Tampere University

Maral Zadehdarrehshoorian

BLE ADVERTISING CHANNEL ANALYSIS

RSS-based Indoor Positioning

Master of Science Thesis
Information Technology and Communication Sciences
Supervisors: Jukka Talvitie, Jaakko Peltonen
April 2023

ABSTRACT

Maral Zadehdarrehshoorian: BLE Advertising Channel analysis
Master of Science Thesis
Tampere University
Master's Degree Programme in Data Science
April 2023

Positioning services are among the important matters of our world. GNSS has proven to provide high quality outdoor localization; contrarily, it results in poor accuracy in indoor cases due to the complexity of indoor environments. Indoor positioning systems can benefit from pre-existent radio signals in the environment. Thus, WIFI and Bluetooth infrastructures that are already set up for other purposes can be utilized for indoor localization purpose as well.

In this thesis, our concentration is on Bluetooth low energy based indoor positioning and one specific challenge that follows it. BLE spectrum consists of 40 different channels: 3 advertising channels and 37 data channels. In positioning, advertising channels are utilized in BLE peripherals to transfer packets to the main device for estimating the location of it. The problem is that the main device receives an aggregation of signals from all three advertising channels instead of receiving them separately alongside their source channel. This causes some level of instability and inaccuracy in indoor positioning systems.

We examine multiple approaches in this thesis to analyze this characteristic of BLE advertising channels and try to mitigate its effect by making use of algorithms to provide us the source channel information of each received signal strength. During the research, data collection with different devices is done for different types of analysis. First, we investigate whether there is any pattern visible in the data which indicates different advertising channels and compare the received signals of two different devices, one is a regular mobile phone and the other one is a special tool that reports channels separately while receiving signals. Next, we use an interval averaging approach to reduce the effect of sudden channel changes in positioning. Then a Gaussian method is used to separate RSS values based on their source channel, make use of channel information in positioning and investigate whether they can improve the performance.

The results we obtained give a good understanding of different aspects of BLE advertising channels, both mathematically and intuitively. We also noticed that using mixture of Gaussian methods can result in reliable channel separation information and can improve the accuracy of positioning to some extent.

Keywords: Indoor positioning, Bluetooth low energy (BLE), RSS, Advertising channels

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ACF	Auto-correlation Function
AOA	Angle of Arrival
BLE	Bluetooth Low Energy
BR	Basic Rate
$\mathbf{c}$	Speed of Light
λ	Carrier Wavelength
f_c	Carrier Frequency
CH	Channel
B_c	Coherence Bandwidth
T_c	Coherence Time
dBM	decibel milliWatt
d	Distance between Antennas
DK	Development Kit
f_D	Doppler Shift
DPSK	Differential Phase Shift Keying
DQPSK	Differential Quadrature Phase-Shift Keying
EDR	Enhanced Data Rate
EM	Expectation Maximization
$\mathbf{G}$	Gain of Antenna
GFSK	Gaussian frequency-shift keying
GNSS	Global Navigation Satellite System
HIRM	HERE Indoor Radio Mapper
IO	Interacting Object
IPS	Indoor Positioning System
IR	Infrared
LOS	Line of Sight
MAC	Media Access Control

OMGP	Overlapping Mixture of Gaussian Processes
P_r	Received Signal Power
P_t	Transmitted Signal Power
PACF	Partial Auto-correlation Function
PDU	Protocol Data Unit
PL	Path Loss
RF	Radio Frequency
σ_T	RMS delay spread
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
TDOA	Time difference of arrival
TOA	Time of arrival
UWB	Ultra Wide Band
VLC	Visible Light Communication
WLAN	Wireless local area network
WPAN	Wireless Personal Area Network

1 INTRODUCTION

1.1 Background

In today's modern world, position-based services are in high demand. GNSS plays an essential role in localization and navigation. Electronic devices such as smart phones on Earth receive the radio signals transmitted by GNSS satellites and the location is determined with an acceptable accuracy. Although the global navigation satellite system (GNSS) is able to fulfill the requirements for almost all outdoor cases, indoor localization is still considered a challenging case. GNSS is not a suitable choice for indoor positioning tasks, since the signal can get easily blocked, attenuated or reflected in the indoor environment [1, 2, 3].

The huge number of sensors and equipment have provided the current mobile devices a significant level of awareness of indoor environment, which is a great opportunity for indoor positioning field [4]. Currently, indoor positioning methods mainly depend on pre-installed infrastructures and result in acceptable precision. Ultra-wideband (UWB), Bluetooth and Wireless Local Area Networks (WLAN) technologies are among the popular indoor positioning methodologies, each having their own advantages and disadvantages. One of the highly used methods is Bluetooth Low Energy (BLE) which is very well-known for its availability, low cost and low power consumption [5].

BLE devices use point-to-point and broadcast modes to communicate via data channels and advertising channels respectively. Out of 40 RF channels in BLE, three are advertising channels (37, 38 and 39) [6, 7]. In positioning use cases, only broadcast mode hence advertising channels are needed. Through the advertising channels, a transmitter broadcasts packets of data and receivers collect the packets for further process. Advertising packets are sent at fixed intervals on all three advertising channels one after the other. On most of mobile phones though, instead of catching signals separately, the aggregation of the signals from all three channels is received [7].

BLE technology can be utilized in different scopes such as range-based and fingerprinting-based. In this thesis we only consider fingerprinting-based scope. Fingerprinting is mostly split into two steps, namely training and positioning. During the training step, signals are collected to create radio maps which are used later in the positioning phase.

1.2 Thesis objectives

Currently, mobile devices used for positioning collect aggregate received signal strength (RSS) values from all three advertising channels. However, positioning would be more stable and reliable if RSS values from separate channels were separately available.

A problematic point of BLE-based positioning is that BLE beacons generate advertising packets through three different advertising frequency channels [6]. However, little attention is drawn to this aspect of BLE channels in the indoor positioning domain, considering the fact that these three channels are uncorrelated and in some cases the difference of RSS values from two channels can be considered high. This leads to fluctuations in positioning performance, since during the offline phase of positioning, RSS values might have been collected from one channel, for instance channel 37, and during the online phase from another channel, for instance channel 38, where the RSS values happen to differ considerably compared to channel 37. Some studies are done to address this issue by adjusting the firmware, to make the BLE beacons advertise only in one channel instead of three [8, 9] or by using Apple's iOS 8 operating system, that happened to provide channel information [10]. These approaches are not ultimate solutions to the problem because the former requires adapting the beacon to operate on one of the three channels, which is not recommended by the manufacturers. This is because, if the selected channel is blocked, device does not work anymore, and the latter depends on a specific mobile device which is now obsolete and Apple has cut out the channel information from currently used devices.

In this thesis, the effect of this phenomenon on the performance of BLE indoor positioning is examined. In addition, some methods are utilized to address this issue. The thesis outline is as follows:

- Chapter 2 discusses multiple properties of radio propagation, different characteristics of wireless channels and BLE technology in order to provide a better understanding of the problem.
- Chapter 3 focuses on the indoor positioning technologies and techniques, including the BLE-based fingerprinting approach which falls within the scope of this thesis.
- Chapter 4 contains the data collection, data pre-processing and some initial data analysis.
- Chapter 5 includes further analysis and approaches used to improve indoor positioning performance.
- Chapter 6 consists of the tables of results obtained from the previous chapter's investigations.

2 RADIO PROPAGATION AND BLE PROPERTIES

2.1 Radio propagation

The simplest case in the propagation of electromagnetic waves is free space propagation, in which two antennas, one in the transmitter end and one in the receiver end, operate in free space [11]. Under more realistic circumstances, the propagation behavior can be altered by Interacting Objects (IOs) such as conducting and dielectric obstacles. In other words, in urban environments, radio propagation is very complicated since it contains reflected and diffracted waves generated by multipath propagation.

In radio propagation, three main phenomena might occur: reflection, diffraction and scattering [12, 13]. Reflection happens just as a radio wave in a medium collides with another medium with dissimilar electromagnetic attributes. The phase and amplitude of the reflected wave are dependent on the medium's properties, the incident angle and the polarization of electric field. The reflected wave is weakened because a part of the wave energy might propagate through the reflecting surface or get absorbed by it. Diffraction occurs when propagating radio waves lean toward another direction or deviate when close to an obstacle. Lastly, scattering is a phenomenon by which the radio wave energy spreads out in all directions in the event of collision with a rough surface or an obstacle with small size. Scattered waves can be considered as another radio source in the receiver side [14]. Figure 2.1 illustrates these three phenomena in a simple manner.

2.1.1 A wireless channel's characteristics

In order to avoid model complexities due to different propagation paths, wireless channels are mainly defined by three dominant effects: path loss that characterizes the dissipation of signal power as the propagation distance increments, shadowing which characterizes the effect of sizable objects in the environment and fading which defines the effect of signal replicas received from multiple propagation paths in the receiver side. Shadowing and path loss are labeled as large-scale propagation effects and fading is labeled as small-scale propagation effect [13]. In this section, large-scale propagation effects are very briefly explained as the main focus is on the fading effect.

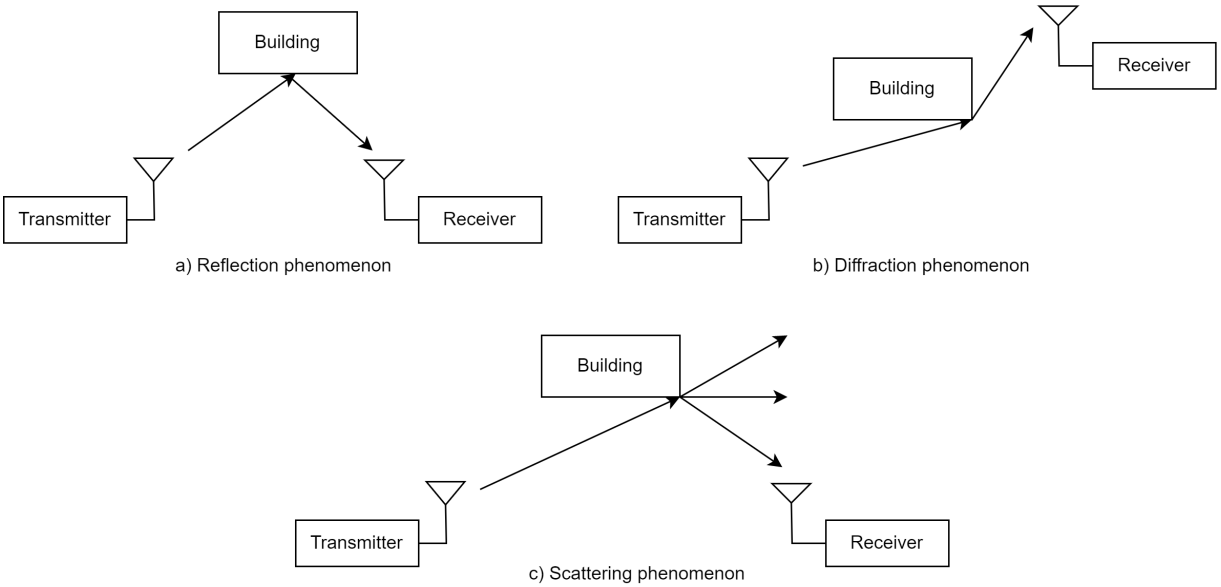


Figure 2.1. *Reflection, diffraction and scattering phenomena in radio propagation*

Path loss

Path loss models describe the decrease in the signal power as the propagation distance increments. These models usually neglect the multipath effects and environmental factors which are taken care of by fading and shadowing models, respectively. The simplest way to define a path loss model is to divide the transmitted signal power by the received signal power.

$$PL = P_t/P_r \tag{2.1}$$

The dB value of path loss can be obtained using the logarithmic definition as follows:

$$PL_{dB} = 10\log_{10}(P_t/P_r) \tag{2.2}$$

So far, different path loss models have been suggested and utilized in calculating the path loss at a particular distance from the source of transmission. One of the known and inaccurate models is the free space path loss model which utilizes Friis equation to calculate the path loss in a certain distance in free space environment. In this model, it is considered that there is only a single direct line-of-sight (LOS) path between the receiver and transmitter, while in the real-world scenario there are several paths between them which make the path loss model more complicated. The path loss model utilizing Friis equation could be written as

$$PL = (4\pi)^2 d^2/G_t G_r \lambda^2 \tag{2.3}$$

and correspondingly in dB-scale as

$$PL_{dB} = 21.98 - 10\log_{10}(G_t G_r) - 20\log_{10}(\lambda) + 20\log_{10}(d) \tag{2.4}$$

where λ is the carrier wavelength, d represents the distance between the antennas and G_r and G_t are the gains of antennas in the receiver and transmitter side, respectively.

Shadowing

Shadowing defines the effect of huge items such as buildings, cars and people in the propagation environment. Due to the existence of sizable items in the propagation area, the received signal power at certain receiving points the same distance from the transmitter might be diverse. Since the size and other features of the objects are usually unknown, shadowing effect is better to be explained in a statistical manner.

Multipath fading

Multipath fading happens when a transmitted signal is reflected by obstacles and is received as multiple signals with random phase offsets due to the fact that reflected signals have taken various routes to the receiver. This could lead to either an amplified or degraded signal in the receiver side as the reflections might have constructive or destructive effects on each other [14].

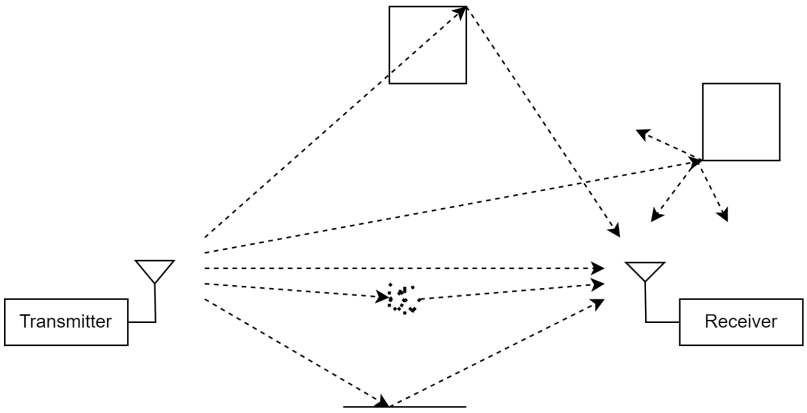


Figure 2.2. *Multipath radio propagation*

Two main characteristics can be observed in multipath fading: time-domain variation and delay spread. Alterations in the propagation environment and the relative motion of the receiver and the transmitter cause time-domain variation, while the delay spread is the result of signal experiencing different amounts of delay due to various propagation paths, hence the received signal is the sum of multiple signal replicas with various delays. Multipath fading could be split into fast/slow fading and flat/selective fading.

Doppler effect

In order to understand the time-domain variation characteristic, a basic knowledge of Doppler effect is required. Doppler effect is the outcome of relative motion between the signal source and the observer. It is important to remember that Doppler effect does not indicate an actual change in the carrier frequency, it just appears to receiver that frequency is different. The following equation presents the quantity of phase change in the signal.

$$\Delta\phi = 2\pi \frac{\nu \cos\theta_i}{\lambda} \Delta t \tag{2.5}$$

In the above-mentioned equation, ν represents the speed of moving receiver (relative motion), λ stands for the carrier wavelength, θ_i is the angle created by the direction of movement and the direction of received signal and Δt is the time period over which the signal transmission is done. Using the equation defined above and $\lambda = c/f_c$ (where f_c and c stand for the carrier frequency and the speed of light, respectively), the Doppler frequency shift is identified as

$$f_{D_i} = \frac{1}{2\pi} \lim_{\Delta t \to 0} \frac{\Delta\phi}{\Delta t} = \frac{\nu \cos\theta_i}{\lambda} \tag{2.6}$$

The frequency of the received signal is the result of adding the carrier frequency and Doppler frequency shift. It is very clear that the maximum Doppler frequency shift happens if $\theta_i = 0$, which denotes that the direction of the incoming signal and the movement are opposite. The maximum Doppler frequency shift which is basically equivalent to the highest speed variation of channel phase and the largest potential spread of signal bandwidth can be calculated as

$$f_D = \frac{\nu}{\lambda} \tag{2.7}$$

The combination of Doppler effect and multipath propagation results in an increase in the signal bandwidth, as the signals received from various paths experience varying degrees of Doppler shift. Now, going back to the time-domain variation property of fading, it is worth to note that the speed of phase shift variation determines the rate of time-domain channel variation.

Multipath delay spread

Referring back to the previous sections, multipath could lead to destructive or constructive interference of the signal instances from various paths at the receiver. Each of the received signals arrives at the receiver with a different delay. The variation of these delays

could be up to thousands of nanoseconds.

The range over which the signal instances are spread along the delay axis is commonly known as the delay spread. In other words, delay spread can be described as the time difference between the arrival of the first signal instance and the last signal instance arrived. One useful metric derived from delay spread, is the RMS (root mean square) delay spread σ_T, which can be considered as the standard deviation of the delay of multipath components, with each delay being weighted by the amount of contribution it has in the energy of the received signal and is calculated as

$$\sigma_T = \sqrt{\frac{\sum_{n=1}^{N} \alpha_n^2 (\tau_n - \mu_T)^2}{\sum_{n=1}^{N} \alpha_n^2}} \tag{2.8}$$

where τ_n and α_n are the delay and amplitude scaling factor of the n^{th} path respectively, and μ_T is the average delay and defined as

$$\mu_T = \frac{\sum_{n=1}^{N} \alpha_n^2 \tau_n}{\sum_{n=1}^{N} \alpha_n^2} \tag{2.9}$$

Coherence time and bandwidth

In order to comprehend the difference between slow and fast fading, coherence time and bandwidth are necessary to be understood. Coherence bandwidth B_c is the frequency band where two frequency components may have a high probability of amplitude correlation [15] and coherence time T_c is the time duration over which the channel response remains strongly correlated [16].

Coherence time and coherence bandwidth can be approximated as below:

$$T_c = \sqrt{\frac{9}{16\pi}} \frac{1}{f_D} \approx \frac{0.4}{f_D} \tag{2.10}$$

$$B_c \approx \frac{1}{2\pi\sigma_T} \tag{2.11}$$

where f_D is the maximum Doppler shift and is calculated by $f_D = \frac{v}{c} f_c$ where f_c is the carrier frequency. In 2.11 σ_T is the RMS delay spread of the signal.

Slow/fast fading

Maximum Doppler shift basically defines the highest rate of channel phase variation, hence it can also be a proper means for measuring the rate of time domain variation

of the fading channel.

As mentioned earlier, coherence time can be approximately calculated using the maximum Doppler shift. In other words, both f_D and T_c can be recognized as quantitative measures of time-domain channel variation rate. Essentially fading channels are divided into two categories, slow and fast. T_c can help us check whether the channel variation is slow or fast. A signal can be considered to experience slow fading if the coherence time T_c is much larger than the symbol duration T_s, or if the symbol bandwidth B_s is much greater than the maximum Doppler shift f_D. Otherwise, we can claim that the fading is fast.

Frequency-flat/selective fading

Using the metric RMS delay spread, the fading channels can be divided to the frequency-flat and frequency-selective. Frequency-selective fading occurs when the symbol duration T_s is shorter than the RMS delay spread σ_T . Assuming that, the propagation will undergo a considerable amount of interference of signal instances. On the contrary, if σ_T is smaller than T_s which means the delay spread is not significant enough, the channel experiences the frequency-flat fading.

2.2 Bluetooth Low Energy

Bluetooth is among the most popular wireless technologies. The huge growth in the new use cases and devices made the Bluetooth Special Interest Group and other corporations realize that Bluetooth does not work very efficiently in some applications. The inefficiencies included too much power consumption and long connectivity time. In 2010, Bluetooth low energy (BLE) was introduced. Although BLE shares some features with Bluetooth Classic, it does not have the complexity and cost problem of the Bluetooth anymore. As mentioned in chapter 1, BLE operates in two different modes, advertisement and connection establishment. In this thesis, our main focus is on advertisement mode.

Knowing how BLE Advertising operates can assist us minimize the power consumption, boost the reliability and performance accuracy of BLE-based positioning. BLE communicates in two different ways. One is through advertisement, in which a BLE beacon or any other BLE peripheral device sends out packets to all the receivers in its vicinity. The receiver then processes this received data or requests for connection to receive further information. The other way of communication is through connecting to another device directly, where both central and peripheral devices transfer and receive packets.

2.2.1 BLE Physical Layer

BLE Physical Layer operates the transmission of radio signals, hence it is beneficial to know more about it before going through how the advertisement process is done.

Bluetooth Low Energy has some common aspects with Bluetooth Classic Basic. Both work in 2.4GHz bandwidth and use GFSK modulation at 1Mbps with different modulation index. On the other hand, Enhanced Data Rate (EDR) operates using a totally different modulation system than GFSK. While Bluetooth Classic has 79 channels, BLE has only 40 channels. Moreover, the channels are spaced differently. There are some devices such as Dual Mode Radio instruments that support both BLE and Classic technologies by changing their parameters of modulation and the channels on which they are operating.

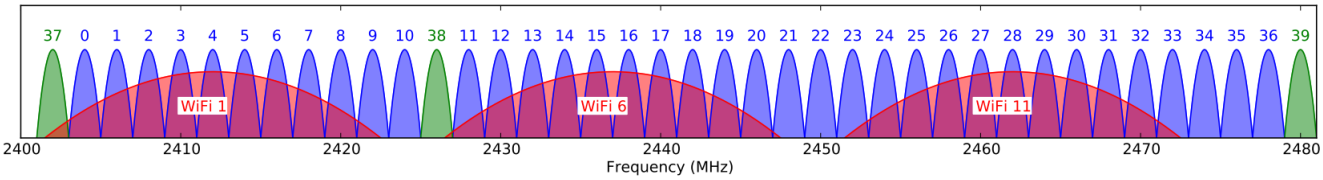


Figure 2.3. *BLE spectrum and channels*

Bluetooth low energy spectrum extends from 2402MHz to 2480MHz. It contains 40 channels, each separated by 2MHz [5, 7]. Channels 0-36 are called data channels and are used for bi-directional connection and data exchange while channels 37, 38, and 39 are called advertising channels and are aimed to send advertisement packets. Our area of focus is on advertising channels, how they operate and how they impact accuracy in BLE-based indoor positioning.

Table 2.1. *Frequency values of advertising channels*

BLE Channel	Frequency Value
37	2402
38	2426
39	2480

During advertisement, a BLE peripheral device transfers the same packet three times each time on one of the advertising channels, one after another. Central devices might be scanning and listening to advertising channels to receive packets and then process them. As shown in the figure 2.3, advertising channels (illustrated by green color) are not located in each other's neighborhood but spread across the 2.4GHz spectrum intentionally. The reason for that is if any of the advertising channels happens to get blocked, the other channels most probably are available to do the broadcast as they have enough distance from each other.

Table 2.2. *BLE and Bluetooth Classic features compared*

Bluetooth Low Energy		**Bluetooth Classic**	
		Basic Rate (BR)	**Enhanced Data Rate (EDR)**
Channels	40	79	79
Modulation	GFSK 0.45 to 0.55	GFSK 0.28 to 0.35	DQPSK / 8DSPK
Spacing	2 MHz	1 MHz	1 MHz
Data Rate	1-2 Mbps	1 Mbps	2-3 Mbps

This property of BLE advertising channels decreases the effect of interference with Wi-Fi, Microwaves, Bluetooth Classic and narrow band signals. Particularly, channel 38 is fixed between two Wi-Fi channels to prevent interference with Wi-Fi signals. This guarantees a successful transmission during advertisement.

2.2.2 BLE Advertisement Interval

In the advertising mode, packets are sent on advertising channels in a periodical manner. The time between each packet set is composed of a fixed time interval which is called the advertisement interval plus a random delay. The fixed time interval can be set to 20ms to 10.24 seconds. The random delay is a value randomly selected from 0ms to 10ms added to the interval. In the advertising mode, it is essential that packets are received without any sort of collision and critical interference. This random delay added to the interval helps decrease the probability of advertisements of different devices collide with each other. This BLE attribute boosts robustness by preventing collisions from happening.

One might assume advertising on only one or two channels instead of three would be more profitable from energy point of view, since transmitting packets on each channel has a certain cost. However, this idea has been frowned at by many companies. The reason behind it is that if the selected channels are blocked due to the interference, the advertising packets are not sent, and the advertisement process would not work properly. Which is why advertising is strongly recommended on all three channels.

2.2.3 BLE Advertising Packets

BLE packets consist of several parts, including preamble, access-address, PDU and CRC. PDU (Protocol Data Unit) is the one that is adjusted depending whether the packet being sent is on advertising mode or data mode. PDU is divided to a 2-byte header and a payload section which varies from 6 to 37 bytes. Several parts build up the header section, from which PDU type is the most generally known. It is useful to know that there are several types of PDU used for advertisement. Table 2.3 contains some of PDU types, their names and physical channels [7]. ADV_IND and ADV_NONCONN_IND are com-

Table 2.3. *PDU types, names and their physical channels*

PDU Type	PDU Name	Physical Channel
0b0000	ADV_IND	Primary Advertising
0b0001	ADV_DIRECT_IND	Primary Advertising
0b0010	ADV_NONCONN_IND	Primary Advertising
0b0011	SCAN_REQ	Primary Advertising
	AUX_SCAN_REQ	Secondary Advertising
0b0100	SCAN_RSP	Primary Advertising
0b0101	CONNECT_IND	Primary Advertising
	AUX_CONNECT_REQ	Secondary Advertising
0b0110	ADV_SCAN_IND	Primary Advertising
0b0111	ADV_SCAN_IND	Primary Advertising
	AUX_ADV_IND	Secondary Advertising
	AUX_SCAN_RSP	Secondary Advertising
	AUX_SYNC_IND	Periodic
	AUX_CHAIN_IND	Secondary Advertising and Periodic
0b1000	AUX_CONNECT_RSP	Secondary Advertising

monly used; thus, it is worthwhile to explain them here. In the following chapters, you will see that in this thesis, our focus is mainly on ADV_NONCONN_IND.

ADV_IND is the most common PDU for advertisement. Its main characteristic is that multiple central devices can connect to the peripheral device. In other words, the peripheral device is not directed towards a specific central device for sending packets. Once the central device receives the advertisement, it can start the connection process. ADV_NONCONN_IND is utilized when there cannot be any connection between peripheral and central devices. It is user's choice to use which PDU based on their application, whether it requires connection between devices or needs to avoid it. The content and size of the payload relies on the PDU type. It consists of a 6-byte advertisement address which could be either the device's public or random address, and 31 bytes of advertising data structure including the type, length and the advertising data itself. Figure 2.4 illustrates the format of a BLE advertising packet [7].

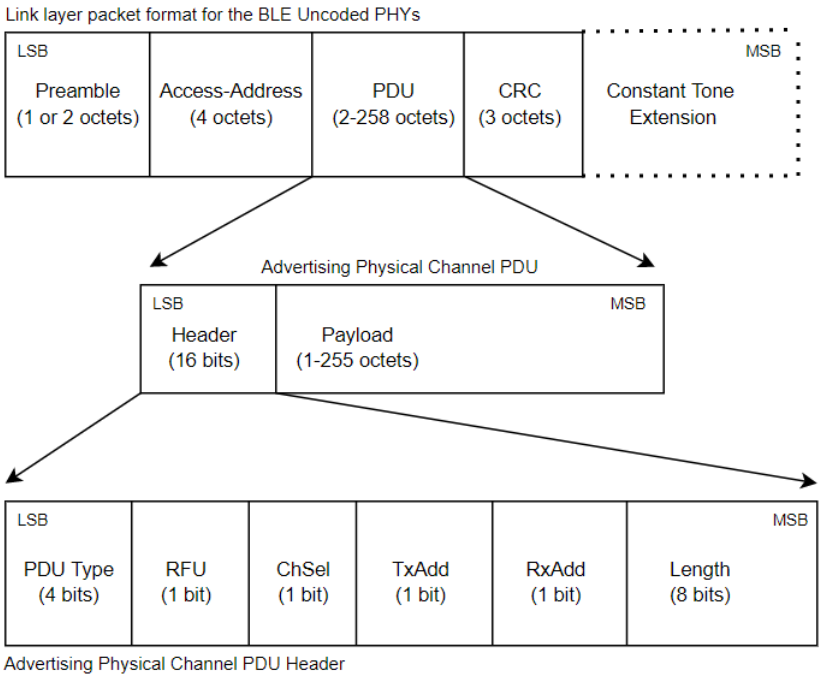


Figure 2.4. *BLE advertising packet format*

3 INDOOR POSITIONING

An Indoor positioning system (IPS) is utilized to establish the position of objects in an indoor environment. Indoor environments exhibit a greater degree of complexity as compared to outdoor environments due to the walls and all other different objects which make the multipath effect more problematic. However, indoor environments have some characteristics that outdoor environments do not, which can ease the positioning. For instance, the fixed infrastructure and less dynamic surroundings features are more compliant with positioning.

So far, various applications have been identified for indoor positioning systems. Being used in the private houses, they can aid disabled or elderly people with their everyday routine or health monitoring in case of emergency [17]. In public venues such as airports, shopping centers and museums, it can provide navigation and tracking systems to guide and keep track of objects. Moreover, in hospitals, IPS can be utilized to track patients and expensive equipment, also provide positioning for rescue teams to save lives [18, 19].

3.1 Performance metrics

Like any other system, IPSs have some metrics to evaluate their performance. In addition to accuracy, scalability, cost, robustness and complexity are the measures that can determine the efficiency of an IPS. This section provides a brief explanation of various performance metrics.

Accuracy

The term accuracy refers to the proximity of compliance between the actual value of a quantity and its measured value. In an IPS, the accuracy can be introduced as the mean Euclidean distance between the true location and the estimated one [20]. Although accuracy is an important metric for IPSs, usually there is a trade-off among metrics. Therefore, there might be some compromise needed between accuracy and another metric.

Scalability

The scalability metric guarantees that an IPS functions normally when the scope of positioning gets larger in either the number of users or the geographical coverage. As the area of coverage or the density of users increases, there might be some extra infrastructure needed or more computation might be required to estimate the location. The IPS should be able to scale in case of necessity [20].

Cost

The cost of an IPS could be attributed to money, energy, space and time. A certain amount of cost is required for system installation and maintenance. However, in some cases, the infrastructure is already installed for a previous purpose and can be reused for positioning purposes as well which saves some resources and effort. Moreover, some systems demand more space and are not suitable in case of having limited space. Among these costs, energy is one of the important ones. Some IPSs are more energy efficient which makes them operate without being dependent on too much energy. BLE-based IPSs are among the energy efficient systems.

Robustness

There are cases where an IPS has to use fragmentary data to locate the position. Not all the signals might arrive to the receiver due to transmitter malfunction or blockage by environmental conditions. A reliable indoor positioning system should be capable of functioning well in any of those cases.

Complexity

Different forms of complexity could be identified for indoor positioning systems, such as complexity in software, hardware or operation [20]. One example of software complexity which is within the scope of this thesis, is computation complexity. The system is computationally complex if a lot of power is required to calculate the position of objects using a positioning algorithm. To measure the computation complexity, we usually consider the time of computation. The more time an IPS spends on locating the items, the more complex the system is. This leads to a positioning system with a non-negligible lag while lag is the delay between the reported location of a moving device and the actual time of arriving to a new spot.

3.2 Indoor positioning technologies

Indoor positioning systems can be categorized into different types based on different criteria. One criterion is the technology they utilize to perform the task of positioning. There are various types of technologies, both radio-based and non-radio-based. In the following sections, some of these technologies are shortly explained. However, our main focus is on BLE technology which is a radio-based technology.

3.2.1 Radio frequency technologies

Radio frequency (RF) technology is among the popular technologies used in indoor positioning systems. They can operate in narrow-band mode which contains a small range of frequencies or in wide-band mode. Some of the popular RF approaches used in IPSs are Wireless Local Area Network (WLAN), Bluetooth and Ultra-wideband (UWB).

Wireless Local Area Network

Wireless local area network (WLAN) is the IEEE 802.11 standard and it is dominantly used in indoor positioning systems. One of the benefits of a WLAN-based IPS is the opportunity to utilize the pre-existing WLAN infrastructure and the standard mobile devices for indoor localization purposes. WLAN technology operates in 2.4GHz and has the coverage range of 50-100m, with the bit rate of 11, 54, or 108Mbps. Another benefit of WLAN-based IPS is that line of sight is not needed. The most widely used WLAN-based positioning approach is RSSI (Received Signal Strength Indicator), since it is easy to extract in 802.11 networks. Approaches like Angle of Arrival (AOA), Time of Arrival (TOA) and Time Difference of Arrival (TDOA) are less popular, due to the complexity in angular and delay measurements. The accuracy of a typical WLAN-based IPS is usually 3-30m and the rate of the update is in the order of few seconds.

Bluetooth

Bluetooth is the IEEE 802.15.1 standard for wireless personal area networks (WPANs) handled by the Bluetooth Special Interest Group (SIG). It operates in 2.4GHz and has a coverage of 10-15m, which is considerably smaller compared to WLAN technology. However, similar to WLAN, Bluetooth is ubiquitously used for different purposes and is already embedded in many devices and infrastructures, which makes it a great option for indoor positioning systems. Bluetooth has some advantages that make it a competitor to WLAN-based positioning systems. Especially, Bluetooth low energy (BLE) which was extensively discussed in the previous chapter benefits from features such as low power consumption, high availability and low cost. This thesis focuses on a BLE-based IPS and aims to solve a problem regarding the BLE advertising channels.

Ultra-wideband

Ultra-wideband (UWB) is a technology based on the transmission of short pulses over a wide frequency bandwidth. The short pulses make this technology resistant to multipath which results in higher accuracy. The reason behind this is that short pulses of UWB are simple to filter in order to decide which signal is the original one and which one is the outcome of multipath. Another benefit of using UWB in indoor positioning is that the signal does not get blocked by the walls or furniture. However, it can cause some interference in liquids or metallic materials which can be fixed by placing UWB readers in the right place [20].

3.2.2 Other technologies

In addition to the radio-based technologies, there are some non-radio-based mechanisms that use light, sound or magnetic fields.

Light-based indoor positioning technologies could use either visible light or infrared. Infrared (IR) IPSs are among the popular ones, since IR is widely available and is used for different purposes. Although IR is cheap and highly accessible, unlike UWB technology it is unable to penetrate through walls and it requires line-of-sight. Similarly, visible light communication (VLC) conveys information through light and depends on line-of-sight. LED lamps are proven to be the best among the lamps [21]. Data can be transferred via encoded light by switching on and off the light source in fixed intervals. These changes in light might not even be noticeable to human eyes. The sensor in the receiver compares the received encoded light with different encoding schemes and selects the dominant one.

Sound-based indoor positioning technologies could use either audible sound or ultrasound. It is possible to use sound like light to encode data in an IPS. The position can be determined using the time it takes for the signal to reach the receiver from the transmitter. The idea of using audible sound might seem not practical since it could disturb the people nearby. However, marking an already available sound in buildings is an appropriate approach which might not even be perceived by humans. Ultrasound technology does not suffer from this problem since its frequency range is higher than audible sound which is not detectable for humans.

There are two ways of using magnetic fields in an IPS, one is to use synthetically generated magnetic fields and another one is to use the Earth's magnetic field. Using a magnetometer, the variations in the magnetic filed can be measured and based on that the location of the object is obtained.

3.3 Indoor positioning techniques

Armed with one or multiple technologies mentioned in the previous section, an indoor positioning system uses one or multiple techniques to estimate the location. There have been several techniques known to be used in IPSs from which some can provide us the proximity position such as the proximity technique where others can provide us the absolute and relative as well as proximity positions such as triangulation, fingerprinting and vision analysis techniques. In this thesis our focus is on the fingerprinting approach.

3.3.1 Triangulation

In the triangulation approach, geometric attributes of triangles are used to calculate the location of an object. Triangulation can be sub-categorized to lateration and angulation [20]. In lateration technique which is also named range measurement sometimes, the location of an object is calculated by evaluating its distance from various reference points. In order to determine the position of an object in two dimensions we need to be able to measure the objects distance from three non-collinear points. Similarly, in three dimension cases, four non-coplanar points are needed. In addition to measuring the distance between the object and reference points using RSS values (signal attenuation-based method), time difference of arrival (TDOA) and time of arrival (TOA) can be measured, based on which the distance can be calculated by estimating the signal attenuation or multiplying the transmission time and the transmission velocity. On the other hand, in the angulation technique, which is known as the angle of arrival as well, the position of an object is estimated by calculating angles relative to various reference points.

Time of arrival

In time of arrival (TOA) approach, the distance is derived from the time over which signal travels from the transmitter (any reference point) to the measuring unit (the object with unknown location). Using the measured time and the transmission speed, the distance is calculated as

$$d = c * TOA \tag{3.1}$$

where d is the distance and c is the speed of light. As shown in figure 3.1 [20], in two dimensional positioning, we need to make use of at at least three reference points to estimate the distance of an object. It is also important to make sure that the receiver and transmitter are synchronized together [22].

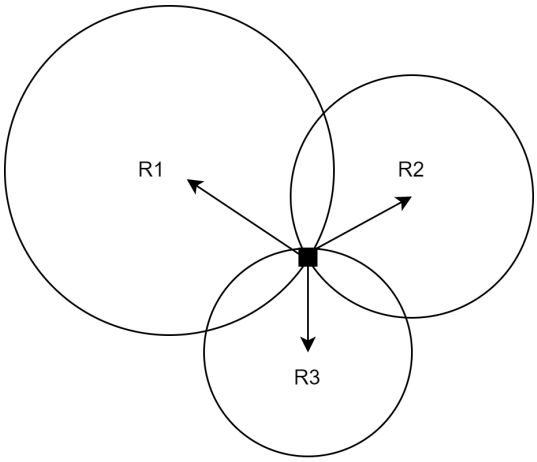


Figure 3.1. *TOA approach in positioning*

Time difference of arrival

Although in TOA approach the synchronization between transmitter and receiver is essential, it is not necessary in time difference of arrival (TDOA) approach. However, all the transmitters should be in sync. That is because TDOA method calculates the distance based on the difference of signal arrival time of signals received from multiple transmitters. Figure 3.2 illustrates a cellular system with TDOA setup. The distance between the receiver and each transmitter can be defined as below [23],

$$d_1 - d_2 = c * TDOA_{12} = c * (TOA_1 - TOA_2) \tag{3.2}$$

$$d_1 - d_3 = c * TDOA_{13} = c * (TOA_1 - TOA_3) \tag{3.3}$$

$$d_2 - d_3 = c * TDOA_{23} = c * (TOA_2 - TOA_3) \tag{3.4}$$

Where, d_1, d_2 and d_3 are the receiver's distance from the transmitters 1, 2 and 3 respectively and c represents the speed of light. In these equations, $TDOA$ of each two transmitters is known. This can be simply solved by systems of equation method with three variables, d_1, d_2 and d_3.

Received signal strength

Using the two methods mentioned above, we might face some problems. Radio propagation in indoor environments is highly affected by multipath effect. Therefore, measuring time and angle of arrival suffers from multipath, which leads to lower positioning accuracy. In addition to that, it might not always be possible to have access to line of sight propagation due to the indoor environment's restrictions. A substitute method could be measuring the attenuation of the transmitted signal. This method tries to explain the signal path loss (the difference between the transmitted and the received signal strengths)

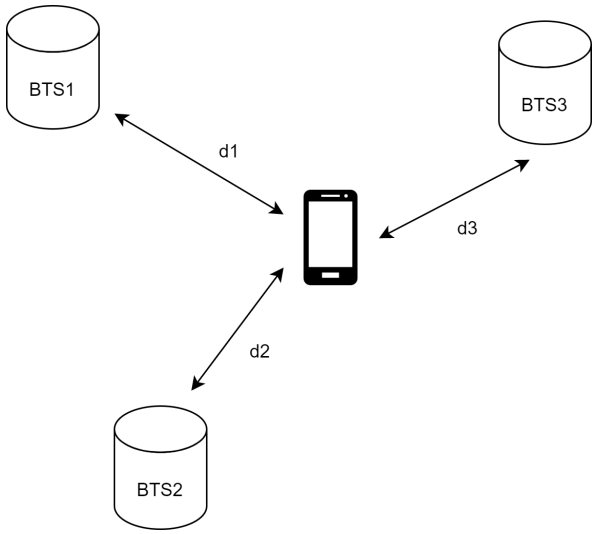


Figure 3.2. *TDOA approach in positioning*

as a range estimate.

Angle of arrival

In angulation approach which is usually referred to as angle of arrival (AOA), the location of an item is determined by the signal arrival angles of at least two transmitters with known location. As illustrated in 3.3, only two reference points are necessary to estimate the position of the target in 2-D case. One of the advantages of this technique is that synchronization between transmitter and receiver is not required. Another benefit is that in 2-D positioning, only two reference points are needed and in 3-D positioning, three reference points. Major disadvantages of this method are the requirement for complex hardware [22] and the decrease in positioning accuracy as the distance between the target and reference points grows.

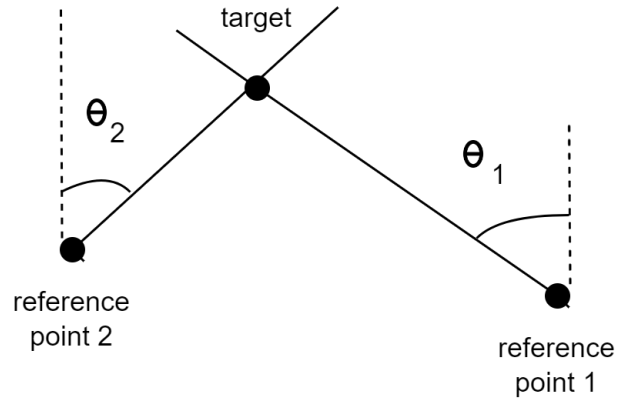


Figure 3.3. *AOA approach in positioning*

3.3.2 Proximity

As mentioned earlier, the proximity technique cannot provide us the absolute or relative location estimation of an object, but only a proximity location. The reason behind this is that this technique determines the position of a target only with respect to a known area or position. In simpler words, certain number of detectors are installed at known positions and once an object is detected in proximity of a detector, the proximity position of the object is obtained. For instance, assume there is a detector D1 fixed in room 3 of a building and it only covers the area of that room, object A is located in room 3 and object B is located in another room. In this case, the proximity location of object A can be obtained as room A as detector D1 detects it, while the position of object B is unknown.

3.3.3 Vision analysis

In the vision analysis technique, the target position can be determined by images taken from one or several points. To make this happen, there needs to be cameras installed in the area of coverage to take real-time pictures for analysis. Then the captured pictures are compared to the pre-measured database for estimating the target location. One benefit of this approach is the valuable information on the location context that can be extracted from the images.

3.3.4 Fingerprinting

Unlike methods mentioned above, fingerprinting is an approach that tries to match the fingerprint of a location-dependent feature of the signal, for instance received signal strength (RSS). This method was first proposed to increase the accuracy in indoor positioning by utilizing pre-measured location-dependent data.

BLE-based fingerprinting approach

The fingerprinting-based method in the field of indoor positioning is composed of two fundamental steps: offline and online. During the offline step which is also known as training step, fingerprints which are a set of RSS values at certain coordinates are collected and stored to build a radio map as a reference [24]. In this process, it is required to divide the floor plan into cells. Figure 3.4 illustrates the offline phase of positioning. 6 beacons are installed in an area which is divided to equal-sized cells.

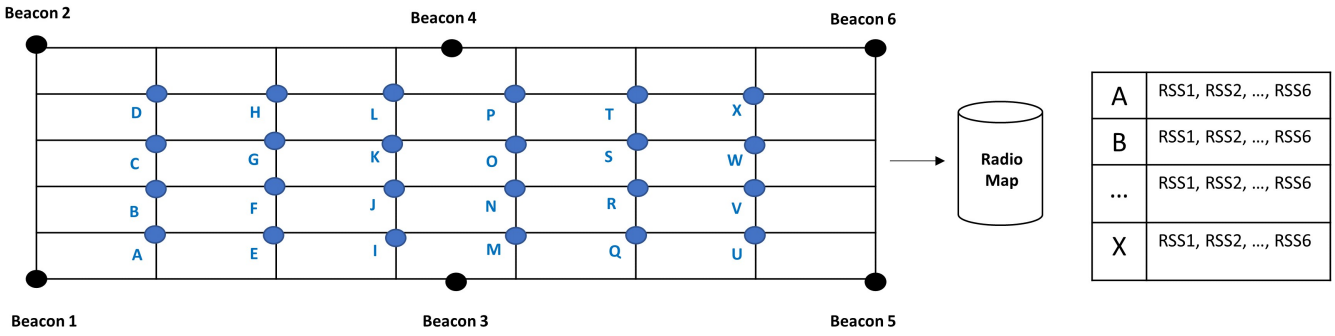


Figure 3.4. *Offline phase of positioning*

In the online step which is also referred to as run time step, new RSS measurements collected at an unknown coordinate are compared to the reference map to find a match for the unknown position [25], as shown in the figure below. The received signals from all the beacons at the unknown position marked with a star in figure 3.5, are compared to the RSS values of all known locations stored in the radio map and the closest known position, in this example position at point I, is determined. This technique could be challenging due to some propagation-related phenomena such as reflection and diffraction [18].

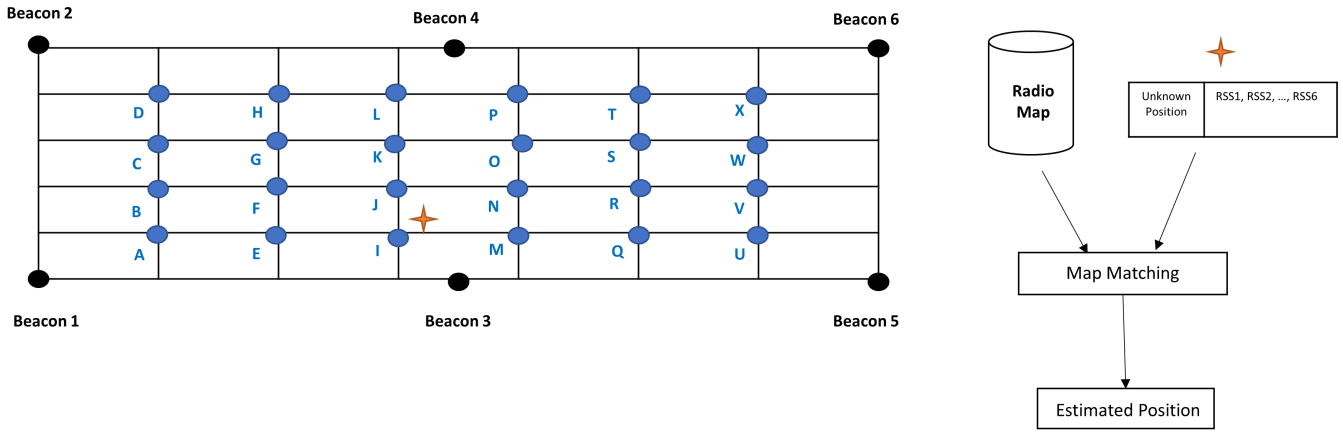


Figure 3.5. *Online phase of positioning*

4 DESCRIPTION OF RESEARCH DATA

This chapter contains data collection methods and the description of data.

4.1 Data collection

All the data used in this thesis is collected using HERE Indoor Radio Mapper (HIRM) mobile application and Nordic Semiconductor nRF52840 Development Kit.

4.1.1 HERE Indoor Radio Mapper application

HERE Indoor Radio Mapper (HIRM) is an Android-based application for collecting geo-referenced radio data on the Bluetooth™ beacon or WiFi access point signal strengths in buildings [26]. To use this tool, one needs to have an indoor map of the venue and Bluetooth beacons or WiFi access points installed in the building.

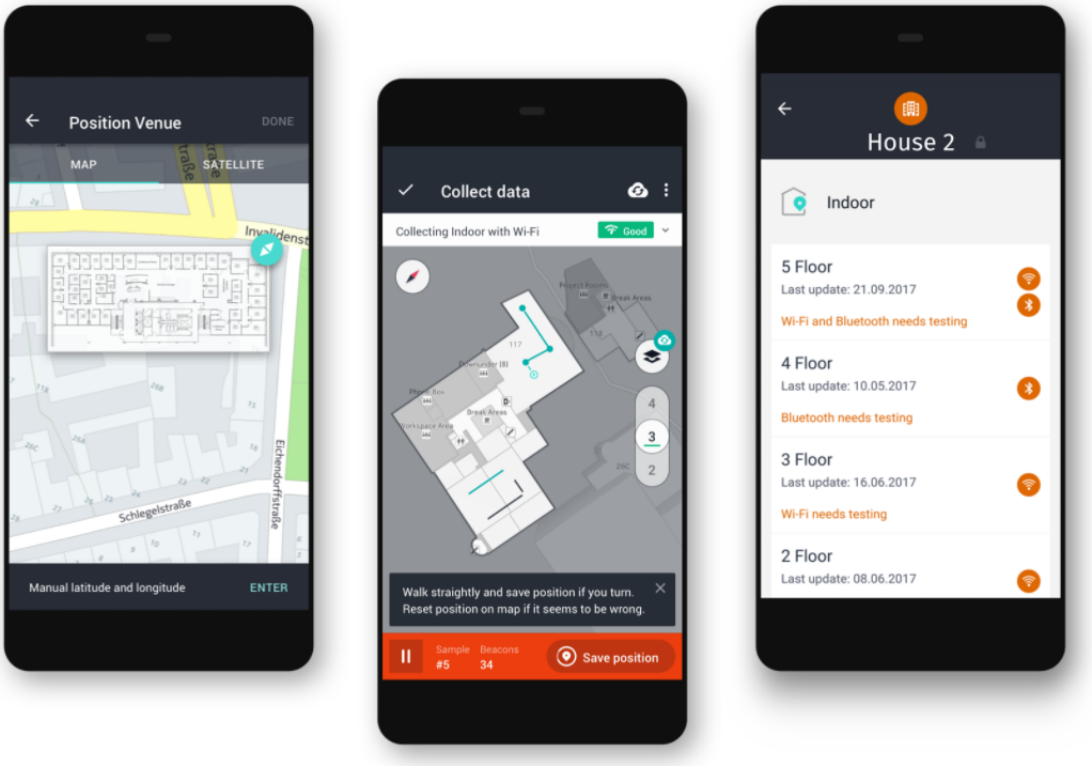


Figure 4.1. *HERE Indoor Radio Mapper application*

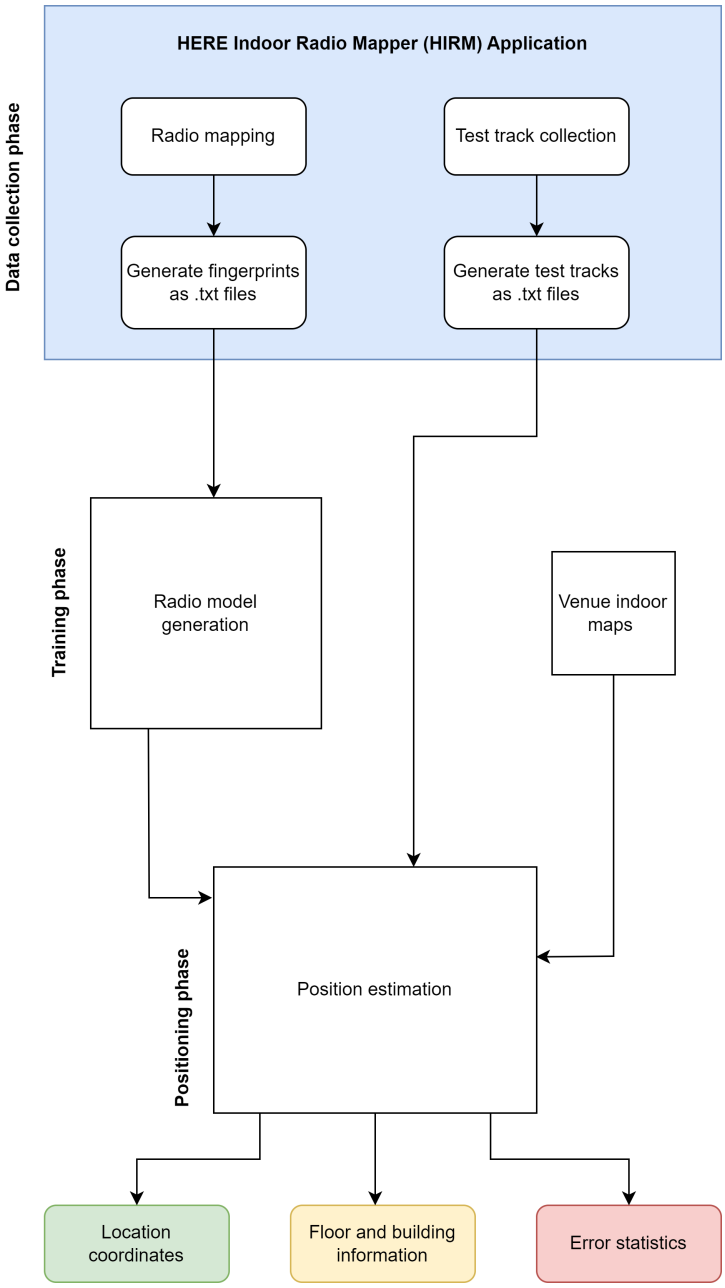


Figure 4.2. *Data collection and processing procedure [27]*

Figure 4.2 illustrates the whole process from data collection to estimation of location coordinates. After HIRM application does the radio mapping and exports fingerprints as text files, in the training phase, Radio model generation is done by MATLAB simulator. In this step, certain data compression methods are used to avoid computational complexity and decrease storage. Finally in the posirioning phase, the generated radio model, the venue indoor maps and collected test tracks using HIRM are fed to the MATLAB simulator to generate the location coordinates, floor and building information and error statistics.

The data collected using HIRM application on a mobile device is in the form of a text file containing rows and sub-rows. Rows contain building ID, latitude, longitude, time and some other information which is not used in this thesis. Each sub-row associated with a parent row contains the beacon Media Access Control (MAC) address, Received Signal Strength (RSS) value, beacon type and some extra information. In each row, the time at which the signal is received is reported. Accordingly sub-rows report the RSS values obtained by beacons identified by their MAC addresses at that specific time. In figure 4.3 you can see a piece of this text file.

```
3;DM_21537;61.4937910,23.7755802,2,0.00,;1.00,1.00,0,-1000.00,0.00;2020-06-08 06:55:25:723;;,,;,,;BLE;-;1;0
        EE010200082A,-60,Unknown device,BLE,0,1591599325577,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082A
        EE0102000816,-67,Unknown device,BLE,0,1591599325659,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000816
        EE0102000820,-84,Unknown device,BLE,0,1591599325211,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000820
        EE0102000815,-91,Unknown device,BLE,0,1591599325977,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000815
        EE0102000817,-92,Unknown device,BLE,0,1591599325618,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000817
        EE0102000821,-92,Unknown device,BLE,0,1591599325947,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000821
        EE010200082B,-94,Unknown device,BLE,0,1591599325404,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082B
        EE010200086A,-98,Unknown device,BLE,0,1591599325601,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000086A
        EE0102000862,-99,Unknown device,BLE,0,1591599325407,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000862
        EE0102000872,-102,Unknown device,BLE,0,1591599325944,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000872
4;DM_21537;61.4937968,23.7755805,2,0.00,;1.00,1.00,0,-1000.00,0.00;2020-06-08 06:55:26:236;;,,;,,;BLE;-;1;0
        EE010200082A,-66,Unknown device,BLE,0,1591599326448,1,EddystoneTLM,0201060303AAFE1116AAFE20000C151A0007A4E7AD4120D499
        EE0102000816,-67,Unknown device,BLE,0,1591599325659,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000816
        EE0102000817,-80,Unknown device,BLE,0,1591599326489,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000817
        EE0102000820,-81,Unknown device,BLE,0,1591599326079,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000820
        EE0102000815,-91,Unknown device,BLE,0,1591599325977,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000815
        EE0102000821,-92,Unknown device,BLE,0,1591599325947,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000821
        EE010200082B,-95,Unknown device,BLE,0,1591599326280,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082B
        EE0102000818,-96,Unknown device,BLE,0,1591599326384,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000818
        EE010200086A,-98,Unknown device,BLE,0,1591599326465,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000086A
        EE0102000869,-98,Unknown device,BLE,0,1591599326366,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000869
        EE0102000872,-102,Unknown device,BLE,0,1591599325944,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000872
5;DM_21537;61.4938025,23.7755808,2,0.00,;1.00,1.00,0,-1000.00,0.00;2020-06-08 06:55:26:743;;,,;,,;BLE;-;1;0
        EE010200082A,-66,Unknown device,BLE,0,1591599326448,1,EddystoneTLM,0201060303AAFE1116AAFE20000C151A0007A4E7AD4120D499
        EE0102000817,-80,Unknown device,BLE,0,1591599326489,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000817
        EE0102000816,-85,Unknown device,BLE,0,1591599326521,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000816
        EE0102000820,-86,Unknown device,BLE,0,1591599326932,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000820
        EE0102000815,-88,Unknown device,BLE,0,1591599326828,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000815
        EE010200082B,-95,Unknown device,BLE,0,1591599326280,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082B
        EE0102000818,-96,Unknown device,BLE,0,1591599326384,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000818
        EE0102000821,-97,Unknown device,BLE,0,1591599326823,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000821
        EE010200086A,-98,Unknown device,BLE,0,1591599326465,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000086A
        EE0102000869,-98,Unknown device,BLE,0,1591599326366,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000869
        EE0102000872,-100,Unknown device,BLE,0,1591599326816,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000872
6;DM_21537;61.4938083,23.7755811,2,0.00,;1.00,1.00,0,-1000.00,0.00;2020-06-08 06:55:27:247;;,,;,,;BLE;-;1;0
        EE0102000816,-64,Unknown device,BLE,0,1591599327386,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000816
        EE0102000817,-77,Unknown device,BLE,0,1591599327351,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000817
        EE010200082A,-77,Unknown device,BLE,0,1591599327301,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082A
        EE0102000820,-86,Unknown device,BLE,0,1591599326932,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000820
        EE0102000815,-88,Unknown device,BLE,0,1591599326828,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000815
        EE010200082B,-92,Unknown device,BLE,0,1591599327128,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082B
        EE010200086A,-96,Unknown device,BLE,0,1591599327331,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000086A
        EE0102000821,-97,Unknown device,BLE,0,1591599326823,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000821
        EE0102000872,-100,Unknown device,BLE,0,1591599326816,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000872
7;DM_21537;61.4938140,23.7755814,2,0.00,;1.00,1.00,0,-1000.00,0.00;2020-06-08 06:55:27:752;;,,;,,;BLE;-;1;0
        EE0102000816,-64,Unknown device,BLE,0,1591599327386,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000816
        EE0102000817,-77,Unknown device,BLE,0,1591599327351,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F3000000000817
        EE010200082A,-77,Unknown device,BLE,0,1591599327301,1,Eddystone,0201060303AAFE1516AAFE00004ADBD94E815DEE7E55F300000000082A
```

Figure 4.3. *A piece of mobile device data collected by HIRM in the form of a text file*

4.1.2 Nordic Semiconductor nRF52840 DK

The nRF52840 DK is a versatile development kit for BLE. I programmed this kit to use it as a BLE sniffer in order to collect RSS values from different sub-bands (channels) of BLE beacons.

The data collected using the BLE sniffer is in the form of an excel file. This file contains

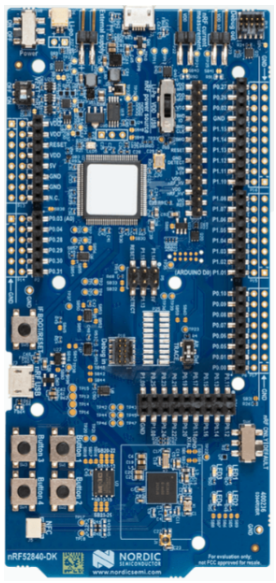


Figure 4.4. *Nordic Semiconductor nRF52840 development kit*

time, MAC addresses, channel numbers, RSS values and PDU types represented as Time, Source, Channel, RSS (dbm) and Info columns respectively in the excel file.

1	No.	Time	Source	PHY	Channel	RSSI (dBm)	Service Data	Info
2	1	0	ee:01:02:00:00:0a	LE 1M	39	-70	00004adbd94e815dee7e55f3000000002011	ADV_NONCONN_IND
3	2	0.003	cb:0d:7b:56:bc:fa	LE 1M	37	-76	00d88b9cc73c3ae747ef65bc0000000002840000	ADV_NONCONN_IND
4	3	0.005	cb:0d:7b:56:bc:fa	LE 1M	38	-61	00d88b9cc73c3ae747ef65bc0000000002840000	ADV_NONCONN_IND
5	4	0.006	cb:0d:7b:56:bc:fa	LE 1M	39	-83	00d88b9cc73c3ae747ef65bc0000000002840000	ADV_NONCONN_IND
6	5	0.019	01:5f:58:0a:4b:d6	LE 1M	37	-37		ADV_NONCONN_IND
7	6	0.02	01:5f:58:0a:4b:d6	LE 1M	38	-48		ADV_NONCONN_IND
8	7	0.021	01:5f:58:0a:4b:d6	LE 1M	39	-48		ADV_NONCONN_IND
9	8	0.028	7a:ba:c2:68:26:ff	LE 1M	37	-87		ADV_NONCONN_IND
10	9	0.029	7a:ba:c2:68:26:ff	LE 1M	38	-87		ADV_NONCONN_IND
11	10	0.03	7a:ba:c2:68:26:ff	LE 1M	39	-87		ADV_NONCONN_IND
12	11	0.033	ee:01:02:00:0a:1c	LE 1M	37	-50	00004adbd94e815dee7e55f3000000000a1c	ADV_NONCONN_IND
13	12	0.034	ee:01:02:00:0a:1c	LE 1M	38	-49	00004adbd94e815dee7e55f3000000000a1c	ADV_NONCONN_IND
14	13	0.035	ee:01:02:00:0a:1c	LE 1M	39	-50	00004adbd94e815dee7e55f3000000000a1c	ADV_NONCONN_IND
15	14	0.041	Shenzhen_22:df:17	LE 1M	37	-75		ADV_IND
16	15	0.042	79:5f:56:fb:21:81	LE 1M	37	-82		SCAN_REQ
17	16	0.042	Shenzhen_22:df:17	LE 1M	37	-75	6427114cb9ac233f22df17	SCAN_RSP
18	17	0.043	Shenzhen_22:df:17	LE 1M	38	-72		ADV_IND
19	18	0.044	Shenzhen_22:df:17	LE 1M	39	-73		ADV_IND
20	19	0.044	70:45:31:5a:b1:b7	LE 1M	39	-30		SCAN_REQ
21	20	0.045	Shenzhen_22:df:17	LE 1M	39	-72		SCAN_RSP[Malformed Packet]
22	21	0.045	76:5f:aa:42:1e:67	LE 1M	37	-84		SCAN_REQ
23	22	0.058	Slave	LE 1M	38	-93		Unknown
24	23	0.059	cf:2e:c7:45:0e:06	LE 1M	39	-86	10fd006865726500	ADV_NONCONN_IND
25	24	0.073	f3:ee:a7:2b:2b:52	LE 1M	37	-85	00e52abe5be3e0b6c33a1b5b1111111111910000	ADV_NONCONN_IND
26	25	0.083	fe:2e:b8:83:fc:a5	LE 1M	38	-86	00d88b9cc73c3ae747ef65bc0000000002680000	ADV_NONCONN_IND
27	26	0.085	fc:d8:83:ce:11:17	LE 1M	39	-81		ADV_NONCONN_IND
28	27	0.086	d1:ef:64:16:d3:27	LE 1M	37	-79	00e52abe5be3e0b6c33a1b5b1111111111850000	ADV_NONCONN_IND
29	28	0.093	f8:e2:15:44:7e:5c	LE 1M	38	-87		ADV_NONCONN_IND
30	29	0.094	f8:e2:15:44:7e:5c	LE 1M	39	-83		ADV_NONCONN_IND
31	30	0.107	ee:01:02:00:00:08	LE 1M	37	-77	00004adbd94e815dee7e55f3000000002009	ADV_NONCONN_IND
32	31	0.108	ee:01:02:00:00:08	LE 1M	38	-68	00004adbd94e815dee7e55f3000000002009	ADV_NONCONN_IND
33	32	0.109	47:d7:18:da:e8:9b	LE 1M	39	-85		ADV_IND[Malformed Packet]
34	33	0.121	f3:ee:a7:2b:2b:52	LE 1M	37	-85	20000b4d12000103f2ec027c9c08	ADV_IND
35	34	0.122	f3:ee:a7:2b:2b:52	LE 1M	38	-72	20000b4d12000103f2ec027c9c08	ADV_IND
36	35	0.123	f3:ee:a7:2b:2b:52	LE 1M	39	-79	20000b4d12000103f2ec027c9c08	ADV_IND
37	36	0.124	70:45:31:5a:b1:b7	LE 1M	39	-30		SCAN_REQ

Figure 4.5. *A piece of BLE sniffer data in the form of an excel file*

4.2 Data pre-processing

The first step in each data mining task is data pre-processing. Data cleansing is a technique that we used to remove unwanted data and then make the whole data into an organized structure. One example of this data cleansing is removing the data rows in which the value of info column is not ADV_NONCONN_IND in the sniffer data. According to table 2.3, ADV_NONCONN_IND is the PDU type utilized for primary advertising channels.

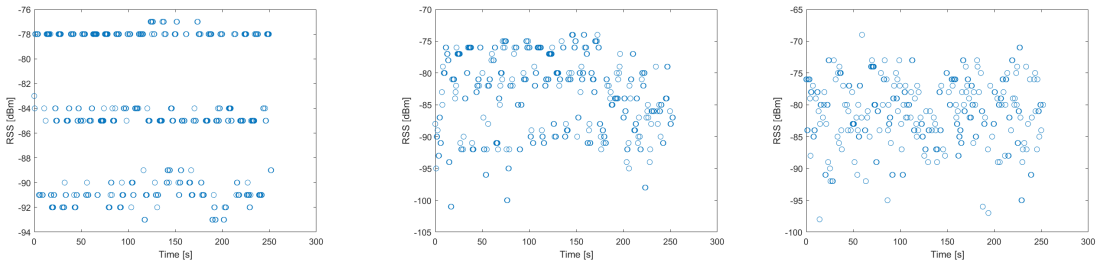
4.3 Data processing and analysis

For this section, data is collected with HIRM in different locations at HERE Technologies office premises in Tampere and for each location under three different circumstances:

- Condition I: Static mode
- Condition II: Semi-static mode
- Condition III: Dynamic mode

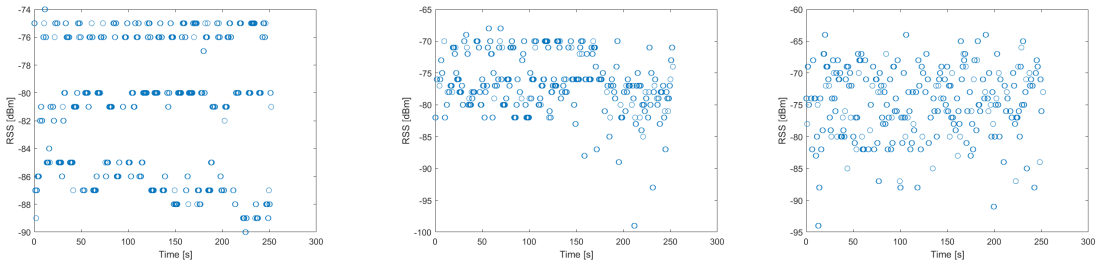
In the static mode, the data collection is done out of working hours while there is no movement in the office and the environment is static. In the semi-static mode, the device is static while a person is moving around and finally, in the dynamic mode, the device is being moved during the measurements.

Below you can see some figures of the RSS values measured over time in the above-mentioned conditions for several BLE beacons. These figures give us the opportunity to compare RSS behaviors in different conditions. It is quite obvious from the figures that the effect of Multipath phenomenon in semi-static and dynamic mode conditions is significant compared to static condition. While a visible pattern of three different levels of RSS values originated from three advertising channels appeared in the static mode figures, there is no pattern in the other two figures.



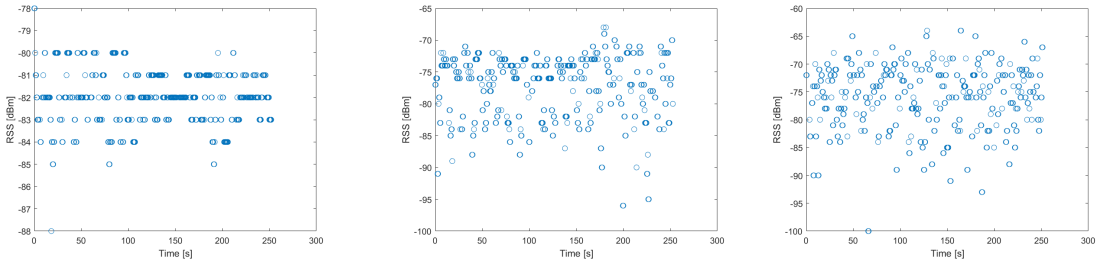
(a) *Static mode* ***(b)*** *Semi-static mode* ***(c)*** *Dynamic mode*

Figure 4.6. *Collected RSS values in three modes: beacon AC233F253AB3*



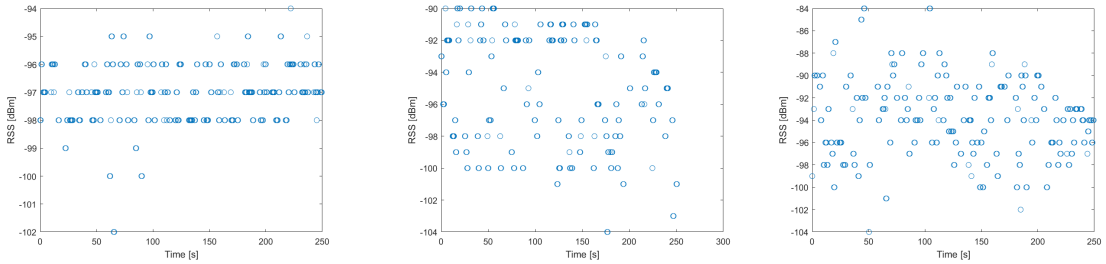
***(a)** Static mode* ***(b)** Semi-static mode* ***(c)** Dynamic mode*

***Figure 4.7.** Collected RSS values in three modes: beacon EE0102000A22*



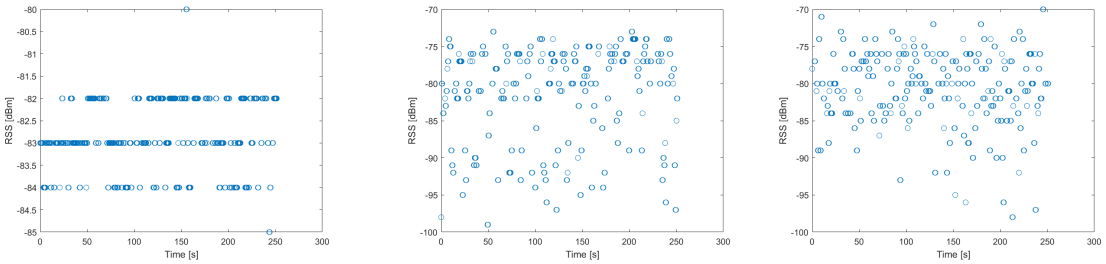
***(a)** Static mode* ***(b)** Semi-static mode* ***(c)** Dynamic mode*

***Figure 4.8.** Collected RSS values in three modes: beacon EE0102000B75*



***(a)** Static mode* ***(b)** Semi-static mode* ***(c)** Dynamic mode*

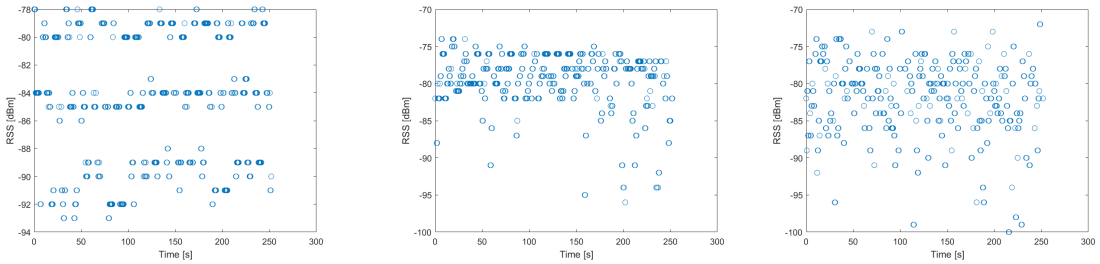
***Figure 4.9.** Collected RSS values in three modes: beacon EE0102000BC8*



***(a)** Static mode* ***(b)** Semi-static mode* ***(c)** Dynamic mode*

***Figure 4.10.** Collected RSS values in three modes: beacon EE0102000013*

As seen in the figures above, collected RSS values using HIRM application are the aggregated RSS values received from all three advertising channels and there is no way to distinguish the channel source associated with each value.



***(a)** Static mode* ***(b)** Semi-static mode* ***(c)** Dynamic mode*

Figure 4.11. *Collected RSS values in three modes: beacon EE010200000A*

However, using BLE sniffer, we have access to the advertising channel source of each RSS value. Below you can see some figures of the RSS values measured over time using HIRM and BLE sniffer simultaneously for a few beacons. RSS value are illustrated on the left and their distributions are illustrated on the right. The measurements were done in a static mode where both the device and surroundings are static.

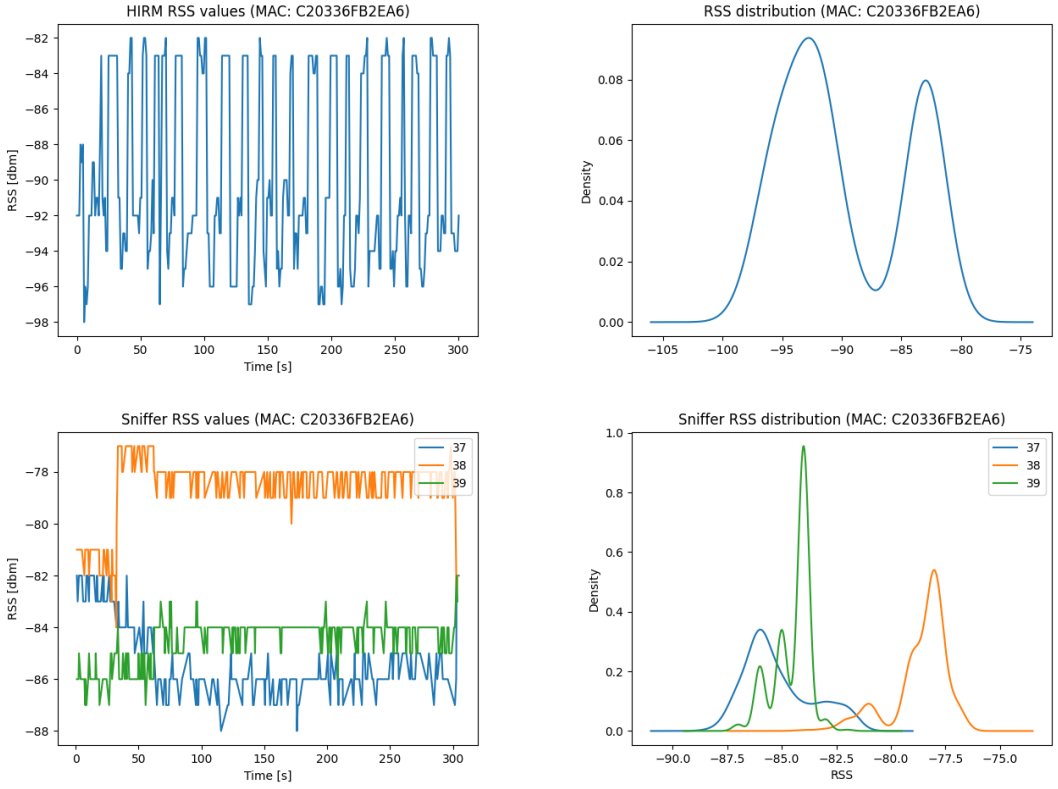


Figure 4.12. *RSS values and their distributions: beacon C20336FB2EA6*

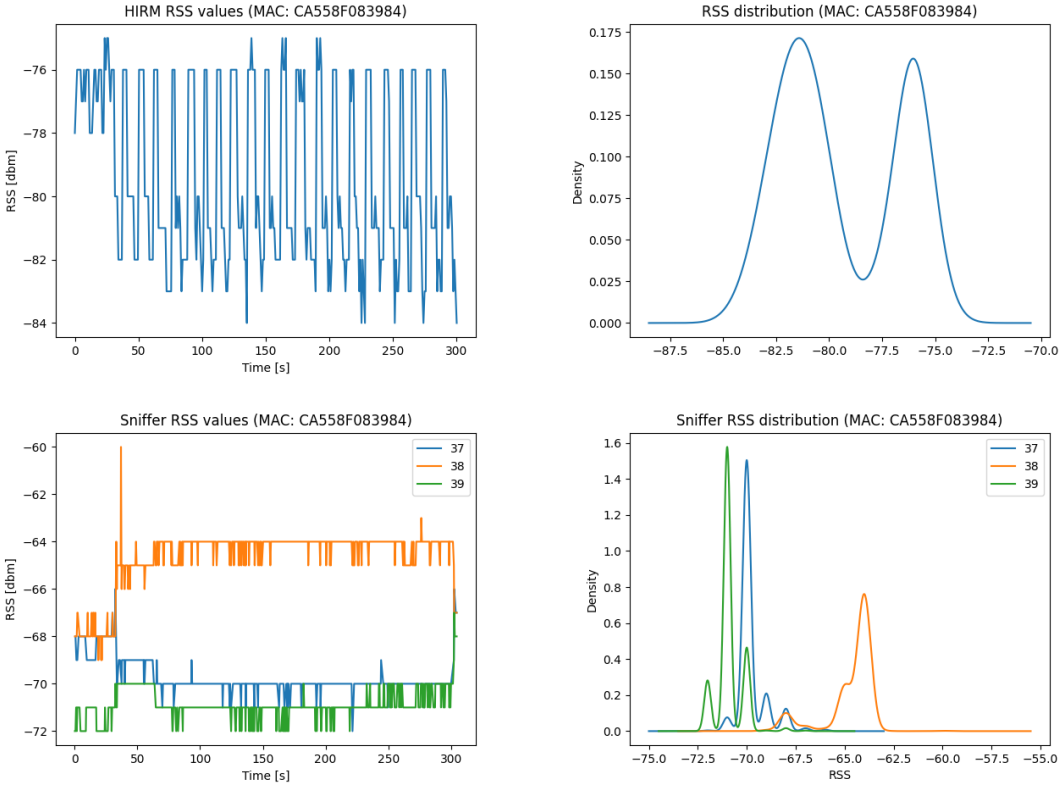


Figure 4.13. *RSS values and their distributions: beacon CA558F083984*

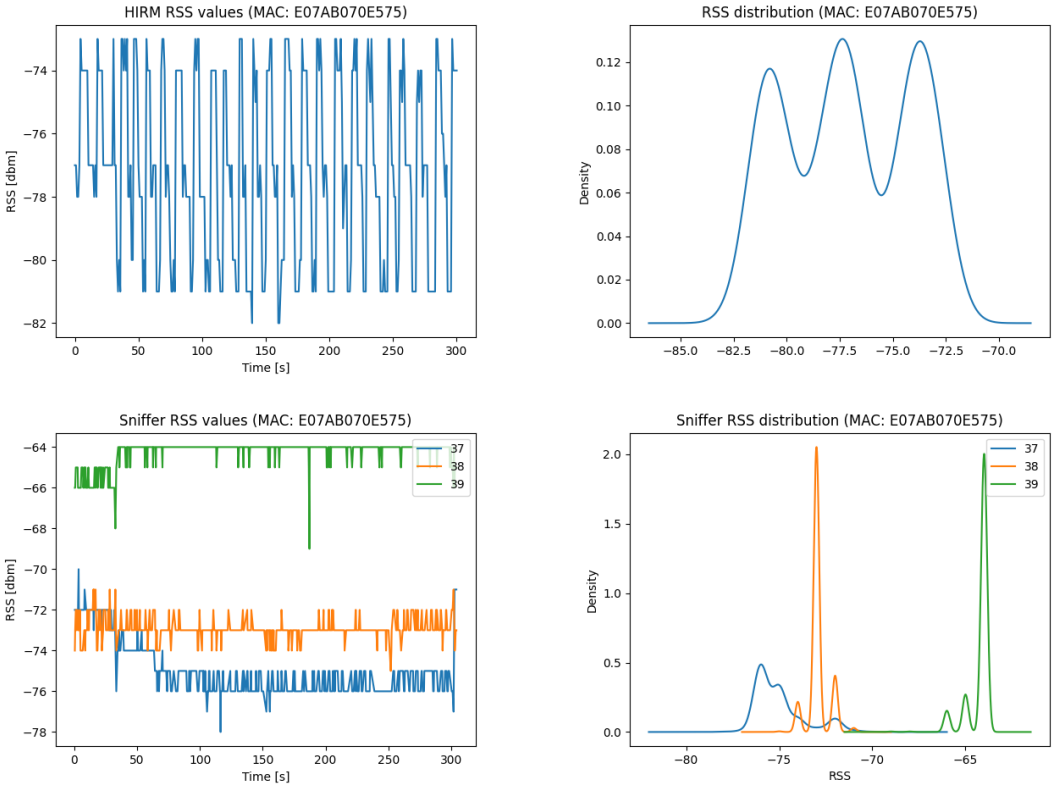


Figure 4.14. *RSS values and their distributions: beacon E07AB070E575*

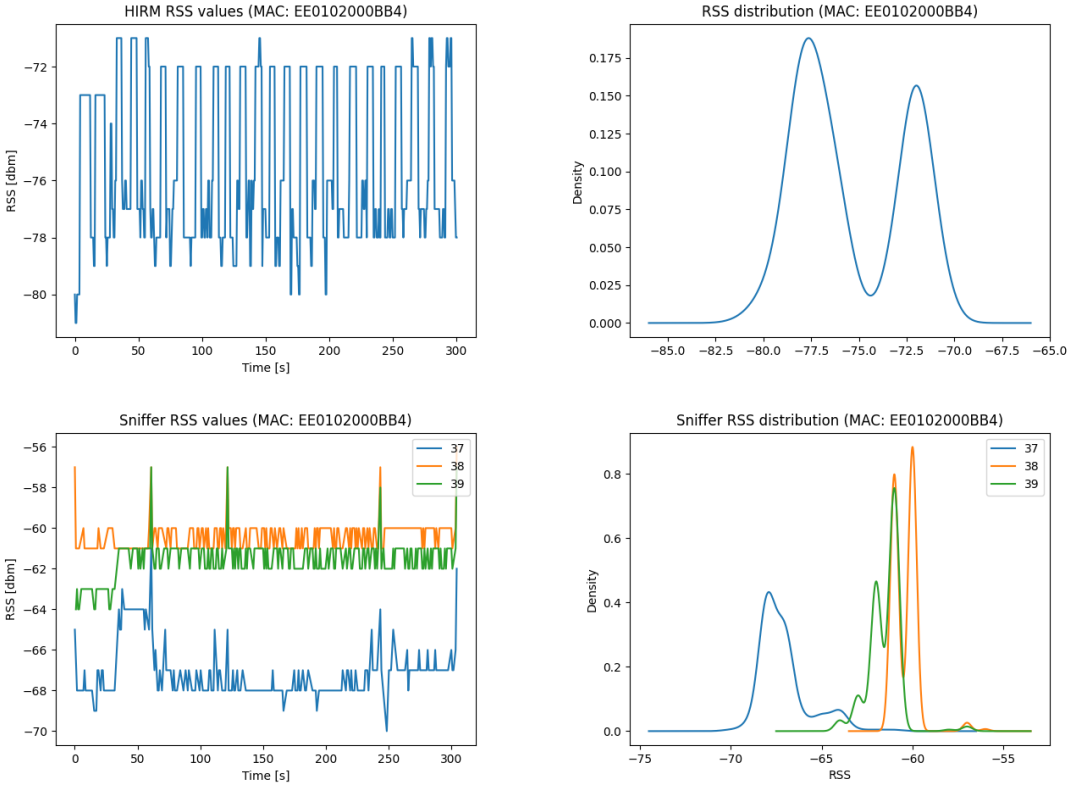


Figure 4.15. *RSS values and their distributions: beacon EE0102000BB4*

Even though, the data was collected with both devices (mobile device with HIRM application and BLE sniffer) at the same time and location, some differences in the level of RSS values can be observed when comparing the figures on the top left corner to the figures on the bottom left corner. This difference was anticipated due to the different antenna gains in different devices.

5 METHODOLOGY

This chapter contains the approaches we used in this thesis to investigate the data further and stabilize the positioning.

5.1 Averaging over RSS measurement intervals

The purpose of this section is to study the effects of averaging RSS measurements over different intervals in the online phase of positioning on positioning performance. If we have an interval of m measurements, when a new measurement arrives, the latest measurement is replace with the average value of the latest measurement and m-1 measurements before itself. Very similar to a shift register, where each time the oldest measurement in the interval is dropped, the latest measurement is added to the interval. This way we try to reduce the effect of sudden channel changes. In figure 5.1, yellow dashed boxes are example of 5-measurement intervals with values from two different channels. As it is clear from the figure, the difference between RSS values of two different channels are not negligible, thus averaging over intervals may improve the performance by mitigating sudden RSS changes.

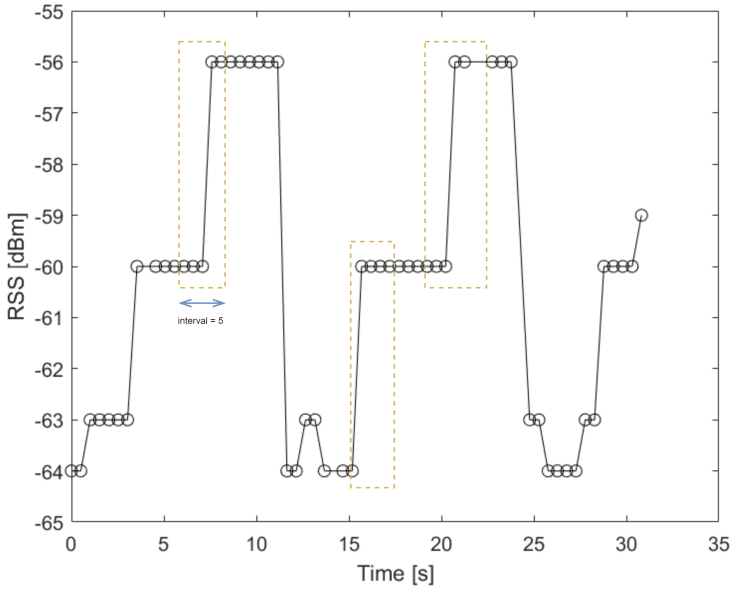
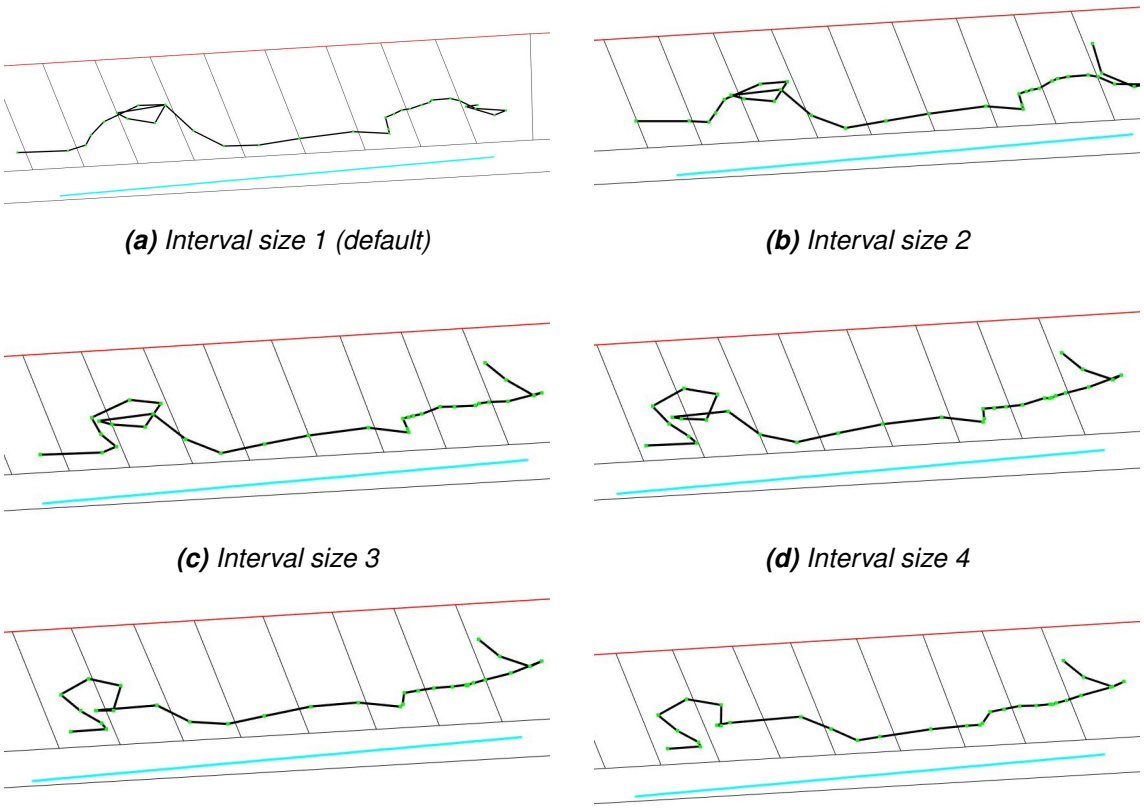


Figure 5.1. *Sudden channel changes in interval positioning*

To evaluate the positioning performance using interval averaging with intervals 2-6 measurements, 25 test tracks were collected and mean and maximum errors of each are reported in chapter 6. This study has been done both with and without Kalman filter in the positioning algorithms.

Figures 5.2 and 5.3 illustrate the positioning performance for two test tracks. Even though it is clear from the figures that averaging RSS values over intervals helps reducing the instability in some test tracks, according to table 6.1 on average no significant improvement in mean or maximum error of test tracks was observed while using Kalman filter in positioning. However, table 6.2 indicates an average of 3 meters improvement in maximum error using 5-measurement intervals without using Kalman filter in positioning.



(a) *Interval size 1 (default)*

(b) *Interval size 2*

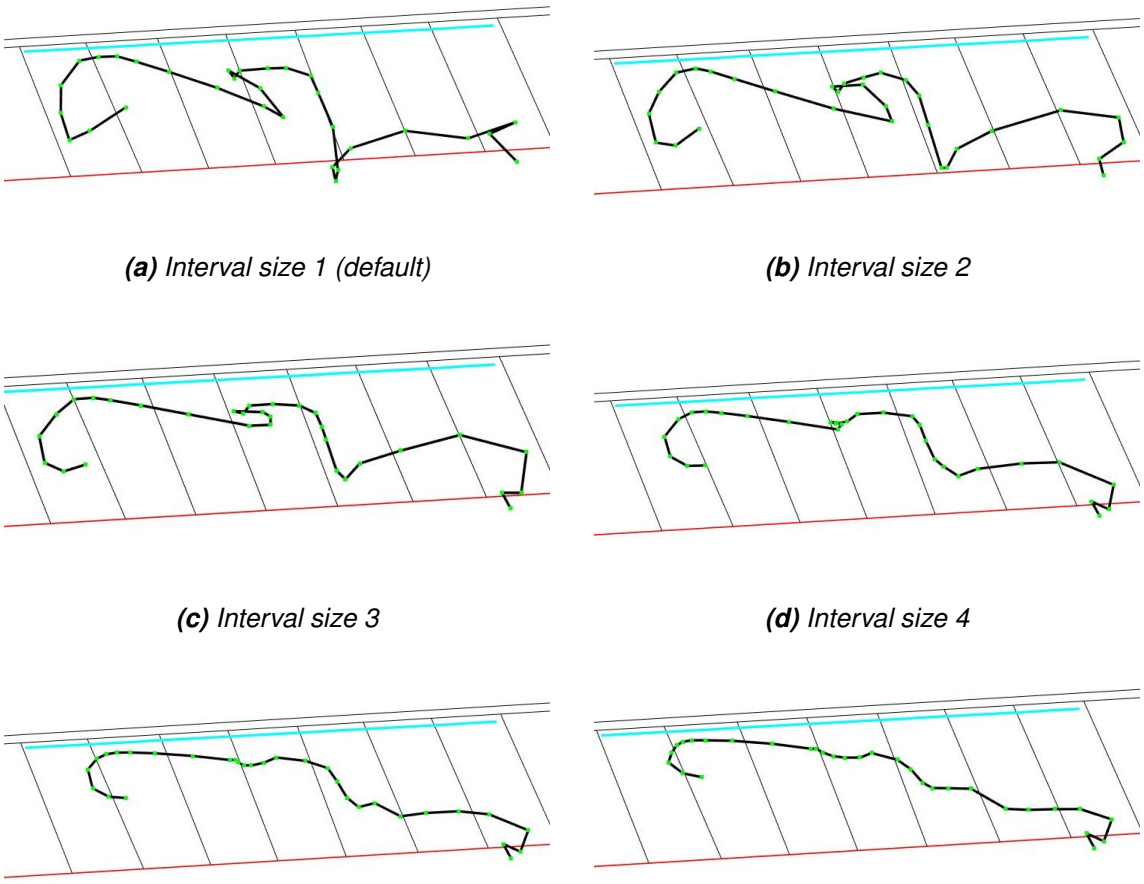
(c) *Interval size 3*

(d) *Interval size 4*

(e) *Interval size 5*

(f) *Interval size 6*

Figure 5.2. *Interval positioning from interval size 1 to 6 for test track 1*



(a) *Interval size 1 (default)*

(b) *Interval size 2*

(c) *Interval size 3*

(d) *Interval size 4*

(e) *Interval size 5*

(f) *Interval size 6*

Figure 5.3. *Interval positioning from interval size 1 to 6 for test track 2*

5.2 Channel separation using OMGP method

Expectation maximization (EM)

Expectation maximization (EM) algorithm is a computational iterative method in which we iterate through E (Expectation) and M (Maximization) steps to get the maximum likelihood estimate of the parameters in a probabilistic model, where latent variables exist [28]. The iterations continue until convergence and the parameters in the parameter set $\theta = \{\pi, \mu, \sigma\}$ are obtained.

E-step. At the t^{th} iteration, the expected value of log-likelihood function, $Q(\theta, \theta^t)$, is calculated with respect to θ^t.

$$Q(\theta, \theta^t) = \mathbb{E}[\ln p(X, Z|\theta^t)] \tag{5.1}$$

M-step. Parameters are updated at θ^{t+1} by maximizing $Q(\theta, \theta^t)$.

$$\theta^{t+1} = \arg\max_{\theta} Q(\theta, \theta^t) \tag{5.2}$$

Overlapping mixtures of Gaussian processes (OMGP)

Overlapping mixtures of Gaussian processes is another method we used in this thesis to separate the RSS values according to their channel source. This mixtures of Gaussian processes approach is known to label observations based on the sources that created them [29]. Below, a short overview of OMGP method is available [29, 30]. In OMGP, each individual data point is formed by applying one of the M latent functions and subsequently introducing Gaussian noise to it.

$$y_i{}^{(m)} = \left\{ f^{(m)}(x_i) + \epsilon_i \right\}_{m=1}^{M} \tag{5.3}$$

where $f^{(m)}(x_i)$ is the latent function m via which variable x_i was generated and ϵ_i is the additive noise. As we do not know from which source each data point was generated, we define a latent variable Z which is a matrix of zeros with a single non-zero element per row. The non-zero entry in row i and column m is referred to as $Z[i, m] \neq 0$ and indicates that x_i was generated using latent function m. The likelihood of OMGP is formulated as,

$$p(y|\left\{ f^{(m)} \right\}_{m=1}^{M}, Z, x) = \prod_{i,m=1}^{N,M} p(y_i|f^{(m)}(x_i))^{Z[i,m]} \tag{5.4}$$

The prior distributions of latent functions and variables are written as,

$$p(Z) = \prod_{i,m=1}^{N,M} \Pi[i,m]^{Z[i,m]} \tag{5.5}$$

$$f^{(m)}(x_i) \sim \mathcal{GP}(\varphi^{(m)}(x_i), K^{(m)}(x_i, x_j)) \tag{5.6}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{5.7}$$

It is worth to note that $\sum_{m=1}^{M} \Pi[i,m] = 1$. In equation 5.5, $\Pi[i,:]$ indicates a histogram over M factors for data point x_i. Functions $\varphi^{(m)}(x_i)$ and $K^{(m)}(x_i, x_j)$ in 5.6 are the mean and kernel functions of Gaussian process distribution, respectively. One important parameter in kernel function K is length-scale Λ which determines how quick the correlation between outputs changes with respect to the changes in the level of separation in corresponding inputs. In equation 5.7, σ is a shared hyper-parameter which defines the noise variance.

As it is not feasible to compute the exact posterior distribution $p(Z, \{f^{(m)}\}_{m=1}^{M} | x, y)$, a variational inference and EM method are used for posterior approximation. First, using Jensen's inequality a lower bound on the marginal likelihood is constructed and the aim is to identify a variational distribution that maximizes this lower bound. This variational distribution happens to be an approximation to the posterior. Iterating through E-step and M-step continues until the lower bound converges.

E-step: Updates to variational distribution (hyper-parameters are fixed).

M-step: Hyper-parameters are optimized (distribution is fixed).

For the analysis in this section, test tracks (data) have been collected using HIRM application over 5 different time intervals, each of them around 35-40 seconds. Data and analysis for 3 beacons are illustrated below. The first set of figures shows the RSS measurements collected from beacon DFA15BED739A during 5 intervals. The second set of figures shows the outcome of OMGP approach over the same measurements from beacon DFA15BED739A.

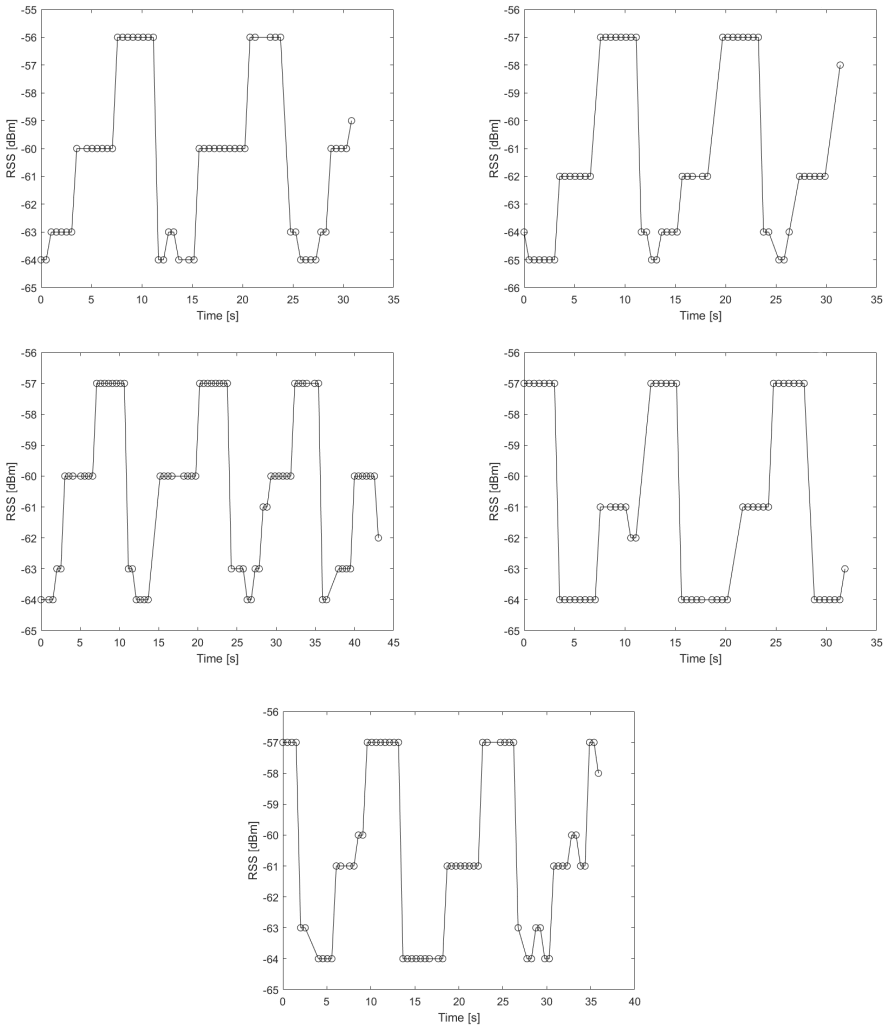


Figure 5.4. *Data collected over 5 different time intervals: beacon DFA15BED739A*

As seen in figure 5.5, OMGP method has separated the RSS values based on the advertising channels that generated them. It is expected to see the average of estimated channels would be rather similar over different trials. Table 5.1 contains the estimated mean values of channels using OMGP method in each test track. It is clear that the mean values for each channel is quite similar in all test tracks.

Table 5.1. *Obtained OMGP mean values [dBm] for channels: beacon DFA15BED739A*

Test tracks	CH 37	CH 38	CH 39
1	-56	-60	-63.3043
2	-57.0588	-62	-64.5
3	-57	-60	-63.2222
4	-57	-61.1538	-63.9583
5	-57.0476	-60.8182	-63.7727

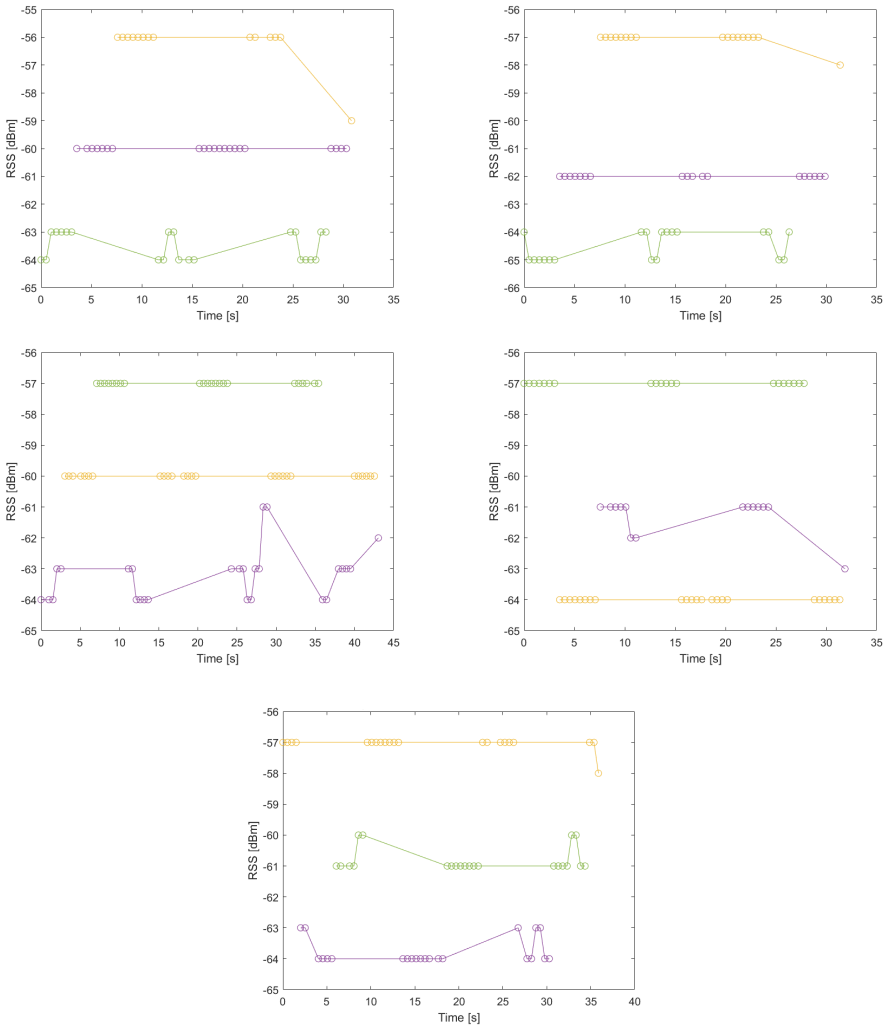


Figure 5.5. *OMGP results over 5 different time intervals: beacon DFA15BED739A*

The range of RSS values for the previous beacon was small ($\sim$ from 56 to 64). The second beacon, EE0102000BC9, has higher range of RSS values ($\sim$ from 70 to 84) and from the set of figures 5.6, it seems that the measurements derived from different channels have almost equal distance from each other.

Figures 5.6 and 5.7 illustrate the RSS measurements and the outcome of OMGP method respectively. As expected and seen in table 5.2, the estimated mean values for each channel are very similar in all the test tracks.

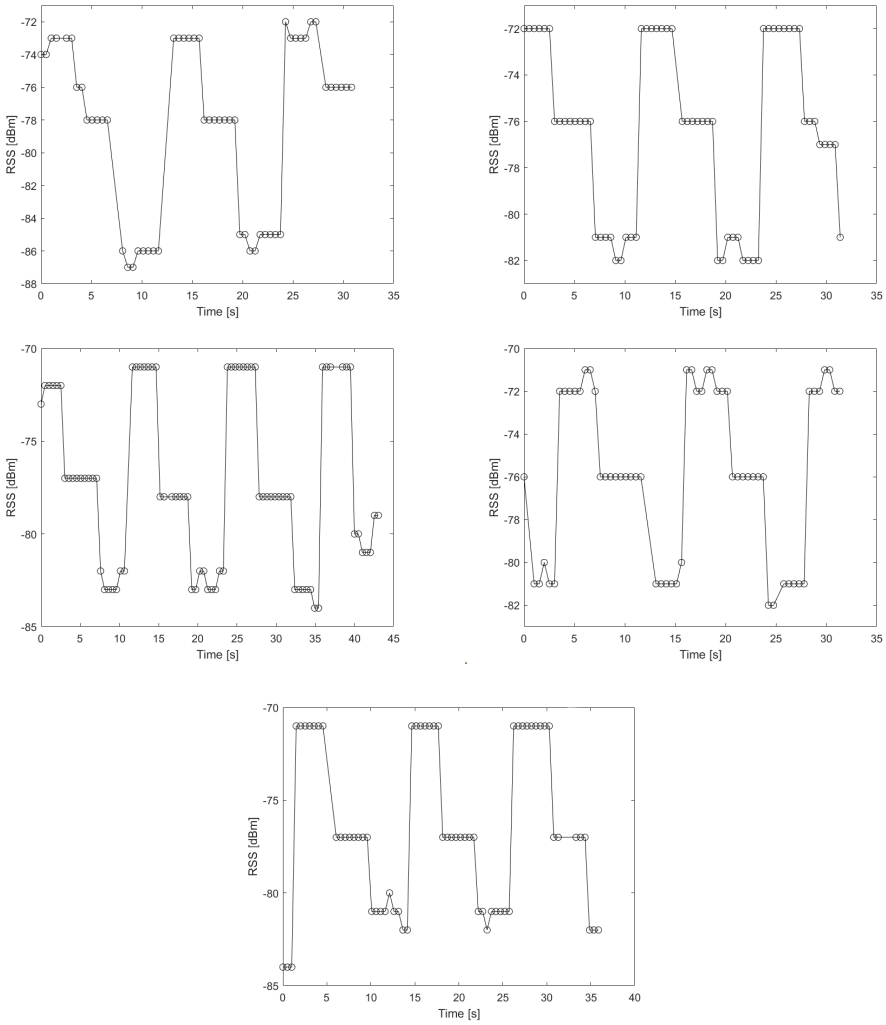


Figure 5.6. *Data collected over 5 different time intervals: beacon EE0102000BC9*

It is also good to note that an important parameter in OMGP model is length scale and should be selected based on the behavior of data. The proper length scale for our data was chosen after trying multiple length scale values and comparing the results.

Table 5.2. *Obtained OMGP mean values [dBm] for channels: beacon EE0102000BC9*

Test tracks	CH 37	CH 38	CH 39
1	-72.9474	-77.2	-85.7059
2	-72	-76.1818	-81.4211
3	-71.2593	-77.8571	-82.4444
4	-71.6522	-75.7778	-81
5	-71	-77	-81.6087

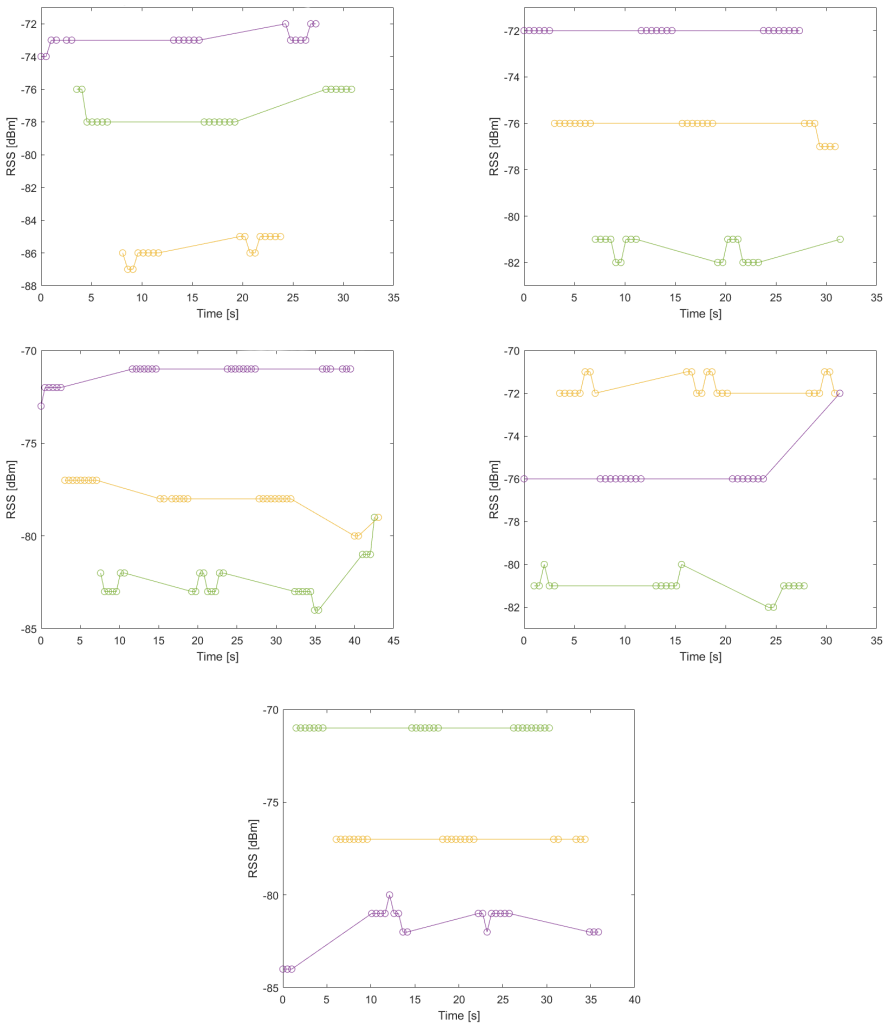


Figure 5.7. *OMGP results over 5 different time intervals: beacon EE0102000BC9*

Beacon EE0102000A18 is the third beacon we present here. Similar to the second beacon, this particular beacon also has a higher range of RSS values comparing to the first beacon. The only difference is that the measurements from one of the channels are far and the measurements from the other two are close to each other.

Figure 5.9 indicates that in some cases including the cases that two channel measurements are closer to each other and further from the other, OMGP may not be able to separate the measurements into three trajectories, but only two. The third and fifth sub-figures in 5.9, are examples of OMGP mistakes when two out of three channel measurements are very close to each other. The estimated mean values for channel 38 in those cases are mentioned as NaN in table 5.3.

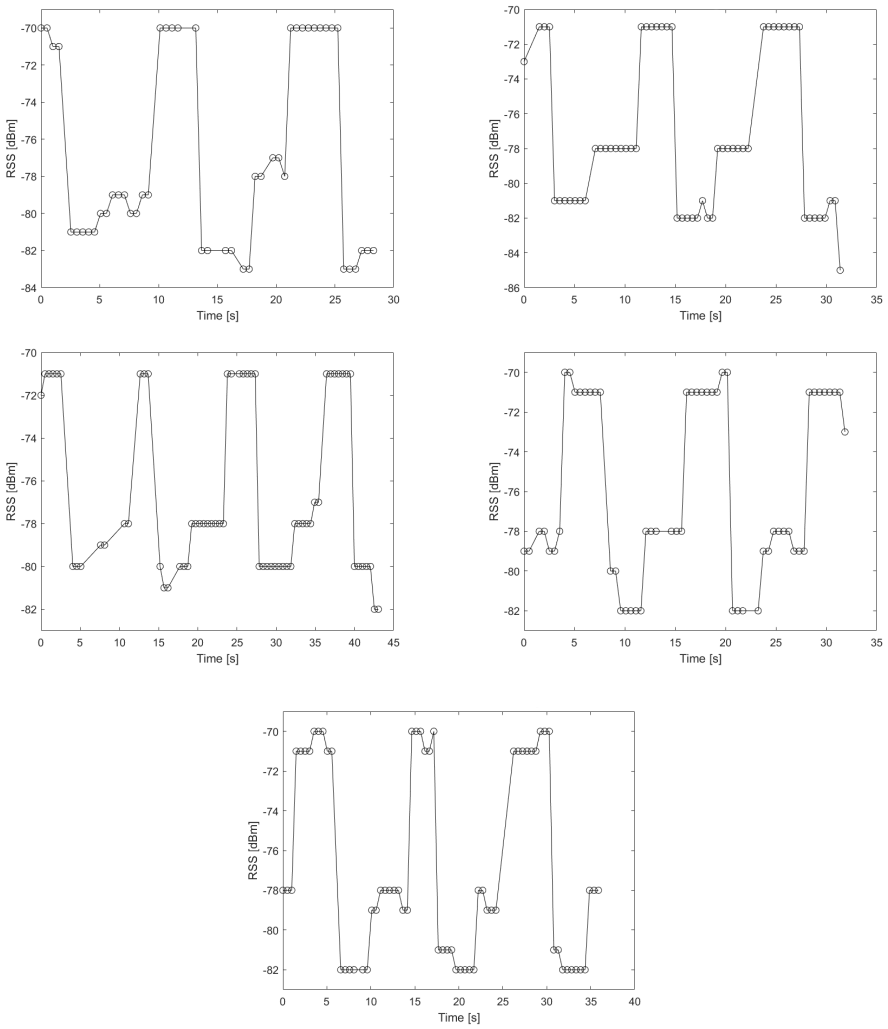


Figure 5.8. *Data collected over 5 different time intervals: beacon EE0102000A18*

Table 5.3. *Obtained OMGP mean values [dBm] for channels: beacon EE0102000A18*

Test tracks	CH 37	CH 38	CH 39
1	-70.1111	-78.7857	-82
2	-71	-77.7059	-81.6957
3	-71.0435	NaN	-79.2444
4	-70.92	-78.5416	-82
5	-70.5833	NaN	-80.1628

After analysing OMGP method on multiple test tracks for a few beacons, we used OMGP in the online phase of positioning to separate the measurements based on their source and utilized source information in two different ways to obtain the position. Below you can find the approaches we used to make use of channel information in positioning described and shown in figures 5.10 and 5.11.

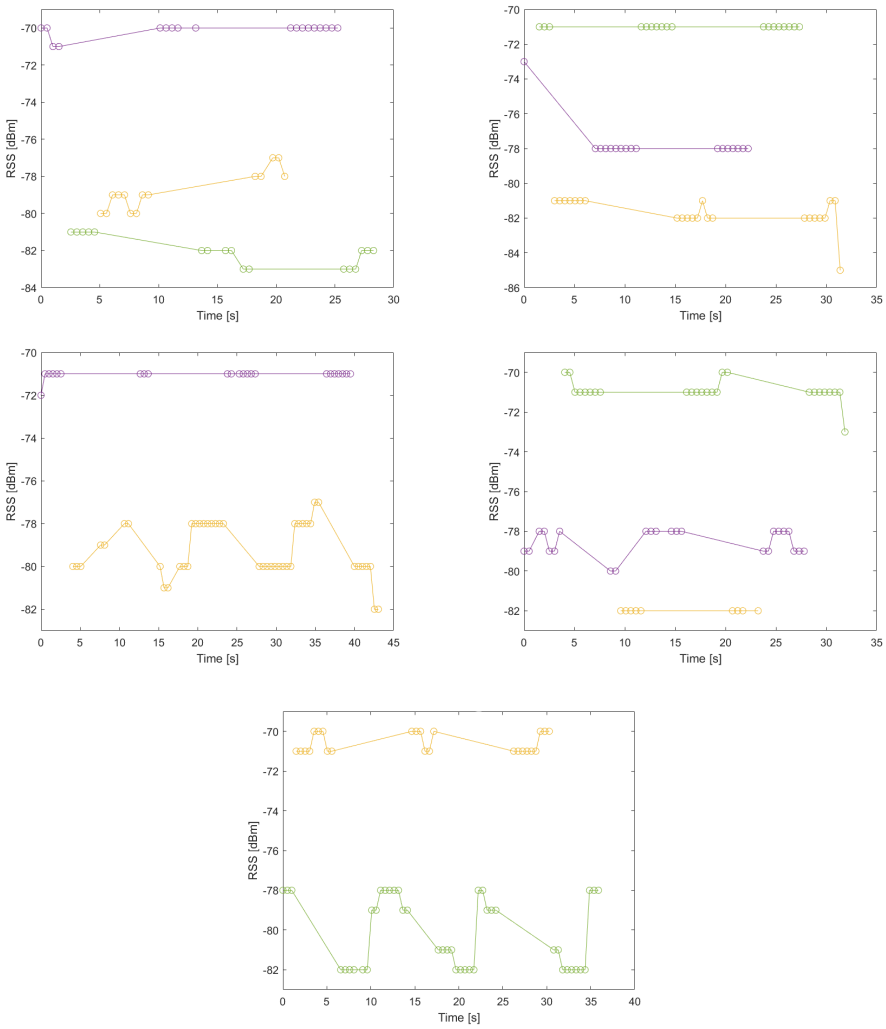


Figure 5.9. *OMGP results over 5 different time intervals: beacon EE0102000A18*

Approach 1

- Use OMGP to separate RSS values by channel.
- At time t the latest measurement is from one of the three channels.
- Obtain the mean value of the other two channels.
- Replace the latest measurement with the average of the latest measurement and the mean values of the other two channels.

The figure below illustrates the idea behind approach 1. As seen from the figure, channels are attained by OMGP and distinguished by three different colors. The latest measurement at time t corresponds to channel indicated with color blue. In approach 1, instead of using that latest measurement in online phase of positioning, we can use the average of that latest measurement with the mean values of the other two channels indicated with dashed lines in green and red (the average of blue, green and red dots). This way our positioning process balances the channels' effect by considering not only the channel from

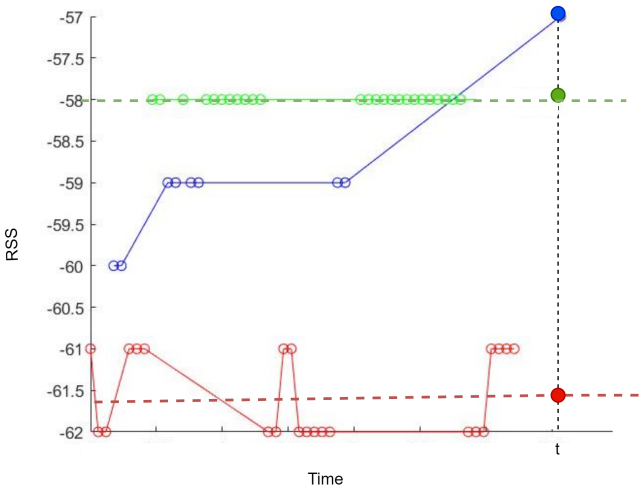


Figure 5.10. *OMGP in positioning, approach 1*

which the latest measurement at time t was received, but also the other two channels.

Approach 2

- Use OMGP to separate RSS values by channel.
- At time t the latest measurement is from one of the three channels.
- Obtain the predicted value of the other two channels.
- Replace the latest measurement with the average of the latest measurement and the predicted values of the other two channels.

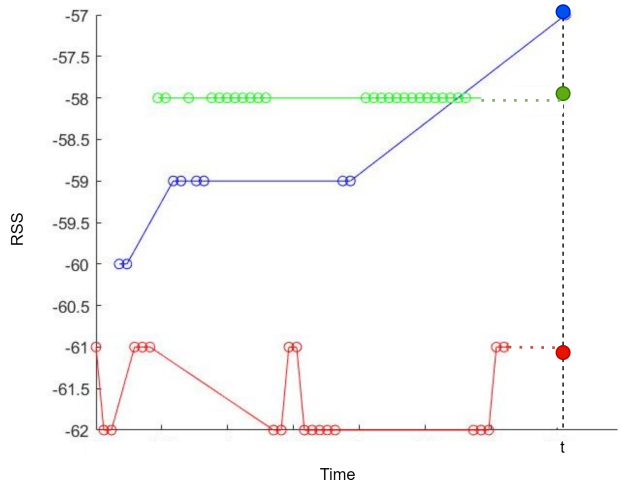


Figure 5.11. *OMGP in positioning, approach 2*

Figure 5.11 shows the idea behind approach 2. Similar to the previous figure, in this figure, the latest measurement at time t corresponds to channel indicated with color blue. In approach 2, instead of using this latest measurement in online phase of positioning, we can use the average of that latest measurement with the predicted values of the other

two channels indicated with dotted lines in green and red (the average of blue, green and red dots). A simpler way would be to use the latest available values from each channel instead of their predicted value which we used in this thesis. This way we can balance the channels' effect on positioning by taking into account the RSS values from all channels at time t.

In order to check the performance of indoor positioning with OMGP, 12 test tracks were collected and their mean and maximum error values were calculated. In both approaches, the average maximum error was reduced up to 60cm. OMGP with a few different length-scales was tested and the best results are provided in table 6.3.

6 RESULTS

In this chapter we present the results of the analysis that were discussed in the previous chapter.

6.1 Positioning performance: Averaging over different intervals

Interval averaging using 5 different intervals is done, once with Kalman filter available in positioning algorithms and once without Kalman filter. The mean and maximum error for 25 test tracks are measured in both cases and there results are reported in tables 6.1 and 6.2. As it is clear from table 6.1, there is no considerable improvement in the mean and maximum positioning error of test tracks on average using interval averaging technique.

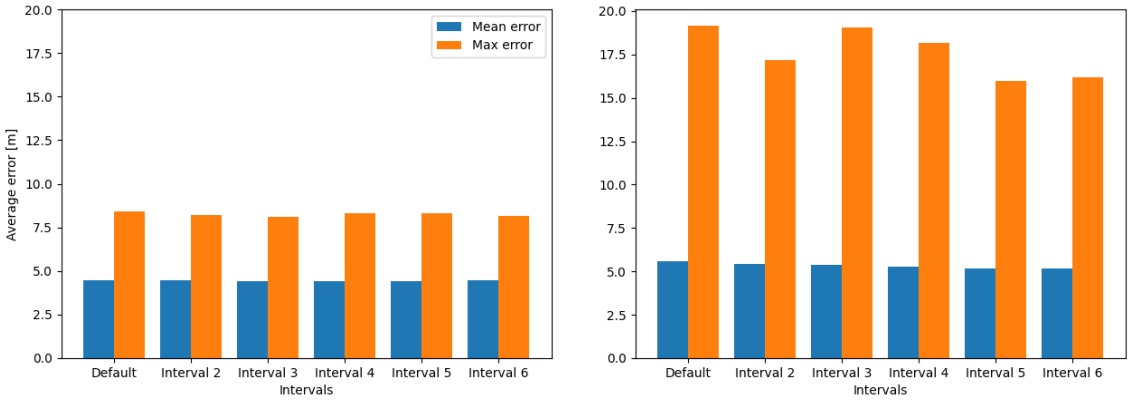
Table 6.1. *Mean and maximum positioning errors [m] of different intervals with Kalman filter*

Logs	Default		Interval 2		Interval 3		Interval 4		Interval 5		Interval 6	
	mean	max	mean	max	mean	max	mean	max	mean	max	mean	max
1	2.504	4.422	2.307	3.831	2.175	3.729	2.304	5.145	2.265	4.776	2.316	4.203
2	4.569	8.569	4.588	8.568	4.583	8.546	4.56	8.511	4.505	8.219	4.494	7.663
3	3.837	6.08	3.642	5.56	3.661	5.622	3.725	5.733	3.779	5.8	3.844	5.741
4	4.899	7.574	4.909	7.414	4.966	7.299	4.945	7.232	4.923	7.233	4.947	7.252
5	5.795	9.01	6.312	10.674	6.059	8.476	5.856	8.159	5.846	8.335	6.017	8.306
6	3.058	5.216	2.855	5.302	2.728	5.063	2.517	3.731	2.485	5.516	2.43	5.295
7	3.206	6.307	2.982	6.121	2.863	5.977	2.909	6.572	3.152	6.588	3.356	6.758
8	2.788	7.857	3.319	7.967	3.17	7.793	2.875	7.797	2.878	7.745	2.71	7.312
9	3.649	5.747	3.669	5.772	3.686	5.673	3.817	5.626	3.873	5.689	3.958	5.424
10	4.341	8.423	4.234	8.284	4.175	8.831	4.194	9.318	4.218	9.527	4.181	9.374
11	3.073	6.241	2.831	5.445	2.58	5.445	2.372	5.524	2.397	5.524	2.539	5.524
12	3.996	7.444	4.037	7.056	3.95	6.85	3.852	6.473	3.787	5.642	3.871	5.978
13	4.163	7.787	4.194	7.791	4.047	7.774	4.056	7.789	4.088	7.783	4.151	7.756
14	4.316	7.491	4.318	7.698	4.248	7.614	4.27	7.91	4.14	7.609	4.262	7.825
15	2.119	3.525	2.112	5.078	2.342	5.011	2.705	5.365	2.774	4.983	3.026	4.971
16	6.099	11.132	5.997	10.948	5.782	10.751	5.436	10.751	5.029	10.751	4.876	10.751
17	5.427	8.796	5.503	8.809	5.956	10.063	6.435	13.803	6.509	14.788	6.772	14.124
18	6.078	8.227	6.313	8.471	6.253	8.696	6.295	9.11	6.31	9.428	6.468	9.707
19	2.924	8.058	3.092	9.013	3.241	9.069	3.287	8.573	3.485	8.141	3.577	8.141
20	5.798	15.332	5.867	15.332	6.082	16.023	6.345	16.023	6.514	16.023	6.64	16.023
21	2.505	5.477	2.102	3.982	2.158	4.05	2.291	4.558	2.475	4.94	2.535	5.127
22	7.468	13.144	7.365	12.839	7.101	12.956	7.055	13.054	6.851	13.049	6.288	12.035
23	4.791	14.756	4.156	9.867	3.798	8.129	3.516	6.705	3.222	5.345	3.014	4.822
24	7.186	11.082	7.165	11.055	7.144	11.055	6.981	11.237	6.731	11.055	6.551	11.055
25	7.376	12.053	7.501	12.669	7.759	12.669	8.08	12.669	8.2	12.669	8.305	13.065
Average	4.479	8.39	4.455	8.222	4.42	8.127	4.427	8.295	4.417	8.286	4.445	8.169

However, table 6.2 indicates something else. If Kalman filter is not available in the positioning algorithms, interval averaging with the interval size 5 can improve the mean error slightly and the maximum error up to 3 meters. Figure 6.1 also illustrate the average of mean and maximum error of 25 test tracks. Almost no changes are visible in 6.1a, while 6.1b shows a clear decrease in the maximum error of intervals 2, 5 and 6, among which interval 5 has the lowest error.

Table 6.2. *Mean and maximum positioning errors [m] of different intervals without Kalman filter*

Logs	Default		Interval 2		Interval 3		Interval 4		Interval 5		Interval 6	
	mean	max	mean	max	mean	max	mean	max	mean	max	mean	max
1	4.878	13.741	4.858	13.741	4.853	13.862	4.736	12.964	4.502	13.651	4.561	14.497
2	6.351	17.988	6.269	18.932	5.997	19.159	5.837	19.159	5.803	19.159	5.728	19.159
3	3.562	7.196	3.397	6.993	3.048	7.1	2.743	7.279	2.597	7.279	2.625	7.279
4	4.652	11.969	4.561	11.599	4.163	9.959	4.042	8.493	3.946	7.401	4.023	7.401
5	6.405	19.096	6.258	18.76	6.069	19.116	5.575	19.277	5.602	19.277	5.901	20.339
6	4.986	17.512	5.023	17.512	4.919	16.48	4.875	16.031	5.602	16.031	4.978	16.085
7	4.502	21.351	4.053	19.176	4.455	33.537	4.019	32.177	3.246	11.023	3.414	10.587
8	6.575	14.052	6.417	12.419	6.118	10.751	5.954	10.751	5.547	10.751	5.39	10.751
9	7.779	32.781	7.603	28.804	7.724	29.556	7.88	29.636	7.966	29.645	8.011	29.655
10	7.663	15.319	7.473	13.92	7.541	13.878	7.718	13.332	7.798	13.34	7.82	13.4
11	3.393	11.876	3.351	11.511	3.305	10.83	3.332	10.515	3.447	12.118	3.499	12.032
12	7.373	60.702	7.511	60.702	7.655	60.141	7.493	54.771	6.904	20.537	6.887	20.12
13	5.143	27.817	4.392	25.153	4.167	24.042	3.983	23.858	3.949	22.579	3.932	21.748
14	9.184	29.512	9.051	29.512	8.735	29.512	8.72	29.512	8.788	29.512	9.064	34.688
15	5.729	52.202	4.73	14.089	5.66	50.037	5.552	49.906	5.234	48.757	5.028	47.697
16	8.389	24.641	8.361	24.641	8.097	24.561	7.788	23.302	7.557	22.967	7.345	24.706
17	7.789	14.192	7.872	13.977	8.007	16.476	8.164	17.365	8.245	17.828	8.232	17.769
18	2.673	5.751	2.363	4.75	2.161	5.154	2.218	5.985	2.188	6.023	2.305	5.644
19	6.165	19.796	6.181	19.796	6.193	19.796	6.1	19.796	6.027	19.796	5.999	19.796
20	4.568	11.454	4.554	11.454	4.554	11.74	4.561	11.727	4.57	11.732	4.634	11.732
21	4.845	8.1	4.873	8.06	4.892	8.039	4.872	7.961	4.833	7.961	4.857	7.961
22	6.515	19.363	6.76	21.065	6.579	20.795	5.982	7.924	5.962	9.962	6.303	9.92
23	3.39	7.168	3.157	7.168	3.041	7.168	2.855	7.168	2.817	7.167	2.81	7.128
24	3.3	6.538	3.112	6.687	2.744	6.548	2.712	6.663	2.968	6.707	3.092	6.71
25	3.472	8.526	3.796	8.504	3.651	8.498	3.502	8.498	3.493	8.498	3.414	8.496
Average	5.571	19.146	5.439	17.157	5.373	19.069	5.249	18.162	5.184	15.988	5.194	16.212



(a) *With Kalman filter* ***(b)*** *Without Kalman filter*

Figure 6.1. *Positioning errors using interval averaging technique*

Focusing on interval averaging without Kalman filter, figure 6.2 compares the distribution of mean and maximum error of our test tracks using averaging on three best intervals and without any interval averaging. Noticeable improvement can be seen in the distribution of maximum error of size-5 and 5 interval averaging. This visualization indicates that interval averaging approach can be useful when the positioning system does not contain Kalman filter.

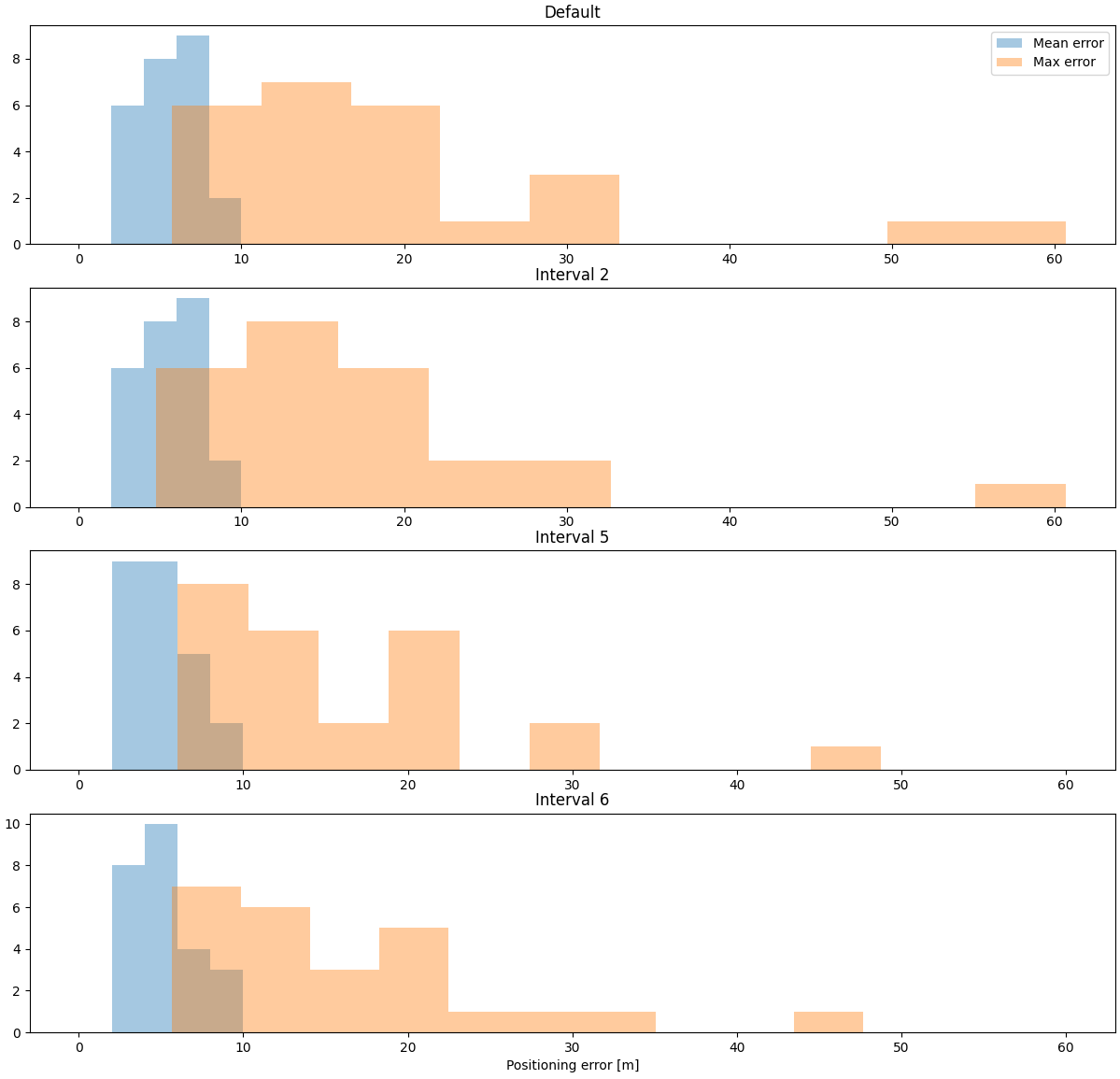


Figure 6.2. *Positioning error improvement in interval averaging without Kalman filter*

6.2 Positioning performance: OMGP method

For evaluating OMGP approaches in the online phase of positioning, 12 test tracks are collected and mean and maximum positioning errors are reported in table 6.3. Evidently, the maximum error values have decreased using OMGP method, while mean errors have not changed drastically. OMGP approach 2 provides us with almost 60 cm improvement in the average of maximum error values. In addition to that figure 6.3 visualizes the mean and maximum error values of the test tracks, without OMGP and with OMGP approaches 1 and 2. The blue and orange curves represent the mean and maximum errors respectively. Comparison of the orange curves in sub-figures demonstrate lower values for maximum error in OMGP approaches 1 and 2.

Table 6.3. *Mean and maximum positioning errors of OMGP-utilized approach 1 and 2*

Logs	No OMGP		OMGP approach 1		OMGP approach 2	
	mean error	max error	mean error	max error	mean error	max error
1	2.673	5.751	2.680	5.888	2.624	5.751
2	6.165	19.796	6.118	19.796	6.081	19.796
3	4.568	11.454	4.568	11.454	4.568	11.454
4	4.845	8.100	4.966	7.891	5.037	7.957
5	6.515	19.363	6.158	10.617	6.384	10.171
6	3.390	7.168	2.749	7.168	2.848	7.163
7	3.300	6.538	2.831	6.450	2.829	6.472
8	3.472	8.526	3.117	8.526	3.375	8.526
9	4.878	13.741	5.917	15.127	6.063	15.795
10	6.351	17.988	5.449	15.572	5.679	18.528
11	3.562	7.196	5.207	10.136	3.627	7.960
12	4.652	11.969	6.269	11.914	4.907	9.949
Average	4.531	11.466	4.669	10.878	4.502	10.794

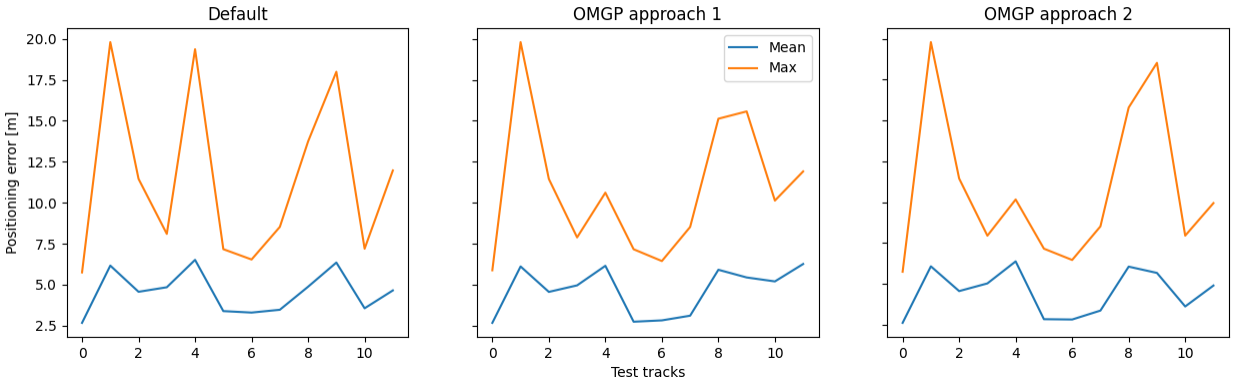


Figure 6.3. *Positioning errors using OMGP approaches*

7 CONCLUSION

GNSS has been playing an important role in global navigation and tracking. It is reliably used for outdoor localization purposes with high accuracy, while it performs poorly in indoor cases as the indoor environment is more complex. It is mostly preferred to use pre-existing radio equipment to obtain acceptable accuracy inside buildings. BLE is one of the popular radio technologies utilized in indoor positioning systems. Low power consumption and easy installation are among its advantages.

In this thesis, our aim is to tackle a problem that comes with BLE RSS-based positioning. BLE-based positioning requires enough beacons installed in the building to broadcast signals and a mobile device to receive the signals and provide us the location of the device. This is done by comparing the received signal values at an known position with the radio map that has been created and stored in the mobile device before. The problem with BLE broadcast is that the signal is transferred through three different channels and in the receiver (mobile device) is only able to receive an aggregation of RSS values from all three channels instead of providing us with the RSS values from different channels separately alongside with their source channels. This leads to some extent of instability and inaccuracy in indoor positioning as this behavior might cause huge difference between the RSS values received at the device at the moment and the RSS values used to create the radio map in the past.

A considerable part of this thesis was allocated to data collection and primary data analysis to provide us a better understanding of BLE broadcast behavior. First, we investigated multipath effect on BLE signals by collecting data under three environmental circumstances. It was clear that in the static environment when multipath effect is at its lowest level, RSS values were easily separable into three categories with human eyes that indicates the three advertising channels through which BLE signals are broadcast.

Starting with a simple approach, we then moved to the main part of our investigation which tried to increase the accuracy of positioning. We tried averaging RSS values over different intervals and the results are available in table 6.1. Even though in some test tracks such as the ones illustrated in 5.2 and 5.3 stability improvements where seen, the average and maximum positioning errors did not show huge improvements when Kalman filter is used in positioning. However, this approach shows up to 3 meters improvement if

Kalman filter is not used in positioning which indicates its effectiveness in BLE-based IPs where Kalman filter is not part of the positioning algorithm.

Using a Mixture of Gaussian Processes method, we attempted to separate the aggregated RSS values to three source channels. Our primary idea was to compare the outcome of our Gaussian method with the real separate channels measured by a special BLE tool kit to see if the channel separation was done correctly. However, the difference in antenna gains of the BLE tool kit and the mobile device used made the comparison difficult. Therefore, having used Mixture of Gaussian Processes in the online phase of positioning, it proved to provide almost the same average positioning error while reduced the maximum error up to 60cm. Table 6.3 contains detailed positioning error measurements for 12 different test tracks.

In this research, the above-mentioned algorithms were only applied in the online phase of positioning. For future research, they can be also integrated in the offline phase of positioning while creating the radio map. Another possibility is to use algorithms that categorize RSS values into three source channels and use channel information in both offline and online phase of positioning.

REFERENCES

[1] Kjærgaard, M. B., Blunck, H., Godsk, T., Toftkjær, T., Christensen, D. L. and Grønbæk, K. Indoor Positioning Using GPS Revisited. *Pervasive Computing*. Vol. 6030. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, 38–56. ISBN: 3642126537.

[2] Mautz, R. Overview of current indoor positioning systems. *Geodesy and cartography (Vilnius)* 35.1 (2009), 18–22. ISSN: 1392-1541.

[3] Puricer, P. and Kovar, P. Technical Limitations of GNSS Receivers in Indoor Positioning. *2007 17th International Conference Radioelektronika*. IEEE, 2007, 1–5. ISBN: 1424408210.

[4] Wang, J., Prasad, R. V., An, X. and Niemegeers, I. G. M. M. A study on wireless sensor network based indoor positioning systems for context-aware applications. *Wireless communications and mobile computing* 12.1 (2012), 53–70. ISSN: 1530-8669.

[5] Faragher, R. and Harle, R. Location Fingerprinting With Bluetooth Low Energy Beacons. *IEEE Journal on Selected Areas in Communications* 33.11 (2015), 2418–2428. ISSN: 0733-8716.

[6] Gentner, C., Gunther, D. and Kindt, P. H. Identifying the BLE Advertising Channel for Reliable Distance Estimation on Smartphones. *IEEE access* 10 (2022), 9563–9575. ISSN: 2169-3536.

[7] Bluetooth SIG. (Dec. 2019). *Specification of the Bluetooth System 5.2.* URL: `https://bluetooth.org`.

[8] Huang, B., Liu, J., Sun, W. and Yang, F. A robust indoor positioning method based on bluetooth low energy with separate channel information. *Sensors (Basel, Switzerland)* 19.16 (2019), 3487–. ISSN: 1424-8220.

[9] Powar, J., Gao, C. and Harle, R. Assessing the impact of multi-channel BLE beacons on fingerprint-based positioning. *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 2017, 1–8.

[10] Zhuang, Y., Yang, J., Li, Y., Qi, L. and El-Sheimy, N. Smartphone-based indoor localization with bluetooth low energy beacons. *Sensors (Basel, Switzerland)* 16.5 (2016), 596–596. ISSN: 1424-8220.

[11] Jakes, W. C. Microwave Mobile Communications. John Wiley Sons, 1994. ISBN: 9780780310698.

[12] Clarke, R. H. A statistical theory of mobile-radio reception. *Bell System Technical Journal* 47.6 (1968), 957–1000. ISSN: 0005-8580.

[13] Rappaport, T. S. *Wireless communications : principles and practice*. 2nd edition. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 2002. ISBN: 0130422320.

[14] Garg, V. K. *Wireless communications and networking*. 1st edition. The Morgan Kaufmann series in networking. Elsevier Morgan Kaufmann, 2007. ISBN: 1-281-11905-9.

[15] Debaenst, W., Feys, A., Cuiñas, I., Sánchez, M. G. and Verhaevert, J. Rms delay spread vs. Coherence bandwidth from 5G indoor radio channel measurements at 3.5 GHz band. *Sensors (Basel, Switzerland)* 20.3 (2020), 750–. ISSN: 1424-8220.

[16] Bykhovsky, D. *Coherence time evaluation in indoor optical wireless communication channels*. Vol. 20. 18. MDPI AG, 2020, 1–13.

[17] Al-Ammar, M. A., Alhadhrami, S., Al-Salman, A., Alarifi, A., Al-Khalifa, H. S., Alnafessah, A. and Alsaleh, M. Comparative Survey of Indoor Positioning Technologies, Techniques, and Algorithms. *2014 International Conference on Cyberworlds*. IEEE, 2014, 245–252. ISBN: 147994677X.

[18] Gu, Y., Lo, A. and Niemegeers, I. A survey of indoor positioning systems for wireless personal networks. *IEEE Communications surveys and tutorials* 11.1 (2009), 13–32. ISSN: 1553-877X.

[19] Al Nuaimi, K. and Kamel, H. A survey of indoor positioning systems and algorithms. *2011 International Conference on Innovations in Information Technology*. IEEE, 2011, 185–190. ISBN: 9781457703119.

[20] Liu, H., Darabi, H., Banerjee, P. and Liu, J. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE transactions on systems, man and cybernetics. Part C, Applications and reviews* 37.6 (2007), 1067–1080. ISSN: 1094-6977.

[21] Komine, T. and Nakagawa, M. Fundamental analysis for visible-light communication system using LED lights. *IEEE transactions on consumer electronics* 50.1 (2004), 100–107. ISSN: 0098-3063.

[22] Cassano, E., Florio, F., De Rango, F. and Marano, S. A performance comparison between ROC-RSSI and trilateration localization techniques for WPAN sensor networks in a real outdoor testbed. *2009 Wireless Telecommunications Symposium*. IEEE, 2009, 1–8. ISBN: 1424425883.

[23] Shrestha, S. *RSS-based position estimation in cellular and WLAN networks*. eng. Tietoliikennetekniikan laitos - Department of Communications Engineering, 2012.

[24] Li, B., Salter, J., Dempster, A. G. and Rizos, C. Indoor positioning techniques based on wireless LAN. *Proceedings of the 1st IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, AusWireless 2006*. 2006, 130–136. ISBN: 9780977520008.

[25] Pu, Y.-C. and You, P.-C. Indoor positioning system based on BLE location fingerprinting with classification approach. eng. *Applied Mathematical Modelling* 62 (2018), 654–663. ISSN: 0307-904X.

[26] HERE Technologies. *HERE Radio Mapper application*. URL: https://play.google.com/store/apps/details?id=com.here.radiomapper.

[27] Haq, Z. U. *Hybrid Spatial Interpolation : RSS based Indoor localization*. eng. Informaatioteknologian ja viestinnän tiedekunta - Faculty of Information Technology and Communication Sciences, 2021.

[28] Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B, Methodological* 39.1 (1977), 1–38. ISSN: 0035-9246.

[29] Lázaro-Gredilla, M., Van Vaerenbergh, S. and Lawrence, N. D. Overlapping Mixtures of Gaussian Processes for the data association problem. eng. *Pattern recognition* 45.4 (2012), 1386–1395. ISSN: 0031-3203.

[30] Dardeno, T. A., Bull, L. A., Dervilis, N. and Worden, K. A generalised form for a homogeneous population of structures using an overlapping mixture of Gaussian processes. *arXiv.org* (2022). ISSN: 2331-8422.