
DD	Decimal Degree
DL	Down Link
DT	Drive Test
DTOA	Difference Time of Arrival
DQPSK	Differential Quaternary PSK
DSSS	Direct Sequence Spread Spectrum
E911	Enhanced 911 Emergency Wireless Service
E-Cell ID	Enhanced Cell ID
ENU	Est Nord Up
ESO	European Standards Organization
ETSI	European Telecommunication Standards Institute
FCC	Federal Communication Commission

FDD	Frequency Division Duplex
FP	Fingerprinting
FS	Free Space
GDOP	Geometrical Dilution of Precision
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
ID	Identification Code
JADE	Joint Angle and Delay Estimation
LAN	Local Area Networks
LBS	Location Based Services
LMSC	LAN/MAN Standards Committee
LOS	Line of Sight
LPP	LTE Positioning Protocol
LS	Least Squares
LTE	Long Term Evolution
MAN	Metropolitan Area Networks
MDT	Minimization of Drive Test
MIMO	Multiple Input Multiple Output
MS	Mobile Station
MSC	Multiple Scan
MU	Multiple User
MUSIC	Multiple Signal Classification
NaN	Not a Number
NLOS	Non Line of Site
NN	Nearest Neighbor
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OSI	Open System Interconnection
OTDOA	Observed Time of Arrival
PARP	Peak to Average Power Ratio
PCC	Poly-phase Complementary Code

PHY	Physical Layer
PL	Path Loss
PLCP	Physical Layer Convergence Protocol
PRS	Positioning Reference Signal
PSK	Phase Shift Code
RB	Resource Block
RMSE	Root Mean Square Error
RS	Reference Signal
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RSTD	Reference Signal Time Difference
RTT	Round Trip Time
RX	Receiver
RXS	Receiver Sensitivity
SC-FDMA	Single Carrier Frequency Division Multiple Access
SHARING	Self-organized Heterogeneous Advanced Radio Networks Generation
SHF	Super High Frequency
SINR	Signal to Interference plus Noise Ratio
SSC	Single Scan
TDD	Time Division Duplex
TOA	Time of Arrival
TTFF	Time to First Fix
TX	Transmitter
TXID	Transmitter Identity
UE	User Equipment
UL	Up Link
U-NII	Unlicensed National Information Infrastructure
UWB	Ultra Wide Band

WGS84	World Geodetic System 1984
WiFi	Wireless Fidelity
WKNN	Weighted K-Nearest Neighbor
WLAN	Wireless Local Area Network
WNIC	Wireless Network Interface Card

SYMBOLS

A	Apparent transmitted power
A_{txid}^{m}	Vector of apparent transmitted power parameters
α	Angle of arrival
c	Speed of light
d	Distance
d_0	Reference distance
d_{ar}	Antenna array elements inter-distance
$\mathrm{d}_{\mathrm{ar}_1}$	Projection of d_{ar} onto E
$\mathrm{d}_{txid,l}$	Euclidean distance between the $txid^{th}$ estimated transmitter location and the l^{th} location point of the $txid^{th}$ location grid
Δt	Time-offset
f_{c}	Carrier frequency
$\mathrm{FSPL}_{\mathrm{dB}}$	Free-space path-loss model expressed in decibel unit
$\mathrm{g}(f_{\mathrm{c}})_{txid,l}^{m}$	Complementing function in matrix H_{txid}^{m}
h	Signal removal flag
H	Distance between transmitter and receiver
H_{txid}^{m}	Matrix of signal propagation frame
I_{txid}^{m}	Vector of re-created RSS
m	PL model-dependent flag
n	Path-loss coefficient parameter
n_{txid}^{m}	Vector of path-loss coefficient parameters
N_{ds}	Number of detected signals at the radio device
$\mathrm{N}_{\mathrm{meas}}$	Number of measurement location points in the considered location grid
N_{rb}	Number of measurement resource blocks
N_{txid}	Number of location points referenced to $txid^{th}$ location grid
P_{r}	Received power
P_t	Transmitted power
P_{txid}	Vector of measured RSS

PL_{dB}	Path-loss model in decibel unit
φ	Phase shift
R_{txid}^{m}	Vector of RSS estimates
RSS_j	RSS measurement of j^{th} measurement points in the considered radio grid
$RSTD_i$	RSTD referenced to the i^{th} neighbor BS
σ_{txid}^{m}	Set of shadowing standard deviation estimates
t	Received signal time
t_0	Transmit signal time
t_1	Received signal time at reference node one
t_2	Received signal time at reference node two
T_{bs}	Time between the reception and transmission of a certain signal sub-frame by the BS
T_{ms}	Time between the transmission and the reception of a certain signal sub-frame by the MS
T_s	RSTD time unite
$\tau_{meas\ err}$	Overall time error of the signal travelling time measurements
$\tau_{txs\ sych}$	Mutual time difference of the transmit time
x_i^{map}	Coordinate x of the i^{th} synthetic location points in the considered power map
x_j^{meas}	Coordinate x of j^{th} measurement location points in the considered location grid
x_l	Coordinate x of the l^{th} location point in the considered location grid
x_{n_i}	Coordinate x of the i^{th} neighbor BS
x_r	Coordinate x of the reference BS
x_{txid}	Coordinate x of transmitter location referenced to $txid^{th}$ TXID
X_{txid}^{m}	Vector of PL parameters
y_i^{map}	Coordinate y of the i^{th} synthetic location points in the considered power map

y_j^{meas}	Coordinate y of j^{th} measurement location points in the considered location grid
y_l	Coordinate y of the l^{th} location point in the considered location grid
y_{n_i}	Coordinate y of the i^{th} neighbor BS
y_r	Coordinate y of the reference BS
y_{txid}	Coordinate y of transmitter location referenced to $txid^{th}$ TXID
w	Random variable modelling shadowing
W_{txid}^m	Vector of shadowing parameters

1. INTRODUCTION

Today, according to the Federal Communication Commission (FCC), the majority of emergency calls come from mobile phones [25]. Because of the importance to provide accurate and reliable location information for safely service, the government agency FCC in the USA established the location accuracy requirements for the enhanced 911 (E911) emergency wireless service [26]: 67 % of calls expected below 50 m and 95 % of calls expected below 150 m for the end-node-based positioning; 67 % of calls expected below 100 m and 95 % of calls expected below 300 m for the network-node-based positioning. Corresponding EU requirements are addressed in Europe for the E112 Enhanced European emergency number. However, the United States is the country leading the regulation in place, thus, our reference rule-maker. Laying the foundation to research solutions performing satisfactory location estimation in wireless devices.

By determining the location of a commercial mobile radio device on Earth's surface, both in indoor and in outdoor environments, network operators can offer location-based services (LBS) [7]. The spectrum of LBS services though being rapidly expanding is limited by the environment-dependent positioning accuracy that the current available technology can offer. Thus, always-increasing market needs may require solutions even more stringent than that declared by E911.

Positioning based on Global Navigation Satellite System (GNSS) is the most effective technology in the outdoor environment. However its performance dramatically degrades in dense urban areas due to the multipath channel. Besides, not all the mobile devices are equipped with GNSS-receiver and users may want to switch it off, to reduce the battery consumption of the device. Furthermore, the acquisition of a GNSS signal is not immediate but lasts several tens of seconds under usual circumstances. Conversely, mobile network positioning represents an energy-efficient and low-latency solution but accuracy is poor when compared to GNSS-based system, for most of the outdoor scenarios.

With the availability of geographical information measured by commercial mobile radio devices, network operators can investigate the performance of their cellular networks, for example to verify the presence of coverage holes, in automatic way; consequently, minimizing cost and also time otherwise needed to manually carry out Drive Test (DT) by engineers and operators in the field. The feature of combining radio measurements of the User Equipment (UE) with location information for the specific purpose of optimizing the cellular network at minimal operation expenditure is referred to as Minimization of Drive Test (MDT) [30], [42].

Collecting data by means of MDT is in turn also convenient from the purely positioning prospective. Indeed, MDT information might allow to overcome the main drawback of the

fingerprinting (FP) positioning technology concerning the periodic updating of the database due to inevitable time variation of the spatial distribution of the signal strength [32]. One such opportunity, adds value and motivates the choice of exploring received signal strength (RSS) based positioning methods to our work.

This thesis focuses on outdoor positioning in Wireless Local Area Network (WLAN) and Fourth Generation (4G) wireless cellular networks and is based on an extensive collection of radio measurements, i.e. Received Signal Strength Indicator (RSSI)/Basic Service Set Identifier (BSSID) from Wireless Fidelity (WiFi) signals, i.e. Reference Signal Received Power (RSRP)/Cell Identifier (CID) from Long Term Evolution (LTE) signals. In addition, radio measurements are supplied with Global Positioning System (GPS) location information and time stamp. The dataset was provided by Magister Solution Ltd after conducting a war-driving-like activity, moving by either car, bicycle or foot, using Samsung Galaxy S3 mobile device for dense sampling. The sampling counts 36648 unique GPS location points in a light building density area of 0.4 square kilometers. The urban characteristics of the measurement area allow to assume the GPS measurement error in a few meters; a small error when compared with our target accuracy which can therefore be neglected to our purposes.

The objective of this scientific work is to explore the performance of three outdoor positioning algorithms designed to achieve horizontal positioning in area covered by radio cellular networks signals. The exploration interests are focused on positioning accuracy and database size. The research aims to address the E911 location accuracy requirements while trying to reduce the number of data needed to compute a location estimate. Fingerprinting is used as benchmark against which to assess the performance of methods based on statistical models. Statistical models need a few parameters per transmitter identification (TXID) and, compared to fingerprinting, appear more suitable for low-cost devices because allow to reduce the database size while offer comparable accuracy.

1.1 Author's contribution

The substantial contributions made by the author of this thesis are outlined as follows:

- The estimated parameters apparent transmit power, path-loss (PL) coefficient and shadowing standard deviation are analyzed with respect to the examined study case. Also, the effect of using the carrier frequency information in the estimation model in relation to estimating is studied.

- The process of splitting data into estimation and training data is explained. A framework block diagram of the process is created for illustrative purpose and for

applicability. The provided methodology allows to develop different indexing algorithms to the same structure.

- In addition to fingerprinting, two positioning methods based on the statistical models, respectively, coverage-area (CA) and path-loss, are implemented. The latter method rely upon the knowledge of the true location of the antennas, thus, in absence of this information, the definition of an estimation method become crucial. Two possible estimation methods, K-strongest RSS-weighted centroid and first strongest selection, are explained and their performance are studied in mutual respect. In this regard, a set up algorithm, required in the performance studies, is developed and its flow chart is shown.

- An algorithm performing transmitter (TX) location estimation via the method of K-strongest RSS-weighted centroid is implemented. The algorithm is suitable for an input data format such as TXID grid and supports a number of estimates as large as the number of the transmitter identities, or number of the TXID grid points. By definition, the algorithm turns from a K-strongest RSS-weighted centroid based estimator into a first strongest selection based estimator by setting the input parameter K to 1.

- Two algorithms with functionality of, respectively, multiple scans (MSCs) removal and TXID grid building are designed and implemented, however, because they are marginal topics of the thesis, are not described in detail.

- A simulation software framework is developed for running of experiments to positioning accuracy. This framework is designed to execute preliminary operations such as the multiple scan removal and the transmitter identity grid forming. The framework also allows to set simulation parameters relevant to the positioning algorithm. The choice of the simulated positioning algorithm as well as the number of experiments is tunable by design.

- The process of creating a power map by collection of signal strength measurements is step-by-step explained.

- Based on an existing framework, the fingerprinting, the coverage-area and the path-loss model based positioning algorithms are developed. Their performance, in a suburban-type outdoor environment, is experimentally measured by the cumulative distribution of the distance error and by the database extent. The results are compared to each other in two separate cases, that is, results either based on LTE or

based on WiFi. In addition, the effect of using the carrier frequency information in the estimation path-loss model, and of using only strong signals (instead of all signals) in the estimation coverage-area model, in relation to positioning are to some extent evaluated.

- A computer-based program set is developed to assist the research in performing of data analysis and interpretation. For instance, the empirical statistic of the received signal strength measurements is shown and discussed.

- The idea of developing an angle-based estimator to cope with path-loss modeling in directional electromagnetic field is conceived. The conception is delineated. It represents an intellectual content made by the author who believes that can play a key role in addressing outdoor positioning advancement.

1.2 Thesis outline

In Chapter 2 we firstly introduce the LTE signal structure and the LTE positioning-dedicated feature. Next, with reference to LTE, we address non RSS-based and RSS-based positioning. The former topic begin with a description of the positioning concept model, after which the main LTE standardized positioning methods are delineated. The latter topic is prefaced with the description of RSRP, Reference Signal Received Quality (RSRQ) and RSSI LTE standard measurements and relative mapping systems, and includes a preliminary description of the general two-stages fingerprinting approach followed by the description of an estimator based on classical fingerprinting algorithm.

In Chapter 3 we present an overview of the main IEEE 802.11 family standards. Next, based on the positioning signal parameters introduced in Chapter 2, we address IEEE 802.11-based non RSS-based positioning technique, pointing out the lack of positioning-dedicated supporting amendments as relevant issue. After that, in the same chapter, we address positioning based on signal strength facing the RSSI-RSS mapping system and the inherent inconsistency among 802.11 chipset manufacturers.

Chapter 4 contains a description of the statistical channel models that are used for positioning purposes and it includes methodology, experimental analysis and results. This chapter ends with an example of 2-dimensional measurement power map and a step-by-step power map building definition.

In Chapter 5 two positioning algorithms based on the statistical model coverage-area and path-loss respectively, are explained in detail.

In Chapter 6 firstly the measurement campaign is described and its output dataset carefully analyzed. Secondly, the topic of transmitter localization estimation is addressed with respect to two possible algorithms. Thirdly, a performance evaluation of the considered outdoor positioning algorithms is addressed in terms of Root Mean Square Error (RMSE) and Cumulative Distribution Function (DCF) of the distance error, as well as in terms of number of parameters inherent in the database.

This thesis concludes in the Chapter 7. Here the main results of the thesis are summarized and the directions for the future work are pointed out.

2. LTE-BASED POSITIONING

2.1 LTE signal structure

The frame structure of a LTE signal is defined by the European Telecommunication Standards Institute (ETSI), an European Standards Organization (ESO) partner with the Third Generation Partnership Project (3GPP), in its technical specification document [23]. The standard prescribes two possible frame structures, namely type 1 and type 2, depending on whether the operational duplexing mode is Frequency Division Duplex (FDD) or Time Division Duplex (TDD). Unlike TDD, FDD uses different frequencies to separate the up and down communication links.

The frame structure type 1 for FDD is shown in Figure 2.1. It consists of frames, sub-frames and slots of length respectively 10 ms, 1 ms and 0.5 ms. Each frame is dived into 10 sub-frames being in turn divided into 2 slots. Slots within a frame are numbered with progressive indexes from 0 to 19. Thus, one frame is composed by 20 slots of which 10 slots are dedicated for up-link (UL) and 10 slots are available for Down Link (DL). The two communication links use different multiplexing schemes being Single Carrier Frequency Division Multiple Access (SC-FDMA) and Orthogonal FDMA (OFDMA). The scheme SC-FDMA is preferred in DL since it has low Peak to Average Power Ratio (PARP) compared to OFDMA [31]. This allows employing inexpensive power amplifiers to end-nodes, which is convenient for mass production of user equipment.

The structure of a slot is somewhat flexible. It can vary both in frequency and in time domain, depending on the deployed system bandwidth and Cycle Prefix (CP). In more details, channel bandwidth ranges from 1.4 MHz to 20 MHz with six possible choices, and cycle prefix can be either short CP or extend CP. Which bandwidth is used depends on the amount of available spectrum whereas the choice of the cycle prefix depends on the size of

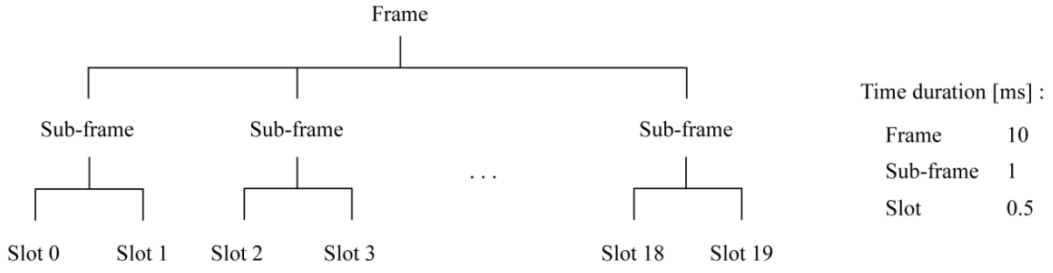

Figure 2.1. The frame structure of type 1 defined for FDD. Adapted from primary source [23].

the coverage cell. The CP is used to cope with signal distortion, or inter-symbolic interference, due to the multi-path propagation channel. The extended CP is used for larger cell size as the amount of delay spread of the radio channel become more significant. Thus, the configuration of the physical resources can be designed in flexible manner, according to the ETSI specifications [23] and cellular radio network requirements such as capacity and coverage.

The term Resource Block (RB) is used to denote the minimum amount of transmission resource, in terms of number of subcarriers and transmit symbols, that can be allocated to the end-node within a time slot. Specifically, one RB consists of 12 subcarriers, equally interspaced with 15 KHz of communication frequencies, and of either 6 or 7 symbols, depending on whether normal CP or extended CP is used. Symbols consist of either OFDM symbols in DL or SC-FDMA symbols in UL, as different multiplexing schemes are adopted for the two communication links. One RB occupies 0.5 ms in time domain and 180 Khz in frequency domain, and forms a resource grid of basic resource elements identified by two indexes associated with subcarrier and symbol, respectively. As the size of a RB is fixed then the overall number of RBs available in one slot increases with the deployed channel bandwidth. For instance, 1.4 MHz of bandwidth might give room to 6 RBs while 20 MHz of bandwidth allows 100 RBs (see [22] annex A.2.2). One user can be provided with a certain amount of RBs in each slot and furthermore, in the DL, several users can be served within the same slot as by OFDMA users are scheduled both in time and frequency domain. Figure 2.2 illustrates the DL resource grid within one sub-frame. It is worth to notice that the UL resource grid would look the same but with SC-FDMA

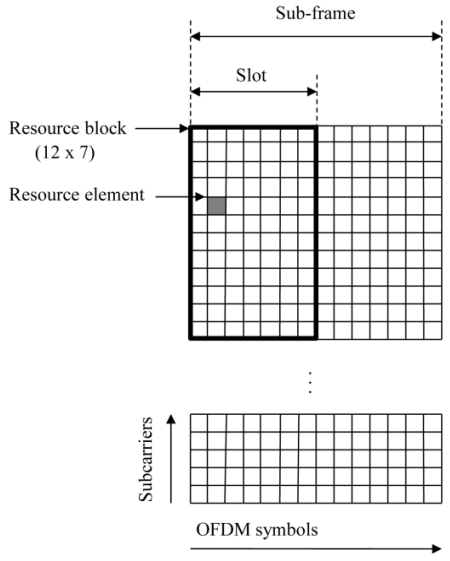

Figure 2.2. The resource grid within one DL sub-frame. Adapted from primary source [23].

symbols instead of OFDM symbols.

The frame structure type 2 which can be applied for TDD is shown in Figure 2.3. Here, one radio frame of 10 ms duration is dived into two half-frames of 5 ms. Each half-frame is split into five sub-frames of duration 1 ms therefore the entire sequence of sub-frames consists of ten sub-frames, numbered as 0,1...9. In the normal case, a sub-frame is divided into two slots each of half ms, and used for data transmission in either DL or UL. Besides, the so called special sub-frame is used to perform duty of switching from DL to UL. In such way, the first sub-frame before it is always used for DL while the first sub-frame sfter it is always reserved for UL. The special sub-frame is obligatory placed to sub-frame number 2 and optionally, depending on the supported configuration, to sub-frame 6. The standard prescribes seven possible configurations to fix the uplink-downlink allocation along the frame as reported in Table 2.1.

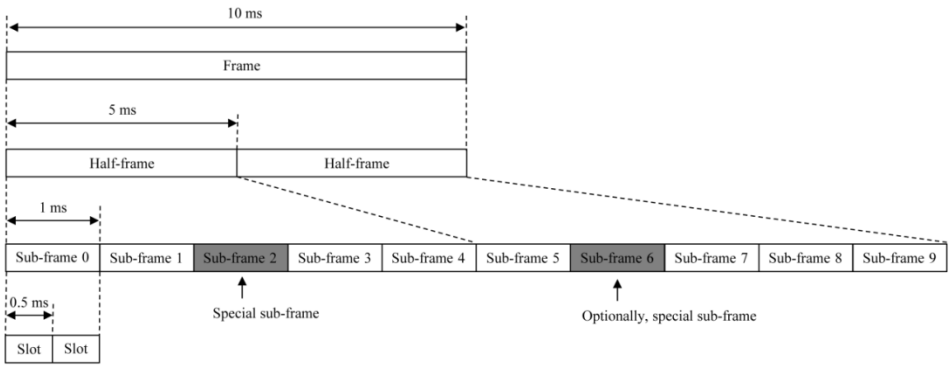

Figure 2.3. The frame structure of type 2 defined for TDD. Adapted from primary source [23].

Table 2.1. Configurations setting for the uplink-downlink resource allocation of frame type 2. Adapted from primary source [23].

Configuration number	Sub-frame number									
	0	1	2	3	4	5	6	7	8	9
0	DL	S	UL	UL	UL	DL	S	UL	UL	UL
1	DL	S	UL	UL	DL	DL	S	UL	UL	DL
2	DL	S	UL	DL	DL	DL	S	UL	DL	DL
3	DL	S	UL	UL	UL	DL	DL	DL	DL	DL
4	DL	S	UL	UL	DL	DL	DL	DL	DL	DL
5	DL	S	UL	DL	DL	DL	DL	DL	DL	DL
6	DL	S	UL	UL	UL	DL	S	UL	UL	DL

2.2 LTE positioning signal

In order to estimate the radio channel characteristics, specific Reference Signals (RSs), also known as pilot symbols, are used in the down-link. These signals are known in advance and mapped into specific resource elements at given time and frequency.

Using RSs was originally intended for the wireless communication need of increasing channel capacity. Nevertheless, it turned out that RSs could also be exploited to positioning purpose, for instance via Observed Time of Arrival (OTDOA), see Section 2.3.3. Implementing OTDOA requires 2 to 4 neighbor Base Stations (BSs), i.e. BSs other than the serving BS, to be detected simultaneously. The detection of RSs coming from neighbor BSs, however, may fail if Signal to Interference plus Noise Ratio (SINR) comes poor, that is below -13 dB [2].

The Cell-specific Reference Signal (CRS), introduced by 3GPP in the LTE standard release 8 [23], exhibits a poor detection probability of the third and fourth best neighbor BSs, even under the ideal scenario of low loading and high processing gain [6]; therefore it is not adequate for OTDOA positioning. Thus, beside to CRS, a new reference signal called Positioning Reference Signal (PRS) was defined in release 9 of the standard [23] and dedicated for OTDOA.

PRS signals along with CRS signals are carried in a special sub-frame which is designed in such a way to increase the reference signal power meanwhile reducing the interference level due to transmission of neighbor BSs. The layout of the resource grid for such a special frame is defined in such a way that PRSs are not overlapping with CRSs, neither with PDCCH; which is a control channel responsible for carrying control information in down-link. Also, PRS can shift in frequency by a number of sub-carrier to some extent depending upon PCI, the physical identity of a transmitter. There are six possible shifting schemes which translate into reuse factor of six preventing, or reducing, inter-cell interference. The construction of a PRS sequence can be found for example in [23], the primary source, and in [41], a secondary source with a clear line of reasoning.

The use of PRS signals can truly support OTDOA positioning by reducing the effect of near-far or receivability problem. As proof, test results on hear-ability of reference signals showed a significant improvement of the performance in terms of detection probability of the fifth best neighbor cell, if PRS is used [6].

To conclude, driven by the interest of increasing the positioning capabilities of LTE signals, positioning reference signals have been introduced within special sub-frames dedicated to positioning purpose. Such positioning reference signals can be detected by neighbor base stations with higher probability compared to the legacy cell-specific reference signal and hence more suitable for OTDOA positioning.

2.3 Non RSS-based LTE positioning

2.3.1 Positioning signal parameters overview

In this section, we present a set of signal parameters that are commonly used in non RSS-based wireless location estimation. We describe how they can be used in principle to determine the user position, i.e. a mobile device within a wireless network coverage area, and which are their major limiting factors conditioning the location finding accuracy.

TOA

The parameter Time of Arrival (TOA) refers to the time a radio signal takes to travel from a communication node to another one. As the signal propagates at the speed of light c then such a time difference translates into distance by $d = c\,(t - t_0)$ where t_0 is the transmit signal time and t is the received signal time. In two dimensions, if the location of one of the two nodes is known then the location of the other one, usually referred as to target node, is defined by a circle. Thus, with three node and hence by three circles, the location of the target node can be in principle uniquely determined as the intersection point. See circular triangulation principle in [7].

The time synchronization that there exists between BS and UE can be suitable for communication purpose but might be not accurate enough for positioning objectives. Indeed, even a very small time offset e.g., a few microseconds, might result in a huge distance error as radio signals propagate at the speed of light (about 300 m per microsecond). Thus, a more practical expression for the distance estimation is given as $d = c\,(t - t_0 + \Delta t)$ where the time-offset Δt carries uncertainty over the location estimation.

A classical way to achieve TOA estimation is cross-correlation-based. This way searches for the maximum of the cross-correlation between the reference signal and the received signal. Due to the propagation channel, the latter is a scaled-delayed and noisy version of the former and hence the peak of the Cross Correlation Function (CCF) corresponds to the sought delay. In practical estimator the resolution of the delay estimate corresponds to the sampling frequency so oversampling the signal at the receiver (RX) is conventionally adopted to improve the performance at expense of processing burden. Further, oversampling factor is constrained by the Nyquist-Shannon sampling theorem [56]. Because the time precision is vital for such TOA-based estimators several advanced methods have been proposed in literature. The reader can see, for instance, the fast parabolic based method for LTE scenario proposed in [13]. The performance of a TOA-based estimator degrade considerably in presence of non-line-of-site (NLOS) and multipath environment. Studies on the impact of the propagation channel to the LTE-based

positioning is carried out in [20] whereas the most commonly adopted mitigation techniques against NLOS accuracy loss are summarized in [29].

DTOA

The parameter Difference Time of Arrival (DTOA) consists of TOA-difference between the target node and two different network nodes. As time translates to distance, by calculating the distance difference $d_{21} = c\,(t_2 - t_0) - (t_1 - t_0)$, the transmit time t_0 cancels itself out; while the received signal times t_2 and t_1 stay on the formula. This in turn eliminates the problem of synchronization between the transmitter and the receiver but it requires aligned transmit time among the reference nodes. To determine the user location at least four base stations provided with their coordinates are required; that is three difference distances. With this information the hyperbolic trilateration criterion [7] geometrically find where user is placed by intersection of the three parabolic equation whose focal points are located at the base station points. DTOA can resolve the clock synchronization problem but still faces the propagation channel limitations just like TOA.

AOA

Angle of Arrival (AOA) refers to the angle of incidence of a radio wave propagating in the up-link from the target UE to the base station. One such parameter can be obtained at the BS provided that it is equipped with specialized adaptive antenna. The antenna array is formed by equally spaced antenna elements with known separation distance $\mathrm{d_{ar}}$ equal to half wavelength. The most simple model considers two antenna elements at the receiver and a single-path propagation channel. In the absence of multipath, the signal coming from the unknown target device reaches the antenna elements by the line of sites or direct path at different time and with different phase shift. Based on the geometrical scheme of Figure 2.4 once the phase shift of carriers is measured the angle of arrival α can be calculated as,

$$\alpha = \cos^{-1}\left(\frac{\mathrm{d_{ar_1}}}{\mathrm{d_{ar}}}\right) = \cos^{-1}\left(\frac{\varphi\, c}{2\pi f_c\, \mathrm{d_{ar}}}\right) \tag{2.1}$$

where $\mathrm{d_{ar_1}}$ is the projection of $\mathrm{d_{ar}}$ onto *H*, φ is the phase shift, f_c is the carrier frequency and *c* is the speed of light. Notice that the condition $H \gg \mathrm{d_{ar}}$, with *H* the distance between transmitter and receiver, is to be satisfied.

The estimation of target location based on AOA requires at least two BSs but no clock synchronization as transmit time is not concerned.

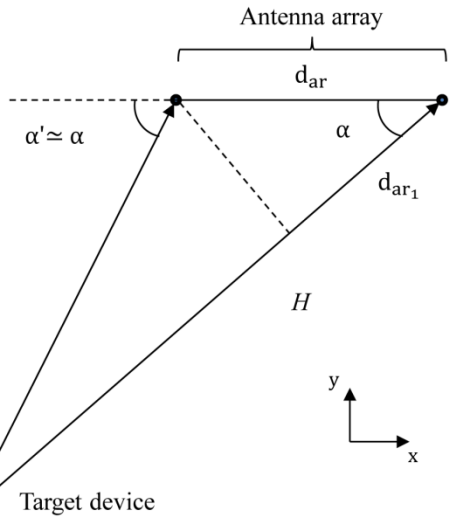

Figure 2.4. AOA geometrical computation. Adapted from [16].

The block diagram of Figure 2.5 shows an AOA system model from the implementation point of view. In this simplified, model only two BSs are concerned and their heights are neglected. Each BS has capability to determine the direction of arrival of the signal transmitted by target device thereby, the intersection of two lines finds the target position. A simulation result is shown in Figure 2.6 for illustration purpose.

The operation principle of AOA based localization was described by emphasizing the geometrical aspects through a completely ideal model. In real case, also this typology of estimators is effected by noise, multipath and other factors diminishing the accuracy. To deal with the communication channel limiting factors several enhanced solutions have been studied (see for instance [67]).

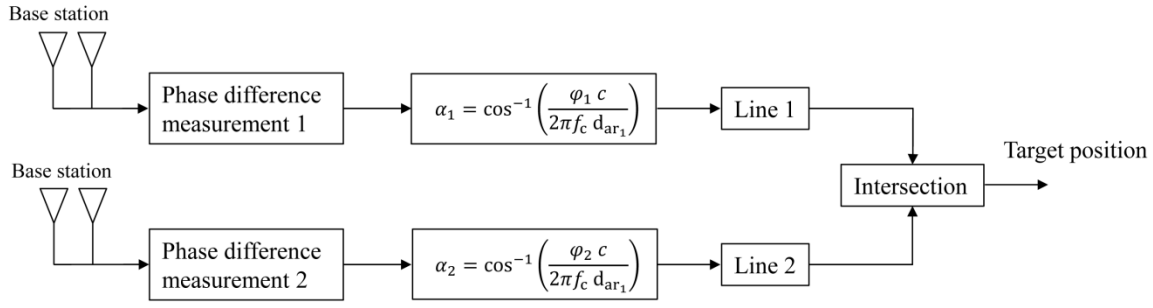

Figure 2.5. AOA block diagram model.

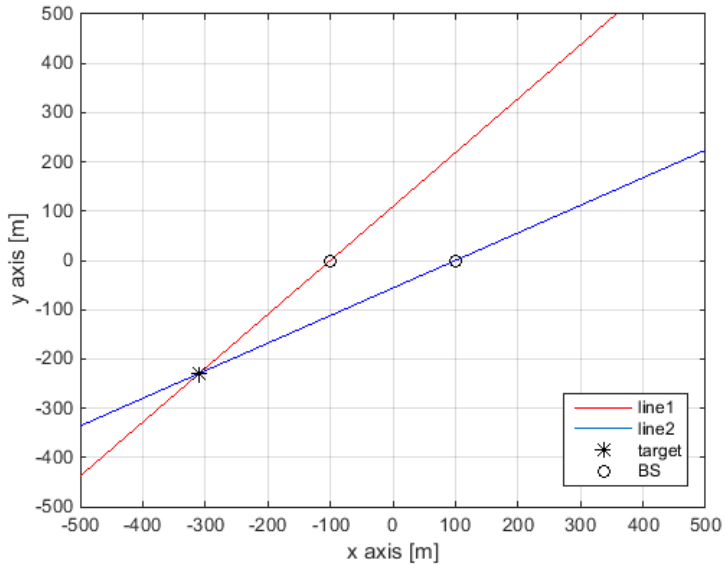

Figure 2.6. Triangulation: a simulation result implemented by block diagram of Fig. 2.5.

2.3.2 LTE positioning concept model

The composition of concepts used in [4] to describe the LTE positioning system is shown in Figure 2.7 for illustrative purpose. There are four fundamental elements involved: two reference sources, namely a GNSS satellite and a LTE base station, a Target Device for instance a mobile phone and a Location Server that is an end node with positioning dedicated hardware and software. They are interconnected to exchanging information and operate with the final objective to locate the target device.

The target device receives positioning radio signals from one or both the reference sources. It obtains positioning measurements and sends them to the location server which determines the location of the target device and sends the location information back to the sender. The exchange of information between the target device and the location server is supported by the LTE Positioning Protocol (LPP) protocol [4]. The LPP supports the hybrid positioning OTDOA/Assisted Global Navigation Satellite System (A-GNSS) and also the delivering of assistance data containing parameters for both reference BS and neighbor BS. Assistant data is send from the local server to the target device and supplies information, such as the identity of the BS and the configuration of the PRS, enabling OTDOA. More details about the LPP procedure including the specification of different type of exchanged messages can be found in the primary source [4].

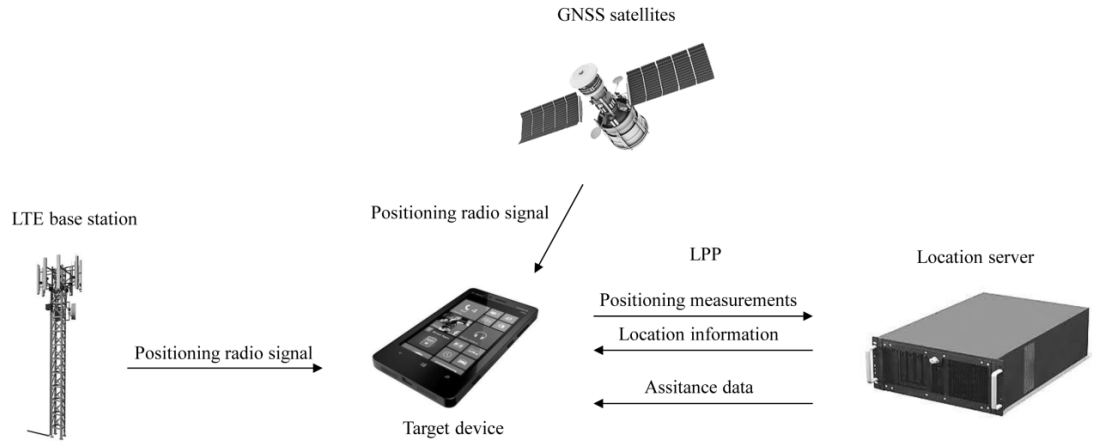

Figure 2.7. Concept model for LTE positioning. Source [4].

2.3.3 LTE positioning methods

The LTE standard specified so far three main positioning methods: Assisted Global Navigation Satellite System (A-GNSS), Observed Time of Arrival (OTDOA) and Enhanced Cell ID (E-Cell ID).

A-GNSS

A-GNSS is the use of a 4G cellular network to provide assistance data to the target device to improve the acquisition time of GNSS satellites. The method requires, first a GNSS-LTE enabled target device; second, at least four satellites to be visible at the same time; third, 4G network coverage.

Depending on the considered GNSS system, for instance GPS, Galileo or GLONASS, the overall number of satellites in the satellite constellation can be different. An number of nearly 30 satellites would be normally required to fully achieve global coverage.

A satellite is said to be visible when its transmit signals can be detected on earth by a GNSS receiver. The number of satellites that are simultaneously visible varies with time, place and the considered satellite constellation. Besides, satellite visibility is influenced by the environment surrounding the receiver.

The detection of GNSS signal is difficult in indoor environments as well as in outdoor environments with high urban density. In contrast, rural and urban areas are beneficial to GNSS signal reception thus, at least four satellites can be typically detected at the same time in such environment.

The greater the size of a satellite constellation the more the processing time for determining which satellites are visible since the number of correlations required at the receiver also

increases [14]. However, the time needed for such operation might be reduced by a coarse knowledge of satellites and target receiver locations.
The basic GNSS-based positioning system is the most accurate positioning method (a few meters accuracy with no augmented systems accounted) but it fails in hostile environments. Furthermore, the acquisition of satellites signals might last several minutes and the time to locate a target device might therefore not be appropriate in many location based service instances, by only GNSS.
In order to find its own location, a GNSS target receiver performs acquisition and tracking of satellite signals and decoding of navigation message. The navigation message contains information, such as the reference time, the state of health of satellites and their orbital position, vital to the position calculation.
The major benefit of combining cellular network capability with GNSS system consists of reducing the Time to First Fix (TTFF): the receiver time processing from positioning request start to position computation output. Indeed, receiving the navigation message from the cellular network along with a coarse location estimate of the receiver, facilitates the determination of which satellite is visible, reducing the frequency window for computing the satellite Doppler shift.

OTDOA

OTDOA relies upon time difference measurement between the RS travelling from the reference BS to the user equipment or target device and the RS travelling from a neighbor BS to the UE. One such time difference referred to as Reference Signal Time Difference (RSTD) [3] is measured in units of $T_s = 0.3255$ ns and can be expressed as:

$$\mathrm{RSTD}_i = c^{-1}\sqrt{\left(x - x_{n_i}\right)^2 + \left(y - y_{n_i}\right)^2} - c^{-1}\sqrt{\left(x - x_r\right)^2 + \left(y - y_r\right)^2} + \tau_{txs\ synch} + \tau_{meas\ err} \quad (2.2)$$

where, in a two dimensional Cartesian system, x and y are the unknown coordinates of the UE, x_{n_i} and y_{n_i} are the coordinates of the i^{th} neighbor BS, x_r and y_r are the coordinates of the reference BS, and where c is the speed of light, $\tau_{txs\ synch}$ is the mutual time difference of the transmit time and $\tau_{meas\ err}$ is the overall time error of the signal travelling time measurements. Notice that the first and second terms are TOAs measurements (see Section 2.3.1) while the third and fourth terms are additional noise factors. Also notice that RSTD_i, neglecting noise, is the equation of a hyperbola with focal points in x_{n_i}, y_{n_i}.
OTDOA requires, in addition to the reference BS, at least two neighbor BSs and UE must be able to detect RS signals from them. In order to enhance the probability of detection, the RS signals involved in the OTDOA measurements are specifically PRS signals, as pointed out in Section 2.2.

The location finding principle of OTDOA is not different from the one used by TDOA which is the fundamental positioning signal parameter described in Section 3.3.1. However, OTDOA specifically refers to the LTE network system and makes use of LTE positioning signals. Thus, provided that the BSs locations are known, a LTE enabled target device is found by the intersection of three or more hyperbolas by using an hyperbolic multilateration algorithm. The intersection, ideally a point of coordinates x and y in a Cartesian coordinate system, becomes an area due to the offset in the transmit time between each pair of BSs participating in the OTDOA calculation; and additionally, due to the measurement errors of TOA which is the signal time of arrival described in Section 2.3.1.

The error in TOA is mostly caused by the multipath propagation which translates into several scaled and delayed copies of the transmit signal, with different delay and scale factors, at the receiver. The extent of the error is influenced by the bandwidth of the signal and, since LTE supports scalable bandwidth, it may vary significantly, depending to the deployed channel bandwidth. For instance, using 20 MHz channel bandwidth the error is much smaller compared to 1.4 MHz bandwidth. As general rule, the larger the bandwidth the smaller the error due to multipath. For this reason, positioning systems based on Ultra Wide Band (UWB) have gained high interest by the scientific community devoted to the study of positioning.

The extent of the area or positioning uncertainty also depends on geometrical factors: the number of BSs, the locations of BSs relative to each other and the location of BSs relative to the target device. The gain of the uncertainty due to the system geometry is usually taken into account by a parameter called Geometrical Dilution of Precision (GDOP).

The parameter GDOP was defined first in GNSS over a tridimensional configuration of satellites [54],[71]. By this parameter, if more than four satellites are visible, the receiver can calculate which satellite configuration has smaller dilution of precision and select such configuration as optimal solution, thus, optimizing the location estimate from the geometrical point of view. In this context, some authors such as [52],[87] and [74], considered the problem of reducing the computational complexity of GDOP.

In contrast to GNSS, in radio cellular network GDOP can be formulated as a two dimensional problem. The GDOP formula remain unchanged but the geometric matrix reduces its complexity of one dimension. Similarly to GNSS, OTDOA GDOP varies with the spatial configuration of base stations which in turn effects the positioning accuracy. In [18] a possible criterion to improving the positioning accuracy by selecting the base station configuration with minimum GDOP is proposed.

The major limiting factor for OTDOA is represented by the clocks used at BSs. Such clocks were designed for communication, not positioning. The latter needs a more stringent requirement of the clock phase to adequately address OTDOA positioning. For instance, using atomic clocks, such as those on board GNSS satellites, at BS would be much more beneficial for OTDOA.

E-Cell ID

The basic principle of E-Cell ID consists of adding information such as RTT, TOA or their combination to the basic Cell ID method, based on just the identification of serving cell.
The basic Cell ID method is very simple but not accurate. It relies on the location of serving BS as UE location estimate. Thus, it can only serve as a coarse estimation of the UE location since the uncertainty of model is the whole serving coverage area. For example, given a LTE macro scenario[1] with inter-cell distance of 500 m and three-sector site per cell, the coverage extent of a cell can be approximate 0.07 km^2 and, by knowing the Cell ID, it can be argued that Mobile Station (MS) places itself at one location point within the area of the identified cell. In more general terms, the size of the cell depends on the deployed cellular network and the positioning accuracy varies accordingly.
The term RTT indicates the round trip time for a radio signal travelling between MS and BS. It is not inclusive of the time spent in the transceiver so it is calculated as the difference $T_{ms} - T_{bs}$ where T_{ms} is the time between the transmission and the reception of a certain signal sub-frame by the MS; T_{bs} is the time between the reception and transmission of the same sub-frame by the BS.
The knowledge of RTT translates into the distance between BS and MS as $d = (\mathrm{RTT}/2) \cdot c$ with c speed of light. Thus, by combining RTT and Cell ID, MS estimated location is constrained by the circle of radius d and center BS location, increasing in accuracy with respect to the sole Cell ID method.
The accuracy of E-Cell ID could further be improved by taking into account also AOA (see Section 2.3.1). In this way the location estimation is found by the intersection between the prior defined circle and the direction of signal transmission from BS to MS as given by AOA.
We have described the ideal case for sake of the simplicity. In real scenario both RTT and AOA are affected by the various error sources and implementation requirements as reported in Section 2.3.1.

2.4 RSS-based LTE positioning

In this section, we first explain how received signal strength is computed in LTE and why it is important to consider RSS-based positioning in LTE systems. Second, we show that fingerprinting approaches positioning by mean of two stages: the off-line and the on-line stages. Third, we differentiate between non-parametric and parametric algorithms and last,

[1] Assuming a network of regular hexagonal shaped cells, split into three regular rhomboidal shaped sectors, and having the apothem of length equal to half an inter-cell distance.

we describe the classical fingerprinting-based estimator and provide an example of implementation.

2.4.1 LTE measurement: computation and mapping

One important feature in cellular radio system concerns user mobility. To maintain seamless connection to mobile users over the whole network coverage area, dedicated procedures such as cell selection, cell reselection and handover have to be executed. By executing mobility procedures network operators provide users with reliable connection.
In LTE, mobility procedures are based on RSRP and RSRQ measurements as reported to BS by UE. RSRQ is calculated by RSSI, therefore, beside RSRP and RSRQ, also RSSI has to be measured by the UE. However RSSI is not reported. The three measurements RSRP, RSRQ and RSSI are specified by the 3GPP LTE standard in [24], referred to simply as the standard from now onwards.

RSRP

The standard defines reference signal received power as the average power of the received signal referred only to resource elements carrying cell-specific reference signal.
The standard indicates that power is to be measured in watts at the antenna connector of UE and that the average power is to be calculated over the number of reference elements within the considered signal frequency bandwidth and the considered number of OFDM symbols. The measurement amount of signal bandwidth as well as the number of measurement OFDM symbols can freely be chosen by the test equipment manufacturers provided that the accuracy requirements of [2] are satisfied. Clearly, the more signal bandwidth the more reference signals, and hence the better trueness. The standard also indicates that RSRP is to be measured primarily in antenna port number zero, and optionally in antenna port number one if this one can be detected in reliable manner.
The prior knowledge about the transmitted sequence of symbols in RS allow to eliminate most of the interference in the received reference signals. The standard however does not provide statements as to whether the interference should be removed or not. Thus, different vendors or test equipment manufacturers may want to adopt different criteria.
RSRP measurements are to be reported by integer values according to a specified mapping table which maps $\{-140, (-140, -139], (-139, -138], \dots, (-46, -45], (-45, -44], -44\}$ into $\{0,1, \dots, 96,97\}$ [2]. The former set is referred to dBm unit whereas the latter set to pure numbers. We believe that the choice of one such set of reporting value was at least in part motivated by the need of minimizing the number of bit to be send to BS. However there is no evidence of the actual reason for adopting such a particular conversion rule.

RSSI

The standard defines received signal strength indicator as the average power of the received signal referred to all resource elements over the considered amount of measurement RBs and measurement OFDM symbols. According to the standard, power is to be measured in watts at the antenna connector of UE. Moreover, the measurement OFDM symbols is to be chosen among those carrying cell-specific reference signals, in antenna port number zero.

It is worthwhile to notice that RSSI includes co-channel interference and adjacent channel interference due to neighboring cell, and thermal noise. Besides, RSSI increases with the bandwidth and varies in time with the communication traffic intensity. Because the definition of RSSI is somewhat flexible, measurements obtained by different UE vendors might not be exactly aligned.

RSRQ

The standard defines reference signal received quality as $N_{rb} \times (RSRP/RSSI)$ where N_{rb} is the number of measurement resource blocks and, RSRP and RSSI are the reference signal received power and the received signal strength indicator obtained over an equivalent number of resource blocks N_{rb}. The recommended N_{rb} is clearly the full set of RBs over the whole supported bandwidth. However the standard leaves free choice to vendors as long as they fulfill the accuracy requirements as in [2].
RSRQ reflects at least in part the quality of connection as it decreases with the interference. However, the decrease in RSRQ is also influenced by the communication traffic thus the assessment of the connection quality by the formula is not straightforward.
Similarly to RSRP, RSRQ is reported by integer values. In this case, the conversion rule is as, $\{-19.5, (-119.5, -19], (-19, -18.5], \dots, (-4, -3.5], (-3.5, -3], -3\} \rightarrow \{0,1, \dots ,33,34\}$ [2], where the set of measuring values (to left) is given in dBm and maps into the set of reporting values (to right) as dimensionless quantities.

2.4.2 RSRP-based positioning

The receiver architecture of UE is designed to measure the power of received radio signals in accordance to the standardized measurement requirements. Among the various LTE measurements, RSRP as described in Section 2.4.1 is suitable for localization purpose. This is because RSRP is not or little contaminated by the interference and furthermore not influenced by the data traffic at all. Therefore, it provides a reliable indicator on the attenuation of the signal level due to the propagation from BS to UE. Thus RSRP, in the way it is defined to the 4G cellular network, enables positioning system through various existing RSS-based methods. Notice that, in this LTE context, the broad term received

signal strength narrows down to RSRP. Even so, we use RSS and RSRP interchangeably in this chapter.

The advantage of RSS-based methods over other types of parameters-based methods is that it relies entirely on the existing network, with no need of additional equipment and with reduced impact of the propagation channel on accuracy loss. On the other hand an extensive and periodic measurement activity is required to record the propagation environment features and to cope with their time variations. By collecting a large quantity of RSS signatures throughout the measurement area, positioning can be achieved. However, conducting an extensive measurement campaign in a manual way might be too expensive and time consuming, therefore not appropriate for large coverage. In this respect, the minimization of drive test [11], [1], [5] has gained interest as an active research area of cellular network positioning, for the last few years [80],[58].

2.4.3 Fingerprinting approach

Positioning methods based on RSS such as fingerprinting and path-loss model are characterized by the two sequential stages of Figure 2.8. The first stage also known as training phase is off-line. Here training data, at least location and radio information, are collected by conducting a measurement campaign, sometimes referred to as drive test. Once

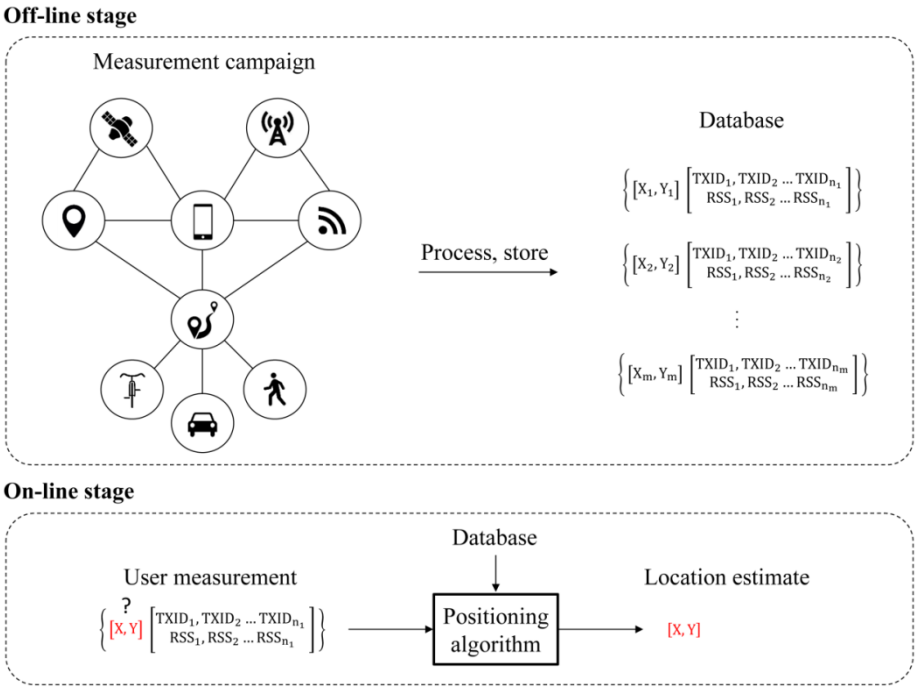

Figure 2.8. Two sequential stages of fingerprinting approach for 2D positioning.

collected, data are possibly processed and stored to be later used in the second stage. The second stage called estimation phase consists of real-time processing of data coming from the previous stage. The on-line stage performs location estimation by executing an estimation algorithm.
Estimation algorithms are classified into two fundamental classes: non-parametric algorithms and parametric algorithms (see [59] and references therein). In brief, non-parametric algorithms require a full database and no modeling of the signal propagation or other elements of system. Conversely, parametric algorithms perform estimation of a few parameters describing the statistical proprieties of data set. The goal is to reduce the size of the database and ease the burden of data transferring to the mobile UE as the algorithm to be executed; meanwhile being competitive with respect to the estimation accuracy. Parametric algorithms are typically implemented with statistical models as presented in Chapter 5. Therein, the two statistical models coverage-area and path-loss model are explained in detail.
Depending on the study case, data may also be pre-processed in the off-line phase. Here pre-processing may need to accomplish tasks such as the estimation of parameters, for example, BS location, transmit power and path-loss coefficient, and the removal of redundant data, i.e. data that have been collected several times for the same location point. In particular, the removal of the so called multiple scans aims to replace the RSS value for instance with the average signal level of the discarded data points, in order to statistically represent a better RRS-image of the radio channel for such location point.

2.4.4 Classical fingerprinting: a deterministic approach

Figure 2.9 shows the block diagram of a classical fingerprinting-based estimator. One such estimator exhaustively compares each radio map point with the radio information heard at the unknown user point and outputs the user location estimate based on the two following processing blocks:

- **Objective function block**, addressing the similarity in terms of radio characteristic between each pair of input data.
- **Decision block**, evaluating the objective function outputs in batch, and making the decision estimate by conventionally adopting criteria of maximizing.

In addition to the processing blocks, a store block and a switch block are used in such a way to deliver to the decision block, the entire set of data at a time.
Typically, the objective function is the inverse function of the signal distance which can be defined in different ways. One way to define the signal distance is by using the Euclidean

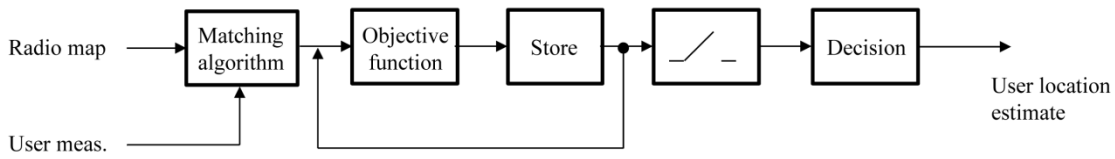

Figure 2.9. Block diagram of an estimator based on classical fingerprinting algorithm.

distance as $d = \left(\sum_{i=1}^{n}\left(\mathrm{x}_{\mathrm{user,i}} - \mathrm{x}_{\mathrm{training},i}\right)^2\right)^{1/2}$ where $\mathrm{x}_{\mathrm{user},i}$ and $\mathrm{x}_{\mathrm{training},i}$ are vectors of length *n* containing RSS values for commonly heard transmitters i.e., transmitters that are heard by both user and training measurement points. Other ways are also possible and some examples are found in [32].

The basic principle of classical fingerprinting consists of assuming the user location to be nearby to the location of the training point minimizing signal distance, or equivalently maximizing objective function. The most simple method, known as Nearest Neighbor (NN), considers the location of this particular training point as the best user location estimate. As the location estimate relies on the location of training points, the estimation accuracy is influenced by the density of training data. Intuitively, the higher density of training data the better accuracy of estimate.

If the measurement campaign is not too extensive, the most accurate way of doing location estimation [32] is to consider the locations associated with the K largest objective functions and computing the weighted centroid by minimizing the sum of square distances. This method known as Weighted K-Nearest Neighbor (WKNN) is generally found best accuracy at K equals to 3 or 4 [49].

WKNN may also be beneficial in case of coverage holes in the measurement area since computing the weighted centroid might result in the location estimate within such uncovered region. The impact of uncovered measurement regions on the positioning accuracy is studied in [79].

An example of script used in classical fingerprinting is shown in Algorithm 2.1 to which the nearest neighbor method is used as an estimation rule. In Algorithm 2.1 *d_vec* represents a vector of signal distances calculated as Euclidean norm, *n_trainingPoints* represents the number of data points obtained from the training phase, *flag* represents a boolean-like variable used to denote whether user position is detectable (*flag=1*) or not (*flag=0*), *data_training* and *data_estim* represent a cell array of location and radio information from training and estimation phases respectively. In more detail, across each row, the first cell consists of one vector of local coordinates $\left[\mathrm{x}_\xi, \mathrm{y}_\xi\right]$ and the second cell consists of one matrix of radio information $\left[\mathrm{tx}_{\xi,1}, \mathrm{tx}_{\xi,2} \ldots \mathrm{tx}_{\xi,\mathrm{n}_\xi}; \mathrm{rss}_{\xi,1}, \mathrm{rss}_{\xi,2} \ldots \mathrm{rss}_{\xi,\mathrm{n}_\xi}\right]$. Within the matrix, received signal strength and transmitter are paired across the columns. For

Algorithm 2.1. Example of program code implementing classical fingerprinting based on NN method.

```
d_vec = nan(1,n_trainingPoints);

for  j = 1:n_trainingPoints
     d = 0;
     flag = 0;
     for  k = 1:length(data_estim{1,2}(1,:))
          for  l = 1:length(data_training{j,2}(1,:))
               if  data_estim{1,2}(1,k) == data_training{j,2}(1,l)
                   d = d + sqrt(data_estim{1,2}(2,k).^2-
                                data_training{j,2}(2,l).^2);
                   flag = 1;
               end
          end
     end
     if  flag == 1;
         d_vec(1,j) = d;
     end
end

[~,pos_min_d_vec] = min(d_vec);

user_loc_est = cell2mat(data_training(pos_min_d_vec,1));
```

example $rss_{\xi,1}$ is associated with $tx_{\xi,1}$, $rss_{\xi,2}$ is associated with $tx_{\xi,2}$ and so on. Subscript ξ is equal to either 1,2,…,*i* or 1,2 … *j*, where *i* and *j* represent the number of samples in estimation and training cell arrays respectively. Subscript n_{ξ} stands for number of transmitters heard by the ξ-*th* sample. Continuing to describe script variables, *d* represents the sum of 2-norm RSS-distance from commonly heard transmitters in *i-th* estimation and *j-th* training points (notice that *i* was set to 1 in this example though any other user point could be chosen within the estimation array), *pos_min_d_vec* represents the position of the smallest element in the *d_vec* vector which is used to address the location estimate across the training data cell array. If user position is not detectable, for instance because no transmitter heard by the user occurs in any point of the training database, then *d_vec* consist of all not-a-number (NaN) elements.

3. WIFI-BASED POSITIONING

3.1 WLAN standard

The IEEE 802 LAN/MAN Standards Committee (LMSC) is in charge of developing the standards for Local Area Networks (LANs) and Metropolitan Area Networks (MAN). The committee is divided into many working groups (WG) of which 802.11 concentrates on WLAN networks related issues and standard definitions.

IEEE 802.11

IEEE 802.11 is the first legacy standard of the IEEE 802.11 family, released in 1997 and no longer in use. It provides the basis for the definition of WLAN networks in a flexible manner, with poor interoperability between vendors due to the lack of strict specifications. It was in fact not very successful and became soon obsolete, after the publication of more effective standards.

IEEE 802.11b

The standard IEEE 802.11b [35] was produced in 1999 by WG 802.11, a working group for WLAN standards. It supports an enhanced data rate compared with prior legacy standard, up to 11 Mbit/s per channel against the former 2 Mbit/s. Data rate decreases with distance from the Access Point (AP) to guarantee reliable connection, by changing throughput dynamically among 4 possible choices: 11 Mbit/s, 5.5 Mbit/s, 2 Mbit/s, 1 Mbit/s. According to the documentation [33],[36] the Physical Layer (PHY) of the standard supports Binary Convolutional Coding (BCC) as well as Phase Shift Code (PSK), for coding and modulation respectively [63]. There exist 3 possible digital modulation schemes: DBPSK (Differential Binary PSK), DQPSK (Differential Quaternary PSK), CCK (Complementary Code Keying) DQPSK, depending on required speed of data transfer. For instance, CCK-DQPSK [60], [9] uses Poly-phase Complementary Code (PCC) [27], [72] to achieve the higher data rates 5.5 Mbit/s and 11 Mbit/s, and allows interoperability with the lower data rates 1 Mbit/s, 2 Mbit/s schemes [9]. PHY amendment of 802.11b prescribes Carrier Sense Multiple Access along with Collision Avoidance (CSMA/CA), a multiple access technique performing spectrum sensing to infer packet traffic level and aiming to avoid collision of data packet on a probabilistic basis. If interested, the reader may find in [88] a theoretical study on CSMA/CA performance, in terms of both throughput [43] and delay. IEEE 802.11b PHY is provided with 83.5 MHz of frequency bandwidth, ranging from 2.4 GHz to 2.4835 GHz. It can accommodate up to 14 frequency channels of breadth

22 MHz, 3 of which are non-overlapping channels, using Direct Sequence Spread Spectrum (DSSS) modulation technique as originally defined in the 802.11 standard. A reader feeling not familiar with spread spectrum communication, may go through [61] to get an introduction into the topic.

IEEE 802.11a

The standard IEEE 802.11a [34] was completed in 1999, at the same time as IEEE 802.11b but manufacturers encountered some issues with the first produced devices. This caused delay to the market entry, which maybe eased the way for its competitor IEEE 802.11b which instead had an immediate success. IEEE 802.11a presents three main differences with respect to IEEE 802.11b (i.e. the closest standard to the original IEEE 802.11). First, it operates in the Unlicensed National Information Infrastructure (U-NII) band which occupies radio frequencies from 5.150 GHz to 5.850 GHz. Because IEEE 802.11a works in U-NII while IEEE 802.11b in the Industrial, Scientific and Medical (ISM) radio band, they cannot interfere to each other but, on the other hand, they cannot inter-operate either. Furthermore, unlike with IEEE 802.11b, IEEE 802.11a is safe from interference coming from Bluetooth devices, cordless phones, microwave ovens and other devices working at ISM frequencies. Second, it adopts the Orthogonal Frequency Division Multiplexing (OFDM) modulation scheme [63]. OFDM exploits orthogonal subcarriers to carry information in a highly spectral efficiency way. Furthermore, as OFDM subcarriers are narrowband, OFDM signals are more resistant against the multipath impairment compared with DSSS signals. Therefore, IEEE 802.11a may deal with heavy multipath environments, for example office space, better than IEEE 802.11b. On the other hand, due to the higher carrier frequency, IEEE 802.11a signals attenuate faster with distance than IEEE 802.11b, especially if walls or other large obstacles are encountered along the propagation path. Third, IEEE 802.11b offers a higher maximum data rate per channel as well as more options for the dynamic management of the data base. The whole set includes 6, 9, 12, 18, 24, 36, 48, 54 Mbits/s.

IEEE 802.11g

The standard IEEE 802.11g [37] presents a mixture of distinctive features from IEEE 802.11b and IEEE 802.11a. The standard IEEE 802.11g operates in the ISM; consequently, it may inter-operate with IEEE 802.11b but not with IEEE 802.11a. The main adopted modulation scheme is OFDM. Nevertheless, DSSS is also supported to be able to operate with IEEE 802.11b nodes. Like IEEE 802.11b, IEEE 802.11g is subject to interference due to wireless systems operating in the same unlicensed ISM frequency band. Compared to IEEE 802.11b, IEEE 802.11g shows better data transferring capability, allowing to switch among a large set of data rate choices: 1, 2, 5.5, 6, 9, 11, 12, 18, 24, 36, 48, 54 Mbits/s.

Since the IEEE 802.11g market entry in 2013, the same WiFi device can typically support 3 modes (a, b, g) and 2 bands (2.4 GHz, 5 GHz).

IEEE 802.11n

The standard IEEE 802.11n [38] introduces Multiple Input Multiple Output (MIMO) with up to 4 spatial streams. MIMO is a transmitter-receiver multiple antennas system that, taking advantage of the multipath propagation channel, can offer significant improvement in the data rate, as well as in the propagation range, compared to the traditional antenna system. Thus, IEEE 802.11n, in the best case, can reach data rate of 600 Mbits/s per channel. This rate is 11 times higher than that prescribed by the prior standards IEEE 802.11a and IEEE 802.11g. IEEE 802.11n operates mainly in the ISM and optionally in the U-NII frequency spectrum. It supports two channel bandwidth modes: 20 MHz and 40 MHz. The former consists of one channel called primary channel while the latter is the sum of two equally wide channels, the primary and the secondary channels. Joining primary and secondary channels not only doubles bandwidth but also doubles the data rate. However, since interference must be prevented this mode may not be practicable, especially in the ISM band where the spectrum is somewhat busy. IEEE 802.11n uses OFDM with 52 subcarriers, and 4 different modulation schemes [63]: BPSK, QPSK, 16QAM or 64QAM. The standard has been developing for 7 years, 2002-2009.

IEEE 802.11ac

The development of standard IEEE 802.11ac started in 2011 and formally ended in 2014 with IEEE official approval on January. The standard is an extension of IEEE 802.11n driven by the need for faster data transferring over WLANs. IEEE 802.11ac works exclusively in the 5 GHz U-NII unlicensed portion of RF spectrum. In addition to the IEEE 802.11n PHY features, IEEE 802.11ac adds 80 MHz and 160 MHz to the channel bandwidth options, 256QAM to the available modulation schemes, and up to 8 spatial streams to the MIMO property. Furthermore, the standard introduces Multiple User (MU) MIMO, a new advanced technology which optionally serves down link stream of data to several users simultaneously.
More details on this and other important IEEE 802.11ac features are presented by Bejarano et al. in their overview article [12]. A complete description on IEEE 802.11ac features is found in the original document [44].

Other IEEE 802.11 standards

The standards so far described in this chapter mainly concern the physical layer, or layer 1, of the Open System Interconnection (OSI) model. Other IEEE standards dealing with the Data Link Layer or layer 2 of OSI exist. In-depth description of all existing IEEE standard,

whether in progress or already completed, is beyond the scope of this thesis. We therefore point out the original amendments, as well as books [66] [73], for coverage of IEEE standards in greater detail.

3.2 Non-RSS-based WiFi positioning

In Chapter 2, we presented a set of signal parameters that are commonly used in non RSS-based wireless location estimation. Positioning techniques based on these parameters have been extensively researched over the existing wireless communication systems. As for WLAN radio network, most of the existing research activity concentrates on indoor positioning since in such environment WLAN infrastructure is mostly deployed. Although with minor extent, WLAN network is currently deployed also in the outdoor environment. Here its deployment is expected to grow in the near future through the implementation of heterogeneous wireless networks [84], [53]. This work contributes to the exploration of WLAN network for the purposes of outdoor positioning.

In first section of this chapter, we presented an overview of the 802.11 standards. These standards were defined for the sole purpose of communication and, to date, are not supported by positioning-dedicated amendments.

Positioning methods based on IEEE 802.11 have been transversally researched over the OSI layers in the OSI reference communication network model. Whether the method is supported with hardware or software, the trade-off always concerns accuracy, cost and complexity aspects but, apparently, there is not yet a valid solution against the multipath and NLOS problems [51]. One straightforward conclusion could be to avoid using such methods in presence of unbearable multipath and NLOS conditions. This in turn would require to identify such critical conditions in a reliable manner. In [8] the NLOS identification problem is addressed. Here, the authors show that the coherence bandwidth represents a more suitable NLOS identification metric compared to the kurtosis and RMS delay spread metrics, for WLAN systems.

The different signal characteristics, in terms of frame structure, frequency band, modulation and channel bandwidth, along the existing IEEE 802.11 standards, makes challenging to find a generic positioning method that suits well in all the cases. In [45] the influence of the channel bandwidth over the mean error, with respect to the coherence bandwidth, is studied under specific study case. Three different IEEE 802.11 standards and relative channel bandwidths are explored and the different behavior and performance is showed, through both simulation and experimental results.

A thorough presentation of the subject is beyond the scope of this thesis. We indicate [51] for anyone interested in a deeper study. In [51] the authors survey many of the time-based positioning methods existing in literature for IEEE 802.11 standards, and report them by

category. As for angle-based localization approach, the current state of the art is identified with the Multiple Signal Classification (MUSIC) [75], [85], [48] and with Joint Angle and Delay Estimation (JADE) [82] though others proposed methods can be found in the literature [86].

3.3 RSS-based WiFi positioning

To perform network-related operations a WLAN node must know the strength of a WiFi signal impinging at the receiver. WiFi signals, if any, may come from multiple wireless channels which in turn may come from the same AP (multiple MAC address) or may not. To distinguish different signals, an Identification Code (ID), called MAC address, is used. Thus, the radio chip of a WiFi device, once inquired, may return a set of radio-related information, such as $\mathrm{R}_i = \{[\mathrm{ID}_{i,1}, \mathrm{RSSI}_{i,1}], [\mathrm{ID}_{i,2}, \mathrm{RSSI}_{i,2}], \dots, [\mathrm{ID}_{i,\mathrm{N}}, \mathrm{RSSI}_{i,\mathrm{N_{ds}}}]\}$ where ID is the MAC address, RSSI is the received signal strength indicator and $\mathrm{N_{ds}}$ is the number of detected signals. One such set of data is a function of time and place, that is the sampling instance *i*. Notice that RSSI cannot be measured as long as the received signal strength stays below the receiver sensitivity (RXS) level, a receiver functionality threshold varying with data transmission rate. RXS is vendor-specific therefore different Wireless Network Interface Card (WNIC) manufacturers may offer different specifications.

The standard IEEE 802.11 defines the RSSI as a dimensionless measure of the energy obtained by the receiver at PHY level. Precisely, RSSI must range from 0 to any positive integer value not greater than 255 and must be captured from the preamble and the header of the Physical Layer Convergence Protocol (PLCP) frame [33], not from the whole packet. For this reason it may be not exploited to fully addressing the communication interference, as stated in [83]. The use of RSSI is recommended, not compulsory. Nevertheless, it is largely adopted by the WNIC manufacturers. RSSI is used for performing 802.11 internal operations such as Clear Channel Assessment (CCA), Roaming and other operations. For instance, in CCA, RSSI is compared with a threshold to establish whether the transmission channel is clear or not. If RSSI stays above the threshold, then the state of the channel is clear, hence a packet is granted for transmitting as there is little or no risk of collision. Conversely, if RSSI stays below the threshold, no packet forwarding is allowed in order to prevent collision. Similarly, by performing Roaming, RSSI is monitored to allow a client to move from one AP to another one, with no impairment to the connection.

Since the standard IEEE 802.11 does not say how RSSI shall be related to the received signal strength, different WNIC manufacturers may adopt different strategies. Consequently, RSSI provided by different nodes could be not consistent to each other if nodes belong to different vendors. This is not an issue as long as nodes perform RSSI-based operations internally, that is the case of 802.11 wireless networking. There are cases,

however, where RSSI-based operations are performed externally to the manufacturers policy. For example, RSS-based WiFi positioning systems rely on signal strength information in absolute sense. In such case, different radio devices may report different positioning results unless RSSI it is correctly mapped into RSS before executing the estimation algorithm.

Most of the WNIC manufacturers published their own RSSI to RSS conversion rule, usually in form of lookup table or mathematical function. For example Atheros has maximum RSSI value equals to 60 and the linear equation $\mathrm{RSS} = \mathrm{RSSI} - 95$ of Figure 3.1 (magenta-dash-circle line) for mapping RSSI into RSS. Beside Atheros, Cisco and Symbol adopt a conversion table, not providing any functional dependence between the two quantities involved in the transformation. The table adopted by Cisco is different from the one adopted by Symbol, not only by content but also by size. Cisco has maximum RSSI equals to 100 while for Symbol the top RSSI value is only 31. Hence the granularity of the RSSI offered by Cisco is higher compared to that one offered by Symbol, which in turn is smaller compered to Atheros. For sake of compactness, the conversion table, both by Cisco and by Symbol, is plotted onto the graph of Figure 3.1.

Other manufacturers, for example Intel, may make use of the simplest conversion rule, giving RSS same as inverse RSSI but dBm unit.

Sometimes the conversion rule could even be unknown if the manufacturer is not interested in publishing it. In this case, however, it can be determined in experimental way. The work

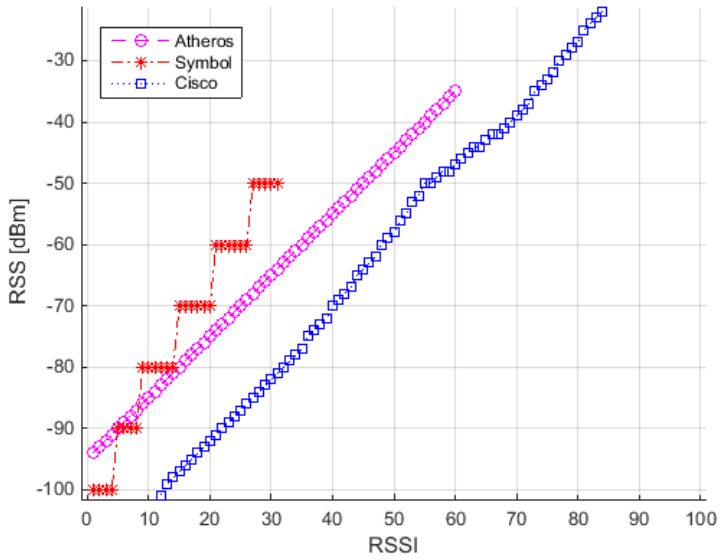

Figure 3.1. Illustration of three different RSSI-RSS mapping systems employed by different vendors: Atheros (magenta dash-circle line), Symbols (red dash-dot star), Cisco (blue dot-square line).

Table 3.1. Parameters of the linear equation $RSS(RSSI) = a \times RSSI + b$ for well-known vendors. Adapted from [17].

WNIC manufacturers	(a, b)
Atheros	(0 dBm, -95 dBm)
Symbol	(1.67 dBm, -103 dBm)
Cisco	(1.09 dBm, -113 dBm)
Intel	(-1 dBm, 0 dBm)

of Buchman et. al [17] indicates that the linear equation $RSS(RSSI) = a \times RSSI + b$ can be used to approximate the relationship between RSSI and RSS, in general manner. The work also shows the parameters a and b as found for the most common vendors, here reported in Table 3.1.

To summarize, most of the manufacturers use the RSSI indicator to performing wireless network-related operations. RSSI could also be exploited to achieve positioning but, since RSSI is not related to RSS in a standardized manner, different measurement devices may provide different readings, at same physical condition. Clearly, the inconsistency of RSSI measurements due to different manufacturer policies translates into error positioning. The extent of the error has not been assessed in this work as we tackle the problem by simply assuming that different measurement devices are consistent to each other. The issue of device diversity has been addressed in literature, for example Christos Laoudias et al. in [47] present a self-calibrated method and demonstrate its effectiveness.

4. STATISTICAL CHANNEL MODELS FOR POSITIONING PURPOSE

This chapter contains a description of the statistical channel models that are used for positioning purposes. This chapter includes methodology as well as experimental analysis and results. To ensure comprehension, reader may go with Chapter 6 as recalled time by time along the reading.

4.1 Path-loss models

According to the electromagnetic theory [10], [76], radio signals attenuate with distance along the propagation path to some extent depending on the environment separating transmitter and receiver. In the Free Space (FS), path-loss between two isotropic antennas is modelled by the equation [57]

$$\mathrm{FSPL_{dB}} = \mathrm{P}_t - \mathrm{P_r} = 20\log_{10} d + 20\log_{10} f_\mathrm{c} + 20\log_{10}\left(\frac{4\pi}{c}\right) \tag{4.1}$$

where P_t and $\mathrm{P_r}$ represent the transmitted power and the received power respectively, expressed in logarithmic unit, d represents the distance expressed in m between the transmitter and the receiver, f_c represents the carrier frequency expressed in Hz and c represents the speed of light expressed in ms^{-1}.

Such model is defined to a free space but in practice it may also be applied to the atmosphere, provided that frequency not be exceeding 10 GHz, that is the lower part of the super-high frequency (SHF) band. Also, it is common practice to consider the line-of-sight (LOS) for this model, as constrained within the Fresnel zone [21].

Beside, to deal with non-line-of-sight and for general applicability, the so called simplified path-loss model makes use of a parameter n to tune the propagation loss to the specific surrounding environment. In its simplest form, the path-loss model can be expressed as [21]

$$\mathrm{PL_{dB}} = \mathrm{P_r}(\mathrm{d_0}) - \mathrm{P_r} = 10\, n \log_{10}\left(\frac{d}{\mathrm{d_0}}\right) \tag{4.2}$$

where n represents the path-loss coefficient parameter and $\mathrm{d_0}$ represents the reference distance.

The environment-related parameter n indicates to what extent the path-loss increases with the distance d. For instance, n in dense urban area will be greater than n in suburban area which in turn will be greater than n in rural area and so on up to the free space to which n is equal to 2 by definition. In other words, n is proportional to the environment complexity thus the smallest value is met in free space as no obstacle is present along the propagation way. Although here n is constant, there exists also a multi-slope path-loss model in which n is a piecewise constant parameter [69]. Nevertheless, a multi-slope path-loss is not treated in this work.

The reference distance d_0 is a relative small distance from the source to which a reference received signal strength is found based on either formula (4.1) or measurement data. It is the distance after which the model can be applied. A common choice for the reference distance in indoor is 1 m however in outdoor, where signal coverage is typically larger, it may be greater than that. For instance, Erceg et al. in [19] use a value of 100 m for d_0 in suburban environments. To sum up, with the simplified path-loss model, the signal power linearly decreases with the logarithmic distance to some extent depending on the number of obstructions in the propagation channel.

Three mechanisms namely reflection, diffraction and scattering [77], [70], [57], which are caused by the presence of objects sparse within the propagation area, produce random variations of the signal level. This phenomenon, also known as shadowing, is to be taken into account for a more realistic modeling of the propagation loss. Thus, by adding the shadowing term to the simplified path-loss model of formula (4.2), one obtains the well-known log-distance model

$$\mathrm{P_r} = \mathrm{P_r}(\mathrm{d_0}) - 10\, n \log_{10}\left(\frac{d}{\mathrm{d_0}}\right) + w \tag{4.3}$$

where the newly added term $w \sim \mathrm{N}(0, \sigma^2)$ is a random variable modelling shadowing, or small fading, as a Gaussian distribution of zero mean and variance σ^2 expressed in decibel unit.

It is common practice to set, in formula (4.3), $\mathrm{d_0}$ to 1 m and, by replacing $\mathrm{P_r}(1)$ with A, to rewrite the expression in a more elegant form as,

$$\mathrm{P_r} = A - 10\, n \log_{10} d + w \tag{4.4}$$

where A is expressed in dBm and reads apparent transmitted power to contrast with the actual radiated power. For sake of simplicity, however, A will be referred to as transmitted power hereafter.

Now, by introducing the free space frequency-dependent term into the log-distance model, it holds

$$\mathrm{P_r} = A - 10\,n\log_{10} d - 10\,n\,\log_{10} f_c - 10\,n\,\log_{10}\left(\frac{4\pi}{c}\right) + w \tag{4.5}$$

where f_c and c carry the same meaning as in formula (4.1).
To summarize, by means of path-loss model one could predict the strength of a radio signal at the particular location in which it is received. The accuracy of the estimate is influenced by the choice of the prediction model, therefore it should suit best to the specific study case. In this work, we concentrate on Log-distance model which is the most widely used model, both in indoor and in outdoor environment. Because in our study case the locations of the transmitters are unknown, there is a need to estimate them (the topic of transmitter localization estimation is wholly treated in Section 6.2). The transmitter-relative path-loss parameters, being the transmit power and the path-loss coefficient, as well as the shadowing variance, have to be estimated in order to be able to apply formulas (4.4) and (4.5).

4.2 Parameters estimation: path-loss and shadowing

The propagation model of (4.4) and (4.5) can be represented in compact form as

$$\mathrm{R}^{m}_{txid} = [\mathrm{H}^{m}_{txid}\mathrm{X}^{m}_{txid} + \mathrm{W}^{m}_{txid}] \tag{4.6}$$

where R^{m}_{txid}, H^{m}_{txid}, X^{m}_{txid} and W^{m}_{txid}, are matrices of real elements. The subscript $txid$ represents the transmitter identity, while the superscript m represents the PL model-dependent flag.
In more details, R^{m}_{txid} is the vector of RSS estimates obtained in dBm for the N_{txid} location points of the $txid^{th}$ observed TXID location grid point. It shows as

$$\mathrm{R}^{m}_{txid} = \left[\, \mathrm{R}^{m}_{txid,1}, \quad \mathrm{R}^{m}_{txid,2}, \quad \ldots \ \mathrm{R}^{m}_{txid,\mathrm{N}_{txid}} \,\right]^{\mathrm{T}} \tag{4.7}$$

where flag m sets the estimation model to be applied, that is model (4.4) or (4.5).

Then, matrix H^{m}_{txid} shows as

$$\mathrm{H}^{m}_{txid} = \begin{bmatrix} 1 & -10\log_{10}\mathrm{d}_{txid,1} - \mathrm{g}(f_c)^{m}_{txid,1} \\ 1 & -10\log_{10}\mathrm{d}_{txid,2} - \mathrm{g}(f_c)^{m}_{txid,2} \\ & \ldots \\ 1 & -10\log_{10}\mathrm{d}_{txid,\mathrm{N}_{txid}} - \mathrm{g}(f_c)^{m}_{txid,\mathrm{N}_{txid}} \end{bmatrix}. \tag{4.8}$$

Here, the distance $d_{txid,l}$ is calculated as

$$d_{txid,l} = ((x_{l-}x_{txid})^2 + (y_{l-}y_{txid})^2)^{\frac{1}{2}}, \quad l = 1,2, \quad \dots, N_{txid} \tag{4.9}$$

whereas, the function $g(f_c)^m_{txid,l}$ is found as,

$$g(f_c)^m_{txid,l} \begin{cases} 10\ \log_{10} f_c + 10\ \log_{10}\left(\frac{4\pi}{c}\right), & \text{if} \quad m = 1 \\ 0\ , & otherwise \end{cases}. \tag{4.10}$$

The parameter $d_{txid,l}$ of equation (4.8) is the distance between the $txid^{th}$ estimated transmitter location of coordinates x_{txid}, y_{txid} and the 1^{st}, 2^{nd}, ... N^{th}_{txid} location point of the $txid^{th}$ location grid point. It worth noticing that as long as the position of the transmitter is statistically determined, the distance $d_{txid,l}$ cannot be exact. Since in our study case transmitter location is unknown, it has to be estimated. Therefore, distance $d_{txid,l}$ calculated as per equation (4.9) will be affected by error to an extent depending upon the accuracy of the used estimation method. To conclude, transmitter localization as presented in Section 6.2 is an essential process, to be executed beforehand with respect to PL parameters estimation.

To continue describing elements of the expression (4.6), the vector of PL parameters X^m_{txid} is

$$X^m_{txid} = [\ A^m_{txid}\ ,\ \ n^m_{txid}\ ]^T \tag{4.11}$$

while the shadowing vector W^m_{txid} corresponds to

$$W^m_{txid} = \left[\ N(0, \sigma^m_{txid,1}),\quad N(0, \sigma^m_{txid,2}),\quad \dots\ N(0, \sigma^m_{txid,N_{txid}})\ \right]^T. \tag{4.12}$$

Both, X^m_{txid} and W^m_{txid} are to be estimated based on measurement data. For instance, the former vector of parameters can be found via the method of least squares (LS) [19],[69],[50]. That is,

$$X^m_{txid} = \left(H^m_{txid}{}^T H^m_{txid}\right)^{-1} H^m_{txid} R^m_{txid}\ . \tag{4.13}$$

With knowledge of X^m_{txid} one may compute the vector $H^m_{txid}X^m_{txid}$ and suppose that it is an re-created image of the received signal strength, for which shadowing is not taken into

account. Furthermore, one may look at the difference between the measured RSS vector P_{txid} and the re-created RSS vector I_{txid}^{m} as an estimate of shadowing [50], [78]. Consequently, the estimated shadowing standard deviation employed in (4.12) could be obtained as

$$\sigma_{txid}^{m} = \left(\frac{1}{\mathrm{N}_{txid}^{\mathrm{mp}}} \sum_{txid=1}^{\mathrm{N}_{txid}} (\mathrm{P}_{txid} - \mathrm{I}_{txid}^{m}) \right)^{\frac{1}{2}} \tag{4.14}$$

where $\mathrm{N}_{txid}^{\mathrm{mp}}$ represents the number of measurement points of the $txid^{th}$ TXID set, $\mathrm{P}_{txid} = \left[\mathrm{RSS}_{txid,1}, \mathrm{RSS}_{txid,2}, \ldots \mathrm{RSS}_{txid,\mathrm{N}_{txid}}\right]^{\mathrm{T}}$, and $\mathrm{I}_{txid}^{m} = \mathrm{H}_{txid}^{m}\mathrm{X}_{txid}^{m}$.

Finally, once both path-loss parameters and shadowing standard deviation are estimated, model (4.6) can be applied and received signal strength modelled accordingly. By applying the model, TXID-related parameter estimates are processed. The parameter estimates are determined in the off-line stage, where they are stored in vectors of length N_{txid} for later use.

In this stage if, along the vector, the $txid^{th}$ path-loss coefficient estimate is found equal or smaller than zero then the $txid^{th}$ path-loss coefficient estimate and $txid^{th}$ transmit power estimate are both replaced with not-a-number values. The not-a-number term refers in general to any numerical result which cannot be defined by a number [39]. Here, NaN is used to indicate such estimates that are unreliable due to unrealistic value of the estimate output.

When the $txid^{th}$ path-loss coefficient estimate is assigned NaN, the $txid^{th}$ transmit power estimate is assigned NaN too, for consistency. The other way around does not hold as no filtering is applied to the transmit power estimate, at this stage. It is worth to mention that, in the on-line stage, the $txid^{th}$ TXID will not be valid if the $txid^{th}$ either path-loss coefficient estimate or transmit power estimate corresponds to NaN.

In our measurement data we observed a relatively high number of unreliable estimates. There are 399 unreliable estimates over 1285 LTE TXID and 353 over 3161 WiFi TXID, that is 31.05 % and 11.17 % respectively. Consequently, the number of estimates after filtering unreliable estimates out is 886 for LTE and 2808 for WiFi. That is 68.95 % and 88.83 % of estimates, respectively, is reliable. The scatter plot and the histogram of the whole set of reliable estimates is shown in Figure 4.1 and Figure 4.2 respectively, the average value of estimates is reported in Table 4.1 for different simulation scenarios.

From Figure 4.1 it can be seen how the estimates vary with respect to the transmitter identity and between two different communication radio systems for the same radio propagation environment. The number of estimates in each observation depends on which radio system is tuned in the estimation model and it matches that one of transmitter

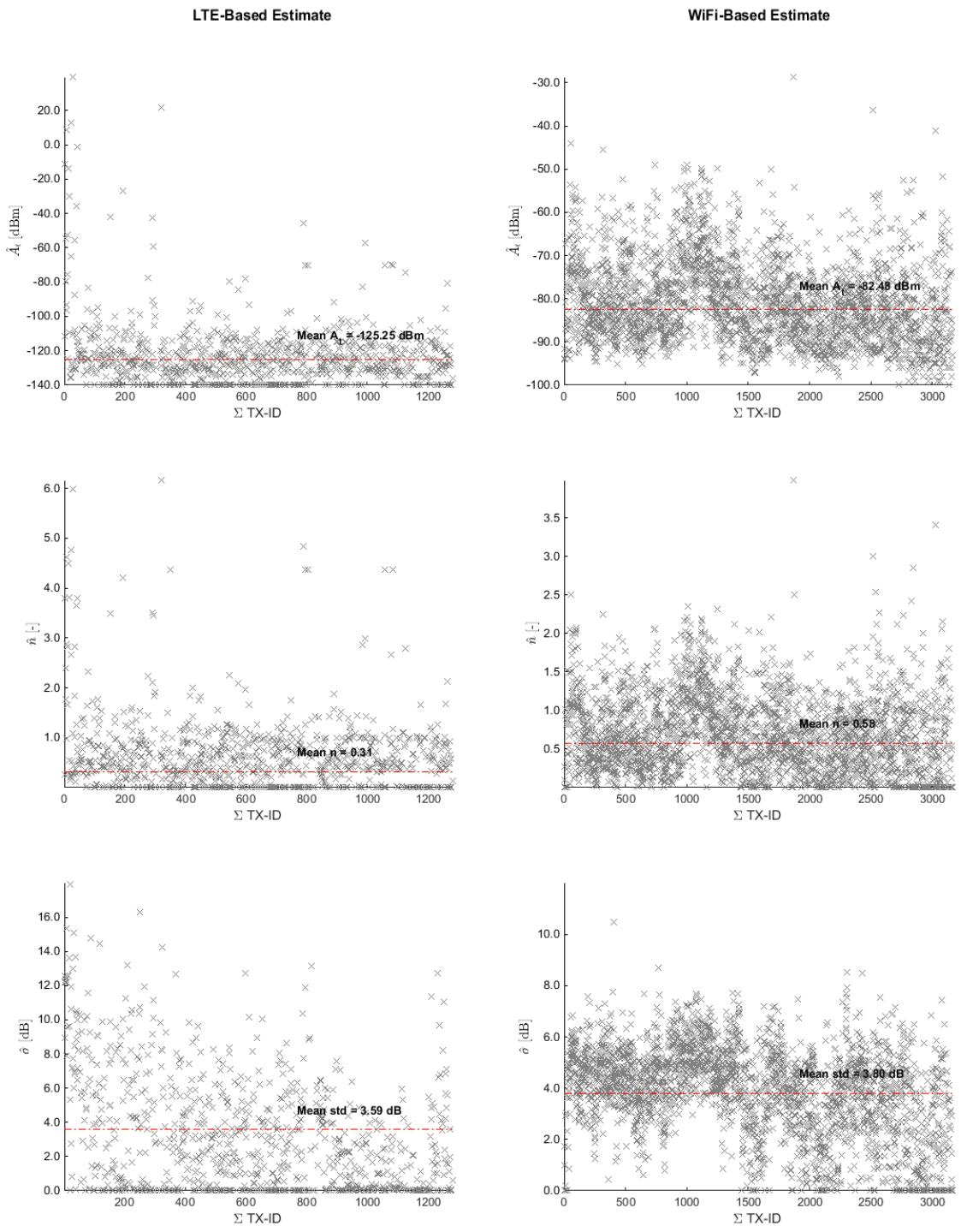

Figure 4.1. The entire set of estimated path-loss parameters and shadowing standard deviation along the transmitter identities sequence based on our measurement data. Apparent transmit power: LTE (top-left) VS WiFi (top-right). Path-loss coefficient: LTE (middle-left) VS WiFi (middle-right). Shadowing standard deviation LTE (bottom-left) VS WiFi (bottom-right).

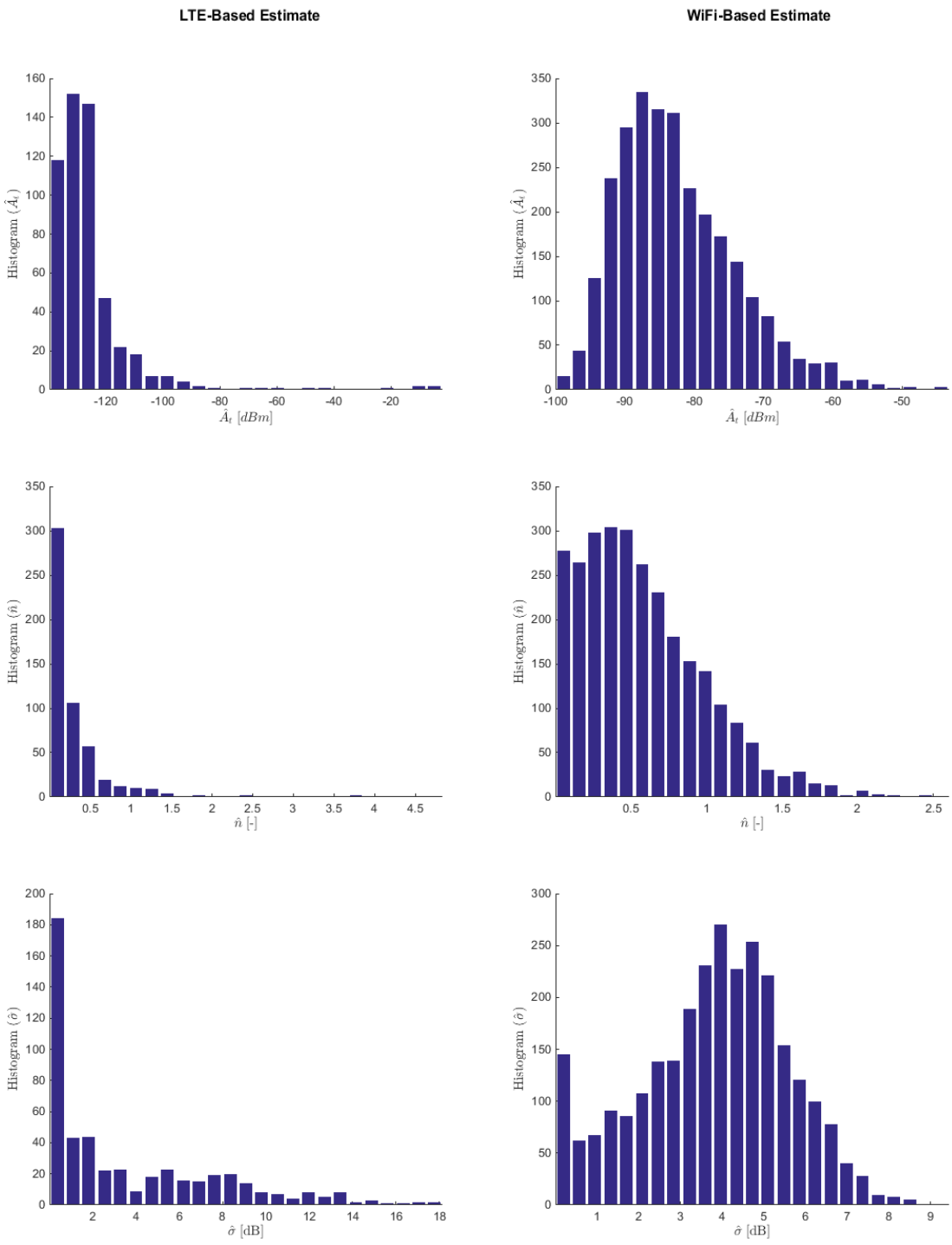

Figure 4.2. Histogram of the estimated path-loss parameters and shadowing standard deviation. Apparent transmit power: LTE (top-left) VS WiFi (top-right). Path-loss coefficient: LTE (middle-left) VS WiFi (middle-right). Shadowing standard deviation LTE (bottom-left) VS WiFi (bottom-right).

Table 4.1. Average value of the estimated parameters: path-loss and shadowing. Estimates are obtained by different combinations of following simulation setting. Frequency (m=0) not counted, (m=1) counted. Signals at lowest recorded power level (h=0) not removed, (h=1) removed.

Data type	m	h	$\hat{A}$ [dBm]	$\hat{n}$ [-]	$\hat{\sigma}$ [dB]
LTE	0	0	-125.25	0.31	3.59
	0	1	-127.17	0.42	4.90
	1	0	-118.82	0.31	3.34
	1	1	-148.82	0.40	4.71
WiFi	0	0	-82.48	0.58	3.80

identities. The number of transmitter identities in the estimation model effects the size of the database as discussed in Chapter 6, Section 6.4. In this sense, the path-loss model based on LTE is worthy because it needs a smaller number of transmitter identities and thus it offers a lower complexity in terms of computation compared to the path-loss model based on WiFi. The potential benefit of processing a small number of transmitter identities is huge considered that only a few antennas are in fact located inside the measurement area. The issue of determining automatically whether a transmitter is located inside the measurement area, is not addressed in this thesis. Thus, all the transmitter identities are counted toward the parameter estimation in our model, with no distinction between transmitters located inside and transmitters located outside the measurement area. The Figure 4.2 shows the histogram of the estimated values for each parameter and for each radio system in interest. The difference in the distribution shapes between the two communication systems is apparent and suggests poor correlation among propagation channels in the LTE radio system.

The Table 4.1 shows both LTE-based and WiFi-based estimates. To the former, estimates are performed by setting either m=0 or m=1 while to the latter, setting can only be $m\neq0$ as carrier frequency is unknown in such case. Furthermore, to the LTE radio system, estimates are also performed after removing signals with the lowest recorded power level, which likely occurs as reported in Section 6.1.3. According to Table 4.1, the choice of the data type, i.e., LTE-based or WiFi-based measurements, produces a notable change in the estimated average apparent transmit power but it has a minor effect on the estimated average path-loss coefficient and on the estimated average shadowing standard deviation. Using frequency information into the estimation model has a low impact to all the three average parameter estimates. By signal removal, average shadowing standard deviation rises about 1.30 dB while average path-loss coefficient increases about 10 units, regardless which propagation model is used. Conversely, average transmit power decreases to a larger extent if carrier frequency is considered by the model; in which case, the decrease is about

23 dBm. All reported changes are slight; hence, the overall impact due to setting state is not particularly significant.

RSS fluctuation in time

A multiple scan is a set of samples taken at the same location over a certain time interval. A multiple scan allows to investigate on the temporal fluctuation of the signal strength at a fixed location. The success of the investigation relies on sampling features such as the number and the temporal separation of consecutive samples. For example, the time scale of the recreated signal fluctuation is affected by the temporal separation of consecutive samples while the number of samples reflects the details in the distribution shape.
The Figure 4.3 shows the fluctuation in time and the probability histogram of the signal strength based on our measurement data, for two different location points against the same 4G transmitter identity. The Figure 4.3 consists of 4 plots, 2 plots for each examined

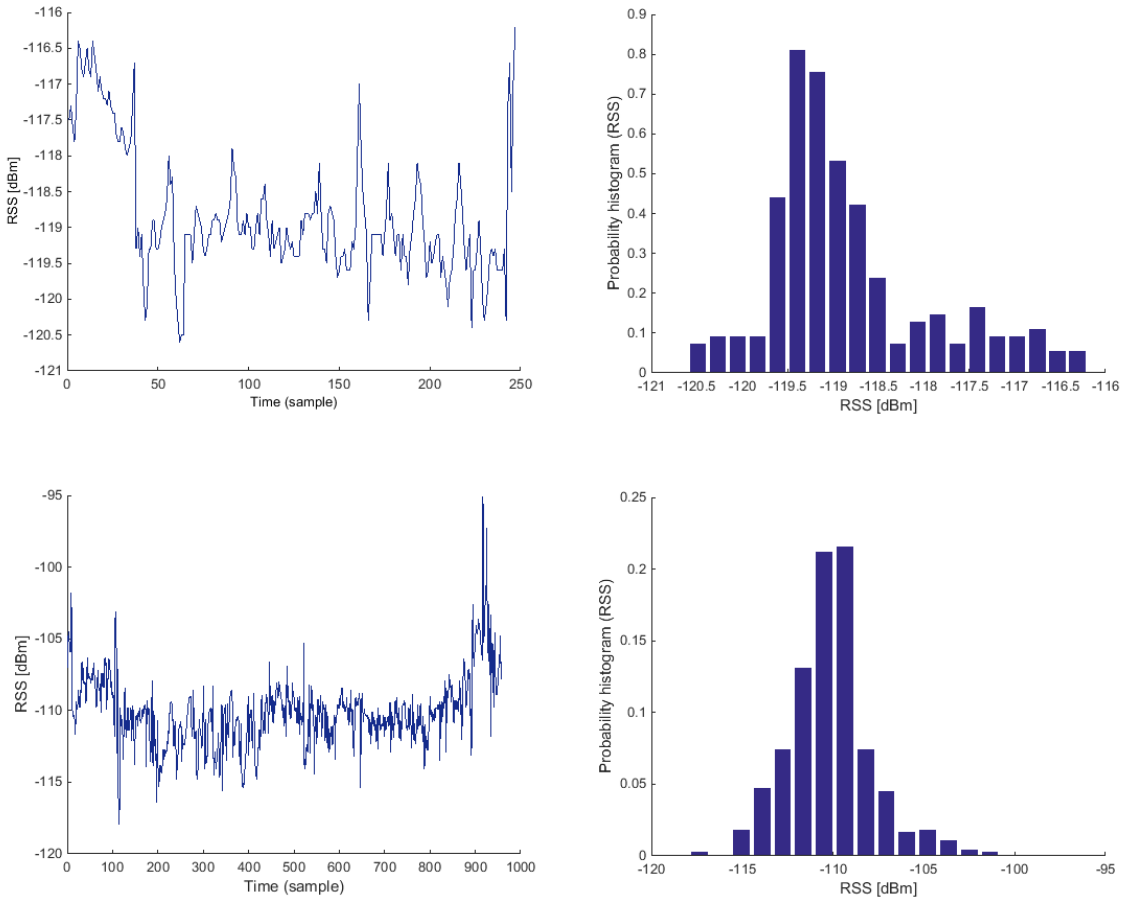

Figure 4.3. LTE RSS fluctuation (left) and RSS probability histogram (right), observed at two different locations. Location test point (a) (up), (b) (down).

Table 4.2. RSS fluctuation over time: statistics summary.

Test point	System/ TXID	Samples	Δt [s]	Mean [dBm]	Std [dB]	Min [dBm]	Max [dBm]	f_c [MHz]
(a)	LTE/105	247	123.24	118.82	0.89	-120.60	-116.20	3350
(b)	LTE/105	958	737.82	-110.12-	2.38	-117.90	-95.10	3350

location point. The location point relative to the top plots is referred to as test point (a) whereas the location point relative to the bottom plots is referred to as test point (b), for example in Table 4.2. Test point (a) includes 247 samples captured by a time period of 2 minutes. Here RSS exhibits a multimodal distribution, ranging from -120.60 dBm to -116.20 dBm and having mean at -118.82 dBm. As for test point (b), the signal fluctuates along 958 samples for 12 minutes. In this case, a near Gaussian RSS distribution is distributed around the mean value -110.12 dBm and ranges 22.8 dBm from minimum to maximum. Carrier frequency is 3350 MHz in both cases as it is reported in the summarizing Table 4.2.

Our empirical investigation into the signal fluctuations leads to the conclusion that shadowing in LTE signals does not follow a log-normal distribution, necessary. Thus, the log-distance path-loss model we adopt in formula (4.6) to predict the received signal strength might not represent the optimal solution for the estimation problem. The log-distance path-loss model is however a low-complexity, yet effective model that suits well for an early stage research; then the possibility of moving the research in the direction of a multi-level path-loss model is not further considered in this thesis.

4.3 Power map

An example of power map coming from measurements is shown in Figure 4.4. This is a two dimensional spatial distribution of the measured signal strength associated with one particular 4G antenna. In this example, the signal strength ranges from -120.9 dBm in dark blue to -42.9 dBm in dark red and exhibits mean in between at -89.2 dBm. The coverage is somewhat directional with maximum distance in the order of several hundred meters. A location estimate of the transmitting base station is shown with a thick black cross marker. It was obtained via the 4-strongest RSS-weighted centroid algorithm as described in Chapter 6.

The power map of Figure 4.4 was created with MATLAB by the following steps:

1. Built a grid of equally spaced synthetic location points set $[\mathrm{x}_i{}^{\mathrm{map}}, \mathrm{y}_i{}^{\mathrm{map}}]$, $i = 1, 2, \ldots, \mathrm{N}_{\mathrm{grid}}$ that ranges in the x and y directions respectively from the minimum *x* coordinate to the maximum *x* coordinate and from the minimum *y*

coordinate to the maximum y, of the measurement location points set $[x_j^{meas}, y_j^{meas}]$, $j = 1, 2, \dots, N_{meas}$. Only measurements relative to the considered antenna-identity are counted.

2. Interpolate the signal strength measurements RSS_j, scattered across the measurement location points $[x_j^{meas}, y_j^{meas}]$, $j = 1, 2, \dots, N_{meas}$. This allows, in the next step 3, to find signal strength values in all the synthetic grid points.
3. Get the interpolated signal strength values at the synthetic grid points and draw a filled contour plot by setting the desired contour properties. The space between contour lines is filled with color for smoothness. The color is associated with signal strength which range from the minimum value to the maximum values of RSS_j.

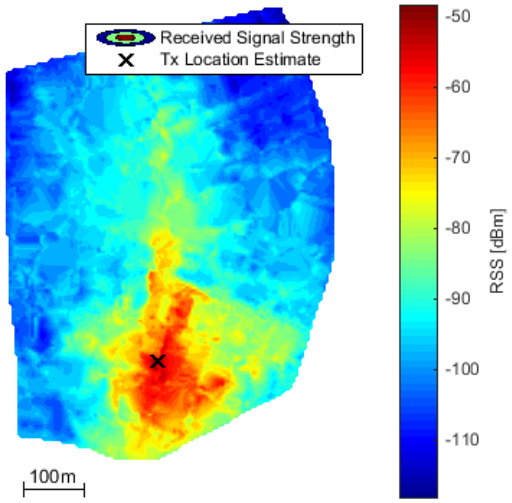

Figure 4.4. An example of power map created by collection of signal strength measurements coming from one particular LTE base station.

5. RSS-BASED POSITION ESTIMATION WITH STATISTICAL INFORMATION

Classical fingerprinting techniques are able to provide fairly accurate positioning, but they require an extremely large amount of data storing, transferring and processing. Due to their inefficiency, in terms of memory space, estimation time, battery consumption and communication complexity, such methods might not be appropriate for low-cost devices. In contrast, statistical models offer a reduced size of the database by using only a few parameters per TXID. In this chapter, two positioning methods based on statistical models, coverage-area and path-loss model respectively, are presented.

5.1 Coverage area based positioning

Coverage area refers to the area within which a user device can receive radio signals from one particular transmitter, with respect to one particular transmitter identity code. For each TXID there exists a set of location points distributed around a source, mostly within an area enclosed by an ellipse (see instances of Figure 5.1). The spatial distribution of such location points can be modeled, typically, as a bivariate normal distribution [46] by a few parameters. Overall five parameters, two for the mean vector and three for the covariance matrix, are sufficient to model the distribution for one single coverage area. That is the probability an observation falls within the considering coverage area. Alternative models are possible, for example bivariate Student-t which is well suited when outliers exist [62],[64].

In our study, coverage area is modeled as a bivariate Normal distribution for the sake of simplicity. Our algorithm implementation, including the coverage area estimate phase and the statistic positioning estimate phase, relies on a prior framework [65], [46] and is outlined in Algorithm 5.1.

In phase A of Algorithm 5.1, as each location point $Z_{i,j}$ is a 2×1 vector, the position mean at step 3 is a 2×1 vector too. The prior covariance at step 4 is a 2×2 diagonal matrix as the constant σ, that changes with input 2 and input 3, is multiplied by the identity matrix I of size 2×2. The position covariance defined as in step 5 turns out a 2×2 matrix.

Based on prior knowledge of the radio maps (see for instance Figure 4.4), σ is found by the coverage distance that exhibits, on average, significant signal level. According to data, for

Algorithm 5.1. Positioning based on coverage area.

A. Coverage area estimate phase

Input:

1 Set of training data grids organized by TXID

2 Data Type: {LTE} or {WiFi}

3 Signal Type: {all signals} or {strong signals}

Offline processing steps:

1 For each i-th TXID, collect signal heard location points Z_i

2 For each i-th TXID, obtain number n_i of collected location points

3 For each i-th TXID, compute position mean by n_i location points:

$$\mu_i = \frac{\sum_{j=1}^{n_i} Z_{i,j}}{n_i}$$

4 Define the prior covarianace B, according to inputs 2,3:

$$B = \sigma^2 I$$

5 For each i-th TXID, compute position covariance Σ_i as:

$$\sum_i = \frac{\left(\sum_{j=1}^{n_i} Z_{i,j} Z_{i,j}^T\right) + B + n_i \mu_i \mu_i^T}{n_i + 1}$$

B. Statistical positioning estimate phase

Input:

1 Estimation data

2 Position mean set calculated at step 3, phase A

3 Position variance set calculated at step 5, phase A

Online processing steps:

1 Obtain TXIDs that are heard by the user

2 Compute the user location estimate covariance

$$\hat{\sum} = \left(\sum_{k \in \text{heard txid}} \sum_k^{-1}\right)^{-1}$$

3 Compute the user location estimate mean

$$\hat{\mu} = \hat{\sum} \sum_{k \in \text{heard txid}} \sum_k^{-1} \mu_i$$

strong signal input 3, we found appropriate to set σ to 200 m and power threshold to -110 dBm, and to set σ to 50 m and power threshold to -90 dBm, for LTE and WiFi respectively. Likewise, for all signal input 3, a good choice for σ was found as 2000 m regardless of the input 2 because the σ value is fairly large in this case. Depending on the input 3, either all signals or strong signals are taken into account when collecting of location points at step 1.

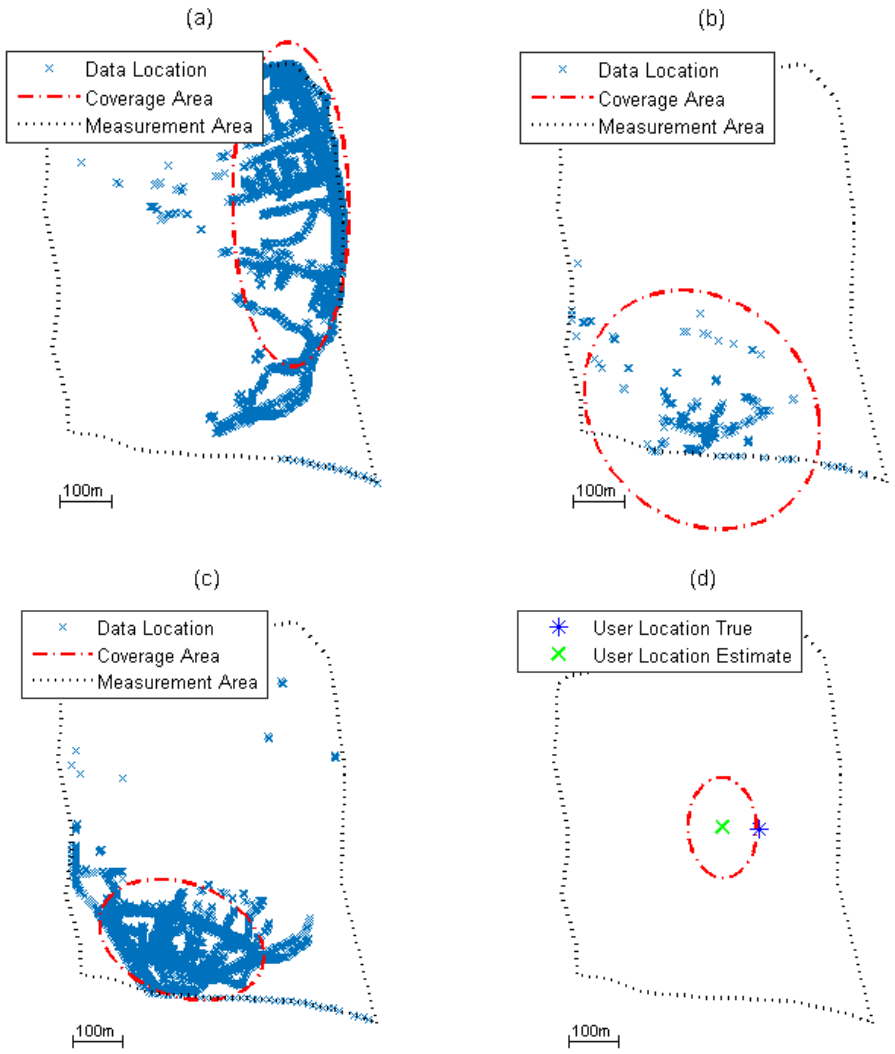

Figure 5.1. An example of plots based on Algorithm 5.1. Coverage area estimates (a), (b), (c), relative to user heard transmitter identities. Positioning estimation (d).

This means that if input 3 corresponds to strong signal then the location points for which signal reception does not exceed power threshold are discarded. Otherwise, if input 3 corresponds to all signal, the location points are fully exploited and a power threshold need not to be addressed.

This model, to the propriety of building coverage area in two possible modes, is referred to as bi-level model. Multi-level models are also possible [64], but they were not considered in this work.

In phase B, the estimate data input 1 is obtained as described is Section 6.1.4. The user location estimate covariance and the user location estimate mean, at step 2 and at step 3, is

a 2×2 matrix and a 2×1 vector respectively. They identify an ellipse whose center exhibits the higher chance of finding the user location. Hence the center of ellipse addresses the user location estimate, see Figure 5.1.
In this section we presented a statistical model based on a bi-level coverage-area method. In the off-line phase, for each model level, and for each detectable transmitter identity, the algorithm estimates 5 parameters in order to model the spatial dissemination of location points in terms of probability. In the on-line phase, as many coverage areas as transmitter identities are heard by the user are combined such that their join probability turns out a two-dimensional normal distribution with mean that addresses the positioning goal. A bi-level model offers enhanced solution at the expense of greater complexity and double database size.

5.2 Path-loss based positioning

The path-loss model based positioning algorithm, developed for the purpose of the thesis work, is described in Algorithm 5.2. Algorithm 5.2 is split into two separate parts for convenience. The first part is the off-line Stage A and the second part is the on-line Stage B.
One such path-loss based algorithm performs positioning with 5 parameters per transmitter: the transmit power, the path-loss coefficient, the horizontal coordinates of the transmitter and the shadowing variance. The off-line stage firstly estimates the locations of the transmitters via RSS-weighted centroid and secondly estimates the path-loss parameters via least squares. Thirdly, after reconstructing the received signal strength via the path-loss model, it also estimates the shadowing parameters. All the estimated parameters are to be used in the on-line stage. Here, the algorithm first of all achieves the RSS-weighted centroid of heard transmitter identities. This is a coarse estimation of the user point around which is built a likelihood grid for a fine re-estimation. Each location point of one such grid, has a certain Gaussian likelihood as a function of the difference between the measured RSS and the RSS reconstructed via path-loss model. If the number of heard TXIDs is greater than one, then the final likelihood is the product of the likelihoods for each heard TXID. Notice that the product in linear scale is equivalent to the sum in the logarithmic scale. Finally, the target position estimate is the likelihood-weighted centroid of the likelihood grid.
Further details are provided in the body of Algorithm 5.2. Before approaching Algorithm 5.2 in deep, though it is self-explanatory, the reader is recommended to go through Chapter 4 and Chapter 6, in Section 6.2, to achieve full understanding of the algorithm.

Algorithm 5.2. Positioning based on path-loss model. ***Stage A.*** Path-loss model estimate.

Input:

1 Set of training-data grids organized by TXID

2 TXIDs set size, N^{txid}

3 Measurement points (mp) per TXID, N_i^{mp}

4 PL Model Type: {freq. not counted} or {freq. counted}

Offline processing steps:

1 **For** i=1 to N^{txid}

1.1 Estimate TXID location $\hat{z}_i^{txid}$ via RSS-W-centroid

$$\hat{z}_i^{txid} = \frac{\sum_{l=1}^{N_i^{mp}} w_{i,l}\bar{z}_{i,l}^{mp}}{\sum_{l=1}^{N_i^{mp}} w_{i,l}}, \quad \bar{z}_{i,l}^{mp} \in \mathbb{R}^{2\times 1} = \left[x_{i,l}^{mp}, y_{i,l}^{mp}\right]^T, \quad \hat{z}_i^{txid} \in \mathbb{R}^{2\times 1} = \left[\hat{x}_i^{txid}, \hat{y}_i^{txid}\right]^T$$

1.2 Build a vector of distances

$$\bar{d}_i = \left[\left(\sum_{\text{by columns}}\left(\left(\hat{z}_i^{txid} * \bar{1}_{1\times N_i^{mp}}\right) - \bar{z}_i^{mp}\right)^2\right)^{1/2}\right]^T, \quad \bar{z}_i^{mp} \in \mathbb{R}^{2\times N_i^{mp}} = \left[\bar{x}_i^{mp}, \bar{y}_i^{mp}\right]^T$$

$$\bar{x}_i^{mp}, \bar{y}_i^{mp} \in \mathbb{R}^{1\times N_i^{mp}}$$

$$\bar{d}_i \in \mathbb{R}^{N_i^{mp}\times 1}$$

1.3 Build a vector of frequencies

$$\bar{f}_i \in \mathbb{R}^{N_i^{mp}\times 1}$$

1.4 Build a vector of RSSs

$$\bar{P}_i \in \mathbb{R}^{N_i^{mp}\times 1}$$

1.5 Estimate path-loss parameters $\hat{P}_i^{tx}$, $\hat{n}_i$ via LS

if input 4 is {freq. not counted}

$$\bar{H}_i = \left[1_{mp\times 1}, \ -10\log_{10}\bar{d}_i\right], \quad \bar{H}_i \in \mathbb{R}^{N_i^{mp}\times 2}$$

else

$$\bar{H}_i = \left[1_{mp\times 1}, \ -10\log_{10}\bar{d}_i - 10\log_{10}\bar{f}_i - C\right], \quad C = 10\log_{10}(4\pi/c) = -73.78$$

$$\bar{L}_i = \left(\bar{H}_i{}^T\bar{H}_i\right)^{-1}\bar{H}_i{}^T\bar{P}_i, \quad \bar{L}_i \in \mathbb{R}^{2\times 1} = \left[\hat{P}_i^{tx}, \ \hat{n}_i\right]^T$$

end if

1.6 Re-create RSS-image by given the set $\left\{\hat{P}_i^{tx}, \hat{n}_i, \bar{d}_i, \bar{f}_i\right\}$

if input 4 is {freq. not counted}

$$\hat{P}_i^{image} = \hat{P}_i^{tx} - 10\hat{n}_i\log_{10}\bar{d}_i, \quad \hat{P}_i^{image} \in \mathbb{R}^{N_i^{mp}\times 1}$$

else

$$\hat{P}_i^{image} = \hat{P}_i^{tx} - 10\hat{n}_i\log_{10}\bar{d}_i - 10\hat{n}_i\log_{10}\bar{f}_i - C$$

end if

1.7 Estimate shadowing

$$\hat{S}_i = \bar{P}_i - \hat{P}_i^{image}, \quad \hat{S}_i \in \mathbb{R}^{N_i^{mp}\times 1}$$

1.8 Estimate Shadowing mean $\hat{\mu}_i$ and std $\hat{\sigma}_i$

$$\hat{\mu}_i = \frac{1}{N_i^{mp}}\sum\hat{S}_i, \quad \hat{\sigma}_i = \left(\frac{1}{N_i^{mp}}\sum\left(\hat{S}_i - \hat{\mu}_i\right)^2\right)^{\frac{1}{2}}, \quad \hat{\mu}_i, \hat{\sigma}_i \in \mathbb{R}^1$$

end for

Algorithm 5.2. Positioning based on path-loss model. ***Stage B.*** Statistic positioning estimate.

Input:

1 Estimation data

2 Set of TXID location calculated in Stage A

3 Set of PL parameters calculated in Stage A

4 Set of shadowing parameters from Stage A

5 PL Model Type: {freq. not counted} or {freq. counted}

6 Number points for the synthetic grid, n^2

Online processing steps:

1 Obtain coarse user location estimate $u_{est,coarse}$ via RSS-W-centroid

2 Create a likelihood grid around the point $u_{est,coarse}$

2.1 Initialize $\bar{L}^{grid} = 0_{n\times n}$ (zero matrix)

2.2 **For** j=1 to $N^{heard} \subseteq N^{txid}$

2.2.1 Hold $TXID_j$

2.2.2 Build a location grid with $n \times n$ synthetic points of coord. $x_{h,w}^{grid}, y_{h,w}^{grid}$, step s and center $u_{est,coarse}$

2.2.3 Build a $n \times n$ matrix of distances as

$$\bar{D}_j = \sqrt{\left(x_{j,h,w}^{grid} - \hat{x}_{txid_j}^{txid}\right)^2 + \left(y_{j,h,w}^{grid} - \hat{y}_{txid_j}^{txid}\right)^2}, \quad \forall\, h,w = 1,2,\dots,n, \quad \bar{D}_j \in \mathbb{R}^{n\times n}$$

2.2.4 Re-create RSS-image based on PL model

if input 5 is {freq. not counted}

$$\hat{P}_j^{rx} = \hat{P}_{txid_j}^{tx} 1_{n\times n} - \hat{n}_{txid_j} 10\log_{10} \bar{D}_j, \quad \hat{P}_j^{rx} \in \mathbb{R}^{n\times n}$$

else

$$\hat{P}_j^{rx} = \hat{P}_{txid_j}^{tx} 1_{n\times n} - \hat{n}_{txid_j} 10\log_{10} \bar{D}_j - \hat{n}_{txid_j} 10\log_{10} \frac{4\pi f_{txid_j}}{c}$$

end if

2.2.5 Re-create shadowing-image as

$$\bar{S}_j = \hat{P}_{txid_j}^{tx} - P_{txid_j}^{rx,user}, \quad \bar{S}_j \in \mathbb{R}^{n\times n}, \quad P_{txid_j}^{rx,user} \in \mathbb{R}^1$$

2.2.6 Build a likelihood grid based on GL model as

$$\bar{L}_j^{grid,ln} = \ln\left(2\pi\, \hat{\sigma}_{txid_j}^2\right)^{-\frac{1}{2}} - \frac{\bar{S}_j{}^2}{2\hat{\sigma}_{txid_j}^2}$$

2.2.7 Update $L^{grid,ln} = L^{grid,ln} + L_j^{grid,ln}$

end for

2.3 $L^{grid} = \exp\left(\frac{L^{grid,ln}}{N_{heard}}\right)$

3 Output a fine user location estimate $u_{est,fine}$ via likelihood-W-centroid

6. MEASUREMENT-BASED ANALYSIS AND RESULTS

6.1 Measurement data description

6.1.1 Introduction

Herein described measurements were provided by Magister Solution Ltd in April, 2015. The Finnish company obtained data from a live cellular network in Tampere, on September 2014. The measurement campaign was conducted as part of the European Celtic-Plus SHARING (Self-organized Heterogeneous Advanced Radio Networks Generation) project, deliverable D6.3 [81].

The measurement campaign was carried out throughout the measurement area of Figure 6.1 moving either by car, bicycle or foot, and using the commercial device Samsung Galaxy S3 as RF measurement tool. Radio measurements consist of RSSI/BSSID and RSRP/CID measurements from WiFi signals and LTE signals respectively, hereinafter simply referred to as RSS/TXID regardless of the signal type. The RSS measurements are reported in dBm whereas the TXID measurement are anonymized, meaning that the original identity was mapped into an arbitrary code. For LTE, RSS and TXID are supplied with the frequency band, that is either 800 MHz, 1800 MHz or 2600 MHz. Conversely, the frequency information is not provided for WiFi.

Measurements also include location information, time stamp, trace reference identifier and other information. Location information was measured by the Global Positioning System [55] receiver that the mobile phone is equipped with. A GPS receiver offers fairly good localization in outdoor environment and distance error depending on the environment complexity [14]. The suburban characteristics of the measurement area allow to assume the GPS measurement error in a few meters: a small error if compared with our target accuracy that will be therefore neglected throughout the analysis.

Measurements were collected in 36648 different location points within an area of 0.4 sq. km, that is 0.0916 samples per square meter. The size of the area was estimated by generating a smooth boundary line enclosing all the measurement points and finding the area of the resulting shape (see the boundary line of Figure 5.1 for instance). For the sake of convenience, the computation was somewhat relaxed to get smoothness in the boundary line.

Sampling is not only highly dense but also fairly homogeneously spread as no significant coverage hole is observed along the outdoor space, see Figure 6.2. Moreover, sampling is redundant since the number of measurement points largely exceeds the number of unique

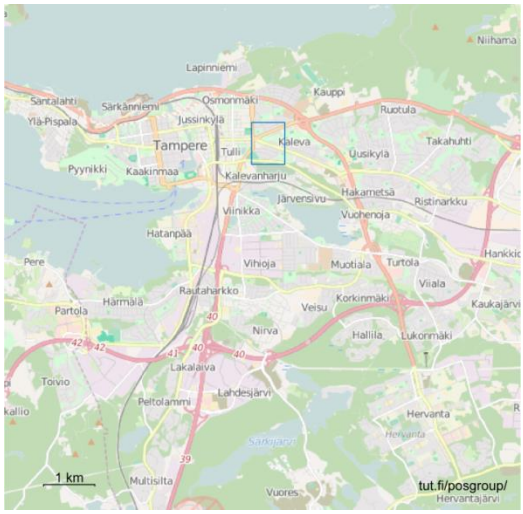

Figure 6.1. Measurement area (blue rectangle) on the Tampere map.

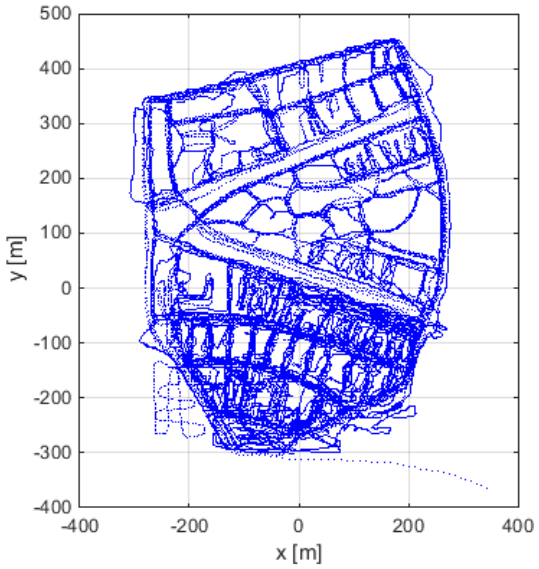

Figure 6.2. Data points (blue dots) in local coordinates.

location points. Precisely, the former is 2.7 times greater than the latter. Different samples having same location will be referred to as multiple scan, shortened MSC, hereafter.

Dataset

Dataset was stored as csv file and provided in such a format. The csv file consists of 99593 rows where each row presents a measurement sample enclosing a set of data separated by commas. The length of row depends on the number of heard TXIDs, that is how many

signals are detected from the device at the time and place of sampling. In detail, 3.52 LTE TXIDs per sample and 22.74 WiFi TXIDs per sample were heard on average. The same statistic applies to RSS as the two measurements, RSS and TXID, come in pairs. The entire dataset contains 2615289 RSS/TXID measurement pairs, that is 350576 LTE RSS/TXID and 2264713 WiFi RSS/TXID.
The overall number of RSS/TXID measurements reduces to 1119920 (165412 for LTE and 954508 for WiFi) as the multiple scans are removed, which corresponds to 2.33 reduction factor. Each sample is provided with GPS geographical coordinates expressed in Decimal Degree (DD). As GPS relies on the World Geodetic System 1984 (WGS84) system we may refer the reader to [15], [68].
Data other than radio information and location information are not too relevant to our analysis and therefore its description is omitted.

CSV File examination

In the provided csv file there are 26 two-consecutive commas, that is no occurrence in place of LTE RSS. The lack of data is perhaps due to an anomaly in the measurement process of which we do not have knowledge. In addition to the empty elements associated to LTE RSS measurements, we found other anomalous elements associated to either the GPS coordinates (34 elements) and the frequency band (2 elements). Such elements consist of strings forming the word *none*. To sum up, 36 words *none*, of which 34 to the latitude-longitude coordinates (17 samples) and 2 to the LTE bands (2 samples), were found. As for the latter set, a word *none* occurs as no LTE signal can be detected in that sample.

6.1.2 Data import and processing

Before importing data into MATLAB, the csv file have to be unwrapped and words *none* replaced with empty elements. Empty elements, once imported, will turn out NaN elements, automatically. In summary, 62 NaN elements, of which 36 coming from *none* elements and 26 coming from empty elements, were obtained by importing the dataset.
If a NaN element is due to the lack of either location information or frequency information, then the whole sample containing NaN is to be removed. Otherwise, if a NaN element is due to a failure in the detection of signal, the NaN element is discarded but other information in the sample holds.
Once that data is imported, some processing activity is required in order to achieve the following tasks:

1. Finding samples where GPS measurements do not occur and removing such samples from the dataset.

2. Converting Geodetic WGS84 Latitude-Longitude to local Cartesian Est-Nord-Up (ENU) coordinates. This involves two preliminary steps:
    a. zero-altitude-padding of location points,
    b. fixing a local system reference point. E.g. the spatial mean of the coordinates.
3. Splitting data into two different sets:
    a. LTE-data-based set
    b. WiFi-data-based set

    Only essential data are retained. The output is a pair of cell arrays with index row scanning samples and index column reporting (1) position, (2) radio fingerprinting, (3) band frequency, (4) trace, (5) trace start. Column number three in the WiFi data cell array is assigned with NaNs since band frequency is unknown for such data type.
4. Removing samples for which no radio transmitter is heard. This operation reduces the row-size but has negligible impact to our analysis as the number of unheard points (2 by LTE and 24 by WiFi) is very small compared to the number measurement points (99593).

At this point, each set of data may require additional processing depending on the task to be achieved. For example removing the lowest recorded signal level, removing the multiple scans which occur in the same geographical location, organizing data into TXID-grids and discarding unwanted grid set.

6.1.3 Empirical RSS statistic

The empirical statistic of the received signal power is shown in Figure 6.3 by the probability histograms of both LTE RSS measurements and WiFi RSS measurements. Each histogram counts all the measurements for the considering radio system, along the entire TXID set.

The shape of the WiFi RSS distribution appears right-skewed with very small skewness and mean around -90 dBm whereas, the LTE RSS distribution is not apparent. The latter exhibits a sharp peak at the value -140 dBm which occurs, as per our knowledge, when the measurement device detects a cell ID but is not able to measure the signal level. Because in our study case the 4G network is low loaded, the detection of poor signals is likely to happen.

Unlike WiFi scenario, most of the LTE signal sources are located outside the measurement area while the minority stands within it. This condition is believed to be the cause of such a shapeless distribution. In fact, the distribution considerably changes when considering the

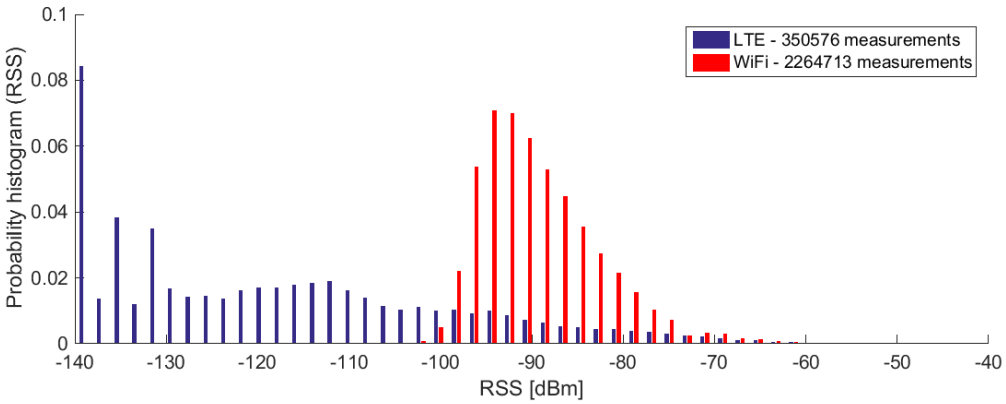

Figure 6.3. Probability histogram of RSS, either by LTE measurements (blue-bars) or by WiFi measurements (red-bars). Both LTE and WiFi counts all measurement points. Supporting statistic LTE, (WiFi): overall meas. 350576, (2264713); min/max -140.00/-42.90, (-106/-44); ave -117.88, (-89.13); # IDs 1285, (3159).

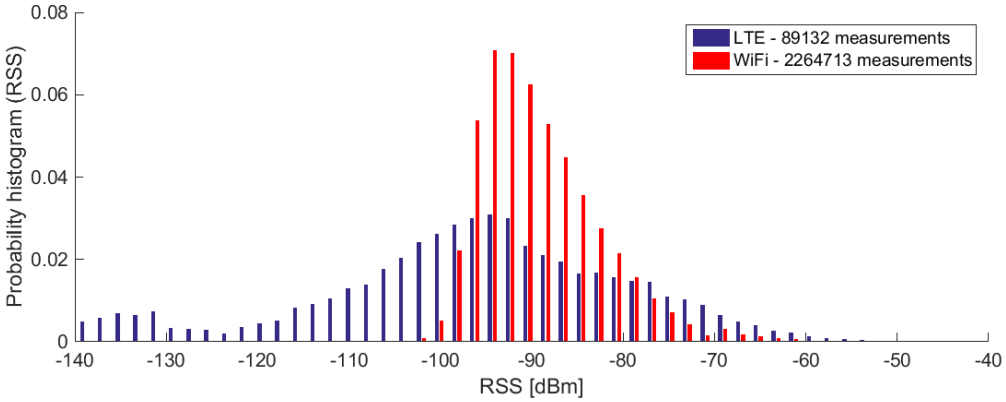

Figure 6.4. Probability histogram of RSS, either by LTE measurements (blue-bars) or by WiFi measurements (red-bars). LTE counts only measurements relative to top 7 TXIDs as for number of measurement points whereas WiFi counts all measurement points. Every measurement equals to -140 dBm was removed. Supporting statistic LTE, (WiFi): overall meas. 89132, (2264713); min/max -139.90/-42.90, (-106/-44); ave -96.20, (-89.13); # IDs 7, (3159).

histogram relative to only transmitters expected inside the measurement area, e.g. TXIDs with highest number of measurements. This way, it comes to a more disciplined shaping of the distribution, see Figure 6.4.

Besides, we might exclude the dynamic antenna power control as other possible cause, because RSRP value is not dynamically controlled according to our knowledge. Different

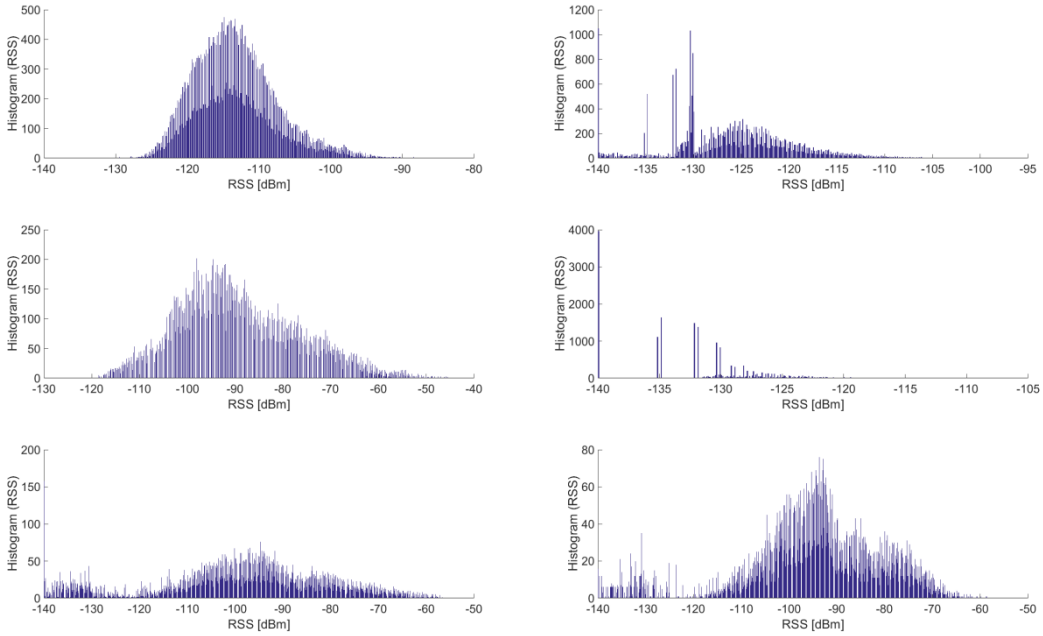

Figure 6.5. Six TXID-specific RSS-LTE histograms.

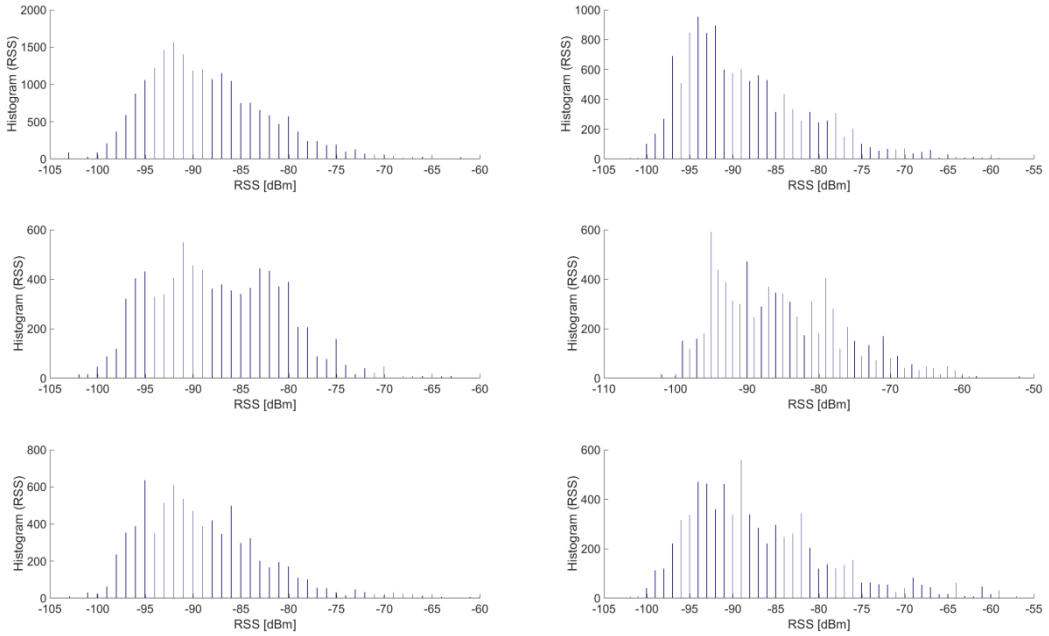

Figure 6.6. Six TXID-specific RSS-WiFi histograms.

Table 6.1. Main RSS [dBm] statistic and data point size relative to the histograms of Figure 6.5 and Figure 6.6. Histograms 1 to 6 match subfigures left to right, up to down.

	LTE					
	Hist. 1	Hist. 2	Hist. 3	Hist. 4	Hist. 5	Hist. 6
Points	47611	29592	26939	17640	15416	11808
Min	-140.0	-140.0	-120.9	-140.0	-140.0	-140.0
Max	-84.6	-99.0	-42.9	-108.3	-54.3	-58.5
Ave	-113.2	-125.9	89.2	-132.3	-98.9	-94.6
	WiFi					
	Hist. 1	Hist. 2	Hist. 3	Hist. 4	Hist. 5	Hist. 6
Points	20249	12201	8352	8092	7827	7367
Min	-104.0	-102.0	-102.0	-102.0	-103.0	-102.0
Max	-62.0	-59.0	-63.0	-52.0	-60.0	-57.0
Ave	-88.5	-88.5	-87.4	-85.3	-89.0	-86.8

sectors, however, may have different transmission powers.

Further, looking at all the RSS histograms, one TXID at a time for both the systems, one notices that while for LTE the distribution varies from skewed to multi-modal, for WiFi it is more stable as typically meets skew shape. In Figures 6.5-6.6, the histograms for the first six top measurement points are shown, each radio system; that is a representative set of histograms, each case. Here, it can be seen that the WiFi RSS histograms are somewhat aligned with each other and there is consistency in the shape. The LTE RSS histograms, in contrast, show poor mutual consistency. In Table 6.1 the main statistics of the histograms, such as the minimum, the maximum, and the average value, are summarized. There is an evident fluctuation of the statistics along the LTE RSS histograms, while the horizontal statistics related to the WiFi RSS histograms are well mutually aligned.

An antenna broadcasting LTE signals is typically designed to offer a tunable radiation pattern, which allow optimizing and planning of the radio network resources. Therefore, different radio patterns are associated to different antennas, normally. One such diversity in the radio pattern, including the effect of side lobes, perhaps yields the inconsistency among the LTE RSS histograms, that was seen before. In addition, the coverage area is large compared to the measurements area (see again figure 4.4), thus, the effect of the boundary proximity on the radio measurements may alter the distribution of the signal strength if an antenna's main lobe crosses the boundary. In fact, the available data is not inclusive of the radio information being part of the signal radiated beyond the limits of the measurement area.

6.1.4 Estimation data

Performing positioning estimation requires using of estimation data. Therefore, collected learning data is to be split into two sets: training data and estimation data. This can be done

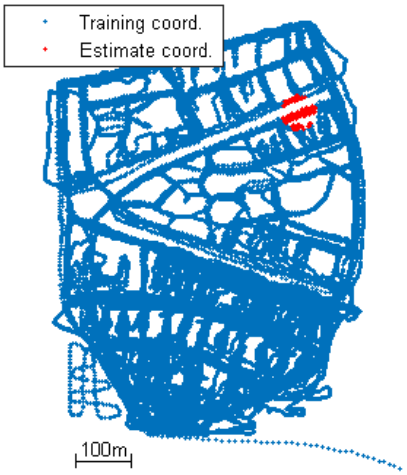

Figure 6.7. Instance of training and estimate LTE data locations obtained by circular portion splitting methods.

for instance via random, or circular portion extraction of data. The latter method was introduced by Talvitie et al. in [79] to simulate possibly existing measurement activity coverage holes and is based on choice of coordinates either in or out a circular area of defined radius (Figure 6.7). Whether the data extraction method is random-based or circular-portion-based, extracted data become estimation data while the reminder data become training data.

The first algorithm of block diagram in Figure 6.8 finds which data point is to be removed and which one preserved, and defines two indexing vectors, accordingly. The sum of the length of the two indexing vectors is the same as the overall number N of data points. The algorithm inputs one N×2 matrix of the data locations and two setting parameters, and outputs two indexing vectors, of length R (removing) and $P = N - R$ (preserving), respectively. The setting parameters specify the amount of estimation data, in terms of

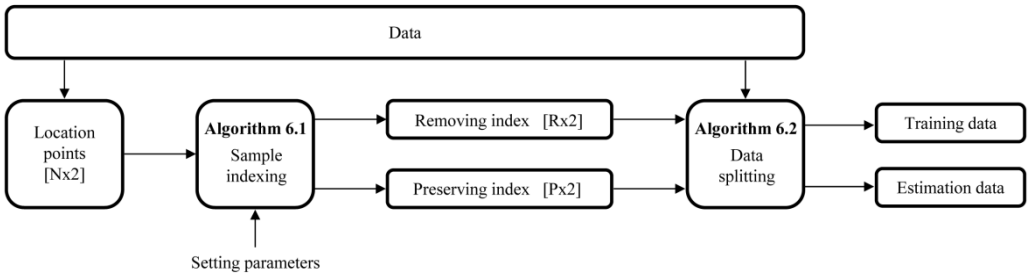

Figure 6.8. Block diagram describing the process of splitting data into estimation data and training data.

percentage, and the radius of a circular region where data is extracted from. There may be more than one circular region up to a number sufficient to contain a needed amount of data. The second encountered algorithm splits input data into two output parts, in accordance with the indexing information obtained by the two input indexing vectors.

Algorithm 6.1. Sample indexing for block removal of data.

```
input: data                          /* data to be split */
       radius, remove percent        /* parameters defining the block of
                                        data to be removed. Block size is
                                        the radius of a circle */

r̄, p̄ ← 0^(N×1)        /* pre-allocate with the removing and the preserving
                          logical indexing vectors */
R ← N × remove percent               /* size of vector r or points to be
                                        removed */
while not removed enough do                    /* indexing loop */
    q ← l ∈ Ω_(N×1) = {data locations}, P_X(q) = 1/N     /* center of circle */
    d_n ← |q, l_n|, with n = 1,2..N  /* set of distances from the n-th
                                        location point to the circle
                                        center */
    r_n = 1 ← d_n < radius           /* ones are index of points to be
                                        removed */
    removed ← Σ_(n=1)^N r_n          /* count removed points */
    if removed > r then              /* removed points exceeding */
        E ← removed - R              /* number of points in excess */
        r_n^ind ← find(r_n)          /* find indices in r same as one */
        r_n^(ind,sort) ← sort(d_n(r_n^ind))  /* sort removed indices by distance
                                                farther to nearest */
        r_q^(ind,back) ← r_q^(ind,sort), q = 1,2..E   /* indices to be get back */
        r_n = 0 ← n = r_q^(ind,back)                  /* get back from one to zero */
    end if
end while

p_n = 1 ← r_n = 0          /* ones are index of points to be preserved */
r_n ← find(r_n)            /* find indices in r same as one */
p_n ← find(p_n)            /* find indices in p same as one */
```

Algorithm 6.2. Splitting data into training and estimation.

```
input:  r̄, p̄                  /* vectors with removed and preserved
                                  indices, respectively */
        data                   /* data to be split */

m, u ← 1                       /* initialize the index for both training
                                  and estimator vector */
for k=1 to N do                /* splitting loop */
    if ismember(k,r) is true then      /* the current index is
                                          contained in r overbar */
        data_m^train = data_k
        m ← m+1                /* unitary increment */
    else
        data_u^est = data_k
        u ← u+1                /* unitary increment */
    end if
end for
```

In Algorithm 6.1, a loop iterates until an amount of removed data $\mathrm{R} = \mathrm{N} \times remove\ \%$ is reached. Before the loop, two indexing vectors of length N are initialized by zeros. Within the loop, to each iteration, a location point is chosen randomly, with uniform probability function, to be the center q of a circle of radius *radius*. Once the circle is defined, the vector of distances $\bar{d}$ is built, each element $\bar{d}_n$ being the distance from q to the n^{th} location point, where $n = 1,2, \dots, \mathrm{N}$. At this point, within the same iteration round, firstly, each element of vector $\bar{d}_n$ is scanned such that if the n^{th} distance is smaller than *radius* then the n^{th} element of the removing indexing vector $\bar{r}$ is changed from 0 to 1. Secondly, ones in the removing indexing vector are counted, that is the number of virtually removed samples. Thirdly, an exceeding amount of removed samples, if any, are put back to data. In doing so, an equivalent number of elements in the indexing vector is changed from 1 to 0. After the loop, firstly, the preserving indexing vector $\bar{p}$ is obtained as a vector of opposite elements with respect to the elements of the removing indexing vector. Secondly, for both vectors, $\bar{r}$ and $\bar{p}$, ones are converted into vector indices while nulling zeros.

Algorithm 6.2 scans input data one sample at a time. If the current sample index is the same as one of the elements in the removing indexing vector, then it is allocated into the training data set; otherwise, it is delivered to the estimation data set.

6.2 Transmitter localization

Transmitter location estimation is an essential function of the path-loss model based positioning algorithm. This section presents two possible estimation methods and shows the histogram of the mutual distance between all the estimated transmitter location pairs. By the histogram we will see also the impact of MSC removal.

6.2.1 Introduction

One way to achieve transmitter estimation is based on the assumption that the location point with strongest recorded signal level is most likely the closest one to the transmitter site.
A second way is slightly more complex. It approaches the estimation by computing the coordinates of the center of mass of a particular set of K mass points. The masses are placed at the location of the K top RSS points and are weighted as their linear scale RSS values. By terminology, we refer to the expression 1st strongest selection to describe the former method whereas the term K-strongest RSS-weighted centroid is used for the latter.
Other ways are obviously possible, for instance [40], but we do not deal with methods other than those described so far; which are compactly summarized as below,

A. 1st strongest selection

- $Arg\ \max_{\vartheta} \mathrm{RSS}_{\vartheta} \leftarrow \overline{\mathrm{T}}^{\mathrm{loc\ est}} = \bar{l}_{\vartheta}$
- $\vartheta \in \Omega = \{1,2,..\theta\}$ where θ is the size of the RSS point set
- $\bar{l}_{\vartheta} \in \mathbb{R}^{2\times 1}$ is the location of the ϑ-*th* RSS point

B. K-strongest RSS-weighted centroid

- $\overline{\mathrm{T}}^{\mathrm{loc\ est}} = \frac{\sum_{k}^{\mathrm{K}} \mathrm{w}_k \bar{l}_k}{\sum_{k}^{\mathrm{K}} \mathrm{w}_k}$
- $k \in \Omega = \{1,2..\mathrm{K}\}$ where K is the size of the RSS point set
- w_k is the weight of *k-th* mass point
- $\bar{l}_k$ is the location of the *k-th* mass point

6.2.2 Estimation algorithm

Herein presented estimation algorithm suits for data delivered in form of TXID grid set. Such format is a cell array equal in length to the size of the TXID set. Each element of the array is a grid consisting of data with common TXID. With this input format, the estimation

algorithm processes all the grids one by one through an iterative loop. To each iteration, the current input TXID grid is processed and the internal output stored. As the cycle ends, the external output is the entire set of location estimates. The number of location estimates equals to the number of transmitter identities as grouping is not involved in the algorithm. Grouping means that different transmitter location estimates, coming from common source signals with separate identity codes, are pooled together with the aim of mapping different estimates to the same location. Defining a mapping rule is not trivial because a TXID identity code, despite its name, does not really say which transmitter the signal belongs to.
For our algorithm, sources of different signals are considered independent to each other. In other words, we suppose that each TXID set of signals is generated by its own source; we do not consider the possibility that some of them may belong to the same transmitter.
A set of fundamental commands to be systematically executed by the algorithm is below reported for both the used methods (see Figure 6.9 for reference and further details).

A. 1st strongest selection

For each TXID grid, find the point with strongest RSS and output its coordinates. If the number of strongest RSS points is greater than one then the selection is done on first-come first-served basis (e.g. top row in the RSS-wise sorted grid).

B. K-strongest RSS-weighted centroid

Before the iteration loop, set the number of strongest neighboring K. As the iterative loop starts, proceed by the following commands:

1. Perform column-wise sorting of the grid, from high to low RSS value - the first row returns the observation with strongest RSS, the second row returns the observation with second strongest RSS and so on up to the last row which returns the weakest RSS observation.
2. Count the number of data points or observations in the grid. If the counted number D is smaller than K, update K = D; the number of strongest neighbors is data point size limited to prevent the absurd K < D.
3. Collect the first K local coordinates pairs and RSS values from the RSS-wise sorted grid, top row to down.
4. Convert RSS from logarithmic to linear scale and perform RSS-weighted centroid - as described in Section 6.2.1. By converting scale, RSS comes to positive values which suits for weight.

To concisely sum up, Algorithm B first sets the initial state for K and second executes an iteration loop. Within the loop, it sequentially: collects K grid points with strongest RSS value, extracts RSSs and local coordinates, computes RSS-weighted centroid and outputs it.

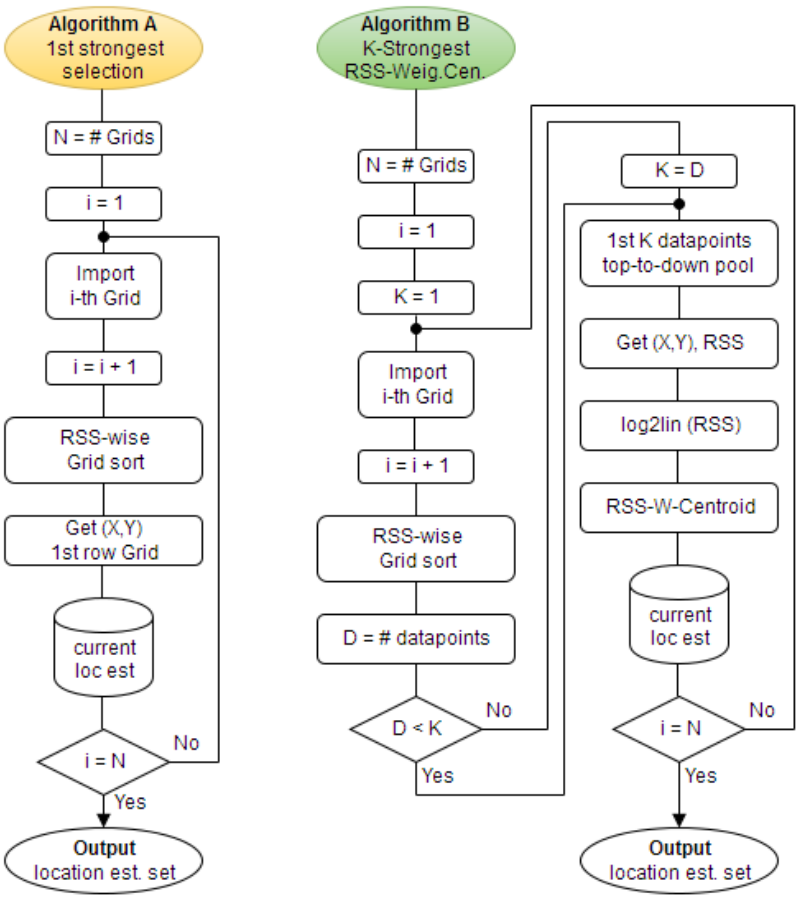

Figure 6.9. Flow chart of Algorithm A (to left) and Algorithm B (to right).

6.2.3 Setup algorithm

At the beginning of the estimation process, there is a set of preliminary operations required to prepare the algorithm start. The logical flow of the operations is carried out according to the settings in the conditionals or diamond shaped blocks of Figure 6.10. The first decision block concerns the data type. It can be either LTE or WiFi and controls the import of data accordingly. The second conditional decides whether to perform multiple scan removal or not. Multiple scan refers to a set of samples with common location information. As MSC is mapped into one single scan (SSC), the multiplicity is removed and the amount of data reduced accordingly. Now, data is split into TXID grids and handled by either Algorithm A or Algorithm B according to the setting of the last two conditional blocks. The adopted technique for removal of the multiplicity and its implementation as well as the implementation of TXID-grid-forming were considered as marginal topics of this work and it is not described in further detail.

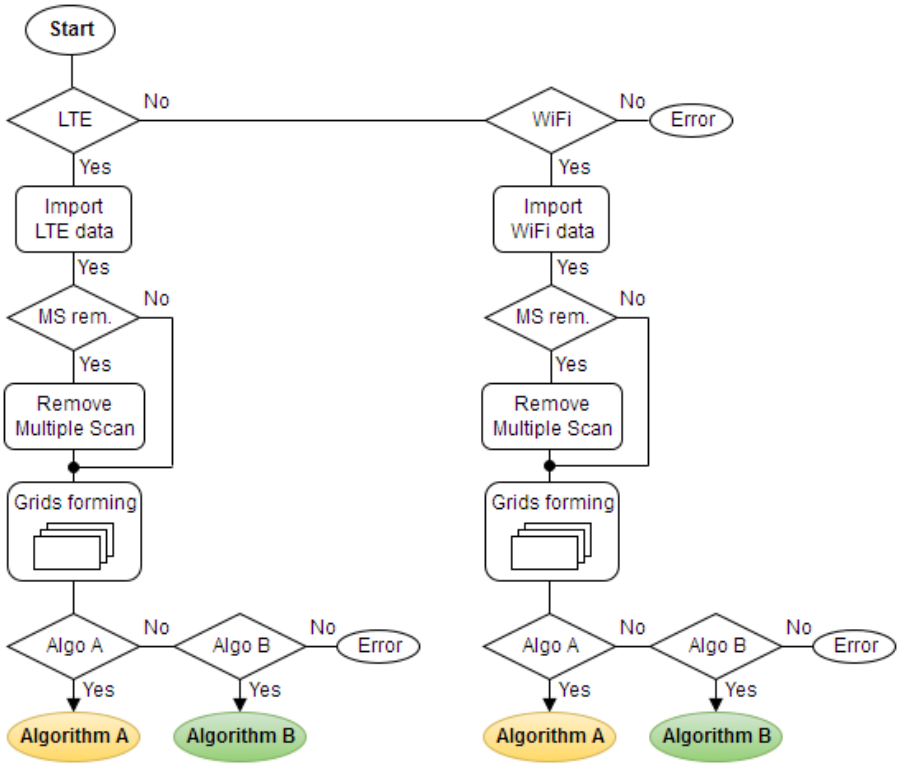

Figure 6.10. Flow chart of preliminary operations: from start to algorithm A/B.

6.2.4 Histogram: mutual distance

To find out which method performs better, the knowledge of the true location of the cellular antennas must be known. Unfortunately, such information remains confidential to the network operators who do not publish it. Thus, in the absence of it, is not possible to assess the performance in absolute sense for neither of the methods.

Whereas we cannot establish which method is most accurate, we can still observe the extent to which the estimates from the two methods differ mutually. We do so by the histogram of the measures of the mutual distance between two transmitter location points obtained by different estimation methods for the same target transmitter. The overall number of mutual distances in the histogram equals to the number of estimation pairs which in turn equals to the number of transmitter to be localized, i.e. 1285 TXIDs for LTE and 3159 TXIDs for WiFi.

The histograms of the mutual distances are shown below in Figure 6.11, for four different setting of the setup algorithm (see Section 6.2.3). The objective is to assess in statistical terms how many meters the transmitter location estimate obtained by one method is far away from the one obtained by the other method, when input is same.

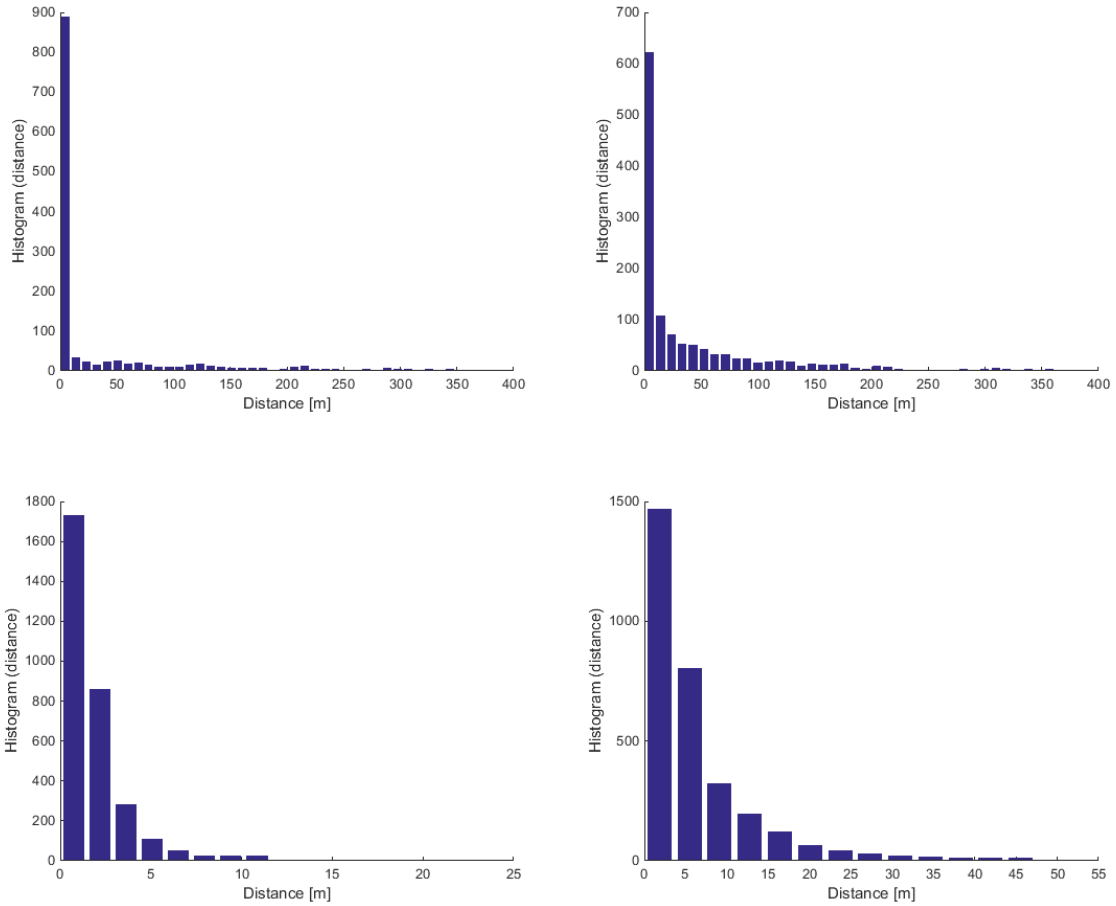

Figure 6.11. Histogram of mutual distance throughout all the (method A - method B) estimated TXID location pairs. Setup simulation: LTE data, MSC removal off (to top left); LTE data, MSC removal on (to top right); WiFi data, MSC removal off (to bottom left); WiFi data, MSC removal on (to bottom left).

By this we aim to establish whether the choice of one method may influence the positioning output in the path-loss model estimator (Section 5.2) or not.

In addition to the illustrations, numerical data is reported to precise knowledge: the mean of the mutual distances results 39.60 m/47.26 m without/with MSC removal by LTE, and 2.10 m/7.20 m without/with MSC removal by WiFi. By removing the MSCs, the mutual distance increases by a factor of 1.19 and 3.42 respectively for LTE and WiFi. It is worth just to mention, without numeric or illustrative response, how the increase of the mutual distance experienced by method A, resulted more sensitive to the MSC removal than that by method B. By the LTE-based data estimation, 70 % of distances fall below 10.30 m, 80 % of distances fall below 63.00 m and 90 % of distances fall below 156.93 m; while by WiFi, 70 % of distances fall below 1.89 m, 80 % of distances fall below 3.89 m and 90 % of distances fall below 3.89 m. Removing MSC causes an increase of 4.50, 1.30, and -1.03

for LTE and, 3.49, 3.74 and 4.04 for WiFi, respectively to 70 %, 80 % and 90 % cumulative distribution of the error.
The reason for which the mutual distance experienced by LTE is greater than that by WiFi is somehow related to the different coverage size offered by the two radio systems. E.g. LTE/WiFi coverage area ratio equals to five, typically. For WiFi, the cumulative distribution of the mutual distance exhibits a faster growth as compared with LTE. LTE measurements include signals from sources outside the measurement area, even far away from the boundary. For those sources, estimation may be not reliable and consequently the mutual distance altered.
Which method to apply is not an issue when localization concerns access points but the choice results more incisive if localization regards base stations. The removal of MSC seems to impair the estimators as mutual distance increases in both AP and BS localization.

6.2.5 Summary

In this section, we described how to achieve the transmitter location estimation through two different approaches. We addressed the topic also from the implementation point of view and shown flow charts for illustration purpose. We adopted two alternative algorithms for the same purpose. We noticed that they have different complexity although without giving scientific evidence. An absolute comparative analysis was not achieved as missing of true reference data. However we showed that the two estimation methods perform closely to each other in terms of distance error for the sole WiFi system, not for LTE.

6.3 Positioning results

To compute the positioning accuracy in terms of RMSE the experiment is repeated 100 times. For sake of the reliability, each run, independently, generates and uses a set of estimation data obtained by circular block removal of 0.5 % data: the circle radius is set to 40 m whereas its center is placed to a random measurement point. Doing so, the overall number of estimation points for an experiment amounts to 184000; that is the number of estimation runs. With reference to RMSE, experimental results are summarized in Table 6.2 for different positioning methods and settings.
The result of the experiment shows that the fingerprinting algorithm outperforms the probabilistic approaches in terms of cumulative distribution function of the distance error, using either LTE or WiFi data. See Figure 6.12 and Figure 6.13.
For WiFi, the coverage-area algorithm performs very close to the fingerprinting one if only strong signals are taken into account otherwise, it performs worse than that the path-loss model algorithm does. The coverage area algorithm via strong signals, however, fail

Table 6.2. RMSE by 100 experiments (184000 evaluation points) for the three different outdoor positioning methods: from top to bottom fingerprinting (benchmark), coverage-area and path-loss model. Simulation setting is reported in Chapter 5.

Fingerprinting method - RMSE [m]			
Evaluation data extraction	Data Type	RMSE [m]	Successful Estimate [%]
Block Removal (5%, 40m)	LTE	183.65	100.00
	WiFi	36.79	100.00

Coverage area method - RMSE [m]					
		All signal counted		Strong signal counted	
Evaluation data extraction	Data Type	RMSE [m]	Successful Estimate [%]	RMSE [m]	Successful Estimate [%]
Random Removal (5%)	LTE	187.79	100.00	183.94	63.24
	WiFi	34.07	100.00	27.49	97.85
Block Removal (5%, 40m)	LTE	183.65	100.00	177.08	65.38
	WiFi	36.79	100.00	30.66	98.08

Path loss method - RMSE [m]					
		Frequency not counted		Frequency counted	
Evaluation data extraction	Data Type	RMSE [m]	Successful Estimate [%]	RMSE [m]	Successful Estimate [%]
Random Removal (0.5%)	LTE	205.55	55.64	206.53	55.72
	WiFi	33.62	99.18	-	-
Block Removal (0.5%, 40m)	LTE	203.21	55.11	205.87	55.31
	WiFi	34.00	97.95	-	-

at rate of 2.15 % for WiFi and 36.76 % for LTE, therefore reliability is not addressed. We point out how the issue could be tackled by developing a bi-mode system able to switching the two of modes according to the received signal level.

Using LTE data, path-loss model and coverage-area algorithms perform similarly to each other to some extent depending on the particular implementation. For instance, using the frequency band information into the path-loss model brings no improvement perhaps due to the OFDM modulation format of the LTE signals: LTE bandwidth, ranging 1.25 MHz to 20 MHz, is designed to accommodate subcarriers spaced 15 KHz to each other. Thus, the proportions are such that a subcarrier could be several MHz apart from the carrier frequency which is the available information associated to it.

In regard of the E911 mandate, the experimental results fully meet the accuracy requirements when using WiFi data; conversely, never do so for LTE. Table 6.3 shows the error distance in reference to the FCC established location accuracy requirement.

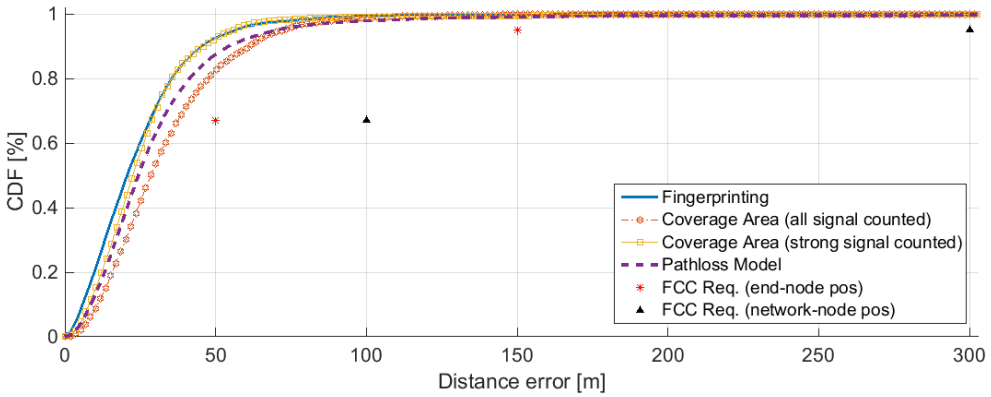

Figure 6.12. WiFi-based data cumulative distribution of the distance error.

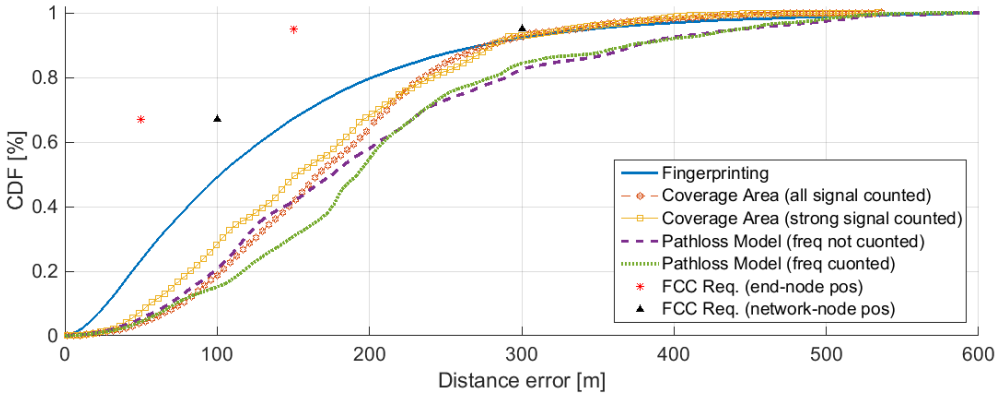

Figure 6.13. LTE-based data cumulative distribution of the distance error.

Table 6.3. Error distance at the FCC established location accuracy requirement.

		Error distance [m]						
		Path loss		Coverage area		Fing.pt.	FCC req.	
Data type	CDF [%]	Freq. not counted	Freq. counted	All signals	Strong signals	- -	End node	Network node
LTE	67	225.58	228.25	204.77	197.39	148.95	50	100
	95	256.12	251.28	235.51	241.32	202.03	150	300
WiFi	67	32.98	-	38.54	28.95	28.23	50	100
	95	42.85	-	48.53	35.71	35.71	150	300

6.4 Database size and reduction factor

With reference to our study case and, the considered positioning algorithm, a necessary quantity of data is to be stored and, when necessary, transferred to the positioning mobile device. The exact number of stored parameters in each of the three examined methods is shown in Table 6.4. It is calculated after mapping multiple scans into a single scans that, consequently, represents the optimal case. The applied formulas are presented in (6.1) for fingerprinting, (6.2) for coverage-area, (6.3) for path-loss model, as following:

$$2 \times \mathrm{N}^{\mathrm{lp,unique}} + \sum_{i}^{\mathrm{N}^{\mathrm{TXID}}} \mathrm{N}^{\mathrm{mp,grid}_i} \tag{6.1}$$

where 2 is the number of parameters used to represent the *x*, *y* geographical coordinates, $\mathrm{N}^{\mathrm{lp,unique}}$ represents the number of unique location points, $\mathrm{N}^{\mathrm{TXID}}$ represents the number of TXIDs, $\mathrm{N}^{\mathrm{mp,grid}_i}$ represents the number of measurement points relative to the i^{th} TXID-grid;

$$(2 + 3) \times \mathrm{N}^{\mathrm{TXID}} \tag{6.2}$$

where 2 and 3 are the number of parameters used to represent the position mean and covariance respectively;

$$(2 + 2 + 1) \times \mathrm{N}^{\mathrm{TXID}} + 2 \tag{6.3}$$

where, by order of encounter, 2, 2, 1 and 2 are the number of parameters used to represent the path-loss model, the transmitter location coordinates, the shadowing standard deviation and the synthetic grid.

Clearly, in our study case, the considered probabilistic approaches offer a significant reduction of the database size. The reduction factors with respect to the considered fingerprinting benchmark are reported in Table 6.4. It can been seen that the two methods, coverage area and the path-loss, can offer an identical reduction factor; that is 148 times by WiFi data and 66 times by LTE data.

Table 6.4. Number of parameters required for the considered positioning algorithms and reduction factors gained by the use of probabilistic approaches.

Data	Fingerprinting database	Coverage area database	Path loss database	Reduction factor (CA)	Reduction factor (PL)
WiFi	2338009	15805	15807	148	148
LTE	423872	6425	6427	66	66

7. CONCLUSION AND FUTURE WORK

In this thesis we explored the performance of three outdoor positioning algorithms based on RSS measurements. We proved that by using the statistical coverage-area and path-loss based models, instead of the deterministic fingerprinting model, the size of the database is reduced significantly while the accuracy loss is negligible with respect to the E911 mandate. We pointed out that the two statistical models are efficient in terms of memory space and consequent benefits are expected in the estimation time, battery consumption and communication complexity, all relevant features in low-cost mobile positioning.

We evaluated three positioning methods using experimental data collected by one commercial mobile device in suburban-type outdoor environment by the cumulative distribution of the distance error. We measured the distance error at the 67% and 95% confidence levels, for each methods, and compared those measurements to the E911 location accuracy requirements. We showed that the E911 location accuracy requirements are fully satisfied through all methods by using WiFi data, not by LTE. For the latter, we experienced that the knowledge of the carrier frequencies might not help to improve the accuracy, yet we are confident that further improvement can be achieved by modeling path-loss in different way.

Future work for the path-loss model should aim to develop an angle-based estimator for directional LTE antennas. By doing so, a plane should be split into sectors and data should be divided in subsets, accordingly. For each sector the path-loss parameters should be estimated via the methods of least squares, or other deconvolution methods, by using the subset of data of the considered sector. Other future work should aim to define a systematic selection criterion of transmitters inside the measurement area. Transmitters outside the measurement area should not be taken into account since their location estimation is error-prone.

With reference to the coverage-area model, further work should aim to implement and simulate a bi-level model able to combine the two existing modes in such way to improve accuracy with no impairment to robustness for the estimator. One such model should perform, firstly under the strong-signal condition and secondly, if location estimation is not successfully achieved, under the all-signal condition.

Finally, future work should also aim to develop LTE/WiFi integrated solutions to get benefit from the coexistence of the two radio access technologies, a noteworthy feature of heterogeneous small cell networks.

BIBLIOGRAPHY

[1] 3GPP TR. "36.805 V9.0.0 (2009-12) LTE; Technical Report." *Study on Minimization of drive-tests in Next Generation Networks*.

[2] 3GPP TS. "36 133 V13.2.0 (2016-01) LTE; Technical Specification." *Evolved Universal Terrestrial Radio Access (E-UTRA)*; *Requirements for support of radio resource management*.

[3] 3GPP TS. "36 214 V13.0.0 (2015-12) LTE; Technical Specification." *Evolved Universal Terrestrial Radio Access (E-UTRA)*; *Physical layer; Measurements*.

[4] 3GPP TS. "36 355 V13.0.0 (2015-12) LTE; Technical Specification." *Evolved Universal Terrestrial Radio Access (E-UTRA)*; *LTE Positioning Protocol (LPP)*.

[5] 3GPP TS. "37.320 V13.0.0 (2015-03) LTE; Technical Specification." *Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio measurement collection for Minimization of Drive Tests (MDT)*.

[6] 3GPP TSG RAN1 #56bis. "Study on hearability of reference signals in LTE positioning support," *3GPP TDoc at meeting*, TDoc R1-091336, Seoul, South Korea, Mar. 2009.

[7] S. A. Ahson and M. Ilyas, *Location-based services handbook: Applications, technologies, and security*. CRC Press, 2010.

[8] E. Almazrouei, N. Al Sindi, S. R. Al-Araji, N. Ali, Z. Chaloupka, and J. Aweya, "Measurement and analysis of NLOS identification metrics for WLAN systems," in *Proc. IEEE International Symposium on Personal, Indoor, and Mobile Radio Communication*, Washington DC, DC, USA, Sep. 2014.

[9] C. Andren and M. Webster, "CCK modulation delivers 11 Mbps for high rate IEEE 802.11 extension," in *Proc. Wireless Symposium/Portable by Design Conference*, Palm Bay, Florida, USA, 1999.

[10] C. A. Balanis, *Antenna theory: analysis and design*, vol. 1. John Wiley & Sons, 2005.

[11] D. Baumann, "Minimization of drive tests (MDT) in mobile communication networks," in *Proc. Seminars on Future Internet and Innovative Internet*

Technologies and Mobile Communications, vol. NET-2014-03-1, pp. 9-1, Munich, Germany, Mar. 2014.

[12] O. Bejarano, E. W. Knightly, and M. Park, "IEEE 802.11 ac: from channelization to multi-user MIMO," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 84-90, Oct. 2013.

[13] F. Benedetto, G. Giunta, and E. Guzzon, "Enhanced TOA-based indoor-positioning algorithm for mobile LTE cellular systems," in *Proc. Workshop on Positioning, Navigation and Communication*, Dresden, Germany, Apr. 2011.

[14] K. Borre, D. M. Akos, N. Bertelsen, P. Rinder, and S. H. Jensen, *A software-defined GPS and Galileo receiver: a single-frequency approach*. Springer Science & Business Media, 2007.

[15] M. J. Boyle, *Department of Defense World Geodetic System 1984 - Its definition and relationship with local geodetic systems*. DMA Technical Report 83502.2., Washington, DC, USA, 1987.

[16] P. Brida, J. Machaj, J. Benikovsky, and J. Duha, "A new complex angle of arrival location method for ad hoc networks," in *Proc. Workshop on Positioning, Navigation and Communication*, Dresden, Germany, Mar. 2010.

[17] A. Buchman and C. Lung, "On the relationship between received signal strength and received signal strength index of IEEE 802.11 compatible radio transceivers," *Carpathian Journal of Electronic and Computer Engineering*, vol. 6, no. 2, pp. 15-20, July 2013.

[18] C. S. Chen, J. M. Lin, W. H. Liu, and C. L. Chi, "Dilution of position calculation for MS location improvement in wireless communication systems," *Journal of Networks*, vol. 6, no. 10, pp. 1452-1458, Oct. 2011.

[19] Y. Chen, Q. Pan, Y. Liang, and Z. Hu, "AWCL: Adaptive weighted centroid target localization algorithm based on RSSI in WSN," in *Proc. IEEE International Conference on Computer Science and Information Technology*, Beijing, China, July 2010.

[20] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci, "Evaluation of the LTE positioning capabilities under typical multipath channels," in *Proc. Conference Advanced Satellite Multimedia Systems*

Conference and Workshop Signal Processing for Space Communications, Vigo, Spain, Sep. 2012.

[21] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 7, pp. 1205-1211, July 1999.

[22] ETSI TS. "136 101 V12.9.0 (2015-10) LTE; Technical Specification." *Evolved Universal Terrestrial Radio Access (E-UTRA)*; *User Equipment (UE) radio transmission and reception*.

[23] ETSI TS. "136 211 V13.1.0 (2016-03) LTE; Technical Specification." *Evolved Universal Terrestrial Radio Access (E-UTRA)*; *Physical channels and modulation*.

[24] ETSI TS. "136 214 V13.0.0 (2016-01) LTE; Technical Specification." *Evolved Universal Terrestrial Radio Access (E-UTRA)*; *Physical layer; Measurements*.

[25] Federal Communication Commission, Wireless E911 Location Accuracy Requirements, *Third Further Notice of Proposed Rulemaking*. Document dates: adopted on February 20, 2014, released on February 21, 2014. Document number: FCC 14-13.

[26] Federal Communication Commission, Wireless Telecommunications Bureau, *Fact Sheet: wireless 911 requirements*, Jan. 2001.

[27] R. Frank, "Polyphase complementary codes," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 641-647, Nov. 1980.

[28] C. Geßner, *Long Term Evolution: A Concise Introduction to LTE and its Measurement Requirements*. Rohde & Schwarz, 2013.

[29] I. Güvenç and C. C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 3, pp. 107-124, Aug. 2009.

[30] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, "Minimization of Drive Tests Solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28–36, June 2012.

[31] J. Heiskala and J. Terry, *OFDM wireless LANs: A theoretical and practical guide*. Sams, 2001.

[32] V. Honkavirta, T. Perala, S. Ali-Loytty and R. Piché, "A comparative survey of WLAN location fingerprinting methods," in *Proc. Workshop on Positioning, Navigation and Communication*, Hannover, Germany, Mar. 2009.

[33] IEEE 802.11 Working Group. *Std. 802.11-2007. Revision of ANSI/IEEE std. 802.11, 1999 edition: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*. Technical report, ANSI/IEEE, 1999.

[34] IEEE 802.11 Working Group. *Std. 802.11a-1999. Supplement to ANSI/IEEE std. 802.11, 1999 edition: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*. Technical report, ANSI/IEEE, 1999.

[35] IEEE 802.11 Working Group. *Std. 802.11b, 1999 edition: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*. Technical report, ANSI/IEEE, 1999.

[36] IEEE 802.11 Working Group. *Std. 802.11b-1999 (R2003). Supplement to ANSI/IEEE std. 802.11, 1999 edition: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*. Technical report, ANSI/IEEE, 1999.

[37] IEEE 802.11 Working Group. *Std. 802.11g-2003. Amendment to ANSI/IEEE std. 802.11, 1999 edition: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*. Technical report, ANSI/IEEE, 1999.

[38] IEEE 802.11 Working Group. *Std. 802.11n-2009. Amendment to ANSI/IEEE std. 802.11, 2007 edition: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*. Technical report, ANSI/IEEE, 2007.

[39] IEEE Standards Committee. "754-2008 IEEE standard for floating-point arithmetic." *IEEE Computer Society Std*, Aug. 2008.

[40] M. Jarvis and B. Tarlow, "Wi-Fi position fix," European patent application, EP 2 574 954 A1, Mar 2013. [Online]. Available: https://register.epo.org.

[41] Y. Jianga, A. Hua, and W. Jinb, "A TDOA Estimation Method of PRS in LTE Systems," *Journal of Information & Computational Science*, pp. 509-517, Jan. 2014.

[42] J. Johansson, W. Hapsari, S. Kelley, and G. Bodog, "Minimization of drive tests in 3GPP release 11," *IEEE Communications. Magazine*, vol. 50, no. 11, pp. 36–43, Nov. 2012.

[43] J. Jun, P. Peddabachagari, and M. Sichitiu, "Theoretical maximum throughput of IEEE 802.11 and its applications," in *Proc. IEEE International Symposium on Network Computing and Applications*, Cambridge, MA, USA, Apr. 2003.

[44] V. Kelly, ”New IEEE 802.11ac specification driven by evolving market need for higher multi-user throughput in wireless LANs”, *International CES*, Las Vegas, NV, USA, Jan. 2014. [Online]. Available: http://standards.ieee.org/news/2014/ieee_802_11ac_ballot.html.

[45] K. Keunecke and G. Scholl, "Deriving 2D TOA/TDOA IEEE 802.11 g/n/ac location accuracy from an experimentally verified fading channel model," in *Proc. International Conference on Indoor Positioning and Indoor Navigation*, Montbeliard-Belfort, France, Oct. 2013.

[46] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen, "Positioning with coverage area estimates generated from location fingerprints," in *Proc. Workshop on Positioning Navigation and Communication*, Dresden, Germany, Mar. 2010.

[47] C. Laoudias, R. Piché, and C. G. Panayiotou, "Device signal strength self-calibration using histograms," in *Proc. IEEE International Conference on Indoor Positioning and Indoor Navigation*, Sydney, Australia, Nov. 2012.

[48] X. Li, X. Ma, S. Yan, and C. Hou, "Super-resolution time delay estimation for narrowband signal," *IET Radar, Sonar & Navigation*, vol. 6, no. 8, pp. 781-787, Oct. 2012.

[49] B. Li, J. Salter, A. G. Dempster, and C. Rizos, "Indoor positioning techniques based on wireless LAN," in *Proc. IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, Sydney, Australia, Mar. 2006.

[50] E. S. Lohan, K. Koski, J. Talvitie, and L. Ukkonen, "WLAN and RFID propagation channels for hybrid indoor positioning," in *Proc. IEEE International Conference on Localization and GNSS*, Helsinki, Finland, June 2014.

[51] A. Makki, A. Siddig, M. Saad, and C. Bleakley, "Survey of WiFi positioning using time-based techniques," *Computer Networks*, vol. 88, pp. 218-233, Sep. 2015.

[52] P. Massatt, W. Rhodus, and K. Rudnick, "2-D and 3-D characterizations of GPS navigation service," in Proc. of *IEEE Position Location and Navigation Symposium*, Las Vegas, NV, USA, Nov. 1986.

[53] A. Meylan, S. A. Ahmadzadeh, S. Veerepalli, A. Gholmieh, G. Giaretta, S. Maheshwari, and U. S. Patel, "Over the top methods for aggregation of WLAN carriers to LTE." U.S. Patent No. 20,160,043,844, Feb. 2016.

[54] R. J. Milliken and C. J. Zoller, "Principle of operation of NAVSTAR and system characteristics," *The Journal of Navigation*, vol. 25, no. 2, pp. 95-106, June 1978.

[55] P. Misra and P. Enge, Global Positioning System: *Signals, Measurements and Performance, Second Edition*. Ganga-Jamuna Press, 2006.

[56] S. K. Mitra and Y. Kuo, *Digital signal processing: a computer-based approach*. McGraw-Hill, 2006.

[57] A. F. Molisch, *Wireless communications*. John Wiley & Sons, 2007.

[58] R. U. Mondal, T. Ristaniemi, and J. Turkka, "Genetic algorithm optimized grid-based RF fingerprint positioning in heterogeneous small cell networks," in *Proc. International Conference on Localization and GNSS*, Gothenburg, Sweden, June 2015.

[59] H. Nurminen, J. Talvitie, S. Ali-Löytty, P. Müller, E. S. Lohan, R. Piché, and M. Renfors, "Statistical path loss parameter estimation and positioning using RSS measurements in indoor wireless networks," in *Proc. International Conference on Indoor Positioning and Indoor Navigation*, Sydney, Australia, Nov. 2012.

[60] B. Pearson, "Complementary code keying made simple," *Application Note*, no. 9850.2, Intersil Corporation, Milpitas, CA, USA, Nov. 2001.

[61] R. L. Peterson, R. E. Ziemer, and D. E. Borth, *Introduction to spread-spectrum communications*, vol. 995. Prentice Hall, 1995.

[62] R. Piché, "Robust estimation of a reception region from location fingerprints," in *Proc. IEEE International Conference on Localization and GNSS*, Tampere, Finland, June 2011.

[63] J. G. Proakis, *Intersymbol Interference in Digital Communication Systems*. John Wiley & Sons, Inc., 2001.

[64] M. Raitoharju, M. Dashti, S. Ali-Löytty, and R. Piché, "Positioning with multilevel coverage area models," in *Proc. IEEE International Conference on Indoor Positioning and Indoor Navigation*, Sydney, Australia, Nov. 2012.

[65] M. Raitoharju, H. Nurminen, and R. Piché, "Kalman filter with a linear state model for PDR+WLAN positioning and its application to assisting a particle filter," *EURASIP Journal on Advances in Signal Processing*, Apr. 2015.

[66] P. Roshan and J. Leary, *802.11 Wireless LAN fundamentals*. Cisco press, 2004.

[67] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24-40, June 2005.

[68] G. Seeber, *Satellite geodesy: foundations, methods, and applications*. Walter de Gruyter, 2003.

[69] S. Shrestha, J. Talvitie, and E. S. Lohan, "Deconvolution-based indoor localization with WLAN signals and unknown access point locations," in *Proc. IEEE International Conference on Localization and GNSS*, Turin, Italy, June 2013.

[70] M. K. Simon and M. S. Alouini, *Digital communication over fading channels*, vol. 95. John Wiley & Sons, 2005.

[71] G. M. Siouris, *Chapter on Inertial Navigation of the Encyclopedia of Physical Science and Technology*, vol. 8, pp. 668-717. Academic Press, 1987.

[72] R. Sivaswamy, "Multiphase complementary codes," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 546-552, Sep. 1978.

[73] W. Stallings, *Wireless communications & networks*. Pearson Education India, 2009.

[74] B. A. Stein and E. Wheaton, "Graphic depiction of dynamic satellite constellation accuracy/coverage over time," *The Journal of Navigation*, vol. 34, no. 3, pp. 260-274, Aug. 2014.

[75] P. Stoica and A Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 5, pp. 720-741, May 1989.

[76] J. A. Stratton, *Electromagnetic theory*. John Wiley & Sons, 2007.

[77] G. L. Stüber, *Principles of mobile communication*. Springer Science & Business Media, 2011.

[78] J. Talvitie, and E. S. Lohan, "Modeling Received Signal Strength measurements for cellular network based positioning," in *Proc. IEEE International Conference on Localization and GNSS*, Italy, June 2013.

[79] J. Talvitie, E. S. Lohan, and M. Renfors, "The effect of coverage gaps and measurement inaccuracies in fingerprinting based indoor localization," in *Proc. International Conference on Indoor Positioning and Indoor Navigation*, Helsinki, Finland, June. 2014.

[80] J. Turkka, "Aspects of Knowledge Mining on Minimizing Drive Tests in Self-Organizing Cellular Networks," *Ph.D. Thesis*, ISBN 978-952-15-3401-0, Tampere University of Technology, Aug. 2014.

[81] J. Turkka, SHARING (deliverable D6.3): "Self-organized Heterogeneous Advanced Radio Networks Generation", Deliverable D6.3, European Celtic-Plus project 2015. [Online]. Available: http://www-sharing.cea.fr.

[82] M. C. Vanderveen, C. B. Papadias, and A. Paulraj, "Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array," *IEEE Communications Letters*, vol. 1, no. 1, pp. 12-14, Jan. 1997.

[83] A. Vlavianos, L. K. Law, I. Broustis, S. V. Krishnamurthy, and M. Faloutsos, "Assessing link quality in IEEE 802.11 wireless networks: which is the right metric?," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, France, Sep. 2008.

[84] X. Wu and Q. Du, "Utility-function-based radio-access-technology selection for heterogeneous wireless networks," *Computers & Electrical Engineering*, vol. 52, pp. 171–182, May 2015.

[85] J. Xiong and K. Jamieson, "ArrayTrack: a fine-grained indoor location system," in *Proc. USENIX Symposium on Networked Systems Design and Implementation*, Lombard, IL, USA, Apr. 2013.

[86] C. Yang and H. R. Shao, "WiFi-based indoor positioning," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 150-157, Mar. 2015.

[87] J. Zhu, "Calculation of geometric dilution of precision," *IEEE Transactions on Aerospace Electronic Systems*, vol. 28, no. 3, pp. 893-895, July 1992.

[88] E. Ziouva and T. Antonakopoulos, "CSMA/CA performance under high traffic conditions: throughput and delay analysis," *Computer Communications*, vol. 25, no. 3, pp. 313-321, Feb. 2002.