RSSI-BASED LOCALIZATION ALGORITHMS IN LORA NETWORKS

LAM KA HO

MPhil

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

RSSI-Based Localization Algorithms in LoRa Networks

Lam Ka Ho

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Philosophy

January 2018

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Lam Ka Ho (Name of student)

Abstract

RSSI-Based Localization Algorithms in LoRa Networks

Localization (positioning) is a very important research topic that has been used in many different applications. Many different wireless technologies such as Bluetooth and ZeeBee have been studied for use in localization in indoor environments. However, the most popular option for outdoor environments is satellite-based localization technology, and Global Positioning System (GPS) is the most popular satellite-based system. LoRa wireless technology has recently been proposed to support M2M (Machine-to-Machine) and IoT (Internet of Things) applications. The key features of LoRa are long range (up to 15 km), low power (five to six-year battery lifetime) and low cost (low cost chipsets and networks). These key features support LoRa technology in becoming an appropriate alternative (other than satellite-based localization technology) for localization in outdoor environments. Based on this new technology, we used Receiver Signal Strength Indicator (RSSI) to develop different localization algorithms using LoRa technology. To the best of our knowledge,

- We are among the first working on localization using LoRa technology;
- We are the first to develop RSSI-based localization algorithms in LoRa networks, and
- We are the first to handle blocking and multi-path (non-Gaussian noise) for localization in LoRa networks.

Different RSSI-based localization algorithms have been proposed to handle blocking and multi-path (non-Gaussian noise) in LoRa networks:

- RSSI-based LoRa Localization with K-mean Clustering (RLL-KC)

- RSSI-based LoRa Localization with Iterative Elimination (RLL-IE)
- RSSI-based LoRa Localization with Minimum MBRE (RLL-MM)
- RSSI-based LoRa Localization with Density-based Clustering (RLL-DC)

The first two algorithms are proposed to eliminate anchor node(s) that are highly affected by noise and then process the localization by using the remaining anchor nodes while the last two algorithms are proposed to select a set of anchor nodes that are not highly affected by noise and then use them to process the localization.

The performance of all proposed localization algorithms was investigated through simulations. We also developed real LoRa localization systems to investigate the performance of the proposed algorithms. Based on these performance investigations, we conclude that the performance of the proposed localization algorithms is much better than a very popular traditional localization algorithm in terms of localization error. Moreover, the performance of the proposed localization algorithms is similar to GPS, the most popular outdoor localization algorithm on system. Finally, the real LoRa localization systems show that the proposed localization algorithms work properly in both outdoor and large-scale indoor environments.

Publications of Candidate

- Ka-Ho Lam, Chi-Chung Cheung and Wah-Ching Lee,"LoRa-based localization systems for noisy outdoor environment", *The 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)(2017)*, pp. 278 - 284, October, 2017.

- Ka-Ho Lam, Chi-Chung Cheung and Wah-Ching Lee,"New RSSI-based LoRa localization Algorithms for Very Noisy Outdoor Environment", accepted by *COMPSAC 2018*, July, 2018.

- Ka-Ho Lam, Chi-Chung Cheung and Wah-Ching Lee,"RSSI-based LoRa localization systems for large-scale indoor and outdoor environments", submitted to *IEEE Transactions on Vehicular Technology*, January, 2018.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincerest gratitude to my chief supervisor, Dr. Wah-Ching Lee, for his guidance, enthusiasm, inspiration, encouragement and support during my studies and research. His help is important for me.

I would also like to express my sincere gratitude to my co-supervisor, Dr. Lawrence Chi-Chung Cheung, for his help in my studies and research. He gave me much advice on my research direction and thesis writing.

I would like to thank Department of Electronic and Information Engineering, The Hong Kong Polytechnic University for its financial support.

I would like to thank Industrial Center and Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University for their support in measurement, especially lending me the GNSS receiver.

I would like to thank my best friends, Miss Janet Oi-Yan Lee and Mr. Francis Ting-Kong Leung for their encouragement and support during my studies, and also for their help in carrying out the real experiments.

I would like to thank Mr. Sean Shensheng Xu and Mr. Alex Wah-Hong Lee for their help in carrying out the real experiments.

I would like to thank Dr. Rex G. Sharman for his proof-reading. Moreover, I would also like to thanks Cloudicty Co. Ltd. for their financial supports in building the LoRa-based localization system.

Finally, I would like to give special thanks to my parents for providing me with the opportunity for a great education as well as for their unfailing support. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

List of Figures **iii**

List of Tables **v**

Chapter 1: Introduction **1**

1.1 Definition of Localization 1

1.2 Location Estimation Approaches 1

1.3 Technologies Used in Localization 2

1.4 LoRa Technology 4

1.5 Previous Work 5

1.6 Clustering Techniques 6

1.7 Conclusions and Organization of the Thesis 7

Chapter 2: Problem Formulation **9**

2.1 System Model 9

2.2 Performance Measures 11

2.3 Optimization Model 13

2.4 Linear Least-Squares (LLS) Position Estimation Model for Localization and its Limitations 14

2.5 Conclusions 17

Chapter 3: RSSI-Based LoRa Localization Algorithms **18**

3.1 Backward RSSI Error (BRE) 18

3.2 RSSI-based LoRa Localization with K-mean Clustering (RLL-KC) . . 20
3.3 RSSI-based LoRa Localization with Iterative Elimination (RLL-IE) . 23
3.4 RSSI-based LoRa Localization with Minimum MBRE (RLL-MM) . . 24
3.5 RSSI-based LoRa Localization with Density-based Clustering (RLL-DC) 25
3.6 Numerical Examples 27
3.7 Conclusions 30

Chapter 4: Performance Investigation **31**
4.1 Design and Implementation of a LoRa-Based Localization System . . 31
4.2 Real Data Measurement 32
4.3 Simulation Model (with real data measurement) 33
4.4 Performance Comparisons 34
4.4.1 Performance Comparisons of RSSI-Based LoRa Localization Algorithms 34
4.4.2 The Effect of K, R and $R[0]$ on RSSI-Based LoRa Localization Algorithms 37
4.4.3 The Effect of The Number of Anchor Nodes on RSSI-Based LoRa Localization Algorithms 38
4.4.4 The Effect of Noise on RSSI-Based LoRa Localization Algorithms 39
4.4.5 Performance Comparisons of RSSI-Based LoRa Localization Algorithms in Real Experiments (Indoor and Outdoor) 40
4.5 Conclusions 42

Chapter 5: Conclusions and Future Work **56**
5.1 Conclusions 56
5.2 Future Work 58

Bibliography **59**

LIST OF FIGURES

2.1 Block diagram of localization using three anchor nodes. 13

3.1 The BRE values of different estimated target nodes. 19

4.1 Anchor Node 32
4.2 Target Node 32
4.3 The location of Tai Po in Google map. 44
4.4 The relationship between the measured distance and its measured RSSI in Tai Po. 44
4.5 The location of Kai Tak in Google map. 45
4.6 The relationship between the measured distance and its measured RSSI in Kai Tak. 45
4.7 The effect of K on the performance of RLL-KC. 46
4.8 The effect of K on the performance of RLL-IE. 46
4.9 The effect of R on the performance of RLL-KC. 47
4.10 The effect of R on the performance of RLL-MM. 47
4.11 The effect of R on the performance of RLL-DC. 48
4.12 The effect of $R[0]$ on the performance of RLL-IE. 48
4.13 The effect of $R[0]$ on the NCC of RLL-IE when $N(\mathbf{\Gamma}) = 12$. 49
4.14 The effect of the number of anchor nodes on the performance of LLS. 49
4.15 The effect of the number of anchor nodes on the performance of RLL-KC. 50
4.16 The effect of the number of anchor nodes on the performance of RLL-KC when fixed number of noisy anchor nodes. 50

4.17 The effect of the number of anchor nodes on the performance of RLL-IE. 51
4.18 The effect of the number of anchor nodes on the performance of RLL-MM. 51
4.19 The effect of the number of anchor nodes on the performance of RLL-DC. 52
4.20 The performance comparison of different algorithms in term of NCC. 52
4.21 The LER of all proposed algorithms with different values of NF. . . . 53
4.22 The location of the first experiment in Google map. 53
4.23 The location of the second experiment in Google map. 53
4.24 The outlook of the GPS Receiver. 54
4.25 The environment of the third experiment. 55

LIST OF TABLES

1.1 The specification comparison of different technologies uses for indoor localization. 3

1.2 The specification comparison of different technologies uses for outdoor localization. 4

1.3 Configuration for LoRa. 5

3.1 The information of anchor nodes 27

3.2 The information of K-mean clusters 28

3.3 The number of occurrences of each anchor node in Cluster 1 and 3 . . 28

3.4 The information of all iterations in RLL-IE 29

4.1 Parameter settings in our performance investigation 35

4.2 The performance comparison of all localization algorithms (NF = 0). 36

4.3 The performance comparison of all localization algorithms (NF = 0.4). 36

4.4 The performance comparison of all localization algorithms in the first experiment. 40

4.5 The coordinates of the anchor nodes in the first experiment. 41

4.6 The specification of the GPS receiver. 41

4.7 The performance comparison of all localization algorithms in the second experiment. 42

4.8 The coordinates of the anchor nodes in the second experiment. 42

4.9 The performance comparison of all localization algorithms in the third experiment. 43

4.10 The coordinates of the anchor nodes in the third experiment. 43

Chapter 1

INTRODUCTION

1.1 Definition of Localization

A localization (positioning) system is a system that provides the location (position) information of an object. It is a very important research topic and it has been used in many different applications such as navigation and tracking. Location information is one of the key factors in analysing human behavior and this information becomes more and more important in providing personalized services like banking and health-care.

1.2 Location Estimation Approaches

There are various location estimation approaches and the most popular ones are angle of arrival (AOA), time of arrival (TOA), time difference of arrival (TDOA) and received signal strength indicator (RSSI).

AOA is used to indicate the direction of the signal sources. The direction of the signal sources can be obtained by measuring the phase difference between received signals from two antennas. By knowing the direction of the signal sources in different sets of antennas (antennas arrays), a triangle can be built up by using two of them and the location of an object can be estimated by applying trigonometric formulas. AOA is effective in localization but it suffers from some problems like multi-path and non-uniform resolution. Moreover, each base station is required to have an equipment upgrade [1].

TOA (sometimes called time of flight (TOF)) is used to calculate the distance between a sender and a receiver by measuring the signal transmission time. The estimated position is the intersection of the three (or more) circles from the different distances. However, this approach requires the time synchronization among senders and receivers, which is very expensive [1–3].

TDOA is used to get the location by measuring the differences of the arrival times in different receivers. The estimated position is the intersection of at least two hyperbolas which requires three antennas. It is less expensive than TOA because it requires time synchronization among receivers only but it is still expensive [1–3]. Moverover, the accuracy of the localization is highly depended on the accuracy of the time difference.

RSSI is used to measure the strength of the received power signals. Then, based on the physical principle of the relationship between the signal strength and the distance between the sender and receiver, we can estimate the distances between the sender and different receivers, and hence locate an object by getting the interesction of the three (or more) circles from the different distances. It is low cost and easy to implementation. Based on this finding, our research focused on RSSI-based localization system [1, 4–10].

1.3 Technologies Used in Localization

For indoor location information, Bluetooth, ZigBee and WiFi are the most popular technologies used for localization [5, 11–18]. Bluetooth uses the Adaptive Frequency Hopping (AFH) technique to split the frequency band into different channels. Bluetooth Basic Rate / Enhanced Data Rate (BR/EDR) (which known as traditional bluetooth) have 79 channels with 1-MHz spacing each. It operates in a master/slave mode and one master can connect to at most seven active slave devices at the same time. Bluetooth Low Energy (LE) has 40 channels with 2-MHz each.

There is a unique broadcasting mode for Bluetooth LE to broadcast message to the nearby peripheral without master/salve connection. This can allow a Bluetooth LE device broadcast message to unlimited nearby peripheral. The specification of a popular Bluetooth module can be found in Table 1.1.

ZigBee uses the Direct Sequence Spread Spectrum (DSSS) for communication.It operates in three license free bands at 2.4 GHz, 915 MHz for North America and 868 MHz for Europe. The modulation technique is different in each band: Offset Quadrature Phase Shift Keying (O-QPSK) for 2.4 GHz band and Binary Phase Shift Keying (BPSK) for 915 and 868 MHz. The specification of a popular ZigBee module can be found in Table 1.1.

WiFi makes use of radio waves for communication across a network. There are four 802.11 standards for WiFi technology. 802.11a uses Orthogonal Frequency-Division Multiplexing (OFDM) to transmit data at 5 GHz and the maximum data rate is 54 Mbps. 802.11b and 802.11g operate, at 2.4 GHz but the data rate of 802.11g (54 Mbps) is much higher than 802.11b (11 Mbps). The data rate of 802.11n is the highest (140 Mbps) and it operates at 5 GHz. The specification of a popular WiFi module can also be found in Table 1.1.

Table 1.1: The specification comparison of different technologies uses for indoor localization.

	Bluetooth	ZigBee	WiFi
Manufacturer	Texas Instruments	Texas Instruments	STMicroelectronic
Module	CC2541	CC2531	SPWF01SA
Sensitivity	-90 dBm	-97 dBm	-95 dBm
Rx Power	17.9 mA	24 mA	105 mA
Tx Power	18.2 mA@0dBm	29 mA@1dBm	243 mA@10dBm
Max. Output Power	0dBm	+4.5dBm	+18.3dBm

From the above table, it is easy to find that all the above technologies cannot be used for outdoor localization because their coverage areas are small (5 to 20 meters).

For outdoor location information, cellular positioning and satellite-based positioning are the options. Cellular positioning was a very active research area before. It is assumed that a user should be connected to a network via the nearest base station. Thus the accuracy of the localization depends on the known range of its base station which may be a few hundred meters in ubran areas and serveral kilometers in rural areas. This accuracy is not acceptable in localization and therefore satellite-based positioning is more popular [19,20]. In satellite-based positioning, a global navigation satellite system provides location information to receivers. Among all satellite-based positioning, GPS (Global Positioning System) is the most popular because of its low cost relative to other navigation systems; it almost covers 100% of the earth and it is regularly updated. However, it cannot be used in indoor environments and its power consumption is high. The specification of a cellular positioning system and GPS can be found in Table 1.2. Based on this finding, we propose an alternative technology to do the localization for outdoor environments.

Table 1.2: The specification comparison of different technologies uses for outdoor localization.

	NB-IoT	GPS
Manufacturer	uBlox	uBlox
Module	SARA-N2	NEO-6m
Sensitivity	-135 dBm	-161 dBm
Rx Power	46 mA	39 mA
Tx Power	220 mA@23dBm	-

1.4 LoRa Technology

LoRa is one of the most important technologies used in Low-Power Wide-Area Network (LPWAN) to support M2M (Machine-to-Machine) and IoT (Internet of Things) applications. It uses the unlicensed spectrum (less than 1 GHz) with chirp

spread spectrum modulation to provide signal detection. The key features of LoRa are long range (up to 15 km), low power (five to six year battery lifetime) and low cost (low cost chipsets and networks). The detailed configuration for LoRa technology can be found in Table 1.3. These key features support LoRa technology as an appropriate alternative (other than satellite-based localization technology) for localization in outdoor environments. The detail information about LoRa Technology can be found in [2, 21–23].

Table 1.3: Configuration for LoRa.

Carrier frequency	437.5MHz (Channel 15)
Signal Bandwidth	125kHz
Output power	+20dBm
Antenna Gain	+2.15dBi

1.5 Previous Work

Since LoRa is a new technology and there are not many researchers working on using it for localization. Through detailed searching of the Internet, [2] is the only paper that works on localization using LoRa. It uses TDOA as the location estimation approach and carries out some real experiments in rural areas. The coverage area is a four-sided polygon around 2 to 3 km. The localization error in such experiments is over 1 km in some cases which is not very satisfactory performance in localization. Even in some cases the localization error is reduced to around 100 m, the number of packets transmitted for the localization is more than 10,000 and the total processing time is too long so that the localization procedure cannot be used in "real-time" [2,24].

One of the most important issues in LoRa localization is how to handle non-Gaussian noise like blocking and multi-path. In GPS, the blocking problem cannot be solved (thus it cannot be used for indoor environments) and the multi-path problem

can be partially solved only. In this thesis, we tackle these two problems effectively by using some mathematical models and clustering techniques.

1.6 Clustering Techniques

As mentioned before, our proposed algorithms will use K-mean clustering and dense clustering. This section describes the basic principle of these two clustering techniques.

In K-mean clustering [17], a set of data points (or vectors) $(a_1, a_2, ..., a_n)$ is given and they will be clustered into K sets $\mathbf{S} = \{S_1, S_2, ..., S_K\}$ to minimize the following objective function:

$$\underset{a}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{x \in S_k} \|a - \mu_k\|^2 \tag{1.1}$$

where μ_k is the centroid in S_k. The clustering algorithm is usually processed by using iterative procedures. At the beginning, an initial set of K centroids is randomly assigned. Then, each data point is assigned to a cluster which is nearest to its centroid. After that, the centroid of each cluster is updated based on the data points inside, i.e.,

$$\mu_k = \frac{1}{|S_k|} \sum_{a \in S_k} a. \tag{1.2}$$

These two steps will be repeatedly executed until the objective function in Equation (1.1) cannot be further reduced.

Dense clustering, which is used to identify dense regions (clusters) in the data space separated by low-density regions [25, 26], requires two important parameters: ϵ and $minPts$. A sample point is said to be a neighbor of another sample point if the distance between these two points is less than ϵ. A point is a core point if it has more than $minPts$ neighbors. Moreover, a point is a border point if it has less than $minPts$ neighbors. Finally, if a point is neither a core nor a border point, it is a noise point. In dense clustering, each sample point is classified as a core, border or noise

point. If it is a core point and it is not assigned to a cluster, we create a new cluster. Then we recursively find all its density connected points and assign them to the same cluster as the core point. If some points do not be assigned to any cluster, they are treated as a noise point.

1.7 Conclusions and Organization of the Thesis

In this chapter, we defined localization and discussed its importance in different applications. We briefly described different location estimation approaches such as AOA and RSSI. We also briefly described different technologies used in localization such as Bluetooth and GPS. After that, we introduced LoRa technology and previous work about localization in LoRa networks. Finally, we briefly described some clustering techniques that will be used in our proposed algorithms.

In the rest of this thesis, it is organized as follows:

Chapter 2 presents the problem formulation of a localization algorithm in LoRa networks. It first describes the system model of our research interest. Then it shows the performance measures of the system model. After that, it presents the optimization model of a localization algorithm. Finally, a well-known linear optimization model is shown (along with its limitations).

Chapter 3 shows our proposed RSSI-based localization algorithms. At the beginning, a new approach is described to identify whether the estimated location of a target node is good or not. Then, by making use of this new approach, four new proposed localization algorithms are presented in detail.

Chapter 4 presents the performance investigation of our proposed localization algorithms. We first compare the performance of our algorithms with the well known linear optimization model and GPS through some simulation results. Then we investigate the effect of different parameters on the performance of different algorithms. Finally, the performance comparisons in some real experiments including

outdoor and large-scale indoor environments are described.

Chapter 5 gives a brief conclusion for all our contributions and summarizes the directions for the future works.

Chapter 2

PROBLEM FORMULATION

Section 2.1 of this chapter defines the system model of our research interest. Section 2.2 describes the performance measures in localization algorithms, and Section 2.3 shows the localization optimization model. Section 2.4 describes the well-known linear optimization model called Linear Least-Squares (LLS) used in localization. Its limitations will also be discussed here.

2.1 System Model

A two-dimensional network is used to represent a scenario for localization. A location in the two-dimensional network is defined as a point $p = (x, y)$ where x and y are the x and y-coordinates of the point p respectively. A network has two kinds of nodes. One kind of nodes are target nodes. They are the nodes that we would like to estimate their locations. There may be more than one target node for localization. For simplicity, in our research, we estimated the location of one target node only, i.e., $p_t = (x_t, y_t)$. Note that, with a sophisticated communication protocol, our proposed localization algorithms can estimate more than one target node at a time. Furthermore, the target node sends signals to other nodes for localization. The other kind of nodes are anchor nodes. They are the receivers to receive signals from the target node. In this system model, we define $\mathbf{\Gamma}$ as the set of anchor nodes such that $\mathbf{\Gamma} = \{p_1^{(\mathbf{\Gamma})}, p_2^{(\mathbf{\Gamma})}, \ldots, p_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})}\}$ where $p_n^{(\mathbf{\Gamma})} = (x_n^{(\mathbf{\Gamma})}, y_n^{(\mathbf{\Gamma})})$ is the n^{th} anchor node in $\mathbf{\Gamma}$ and $N(\mathbf{\Gamma})$ is the total number of anchor nodes in $\mathbf{\Gamma}$. In our research, Received Signal Strength Indicator (RSSI) is a measurement of the power received by an anchor node.

It is used to estimate the distance between the target node and an anchor node. The estimation is based on the free space propagation model [1, 11, 27]:

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2, \tag{2.1}$$

where P_R is the received power in mW, P_T is the transmitted power in mW, G_T is the antenna gain for the sender, G_R is the antenna gain for the receiver, λ is the wavelength of the carrier and d is the distance between the sender and the receiver. Let P_0 and d_0 be the reference received power and distance respectively, then we have

$$P_0 = P_T G_T G_R \left(\frac{\lambda}{4\pi d_0}\right)^2. \tag{2.2}$$

From Equation (2.1) and (2.2), we have

$$\frac{P_R}{P_0} = \left(\frac{d_0}{d}\right)^2. \tag{2.3}$$

Equation (2.3) can be represented by RSSI, i.e.,

$$Z = Z_0 - 20\log_{10}\left(\frac{d}{d_0}\right), \tag{2.4}$$

where Z is the RSSI value measured by the receiver ($Z = 10\log_{10}\left(\frac{P_R}{1mW}\right)$) and Z_0 is the reference RSSI value measured at the reference distance d_0 ($Z_0 = 10\log_{10}\left(\frac{P_0}{1mW}\right)$). In this system model for localization, we define $\tilde{Z}(p_1, p_2)$ and $d(p_1, p_2)$ as the measured RSSI value and the distance between two points p_1 and p_2 respectively. Moreover, in real measurement, the reference RSSI value is measured for one meter, i.e., $d_0 = 1m$ and $Z_0 = \tilde{Z}_0$. Furthermore, the constant 20 in Equation (2.4) should be replaced by $10\tilde{\alpha}$ where $\tilde{\alpha}$ is the measured path loss exponent to represent path loss in such environments. Usually $\tilde{\alpha}$ should be between 2 and 5, and it is equal to 2 in free space. Finally, the noise ω should be considered in the real measurement and usually ω is

represented by a zero-mean Gaussian random variable. Thus, we now have

$$\tilde{Z}(p_1, p_2) = \tilde{Z}_0 - 10\tilde{\alpha} \log_{10} d(p_1, p_2) + \omega, \tag{2.5}$$

where

$$d(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \tag{2.6}$$

$p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$. In localization, we define $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$ as a set of all measured RSSI values from the target node to each anchor node in $\mathbf{\Gamma}$ with the reference measured RSSI values and the measured path loss exponent, i.e.,

$$\tilde{\mathbf{Z}}(\mathbf{\Gamma}) = \left\{ \tilde{Z}(p_t, p_1^{(\mathbf{\Gamma})}), \tilde{Z}(p_t, p_2^{(\mathbf{\Gamma})}), \ldots, \tilde{Z}(p_t, p_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})}), \tilde{Z}_0, \tilde{\alpha} \right\}. \tag{2.7}$$

2.2 Performance Measures

This subsection introduces the parameters used to measure the performance of a localization algorithm. The most important performance measure in localization is called Localization Error (LE). LE is defined as the distance between the estimated location and the actual location of the target node (units in meters). When LE of a localization algorithm is small, the performance (accuracy) of the algorithm is good, and vice versa. Additionally, since Linear Least-Squares (LLS) position estimation model is the most popular model for localization (it will be described later in this chapter), we will compare the performance of our proposed localization algorithms with LLS in terms of LE. To show the performance comparison with LLS more effectively, we introduce a performance measure called Localization Error Ratio (LER). LER is defined as the ratio of the LE of a proposed localization algorithm to that of LLS. If the LER of a proposed localization algorithm is large, it means its performance is much better than LLS in terms of LE, and vice versa.

Another important performance measure in localization is computational

complexity. It is an important factor in determining whether we can get the location in real-time or not. Moreover, a localization algorithm usually takes more computation time for calculation to get a better estimation in localization. Later, in the Performance Investigation (Chapter 4), we will show the computational complexity of all proposed localization algorithms is still low so that they can be used in real-time applications. Thus we will focus on LE only in the problem formulation.

Based on the above observation, we propose to use the following two performance measures to investigate the performance of a localization algorithm:

- Localization Error (LE): The LE of a localization algorithm X is defined as,

$$LE(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})) = d\left(p_t, \hat{p}_t^{(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))}\right) \tag{2.8}$$

  where $\hat{p}_t^{(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))}$ is the estimated location of the target node by using the localization algorithm X in a set of anchor nodes $\mathbf{\Gamma}$ with its measured RSSI values $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$, i.e.,

$$\hat{p}_t^{(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))} = X(\Gamma, \tilde{\mathbf{Z}}(\mathbf{\Gamma})). \tag{2.9}$$

  Note that the localization error is a positive floating point number.

- Localization Error Ratio(LER): The ratio of the LE of a localization algorithm X to LLS is defined as,

$$LER(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})) = \frac{LE(LLS, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))}{LE(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))}. \tag{2.10}$$

- Normalized Computation Complexity (NCC): For ease of comparison, we normalize the computational complexity by LLS in localization. This case will be described in the Performance Investigation (Chapter 4).

2.3 Optimization Model

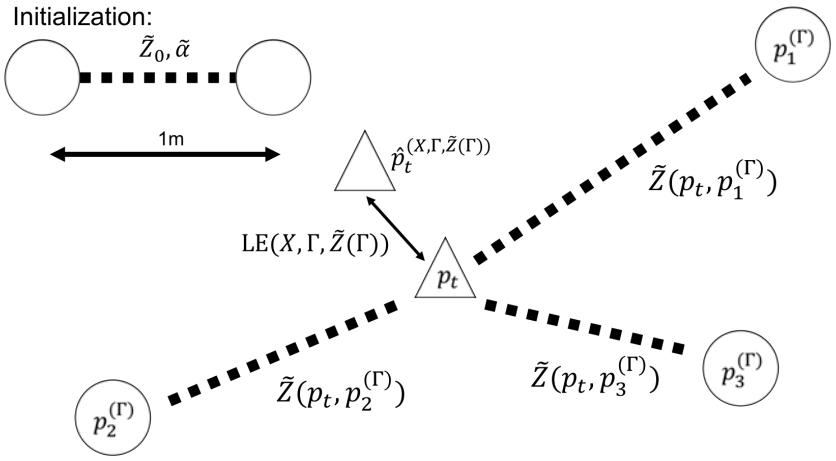


Figure 2.1: Block diagram of localization using three anchor nodes.

Fig. 2.1 shows the block diagram for real measurement localization by using three anchor nodes and the optimization model for localization is formulated as follows:

Given a set of anchor nodes $\mathbf{\Gamma}$ with their locations, and a set of measured RSSI values of all anchor nodes in $\mathbf{\Gamma}$ with the measured reference RSSI values and the measured path loss exponent (i.e., $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$),

Find the optimal localization algorithm X^* that minimizes the localization error,

$$LE(X^*, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})) = \min_X LE(X, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})). \tag{2.11}$$

Three points need to be clarified. (1) Localization is dynamic but not static, meaning a localization algorithm can locate the target node even if it is moving and the estimation can be updated from time to time. Although the measured RSSI value

of a moving object may be different to that of a fixed object, the overall localization procedure should still be valid for an moving object. (2) The next subsection describes a mathematical localization model. The mathematical model includes the calculation to inverse a matrix. If all anchor nodes lie on the same straight line, the determinant of the matrix will become zero and thus there will be no solution to get the inverse of the matrix. Therefore, all anchor nodes must not be located on the same straight line. (3) In real measurement, some errors (e.g., random noise, blocking and multi-path) may exist and thus some measured RSSI values may be wrong due to the above errors. This situation cannot be avoided; thus, this possibility should not be excluded in the optimizations. It means we expect that the optimal localization algorithm should take care of this situation but still can minimize the localization error.

2.4 Linear Least-Squares (LLS) Position Estimation Model for Localization and its Limitations

The Linear Least-Squares (LLS) position estimation model is the most popular model for localization [1, 4, 28] because it provides a close-form solution and its computational complexity is low.

From Equation (2.5), we have

$$\tilde{Z}(p_t, p_n^{(\mathbf{\Gamma})}) = \tilde{Z}_0 - 10\tilde{\alpha}\log_{10} d(p_t, p_n^{(\mathbf{\Gamma})}) + \omega, \tag{2.12}$$

where

$$d(p_t, p_n^{(\mathbf{\Gamma})}) = \sqrt{(x_t - x_n^{(\mathbf{\Gamma})})^2 + (y_t - y_n^{(\mathbf{\Gamma})})^2}, \tag{2.13}$$

and $n = 1, 2, ..., N(\mathbf{\Gamma})$.

For simplicity, the random noise ω is neglected and we assume that it is a noiseless

environment. By combining these two equations, we have

$$-2x_t x_n^{(\mathbf{\Gamma})} - 2y_t y_n^{(\mathbf{\Gamma})} + (x_t^2 + y_t^2) = 10^{\frac{2}{10\tilde{\alpha}}\left[\tilde{Z}_0 - \tilde{Z}(p_t, p_n^{(\mathbf{\Gamma})})\right]} - \left[\left(x_n^{(\mathbf{\Gamma})}\right)^2 + \left(y_n^{(\mathbf{\Gamma})}\right)^2\right], \quad (2.14)$$

for $n = 1, 2, ..., N(\mathbf{\Gamma})$. In a vector form, we have

$$\mathbf{A}\theta = \mathbf{b} \quad (2.15)$$

where

$$\mathbf{A} = \begin{bmatrix} -2x_1^{(\mathbf{\Gamma})} & -2y_1^{(\mathbf{\Gamma})} & 1 \\ -2x_2^{(\mathbf{\Gamma})} & -2y_2^{(\mathbf{\Gamma})} & 1 \\ \vdots & \vdots & \vdots \\ -2x_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})} & -2y_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})} & 1 \end{bmatrix}, \quad (2.16)$$

$$\theta = \begin{bmatrix} x_t \\ y_t \\ x_t^2 + y_t^2 \end{bmatrix} \quad (2.17)$$

and

$$\mathbf{b} = \begin{bmatrix} 10^{\frac{2}{10\tilde{\alpha}}\left[\tilde{Z}_0 - \tilde{Z}(p_t, p_1^{(\mathbf{\Gamma})})\right]} - \left[\left(x_1^{(\mathbf{\Gamma})}\right)^2 + \left(y_1^{(\mathbf{\Gamma})}\right)^2\right] \\ 10^{\frac{2}{10\tilde{\alpha}}\left[\tilde{Z}_0 - \tilde{Z}(p_t, p_2^{(\mathbf{\Gamma})})\right]} - \left[\left(x_2^{(\mathbf{\Gamma})}\right)^2 + \left(y_2^{(\mathbf{\Gamma})}\right)^2\right] \\ \vdots \\ 10^{\frac{2}{10\tilde{\alpha}}\left[\tilde{Z}_0 - \tilde{Z}(p_t, p_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})})\right]} - \left[\left(x_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})}\right)^2 + \left(y_{N(\mathbf{\Gamma})}^{(\mathbf{\Gamma})}\right)^2\right] \end{bmatrix}. \quad (2.18)$$

By applying Linear Least-Squares (LLS) method [29], we have

$$\hat{\theta} = \mathbf{A}^{\dagger}\mathbf{b} \quad (2.19)$$

where

$$\hat{\theta} = \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \\ (\hat{x}_t)^2 + (\hat{y}_t)^2 \end{bmatrix} \tag{2.20}$$

and

$$\mathbf{A}^{\dagger} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}\right)^{-\mathbf{1}}\mathbf{A}^{\mathbf{T}}. \tag{2.21}$$

Note that $\hat{p}_t^{(LLS,\mathbf{\Gamma},\tilde{\mathbf{Z}}(\mathbf{\Gamma}))} = (\hat{x}_t, \hat{y}_t)$ where $\hat{p}_t^{(LLS,\mathbf{\Gamma},\tilde{\mathbf{Z}}(\mathbf{\Gamma}))}$ is the estimated location of the target node by using LLS with $\mathbf{\Gamma}$ and $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$. To make it short, we formulate the whole process as below:

$$\hat{p}_t^{(LLS,\mathbf{\Gamma},\tilde{\mathbf{Z}}(\mathbf{\Gamma}))} = LLS\left(\mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})\right). \tag{2.22}$$

Although the LLS method was developed to estimate the location of the target node in noiseless environments, it can also handle estimation in a noisy environment if it is not too noisy (e.g., Gaussian noise only). In noiseless environments, the solution obtained from Equation (2.22) can satisfy all equations in Equation (2.12), i.e., $\hat{p}_t^{(LLS,\mathbf{\Gamma},\tilde{\mathbf{Z}}(\mathbf{\Gamma}))}$ is exactly equal to p_t. However, in noisy environments, the noise affects the measurement and the measured RSSI values cannot be exactly equal to the expected values. Therefore, there is no single solution which can exactly satisfy all equations in Equation (2.12) but LLS can give the best estimation, i.e., $\hat{p}_t^{(LLS,\mathbf{\Gamma},\tilde{\mathbf{Z}}(\mathbf{\Gamma}))}$ can minimize the differences in Equation (2.12). Therefore, LLS can still be applied to get the best estimation in real measurement.

However, if the environment is too noisy (e.g., non-Gaussian noise like blocking and multi-path) such that the measured RSSI value(s) of one or more anchor node(s) is/are far away from the expected value(s), the LLS solution will be not close enough to the real location of the target node. Consider the following example: there are four anchor nodes and their locations are $p_1 = (0, 0)$, $p_2 = (0, 100)$, $p_3 = (100, 100)$ and $p_4 = (100, 0)$ (units in meter). The location of the target node is $p_t = (35, 60)$. Given that $\tilde{\alpha}$ and $\tilde{Z}_0$ are 2.5 and -25 dB respectively, the mean measured RSSI

readings from these four anchor nodes are -71.04 dB, -68.14 dB, -72.07 dB and -73.67 dB respectively. Based on this setting, LLS gives the estimated location as $\hat{p}_t^{(LLS,\mathbf{\Gamma},\tilde{\mathbf{Z}}(\mathbf{\Gamma}))} = (39.45, 58.49)$, which is quite close to the actual location (localization error = 4.70 m). However, if p_4 is affected by the noise and the measured RSSI reading is -82.63 dB but not -73.67 dB, the estimated location becomes $(46.30, 144.24)$, which is very far away from the actual location (localization error = 117.08 m). If p_4 can be identified as being strongly affected by the noise and it is removed from the localization (i.e., it means we use p_1, p_2 and p_3 only to estimate the location of the target node), the estimated location becomes $(27.69, 70.26)$ and this estimation is still acceptable good (localization error = 12.59 m). This example shows that, if the environment is too noisy, LLS is not good enough and a better solution can be obtained if some anchor nodes that are strongly affected by the noise can be removed during localization, or some anchor nodes that are not strongly affected by the noise can be identified to process the localization. Based on this observation, in this thesis, we propose two new algorithms for the elimination of some anchor nodes during localization and two new algorithms for selecting some anchor nodes to process localization.

2.5 Conclusions

In this chapter, we presented the system model and the problem measures for this model. Then we formulated the optimization model. Finally, we briefly described the LLS position estimation model for localization and discussed its limitations.

Chapter 3

RSSI-BASED LORA LOCALIZATION ALGORITHMS

In this chapter, we propose four new RSSI-based LoRa Localization algorithms to (a) eliminate some anchor nodes that are strongly affected by the noise or (b) select some anchor nodes which are not strongly affected by the noise for localization purposes. For simplicity, we declare an anchor node is "good" if the effect of the noise on the measured RSSI value from the target node to this node is small. Conversely, an anchor node is "bad" if the effect is strong. To identify whether an anchor node is "good" or "bad", we use the concept of Backward RSSI Error (BRE), which will be described in Section 3.1. The four new algorithms will be described afterwards in Section 3.2 to 3.5.

3.1 Backward RSSI Error (BRE)

The advantage of using the LLS approach is that it fully utilizes all information in the measurement (the path loss exponent, the reference measured RSSI value, the location of anchor nodes and their corresponding measured RSSI values) and derives the optimal solution from it. However, it may receive some wrong information from "bad" anchor nodes and thus may sometimes obtain a solution far away to the optimal one. Thus, we propose using a heuristic approach to identify the "good" and "bad" anchor nodes. Consider $Z(\hat{p}_t, p_n)$ as the calculated RSSI value between the estimated location of the target node $\hat{p}_t$ and the anchor node p_n, i.e.,

$$Z(\hat{p}_t, p_n) = \tilde{Z}_0 - 10\tilde{\alpha} \log_{10} d(\hat{p}_t, p_n). \tag{3.1}$$

Additionally, $\tilde{Z}(p_t, p_n)$ as the measured RSSI value between the actual location of the target node p_t and the anchor node p_n. Then we define Backward RSSI Error (BRE) of the estimated target node $\hat{p}_t$ and the anchor node p_n as the absolute difference between these two values i.e.,

$$BRE(\hat{p}_t, p_n) = \left| Z(\hat{p}_t, p_n) - \tilde{Z}(p_t, p_n) \right|. \tag{3.2}$$

Obviously, if $\hat{p}_t$ is close to p_t, $BRE(\hat{p}_t, p_n)$ is close to zero. Thus, heuristically, if $\hat{p}_t$ is not too far away from p_t, $BRE(\hat{p}_t, p_n)$ can show how close they are. Fig. 3.1 shows the relationship between p_t, $\hat{p}_t$, p_n and $\hat{p}_{t'}$ (its location is far away from p_t but $Z(\hat{p}_{t'}, p_n) \sim \tilde{Z}(p_t, p_n)$). For $\hat{p}_t$, since it is quite close to p_t, the value of $BRE(\hat{p}_t, p_n)$ is highly related to the error of the estimation. However, for $\hat{p}_{t'}$, the value of $BRE(\hat{p}_{t'}, p_n)$ is meaningless (even it is equal to zero) because $\hat{p}_{t'}$ and p_t are far away from each other.

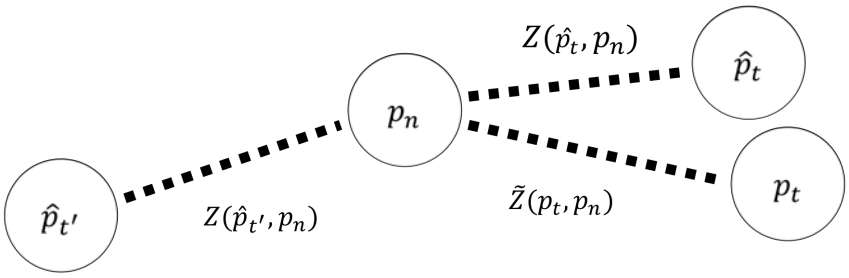


Figure 3.1: The BRE values of different estimated target nodes.

Based on the above heuristic, we propose using Mean BRE (MBRE) to heuristically evaluate the accuracy of the estimated location of the target node $\hat{p}_t$ by using a set of anchor nodes $\mathbf{\Gamma}$ and their measured RSSI values $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$. We define $MBRE(\hat{p}_t, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))$ as the MBRE of the estimated location of the target node $\hat{p}_t$ in

$\mathbf{\Gamma}$ with $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$, i.e.,

$$MBRE(\hat{p}_t, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})) = \frac{1}{N(\mathbf{\Gamma})} \sum_n BRE(\hat{p}_t, p_n^{(\mathbf{\Gamma})}). \tag{3.3}$$

Let $\hat{p}_t^{(X)}$ and $\hat{p}_t^{(Y)}$ be the estimated locations of the target node by using two different approaches X and Y respectively. If $MBRE(\hat{p}_t^{(X)}, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma})) < MBRE(\hat{p}_t^{(Y)}, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))$, it means the approach X gives a better estimation of the target node than the approach Y based on the information provided by $\mathbf{\Gamma}$ and $\tilde{\mathbf{Z}}(\mathbf{\Gamma})$. MBRE is a good heuristic because the MBRE of an estimated location will be small only when the estimated location is not far away from the actual location; otherwise, at least one anchor node will give a large value for its BRE and thus the overall MBRE will also be large. Note that MBRE is still a heuristic because, occasionally, when $MBRE(\hat{p}_t^{(X)}, \mathbf{\Gamma}, \tilde{\mathbf{Z}}(\mathbf{\Gamma}))$ is close to zero, $\hat{p}_t^{(X)}$ may not be very close to p_t. It happens when the measurement error in $\tilde{Z}(p_t, p_n^{(\mathbf{\Gamma})})$ is compensated by the error in $Z(\hat{p}_t^{(X)}, p_n^{(\mathbf{\Gamma})})$ (some parameters used in Equation (3.1) are based on the measurement).

3.2 RSSI-based LoRa Localization with K-mean Clustering (RLL-KC)

In this subsection, we propose a new LoRa localization algorithm called RSSI-based LoRa Localization with K-mean Clustering (RLL-KC). It uses K-mean clustering to eliminate a "bad" anchor node and process the localization by using the remaining of anchor nodes. K-mean clustering is an unsupervised learning to group sample points (that are estimated locations of the target nodes in this research) into different clusters [17]. It works iteratively to assign each sample point into K clusters. At the beginning in this algorithm, we generate M estimated locations of the target nodes by using LLS in M different sets of anchor nodes. Then, through K-mean clustering, we group them into K clusters and locate the best cluster by using MBRE. After that, we check the number of occurrences of each anchor node

in all clusters except the best cluster (we call them "other clusters" elsewhere in the thesis). We believe that the measured RSSI values of some anchor nodes are affected by the noise and they should be found in "other clusters". Furthermore, the number of occurrences of an anchor node in "other clusters" should be related to how much it is affected by the noise and we claim that an anchor node can be declared to be "bad" if it has the largest number of occurrences. Finally, the "bad" anchor node is removed from the set of anchor nodes and LLS is processed again to estimate the target node. Through this procedure, the accuracy of the localization is expected to be improved because the effect of the noise is reduced significantly. The description of the algorithm is shown below:

- **Step 1**: Estimate M locations of the target node by using M different sets of anchor nodes. Later we will use such locations to identify the "bad" anchor node. We define $\mathbf{AN}$ and $\mathbf{AN}_m$ as a set of all anchor nodes and a subset of anchor nodes, respectively, such that

$$\mathbf{AN} = \left\{ p_1^{(\mathbf{AN})}, p_2^{(\mathbf{AN})}, \ldots, p_{N(\mathbf{AN})}^{(\mathbf{AN})} \right\}, \tag{3.4}$$

and

$$\mathbf{AN}_m = \left\{ p_1^{(\mathbf{AN}_m)}, p_2^{(\mathbf{AN}_m)}, \ldots, p_R^{(\mathbf{AN}_m)} \right\}, \tag{3.5}$$

where $N(\mathbf{AN}_m) = R$ (R is a constant for all m and $R \geq 3$), $\mathbf{AN}_m \subset \mathbf{AN}$ for all m and $\bigcup_m \mathbf{AN}_m = \mathbf{AN}$ for $m = 1, 2, \ldots, M$. Thus, we have

$$\hat{p}_t^{(LLS, \mathbf{AN}_m, \tilde{\mathbf{Z}}(\mathbf{AN}_m))} = LLS\left(\mathbf{AN}_m, \tilde{\mathbf{Z}}(\mathbf{AN}_m)\right), \tag{3.6}$$

and

$$\hat{\mathbf{P}}_{\mathbf{t}}^{(\mathbf{M})} = \left\{ \hat{p}_t^{(LLS, \mathbf{AN}_m, \tilde{\mathbf{Z}}(\mathbf{AN}_m))} \right\} \tag{3.7}$$

for $m = 1, 2, \ldots, M$.

- **Step 2**: Use K-mean clustering to group M locations into $K^{(KC)}$ clusters and find the centroid of each cluster. We define $\mathbf{C^{(KC)}}$ as a set of all clusters such that

$$\mathbf{C^{(KC)}} = \left\{ \mathbf{C}_1^{\mathbf{(KC)}}, \mathbf{C}_2^{\mathbf{(KC)}}, \ldots, \mathbf{C}_{K^{(KC)}}^{\mathbf{(KC)}} \right\} = KC\left(\hat{\mathbf{P}}_{\mathbf{t}}^{\mathbf{(M)}} \right) \tag{3.8}$$

where $\mathbf{C}_k^{\mathbf{(KC)}}$ is a set of $\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))}$ generated by K-mean clustering method (i.e., KC) for $k = 1, 2, \ldots, K^{(KC)}$. Note that $\mathbf{C}_k^{\mathbf{(KC)}} \subset \hat{\mathbf{P}}_t^{\mathbf{(M)}}$, $\bigcup_k \mathbf{C}_k^{\mathbf{(KC)}} = \hat{\mathbf{P}}_t^{\mathbf{(M)}}$ and $\mathbf{C}_j^{\mathbf{(KC)}} \cap \mathbf{C}_k^{\mathbf{(KC)}} = \phi$ if $j \neq k$. Thus, we have

$$p_c^{(KC,k)} = E\left(\mathbf{C}_k^{\mathbf{(KC)}} \right) \tag{3.9}$$

where E is the expected mean and $p_c^{(KC,k)}$ is the centroid of the k^{th} cluster. It is expected that the centroid of a cluster represents the cluster and its MBRE is strongly related to the effect of the noise on the anchor nodes in such a cluster.

- **Step 3**: Find the best cluster by examining the MBRE of the center of each cluster and locate the "bad" anchor node by checking the number of occurrences of all anchor nodes in "other clusters". We define that a cluster is the best if its center has the smallest MBRE among all centers, i.e.,

$$p_c^{(KC,k^*)} = \min_k MBRE\left(p_c^{(KC,k)}, \mathbf{AN}, \tilde{\mathbf{Z}}(\mathbf{AN}) \right). \tag{3.10}$$

We assume that there is/are no "bad" anchor node(s) in the best cluster or the number of occurrences of "bad" anchor nodes in this cluster is small. Thus, the number of occurrences of "bad" anchor nodes in "other clusters" should be larger. We define

$$O(p_n) = \sum_{k \neq k^*} O^{(k)}(p_n) \tag{3.11}$$

where $O^{(k)}(p_n)$ is the number of occurrences of the n^{th} anchor node in the k^{th} cluster. We find an anchor node with the largest number of occurrences in

"other clusters" and we claim that anchor node is the "bad" anchor node, i.e.,

$$O(p_{n^*}) = \max_n O(p_n) \tag{3.12}$$

- **Step 4**: Estimate the location of the target node again but this time the "bad" anchor node p_{n^*} is removed from the set of all anchor nodes, i.e.,

$$\hat{p}_t^{(KC,\mathbf{AN},\tilde{\mathbf{Z}}(\mathbf{AN}))} = \hat{p}_t^{(LLS,\mathbf{AN}^*,\tilde{\mathbf{Z}}(\mathbf{AN}^*))} = LLS\left(\mathbf{AN}^*, \tilde{\mathbf{Z}}(\mathbf{AN}^*)\right) \tag{3.13}$$

where $\mathbf{AN}^* = \mathbf{AN} \setminus \{p_{n^*}\}$.

3.3 RSSI-based LoRa Localization with Iterative Elimination (RLL-IE)

In this subsection, we propose another LoRa localization algorithm called RSSI-based LoRa localization with Iterative Elimination (RLL-IE) to eliminate more "bad" anchor nodes by using RLL-KC iteratively. In RLL-KC, only one "bad" anchor node is eliminated to improve the accuracy of the estimation. However, if there is more than one "bad" anchor node, RLL-KC is not good enough to eliminate all of them. To address this issue, RLL-IE applies RLL-KC iteratively to eliminate "bad" anchor nodes until no more anchor nodes can be removed and thus this algorithm is good if the effect of the noise is high for many anchor nodes. The description of RLL-IE is shown below:

- **Step 1**: Process RLL-KC iteratively to get the new set of anchor nodes and get the new estimated location of the target node until the number of anchor nodes in each subset is three, i.e.,

$$\left(\hat{p}_t^{(KC,\mathbf{AN[i+1]},\tilde{\mathbf{Z}}(\mathbf{AN[i+1]}))}, \mathbf{AN[i+1]}\right) = RLL\text{-}KC\left(\mathbf{AN[i]}, \tilde{\mathbf{Z}}(\mathbf{AN[i]})\right), \tag{3.14}$$

where $\mathbf{AN[i+1]} = \mathbf{AN[i]} \setminus \{p_{n^*}\}$ and $\mathbf{AN[0]} = \mathbf{AN}$. Note that R in $\mathbf{AN[i]}$ is $R[i]$. Moreover, $R[i+1] = R[i] - 1$. The setting of $R[0]$ will be discussed later in the Performance Investigation (Chapter 4). The above iteration will be repeated until $i = R[0] - j$ (it means $R[i] = j$) where $3 \leq j \leq N(\Gamma) - 1$.

- **Step 2**: Select the best estimated location of the target node among all $\hat{p}_t^{(\mathbf{AN[i]})}$ by using MBRE, i.e.,

$$\hat{p}_t^{(KC,\mathbf{AN[i^*]},\tilde{\mathbf{Z}}(\mathbf{AN[i^*]}))} = \min_i MBRE\left(\hat{p}_t^{(KC,\mathbf{AN[i]},\tilde{\mathbf{Z}}(\mathbf{AN[i]}))}, \mathbf{AN}, \tilde{\mathbf{Z}}(\mathbf{AN})\right) \quad (3.15)$$

and

$$\hat{p}_t^{(IE,\mathbf{AN},\tilde{\mathbf{Z}}(\mathbf{AN}))} = \hat{p}_t^{(KC,\mathbf{AN[i^*]},\tilde{\mathbf{Z}}(\mathbf{AN[i^*]}))}. \quad (3.16)$$

3.4 RSSI-based LoRa Localization with Minimum MBRE (RLL-MM)

In this subsection, we propose a new LoRa localization algorithm called RSSI-based LoRa localization with Minimum MBRE (RLL-MM) to select the best set of anchor nodes based on MBRE. The idea is simple: we get M locations of the target node by using M different sets of anchor nodes and use MBRE to find the best among all of them.

- **Step 1**: Estimate M locations of the target node by using M different sets of anchor nodes (Same as Step 1 in RLL-KC). Thus we have

$$\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))} = LLS\left(\mathbf{AN}_m, \tilde{\mathbf{Z}}(\mathbf{AN}_m)\right), \quad (3.17)$$

and

$$\hat{\mathbf{P}}_\mathbf{t}^{(\mathbf{M})} = \left\{\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))}\right\} \quad (3.18)$$

for $m = 1, 2, \ldots, M$.

- **Step 2**: Use MBRE to find the best among them, i.e.,

$$\hat{p}_t^{(LLS,\mathbf{AN}_{m^*},\tilde{\mathbf{Z}}(\mathbf{AN}_{m^*}))} = \min_m MBRE\left(\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))}, \mathbf{AN}, \tilde{\mathbf{Z}}(\mathbf{AN})\right) \quad (3.19)$$

and

$$\hat{p}_t^{(MM,\mathbf{AN},\tilde{\mathbf{Z}}(\mathbf{AN}))} = \hat{p}_t^{(LLS,\mathbf{AN}_{m^*},\tilde{\mathbf{Z}}(\mathbf{AN}_{m^*}))}. \quad (3.20)$$

3.5 RSSI-based LoRa Localization with Density-based Clustering (RLL-DC)

In this subsection, we propose a new LoRa localization algorithm by using Density-based spatial clustering of applications with noise (DBSCAN) to identify "good" anchor nodes and process the localization using them. DBSCAN groups sample points into different clusters such that clusters are dense regions in the data space separated by low-density regions [25, 26]. Moreover, some outliers will be excluded from the dense regions and they will not be clustered, which is different to what we have in K-mean clustering. We assume that, in localization, the estimated locations of the target nodes (by using "good" anchor nodes) should be similar and they should be grouped into a cluster by using density-based clustering. Furthermore, we assume that the number of occurrences of an anchor nodes in such a cluster indicates how "good" an anchor node is. It means that the number of occurrences of an anchor node is found in the cluster, the greater possibility that the anchor node is "good". Based on this heuristic, we select the best three anchor nodes from each cluster by examining their number of occurrences, and then use them to estimate the location of the target node. Finally, we use MBRE to get the best estimation among all of them. This algorithm is good if the effect of the noise is strong for many anchor nodes and it is difficult to identify "good" anchor nodes. The description of the algorithm is shown below:

- **Step 1**: Estimate M locations of the target node by using M different sets of anchor nodes (Same as Step 1 in RLL-KC). Thus we have

$$\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))} = LLS\left(\mathbf{AN}_m, \tilde{\mathbf{Z}}(\mathbf{AN}_m)\right), \tag{3.21}$$

and

$$\hat{\mathbf{P}}_{\mathbf{t}}^{(\mathbf{M})} = \left\{\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))}\right\} \tag{3.22}$$

for $m = 1, 2, \ldots, M$.

- **Step 2**: Use density-based clustering to group M locations into $K^{(DC)}$ clusters. We define $\mathbf{C}^{(\mathbf{DC})}$ as a set of all clusters such that

$$\mathbf{C}^{(\mathbf{DC})} = \left\{\mathbf{C}_1^{(\mathbf{DC})}, \mathbf{C}_2^{(\mathbf{DC})}, \ldots, \mathbf{C}_{K^{(DC)}}^{(\mathbf{DC})}\right\} = DC\left(\hat{\mathbf{P}}_{\mathbf{t}}^{(\mathbf{M})}\right) \tag{3.23}$$

where $\mathbf{C}_k^{(\mathbf{DC})}$ is a set of $\hat{p}_t^{(LLS,\mathbf{AN}_m,\tilde{\mathbf{Z}}(\mathbf{AN}_m))}$ generated by density-based clustering method (i.e., DC) for $k = 1, 2, \ldots, K^{(DC)}$. Note that $\mathbf{C}_k^{(\mathbf{DC})} \subset \hat{\mathbf{P}}_{\mathbf{t}}^{(\mathbf{M})}$, $\bigcup_k \mathbf{C}_k^{(\mathbf{DC})} = \hat{\mathbf{P}}_{\mathbf{t}}^{(\mathbf{M})}$ and $\mathbf{C}_k^{(\mathbf{DC})} \cap \mathbf{C}_j^{(\mathbf{DC})} = \phi$ if $j \neq k$.

- **Step 3**: Select the best three anchor nodes by examining the number of occurrences, i.e.,

$$\mathbf{AN}_k^{(\mathbf{DC})} = \left\{p_1^{(DC,k)}, p_2^{(DC,k)}, p_3^{(DC,k)}\right\} \tag{3.24}$$

such that

$$O^{(k)}(p_i) \geq O^{(k)}(p_j) \tag{3.25}$$

where $p_i \in \mathbf{AN}_k^{(\mathbf{DC})}$ and $p_j \in \mathbf{C}^{(\mathbf{DC})} \setminus \mathbf{AN}_k^{(DC)}$.

- **Step 4**: Estimate the locations of the target node by using the best three anchor nodes in each cluster and select the best one by using MBRE, i.e.,

$$\hat{p}_t^{(LLS,\mathbf{AN}_k^{(\mathbf{DC})},\tilde{\mathbf{Z}}(\mathbf{AN}_k^{(\mathbf{DC})}))} = LLS\left(\mathbf{AN}_k^{(\mathbf{DC})}, \tilde{\mathbf{Z}}(\mathbf{AN}_k^{(\mathbf{DC})})\right) \tag{3.26}$$

and

$$\hat{p}_t^{(LLS,\mathbf{AN}_{k*}^{(\mathbf{DC})},\tilde{\mathbf{Z}}(\mathbf{AN}_{k*}^{(\mathbf{DC})}))} = \min_k MBRE\left(\hat{p}_t^{(\mathbf{AN}_k^{(\mathbf{DC})})}, \mathbf{AN}, \tilde{\mathbf{Z}}(\mathbf{AN})\right) \tag{3.27}$$

where $\hat{p}_t^{(DC,\mathbf{AN},\tilde{\mathbf{Z}}(\mathbf{AN}))} = \hat{p}_t^{(LLS,\mathbf{AN}_{k*}^{(\mathbf{DC})},\tilde{\mathbf{Z}}(\mathbf{AN}_{k*}^{(\mathbf{DC})}))}$.

3.6 Numerical Examples

To clearly describe our proposed algorithms clearly, here is a numerical example to show how our proposed algorithms processed the given measured data. We consider a target node at (52.00, 49.00) and the information of anchor nodes is shown in Table 3.1. Note that $\tilde{\alpha}$ and $\tilde{Z}_0$ are 2.32 and -33.01dB respectively. To show the effect of noise on the performance of localization algorithms, three anchor nodes b, c and h are noisy nodes and their measured RSSI values are added by the noise (see Table 3.1). For example, in anchor node b, the original measured RSSI value is -70.50 dB. Now we add the noise of 6.40 dB and thus the RSSI value used for localization is -64.10 dB.

Table 3.1: The information of anchor nodes

Node name	a	b	c	d	e	f	g	h
Measured RSSI	-71.00	-64.10	-85.03	-71.00	-71.50	-71.00	-70.50	-73.96
Noisy node?	No	Yes	Yes	No	No	No	No	Yes
Noise added	-	6.40	-14.03	-	-	-	-	-2.86

In this scenarios, the estimated location of the target node using LLS is (50.31,

-18.67). Its localization error (LE) and localization error ratio (LER) is 67.69 and 1.00 respectively (the definitions of LE and LER can be found in Equation (2.8) and (2.10) respectively). The LE value is large, which means the performance of the localization is greatly affected by the noise.

When RLL-KC is applied with $K = 3$, $R = 3$ and $M = C_3^8 = 56$, 56 locations are estimated and the information of three K-mean clusters is shown in Table 3.2. It shows that Cluster 2 is the winner because its MBRE is the smallest. Then the number of occurrences of each anchor node in Cluster 1 and 3 (the winner is excluded) and node c is the winner because it has the highest number of occurrences (see Table 3.3). Therefore, node c is removed in localization and the remaining nodes are used to process the localization again. This time the estimated location of the target node is (50.31, 56.53). The LE and LER of RLL-KC are 7.72 and 8.77 respectively. From the result, we find that RLL-KC works properly because it can effectively remove a noisy node. Moreover, its LER is large and it means the performance is improved significantly when it compares with LLS.

Table 3.2: The information of K-mean clusters

	Cluster 1	Cluster 2	Cluster 3
Number of locations inside the cluster	5	46	5
Centroid	(-3.52, -462.85)	(15.17, 29.47)	(427.30, 1.67)
MBRE of the centroid	23.58	6.44	20.52

Table 3.3: The number of occurrences of each anchor node in Cluster 1 and 3

Node name	a	b	c	d	e	f	g	h
Number of occurrences	4	6	10	3	2	1	1	3

When RLL-IE is applied with $K = 3$, $R[0] = 3$ and $M = 56$, RLL-KC is applied iteratively in the localization and the result can be found in Table 3.4. Through five

iterations, five locations are found and the best estimation is (52.91, 49.12) because its MBRE is the smallest. The anchor nodes used in the localization are e, f and g. It means all noisy nodes are not selected and it shows that RLL-IE can remove all noisy nodes properly. Moreover, the LE and LER of RLL-IE are 0.92 and 73.61 respectively. Thus, the performance of RLL-IE is better than RLL-KC.

Table 3.4: The information of all iterations in RLL-IE

Iteration	Node to be removed	Estimated location	MBRE
1	c	(50.31, 56.53)	3.771
2	a	(50.64, 56.60)	3.782
3	h	(58.31, 52.18)	4.027
4	b	(53.77, 50.09)	3.772
5	d	(52.91, 49.12)	3.734

When RLL-MM is applied with $R = 3$ and $M = 56$, 56 locations are found and the location with the smallest MBRE (3.61) is (49.68, 48.51). It uses anchor nodes a, d and g for localization. It means no noisy anchor nodes are selected for localization and RLL-MM can select anchor nodes properly. The LE and LER of RLL-MM are 2.37 and 28.58 respectively. Thus, its performance is much better than LLS.

When RLL-DC is applied with $\epsilon = 15$, $minPts = 3$, $R = 3$ and $M = 56$, the number of outliers is 22 and there is only one cluster at the end. Inside the cluster, anchor nodes d, e and f are the largest three number of occurrences. Thus they are used for localization and the estimation location is (54.97, 50.00). The LE and LER of RLL-DC are 3.14 and 21.58 respectively. It means no noisy nodes are selected and RLL-DC works properly. Moreover, the performance of RLL-DC is much better than LLS.

3.7 Conclusions

This chapter described four proposed RSSI-based LoRa localization algorithms and provided numerical examples to illustrate the operations of each algorithm.

Chapter 4

PERFORMANCE INVESTIGATION

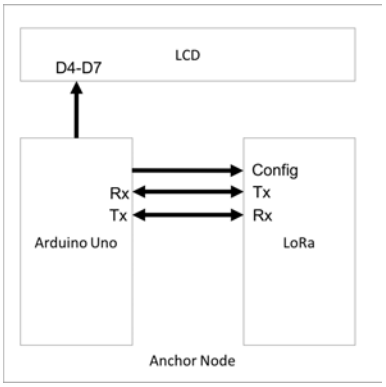
In this chapter, we investigate the performance of our proposed algorithms in a simulation model and in real experiments including outdoor environments and a large-scale indoor environments. Section 4.1 shows the design and implementation of a RSSI-based LoRa localization system which is used to do real data measurement and carry out real experiments. Section 4.2 describes the real data measurement by using the RSSI-based LoRa localization system. Section 4.3 presents our simulation model with real data. Section 4.4 compares the performance of our proposed localization algorithms with LLS through our simulation model. Moreover, the effect of different parameters on the performance of our algorithms will be described. Finally, the performance comparison in real experiments can be found in this section.

4.1 Design and Implementation of a LoRa-Based Localization System

A RSSI-based LoRa localization system is a LoRa network measuring RSSI values for localization. It is developed for real data measurement and localization. Fig. 4.1a and Fig. 4.1b show the real diagram and the block diagram of an anchor node respectively. In this anchor node, an Arduino microcontroller ATmega328p sends and receives packets through a LoRa module HM-TRLR-D-TTL-433. Fig. 4.2a and Fig. 4.2b show the real diagram and the block diagram of a target node. Here, a Raspberry Pi 3 is used to send and receive packets through the same LoRa module. At the beginning, the target node sends a packet to all anchor nodes. When an anchor node receives a packet from the target node, it sends back an acknowledgement with

its own node number and the measured RSSI value from the target node to the anchor node. Note that all anchor nodes take turns to send back their acknowledgements so that there will be no collisions among them.

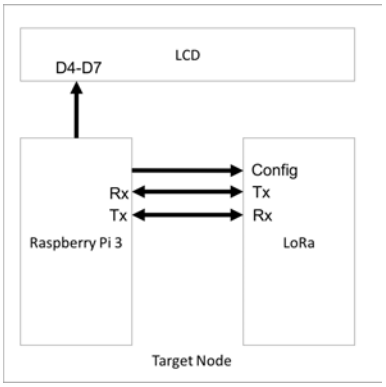
(a) The outlook of an anchor node.



(b) The block diagram of an anchor node.

Figure 4.1: Anchor Node

(a) The outlook of a target node.



(b) The block diagram of a target node.

Figure 4.2: Target Node

4.2 Real Data Measurement

We carried out two experiments to investigate the relationship between measured distances and their corresponding measured RSSI values. One target node and one

anchor node were used. For each measured distance (point), 40 readings were taken and the time interval between two successive readings was three seconds. The first experiment was carried out in Tai Po on 2 May 2016 (the location is shown in Fig. 4.3). It was a sunny day with a few showers (average temperature 25.6 °C, average humidity 89%). Each measured point was located by a mobile app with GPS localization. The relationship between measured distances and their corresponding measured RSSI values shown in Fig. 4.4 shows that it roughly matches the path-loss propagation model. Note that R^2 is the coefficient of determination, which shows the error between the best-fit line and the given points. In this figure, it is close to one, which means the line fits the given points.

The second experiment was carried out in Kai Tak on 26 June 2017 (the location is shown in Fig. 4.5). It was a sunny day with a few showers (average temperature 29.8 °C, average humidity 78%). All distances were measured using an infra-red meter (model number: SW-M100, range: up to 100 m) with a measurement error of 1.5 mm. The relationship between measured distances and their corresponding RSSI values shown in Fig. 4.6 shows that it is also roughly matches the path-loss propagation model and R^2 is even close to one, which means the line fits to the given points better. Due to the high accuracy of the measured data (an infra-med meter was used in measurement), it was used in the simulation model (which will be described in Section 4.3 below).

4.3 Simulation Model (with real data measurement)

We developed a simulation model with real data from the second experiment. In the simulation model, all anchor nodes were evenly distributed on the circumference of a circle with radius 50 m and the target node was randomly generated inside the circle. For each measurement point, 20 readings were randomly selected from 40 readings (duplicated readings were allowed, provided that the measured distance in

the simulation model matched the corresponding measured data in the real data). An outlier filter was used to filter out some measured RSSI values which were too far away from the measured mean [30]. Localizations were carried out 2000 times with different locations of the target node.

To introduce a noisy environment, we randomly selected whether an anchor node was noisy or not with a specified probability. If an anchor node was selected to be a noisy node, its measured RSSI value was randomly added a floating point number between -20 to 20. The specified probability is defined as the noise factor, NF.

Note that the performance investigation in the coming section is based on simulation results from the simulation model using real data. We then will show the performance investigation in real experiments with (a) noisy outdoor environments and (b) a noisy large-scale indoor environment.

4.4 Performance Comparisons

In this section, we compare how all the proposed localization algorithms and LLS performed in terms of their localization error (LE), localization error ratio (LER) and normalized computational complexity (NCC) in a general scenario. Then we will investigate how different parameters affect how different localization algorithms performed in different environments. Finally, we show how the different localization algorithms performed in some real experiments including outdoor environments and a large-scale indoor environment.

4.4.1 Performance Comparisons of RSSI-Based LoRa Localization Algorithms

In this subsection, we compare how all localization algorithms performed in a general scenario. To simulate the real environment, NF = 0 and 0.4 to represent the environment with Gaussian noise only, and the environment with Gaussian noise and also non-Gaussian noise sources such as blocking and multi-path. Note that the latter

simulation is closer to the real environment rather than the former. The parameters of each localization algorithm are summarized in Table 4.1 and the parameters are optimal in our performance investigation.

Table 4.1: Parameter settings in our performance investigation

	$N(\mathbf{\Gamma})$	R	R_0	K
LLS	12	-	-	-
RLL-KC	12	4	-	3
RLL-MM	12	3	-	-
RLL-IE	12	-	9	3
RLL-DC	12	3	-	-

Note that RLL-DC has two parameters in density-based clustering. One is the minimum number of sample points in a cluster. It is set to 5%, which is directly proportional to the total number of estimated locations. This setting is related to the maximum distance among sample points, ϵ (it was described in **??**, Chapter 1). When the number of sample points in a cluster is below the setting, it can either reduce the minimum number of sample points or increase the maximum distance among sample points to increase the number of sample points in a cluster. In our proposed algorithm, the latter approach is selected because the number of occurrences of anchor nodes is counted after clustering and it will be meaningless when the number of sample points in a cluster is too small. The other parameter is the maximum distance among sample points, ϵ. It is set to 15 m initially. If no clusters can be formed with this setting, it will be extended 5 m and reprocess the clustering until at least one cluster is formed. The above setting is a general policy if the number of sample points is not a constant in different cases.

Table 4.2 and 4.3 show the performance comparison of four proposed localization algorithms and LLS in terms of LE, LER and NCC. Note that we compare with LLS only because it is the well known RSSI-based localization algorithm and there

is no other existing RSSI-based LoRa localization algorithm for the comparison. In Table 4.2, where there is Gaussian noise only, all proposed algorithms outperformed LLS in terms of LE and LER except RLL-DC (it was very close to LLS). This means our proposed algorithms can handle noise more efficiently than LLS, even when it is not too noisy. In Table 4.3, when NF = 0.4 (which is a noisy environment), the improvement was much larger and it shows that our proposed algorithms can address the noise issue properly. Among all proposed algorithms, RLL-IE performed the best among all the cases. In NCC, it is normalized by LLS. Note the computational time of LLS was 4.31 ms (computer model: iMac Mid-2011, CPU: 2.8 GHz Intel Core i7). These tables show that the NCC of all four proposed algorithms were much higher than LLS, but their computational complexity is still low and thus still suitable for locating the target node in real time. Finally, according to [19], the localization error for general GPS applications is between 10 m to 70 m, which is similar to the performance of our proposed localization algorithms.

Table 4.2: The performance comparison of all localization algorithms (NF = 0).

	LLS	RLL-KC	RLL-MM	RLL-IE	RLL-DC
LE	23.99 ± 6.63	23.30 ± 0.49	19.64 ± 0.76	21.77 ± 0.98	25.14 ± 0.82
LER	1.00	1.03	***1.22***	1.10	0.95
NCC	1.00	43.81	***14.37***	97.78	15.22

Table 4.3: The performance comparison of all localization algorithms (NF = 0.4).

	LLS	RLL-KC	RLL-MM	RLL-IE	RLL-DC
LE	213.76 ± 6.63	71.88 ± 0.49	30.82 ± 0.76	28.01 ± 0.98	31.10 ± 0.82
LER	1.00	2.97	6.94	***7.63***	6.87
NCC	1.00	43.81	***14.37***	97.78	15.22

Through the numerical data, we found that, in most cases, the 95% confidence interval of each data point was very small so the size of a data point was larger than that. Thus, for simplicity, we will not show the 95% confidence interval in the rest of

this section.

4.4.2 The Effect of K, R and $R[0]$ on RSSI-Based LoRa Localization Algorithms

In this subsection, we investigate the effect of the number of clusters K on the performance of RLL-KC and RLL-IE. Moreover, the effect of the size of the subsets of anchor nodes R on the performance of RLL-KC, RLL-MM and RLL-DC was investigated. Note that to make use of each anchor node evenly, we distributed anchor nodes in all different possible subsets that can be made from the whole set. This means, for the set of anchor nodes is $\mathbf{\Gamma}$, the total number of subsets is ${}_{N(\mathbf{\Gamma})}C_R$. Finally, we also investigated the effect of $R[0]$ on the performance of RLL-IE.

Fig. 4.7 shows the effect of K on the performance of RLL-KC where $N(\mathbf{\Gamma}) = 12$ and $R = 3$. For different values of NF, LE remained almost constant when K increases. This is because the RLL-KC mechanism is independent of the number of clusters. Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$ and R. Fig. 4.8 shows the effect of K on the performance of RLL-IE where $N(\mathbf{\Gamma}) = 12$ and $R[0] = 11$. For different values of NF, when K increases, LE increases. This is because the number of sample points in each cluster decreases iteratively. Thus, a large number of clusters will get a smaller and smaller number of sample points when the number of iterations increases. Therefore, there will be too small number of sample points in a cluster and the advantage of clustering is significantly reduced, which degrades the performance of RLL-IE. Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$ and when the value of $R[0]$ is not small. When the value of $R[0]$ is small, the performance of RLL-IE is similar to RLL-KC because the number of iterations in RLL-IE is small.

Fig. 4.9 shows the effect of R on the performance of RLL-KC where $N(\mathbf{\Gamma}) = 12$ and $K = 3$. For different values of NF, LE remains almost constant when R increases. This is because the RLL-KC mechanism is independent of the number of subsets of anchor nodes. Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$ and K. Fig. 4.10 shows the effect of R on the performance of RLL-MM where $N(\mathbf{\Gamma}) = 12$.

For different values of NF, LE remains almost constant when R is not too large. This is because the effect of the number of the subsets of anchor nodes is insignificant to the RLL-MM mechanism. However, when R is sufficiently large, LE is large. This is because, when R is large, some noisy anchor nodes may be included in some subsets of anchor nodes and cannot be excluded even in the best subset of anchor nodes, especially when NF is high (it means there are too many noisy anchor nodes). Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$.

Fig. 4.11 shows the effect of R on the performance of RLL-DC where $N(\mathbf{\Gamma}) = 12$. For different values of NF, LE remains almost constant when R is not too large. This is because the effect of the number of the subsets of anchor nodes is insignificant to the RLL-DC mechanism. However, when R is sufficiently large, LE is large. This is because, when R is large, there are too many noisy anchor nodes in each subset of anchor nodes and clustering cannot totally eliminate all of them in each cluster. Thus the advantage of clustering is significantly reduced, which degrades the performance of RLL-DC significantly. Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$.

Fig. 4.12 shows the effect of $R[0]$ on the performance of RLL-IE where $N(\mathbf{\Gamma}) = 12$ and $K = 3$. For different values of NF, LE remains almost unchanged when $R[0]$ is not too small or large (close to the total number of anchor nodes). However, the computational complexity is the highest when $R[0]$ is close to $N(\mathbf{\Gamma})/2$ (refer to Fig. 4.13). Thus, it is preferable to set $R[0] = N(\mathbf{\Gamma}) - 3$. Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$ and K.

4.4.3 The Effect of The Number of Anchor Nodes on RSSI-Based LoRa Localization Algorithms

In this subsection, we investigate the effect of the number of anchor nodes $N(\mathbf{\Gamma})$ on the performance of different localization algorithms. Fig. 4.14, 4.15, 4.16, 4.17, 4.18 and 4.19 show the effect of $N(\mathbf{\Gamma})$ on the performance of LLS, RLL-KC, RLL-IE,

RLL-MM and RLL-DC respectively in terms of LE where $K = 3$,$R = 3$ and $R[0] = N(\mathbf{\Gamma}) - 2$. Note that, except RLL-MM, all other proposed localization algorithms are using clustering and thus $N(\mathbf{\Gamma})$ cannot be too small; otherwise, there are not enough sample points for clustering and the performance will be unstable.

For different values of NF, the performance of a localization algorithm in terms of LE and LER is generally better when the number of anchor nodes increases, except for RLL-KC. This is because LLS can provide a better solution when more anchor nodes are involved in the calculation. However, the improvement of LE is not very great in LLS because the effect of noise is only reduced slightly. For our proposed localization algorithms, they can significantly reduce the effect of the noise, and this reduction is very significant when the number of anchor nodes is small. For RLL-KC, since there is only one anchor node to be eliminated, when $N(\mathbf{\Gamma})$ increases, the ratio of noisy anchor nodes to the total number of anchor nodes increases slightly and thus, the overall performance in terms of LE degrades slightly. Fig. 4.16 shows the performance of RLL-KC when the number of noisy anchor nodes are fixed. When $N(\mathbf{\Gamma})$ increases, the performance of RLL-KC in terms of LE decrease, this means RLL-KC work properly in such cases. Fig. 4.20 show the effect of $N(\mathbf{\Gamma})$ on the performance of different localization algorithms in terms of NCC. It is expected that the difference between our proposed localization algorithms and LLS increases significantly when $N(\mathbf{\Gamma})$ increases, and the difference can be up to 200 or even higher. However, this drawback is not important because the overall computational time is still small and can be used in real-time applications. Note that this finding is consistent for different values of K, R and $R[0]$.

4.4.4 The Effect of Noise on RSSI-Based LoRa Localization Algorithms

In this subsection, we investigate the effect of noise in different localization algorithms. Fig. 4.21 shows the effect of NF in different localization algorithms. The performance of our proposed algorithms outperformed LLS significantly and

the differences are greater when NF increases. This means our proposed algorithm can effectively reduce the effect of noise much better than LLS. For RLL-KC, the improvement is not significant when NF increases because it can eliminate one noisy anchor node only. When NF is too large, the advantage of the elimination is limited and thus also limits the improvement. Note that this finding is consistent for different values of $N(\mathbf{\Gamma})$, K, R and $R[0]$.

4.4.5 Performance Comparisons of RSSI-Based LoRa Localization Algorithms in Real Experiments (Indoor and Outdoor)

In this subsection, we investigate the performance of all localization algorithms in outdoor environments and a large-scale indoor environment. The first experiment was carried out in Kowloon Tong Sport Complex (outdoor environment) on 20 July 2017 (the location is shown in Fig. 4.22). It was a sunny day (average temperature 28.6 °C, average humidity 85%). All distances were measured using the infra-red meter. Table 4.4 shows the performance of all localization algorithms in terms of LE and LER and Table 4.5 shows the distribution of the anchor nodes. RLL-MM was found to be the best among all, and all proposed localization algorithms significantly outperformed LLS. Furthermore, the proposed localization algorithms performance was similar to GPS.

Table 4.4: The performance comparison of all localization algorithms in the first experiment.

	LE				
p_t	LLS	RLL-KC	RLL-IE	RLL-MM	RLL-DC
$(22.00, 8.40)$	29.40	43.35	***14.52***	***14.52***	26.03
$(50.00, 22.20)$	32.79	38.01	23.11	***22.07***	24.21
$(66.30, 32.20)$	***33.03***	34.71	34.09	34.09	48.33
Average	31.74	38.69	23.91	***23.55***	32.86

Table 4.5: The coordinates of the anchor nodes in the first experiment.

	X	Y
$p_1^{(\Gamma)}$	0.00	0.00
$p_2^{(\Gamma)}$	0.00	64.00
$p_3^{(\Gamma)}$	50.00	0.00
$p_4^{(\Gamma)}$	50.00	64.00
$p_5^{(\Gamma)}$	110.00	0.50
$p_6^{(\Gamma)}$	110.00	64.50

The second experiment was carried out in Sun Yat-sen Memorial Park (outdoor environment) on 8 December 2017 (the location is shown in Fig. 4.23). It was a sunny day (average temperature 15.1 °C, average humidity 39%). All measured points were located by a GPS receiver (Model number: Trimble R10 GNSS System, see Fig. 4.24) with a measurement error of 8 mm. The specification of the GPS receiver can be found in Table 4.6. This was used because of its high accuracy and also the range of its measurement. The results are shown in Table 4.7 and Table 4.8 shows the distribution of the anchor nodes. It was similar to the first experiment.

Table 4.6: The specification of the GPS receiver.

Horizontal	8 mm + 0.5 ppm RMS (with Network RTK)
Vertical	15 mm + 0.5 ppm RMS (with Network RTK)
Dimensions (W x H)	11.9 cm x 13.6 cm
Weight	1.12 kg
Operating times	5.0 hours (Cellular receive option)
Price	USD $12500 [31]

The third experiment was carried out in Podium of The Hong Kong Polytechnic University (a large-scale indoor environment) on 22 December 2017 (the outlook is shown in Fig. 4.25). It was a sunny day (average temperature 16.8 °C, average humidity 68%). Since the localization was in a large-scale indoor environment, the

Table 4.7: The performance comparison of all localization algorithms in the second experiment.

	LE				
p_t	LLS	RLL-KC	RLL-IE	RLL-MM	RLL-DC
$(19.23, 38.89)$	14.73	***8.12***	16.14	14.87	34.26
$(44.93, 30.45)$	6.73	***4.92***	9.93	8.05	29.48
$(71.15, 37.89)$	14.92	27.17	13.06	***12.28***	13.72
$(71.98, 65.56)$	35.04	10.04	8.94	8.92	***4.20***
$(37.22, 70.23)$	36.09	26.21	15.60	7.93	***4.61***
$(27.04, 51.67)$	***16.65***	19.31	16.87	19.08	18.25
Average	20.63	15.96	13.42	***11.86***	17.42

Table 4.8: The coordinates of the anchor nodes in the second experiment.

	X	Y
$p_1^{(\Gamma)}$	0.00	39.45
$p_2^{(\Gamma)}$	3.19	27.45
$p_3^{(\Gamma)}$	15.11	11.45
$p_4^{(\Gamma)}$	46.27	0.00
$p_5^{(\Gamma)}$	72.70	13.11
$p_6^{(\Gamma)}$	20.05	81.01
$p_7^{(\Gamma)}$	71.77	81.57
$p_8^{(\Gamma)}$	89.25	56.67

GPS meter could not be used so all distances were measured using the infra-red meter. Table 4.10 shows the distribution of the anchor nodes and Table 4.9 shows the results were as good as the experiments carried out in outdoor environments.

4.5 Conclusions

This chapter briefly described the design and implementation of a LoRa-based localization system. We showed how to do real data measurements and build up a simulation model using those real data measurements. We compared the performance

Table 4.9: The performance comparison of all localization algorithms in the third experiment.

	LE				
p_t	LLS	RLL-KC	RLL-IE	RLL-MM	RLL-DC
$(8.19, 3.83)$	40.38	49.04	41.92	***12.62***	***12.62***
$(18.91, 7.82)$	9.43	30.19	***8.74***	27.39	18.52
$(33.57, 5.14)$	308.53	715.52	224.20	***1.84***	***1.84***
Average	119.45	264.92	91.62	13.95	***10.99***

Table 4.10: The coordinates of the anchor nodes in the third experiment.

	X	Y
$p_1^{(\Gamma)}$	0.00	0.00
$p_2^{(\Gamma)}$	0.00	9.93
$p_3^{(\Gamma)}$	14.63	0.00
$p_4^{(\Gamma)}$	14.63	9.93
$p_5^{(\Gamma)}$	35.44	0.50
$p_6^{(\Gamma)}$	35.44	10.43

of our proposed algorithms with LLS algorithm through simulations. We also showed the effect of different parameters on our proposed algorithms. Finally, we compared the performance of our proposed algorithms with LLS algorithms again through real indoor and outdoor experiments.

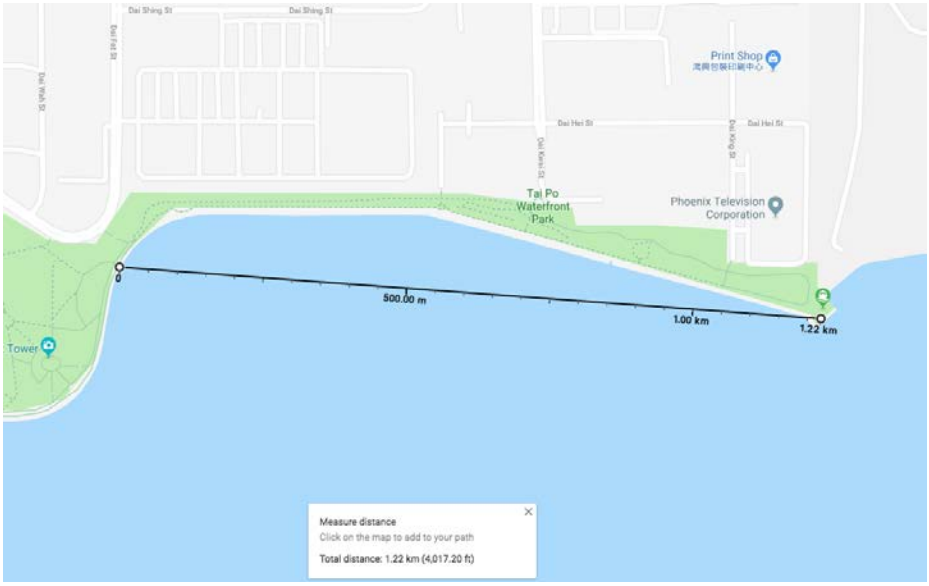


Figure 4.3: The location of Tai Po in Google map.

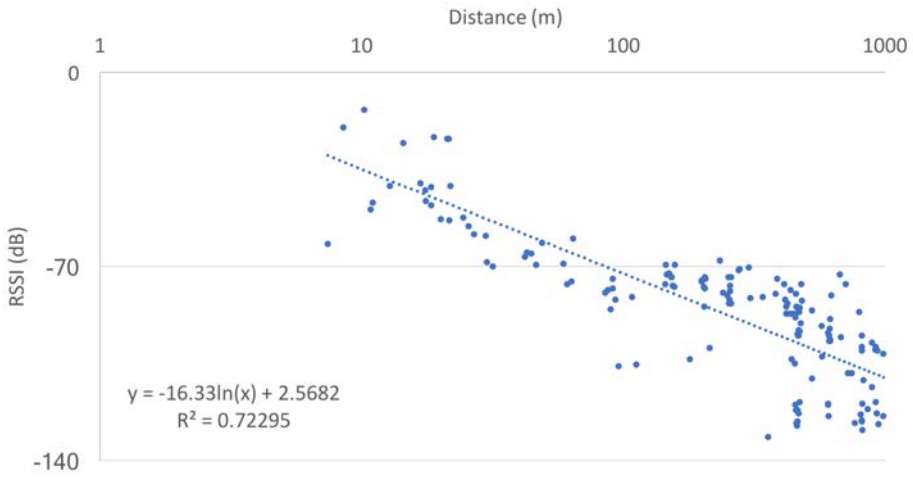


Figure 4.4: The relationship between the measured distance and its measured RSSI in Tai Po.

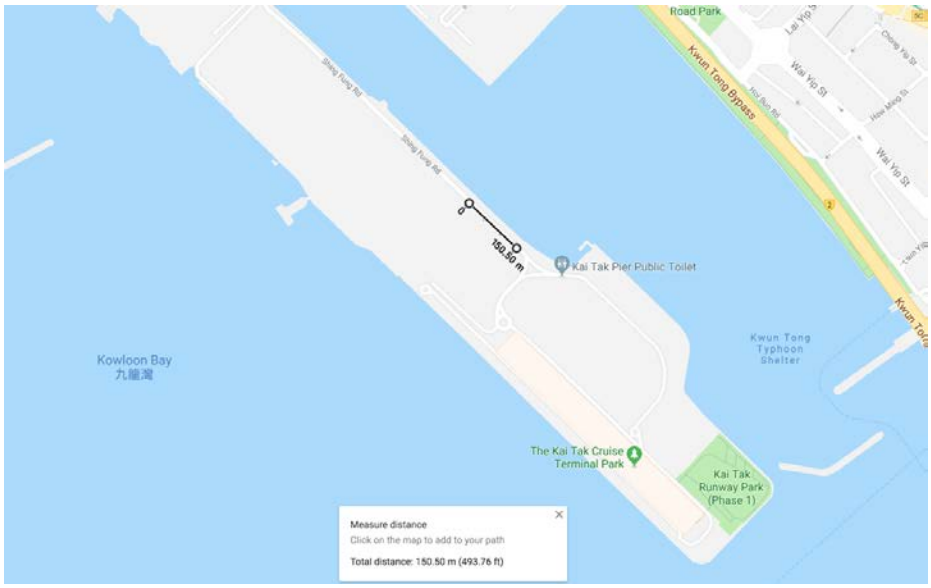


Figure 4.5: The location of Kai Tak in Google map.

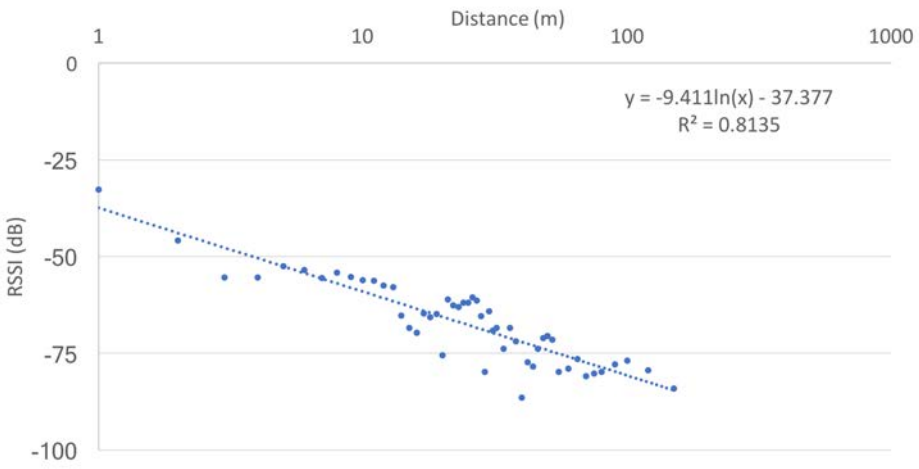


Figure 4.6: The relationship between the measured distance and its measured RSSI in Kai Tak.

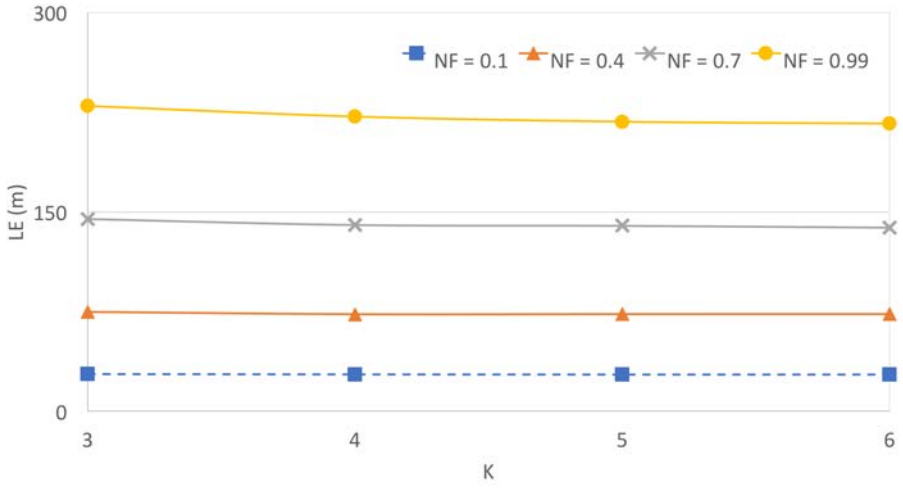


Figure 4.7: The effect of K on the performance of RLL-KC.

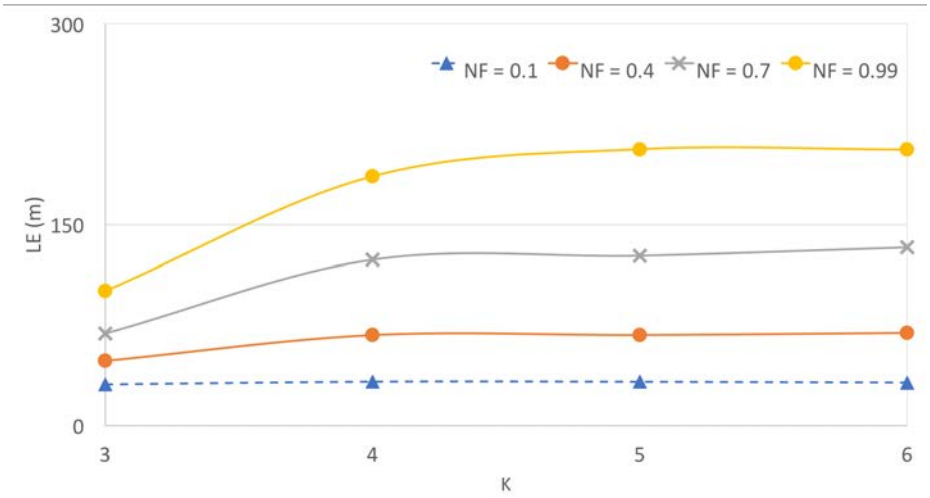


Figure 4.8: The effect of K on the performance of RLL-IE.

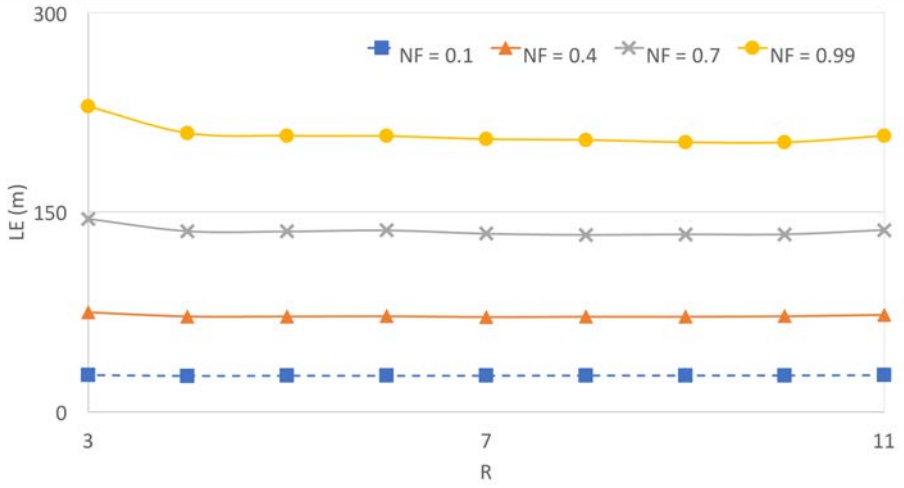


Figure 4.9: The effect of R on the performance of RLL-KC.

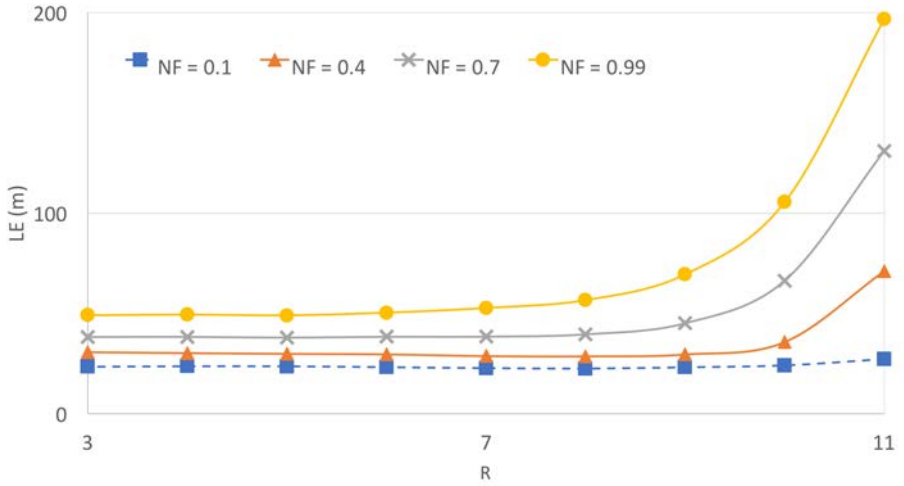


Figure 4.10: The effect of R on the performance of RLL-MM.

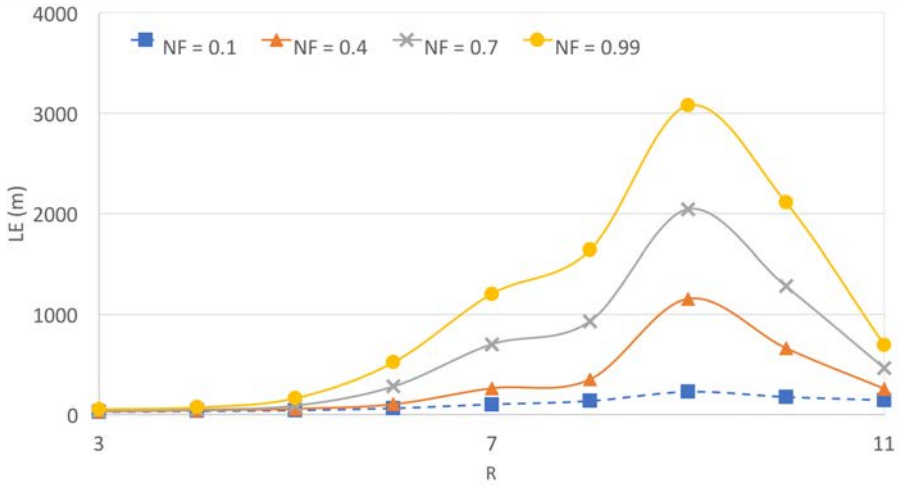


Figure 4.11: The effect of R on the performance of RLL-DC.

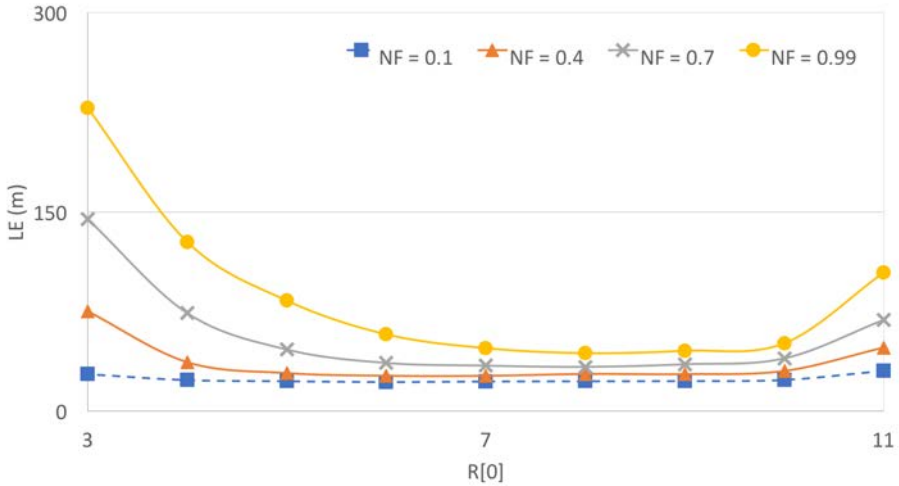


Figure 4.12: The effect of $R[0]$ on the performance of RLL-IE.

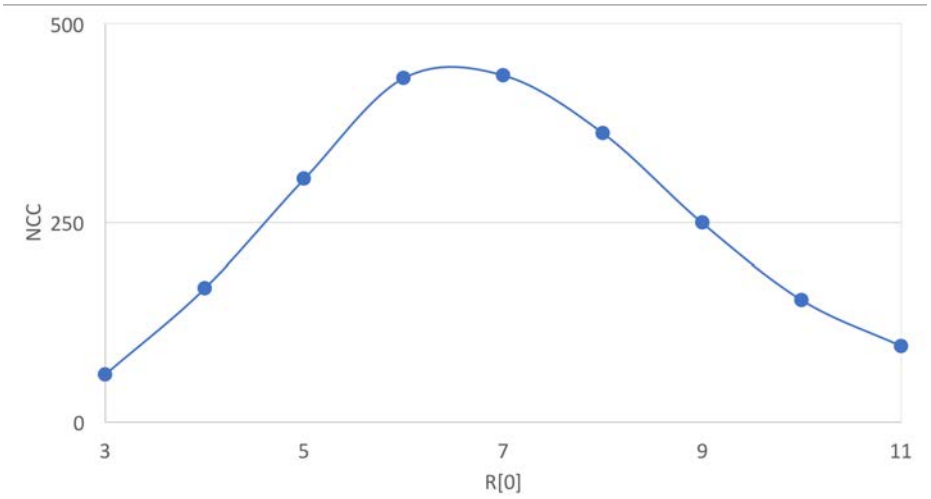


Figure 4.13: The effect of $R[0]$ on the NCC of RLL-IE when $N(\mathbf{\Gamma}) = 12$.

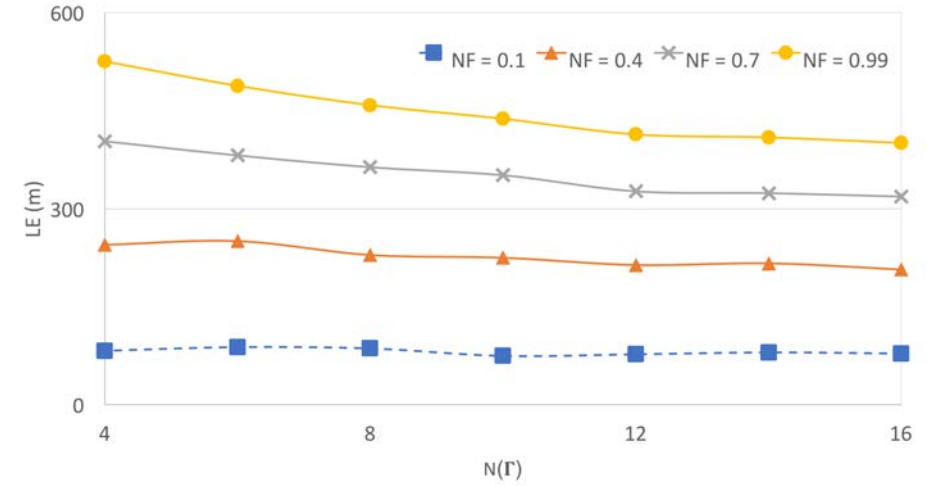


Figure 4.14: The effect of the number of anchor nodes on the performance of LLS.

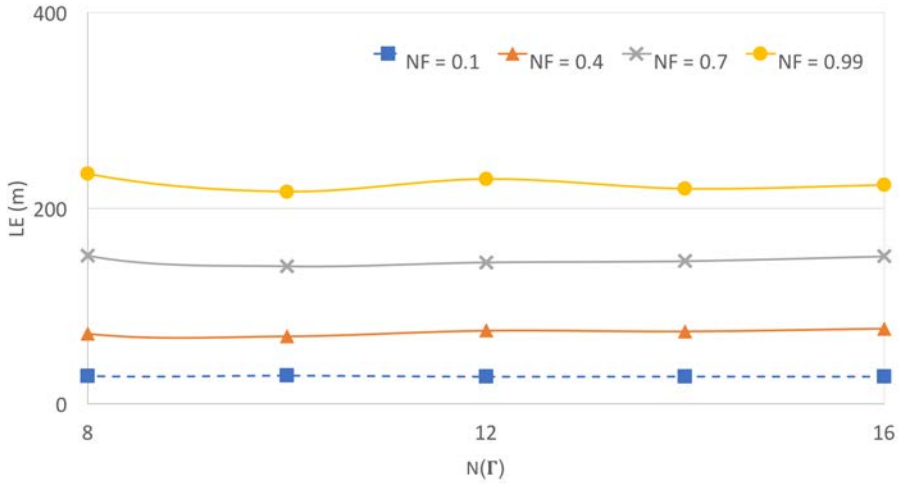


Figure 4.15: The effect of the number of anchor nodes on the performance of RLL-KC.

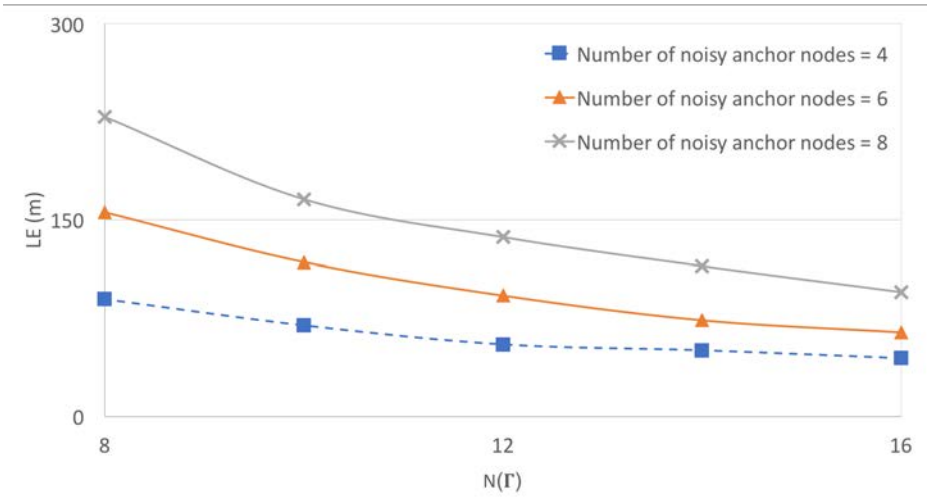


Figure 4.16: The effect of the number of anchor nodes on the performance of RLL-KC when fixed number of noisy anchor nodes.

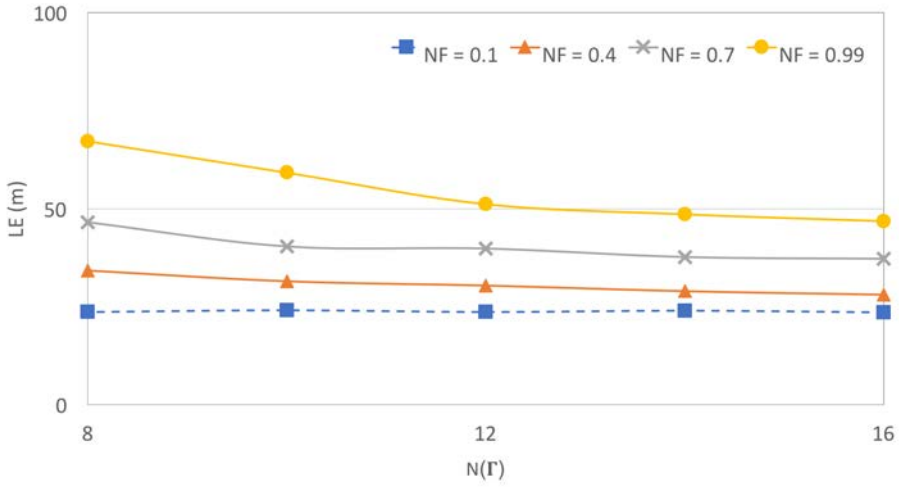


Figure 4.17: The effect of the number of anchor nodes on the performance of RLL-IE.

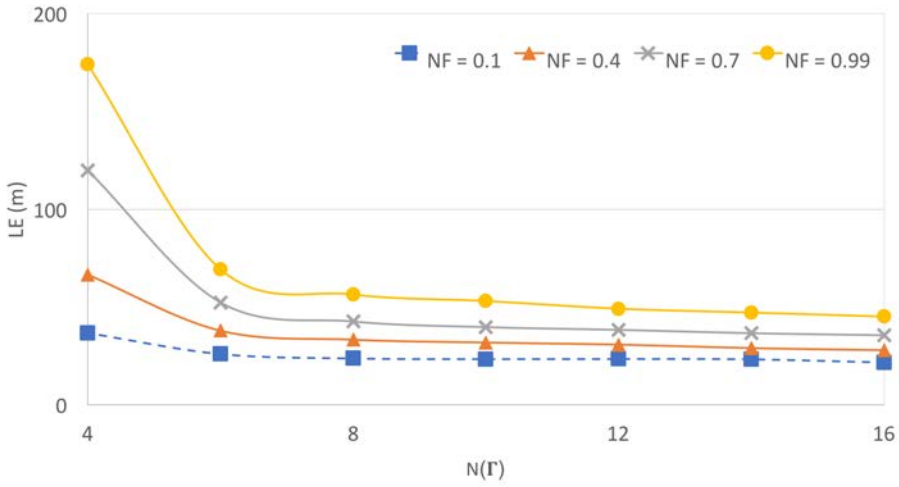


Figure 4.18: The effect of the number of anchor nodes on the performance of RLL-MM.

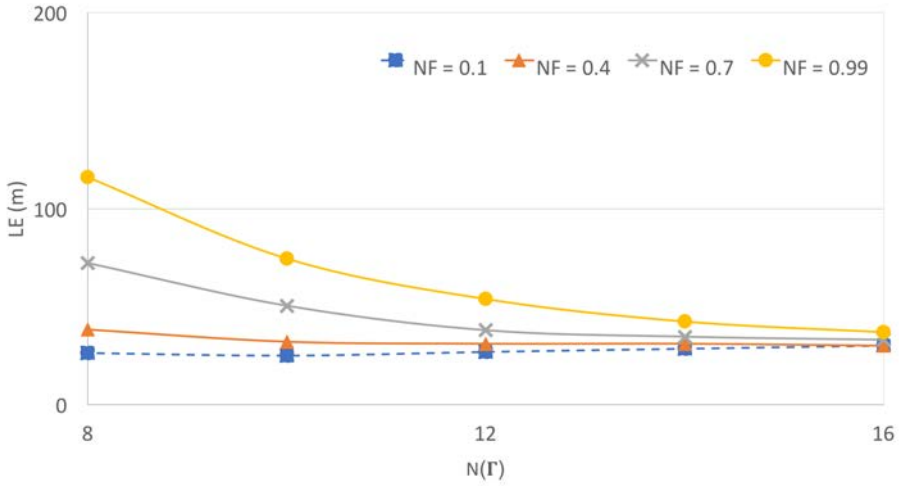


Figure 4.19: The effect of the number of anchor nodes on the performance of RLL-DC.

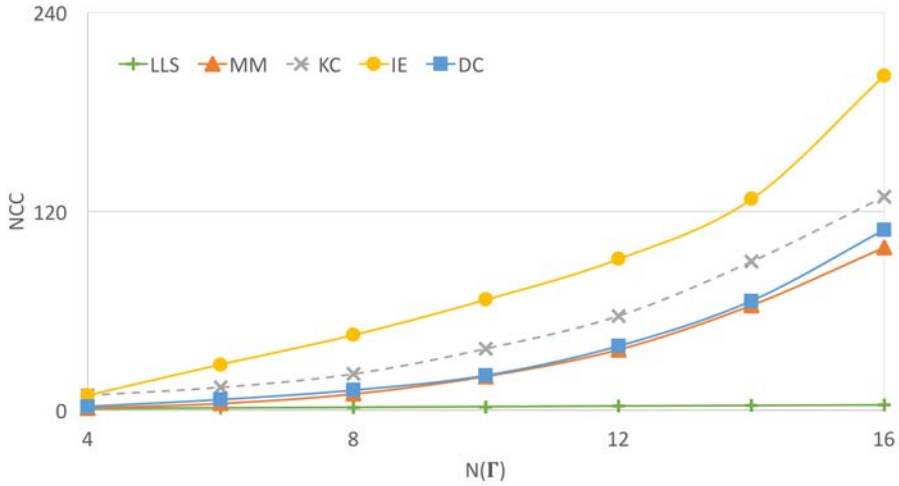


Figure 4.20: The performance comparison of different algorithms in term of NCC.

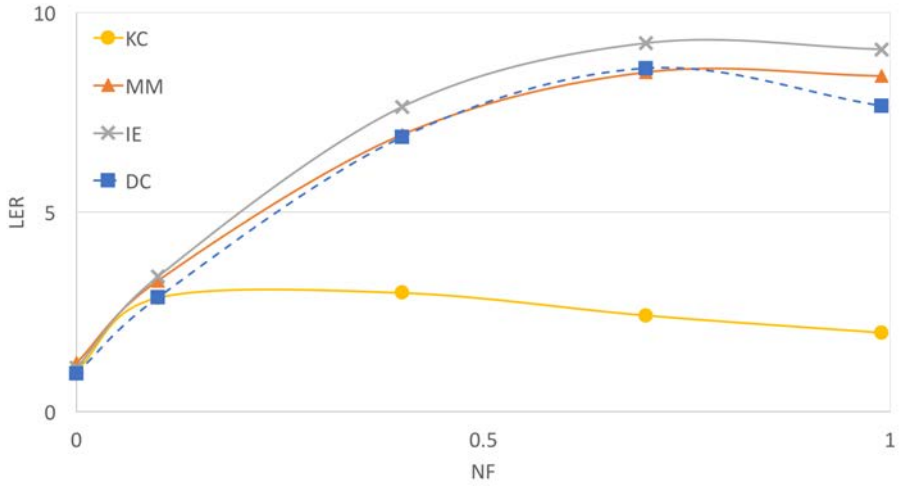


Figure 4.21: The LER of all proposed algorithms with different values of NF.

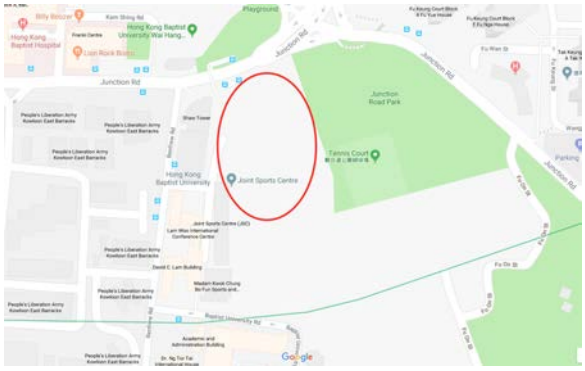


Figure 4.22: The location of the first experiment in Google map.

Figure 4.23: The location of the second experiment in Google map.

Figure 4.24: The outlook of the GPS Receiver.

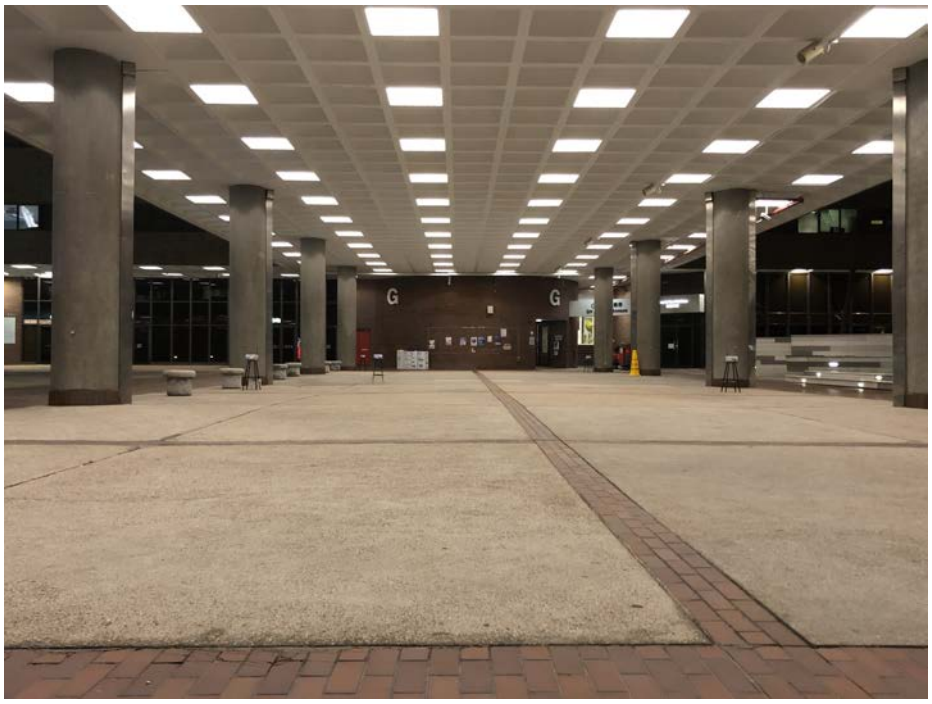


Figure 4.25: The environment of the third experiment.

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Localization is a very important research topic and it has been used in many different applications. Given the popularity of localization, we studied the problem of localization for outdoor environments.

We formulated the localization problem for outdoor environments and studied the limitations of existing satellite-based localization systems to implement the optimal solution. Based on the limitations and the development of LoRa technology, we proposed using Receiver Signal Strength Indicator (RSSI) to develop some localization algorithms in LoRa networks.

To the best of our knowledge,

- We are among the first working on localization using LoRa technology;
- We are the first to develop RSSI-based localization algorithms in LoRa networks, and
- We are the first to handle blocking and multi-path (non-Gaussian noise) for localization in LoRa networks.

We propose the following RSSI-based localization algorithms to handle blocking and multi-path (non-Gaussian noise) in LoRa networks:

- RSSI-based LoRa Localization with K-mean Clustering (RLL-KC): This makes use of K-mean clustering to heuristically eliminate a noisy anchor node from

the set of anchor nodes and then reprocess the localization. This methodology can significantly reduce the effect of noise.

- RSSI-based LoRa Localization with Iterative Elimination (RLL-IE): This applies RLL-KC iteratively to eliminate anchor nodes until three anchor nodes remain to process the localization. This methodology can significantly reduce the effect of noise if there are more than one noisy anchor node.

- RSSI-based LoRa Localization with Minimum MBRE (RLL-MM): This uses of the calculated RSSI values to heuristically select the best subset of anchor nodes among all to process the localization. This methodology can effectively select appropriate anchor nodes for localization to avoid the effect of noise.

- RSSI-based LoRa Localization with Density-based Clustering (RLL-DC): This uses density-based clustering to heuristically select the best subset of anchor nodes among all to process the localization. This methodology can effectively select appropriate anchor nodes in localization when there are many noisy anchor nodes.

The performance of all localization algorithms was investigated using a simulation model with real-data measurement and real experiments with our developed LoRa localization system. Based on the performance investigation, we conclude that the performance of the proposed localization algorithms is similar to the most popular outdoor localization system (GPS) in terms of localization error and much better than the traditional RSSI-based localization algorithm, Linear Least-Squares (LLS) position estimation model. Moreover, through the real LoRa localization systems, we found that the proposed localization algorithms work properly in both outdoor and large-scale indoor environments.

5.2 Future Work

This research focused on applying LoRa technology to develop localization algorithms for outdoor and large-scale indoor environments. To extend the current research in other directions, the following paths should be considered as possible further research:

- At the beginning of this research study, LoRa was the only technology to support long range and low power wireless communication networks. Now, we have Ultra Narrow Band and other technologies exist to support these kinds of networks. Thus, it is possible to investigate the possibility of applying our proposed localization algorithms in such new technologies.

- In this research, we focused on one target node only not only for simplicity, because it was also important to focus on improving the accuracy of localization rather than handling more than one target node. The study of accuracy has now been completed and we may study how to apply our proposed algorithms for many target nodes.

- The value of the path-loss exponent and the reference measured RSSI value is assumed to be the same for all anchor nodes. However, they may be different, especially if their locations are quite far away from each other. Thus, it is possible to obtain further improvements if such values can be measured individually.

BIBLIOGRAPHY

[1] R. Zekavat and R.M. Buehrer, *Handbook of Position Location: Theory, Practice and Advances*, IEEE Series on Digital & Mobile Communication. Wiley, 2011.

[2] B. C. Fargas and M. N. Petersen, "Gps-free geolocation using lora in low-power wans," in *2017 Global Internet of Things Summit (GIoTS)*, June 2017, pp. 1–6.

[3] J. J. Caffery, "A new approach to the geometry of toa location," in *Vehicular Technology Conference Fall 2000. IEEE VTS Fall VTC2000. 52nd Vehicular Technology Conference (Cat. No.00CH37152)*, 2000, vol. 4, pp. 1943–1949 vol.4.

[4] N. Salman, A. H. Kemp, and M. Ghogho, "Low complexity joint estimation of location and path-loss exponent," *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 364–367, August 2012.

[5] Z. Situ, I. Wang-Hei Ho, H. C. B. Chan, and D. P. K. Lun, "The impact of node reliability on indoor cooperative positioning," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, Oct 2016, pp. 300–304.

[6] M. Dashti, S. Yiu, S. Yousefi, F. Perez-Cruz, and H. Claussen, "Rssi localization with gaussian processes and tracking," in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.

[7] J. Jung, K. Kim, S. Yoo, M. Bae, S. K. Lee, and H. Kim, "Rssi localization with db-assisted least error algorithm," in *2015 Seventh International Conference on Ubiquitous and Future Networks*, 2015, pp. 338–343.

[8] J. Xiong, Q. Qin, and K. Zeng, "A distance measurement wireless localization correction algorithm based on rssi," in *2014 Seventh International Symposium on Computational Intelligence and Design*, 2014, vol. 2, pp. 276–278.

[9] S. Shue, L. E. Johnson, and J. M. Conrad, "Utilization of xbee zigbee modules and matlab for rssi localization applications," in *SoutheastCon 2017*, 2017, pp. 1–6.

[10] L. Yi, L. Tao, and S. Jun, "Rssi localization method for mine underground based on rssi hybrid filtering algorithm," in *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, 2017, pp. 327–332.

[11] W. C. Lee, F. H. Hung, K. F. Tsang, C. K. Wu, and H. R. Chi, "Rss-based localization algorithm for indoor patient tracking," in *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 2016, pp. 1060–1064.

[12] N. Mair and Q. H. Mahmoud, "A collaborative bluetooth-based approach to localization of mobile devices," in *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2012, pp. 363–371.

[13] X. Shen, S. Yang, J. He, and Z. Huang, "Improved localization algorithm based on rssi in low power bluetooth network," in *2016 2nd International Conference on Cloud Computing and Internet of Things (CCIOT)*, Oct 2016, pp. 134–137.

[14] T. Ishihara, K. M. Kitani, C. Asakawa, and M. Hirose, "Inference machines for supervised bluetooth localization," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 5950–5954.

[15] Z. Yang, J. Schafer, and A. Ganz, "Disaster response: Victims' localization using bluetooth low energy sensors," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, April 2017, pp. 1–4.

[16] Y. Wang, Q. Ye, J. Cheng, and L. Wang, "Rssi-based bluetooth indoor localization," in *2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, Dec 2015, pp. 165–171.

[17] K. Subaashini, G. Dhivya, and R. Pitchiah, "Zigbee rf signal strength for indoor location sensing - experiments and results," in *2013 15th International Conference on Advanced Communications Technology (ICACT)*, 2013, pp. 50–57.

[18] B. Amer and A. Noureldin, "Rss-based indoor positioning utilizing firefly algorithm in wireless sensor networks," in *2016 11th International Conference on Computer Engineering & Systems (ICCES)*, 2016, pp. 329–333.

[19] L. Lee, M. Jones, G. S. Ridenour, S. J. Bennett, A. C. Majors, B. L. Melito, and M. J. Wilson, "Comparison of accuracy and precision of gps-enabled mobile devices," in *2016 IEEE International Conference on Computer and Information Technology (CIT)*, 2016, pp. 73–82.

[20] M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen, "Robust tracking in cellular networks using hmm filters and cell-id measurements," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1016–1024, 2011.

[21] "Lora-alliance," https://www.lora-alliance.org/technology, accessed 31 December, 2017.

[22] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the iot and smart city scenarios," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, 2016.

[23] L. Vangelista, "Frequency shift chirp modulation: the lorat*m* modulation," *IEEE Signal Processing Letters*, vol. PP, no. 99, pp. 1–1, 2017.

[24] "Lora localization," https://www.link-labs.com/blog/lora-localization, accessed 31 December, 2017.

[25] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, 1996, vol. 96, pp. 226–231.

[26] K. K. Almuzaini and T. A. Gulliver, "Range-based localization in wireless networks using the dbscan clustering algorithm," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–7.

[27] H.L. Bertoni, *Radio Propagation for Modern Wireless Systems*, Wireless communications. Prentice Hall PTR, 2000.

[28] K. Madsen, H. B. Nielsen, and O. Tingleff, "Methods for non-linear least squares problems (2nd ed.)," 2004.

[29] S. M. Kay, *Fundamentals of statistical signal processing : estimation theory*, Englewood Cliffs, N.J. : PTR Prentice-Hall, Englewood Cliffs, N.J., 1993.

[30] F. E. Grubbs, "Sample criteria for testing outlying observations," *The Annals of Mathematical Statistics*, pp. 27–58, 1950.

[31] "Trimble r10 gnss receiver," https://www.ebay.com/itm/Trimble-R10-GNSS-GPS-Base-Rover-Receiver-w-450-470-MHz-Radio-Galileo-GLONASS-/272988575233, accessed 31 December, 2017.