

Chapter 1

Research Design and Methodology

"I have been impressed with the urgency of doing. Knowing is not enough; we must apply. Being willing is not enough; we must do."

~ Leonardo da Vinci

1.1 Introduction

Based on the review of the literature in the previous chapter, Chapter ??, it is clear that the potential and possibilities of using BLE for measuring pedestrians are significant. However, to facilitate the conduct of research within the stipulated deadline for the completion of PhD, a clear demarcation of the scope and research constraints is necessary.

From the literature review, it is clear that the presented research is focused on pedestrians, specifically, movement characteristics and activities of pedestrians. It is also evident from Chapter ?? and Chapter ?? that the environment under scrutiny in this research is urban outdoors. The range of outdoor urban canyons is diverse – a quiet pedestrianised street, a pathway in a park, or even a city shopping district – each with its own distinct features. Moreover, the topologies of outdoor urban environments are also lively and predisposed to frequent changes due to factors such as road works, building construction, and festivals and rallies. However, since this research is largely a feasibility study, a space with limited external influence is suitable for the first stage of critically assessing the suitability of the technology. Therefore, an area of open ground which is not shared by passing pedestrians or vehicular traffic was selected for the presented study. The use of the open ground allows different linear pathways to be marked out for detailed evaluations, as will be presented in Section 1.3.2.3. Similarly, other complex factors such as change in direction during a walk, straying and turning by pedestrians could impede the feasibility study of the technology for understanding pedestrian dynamics. Therefore, a linear pathway is selected for the study, which mitigates additional fluctuations in the measurements caused due to complex pedestrian movement and allows exploration of the suitability of the technology. Thus, this work seeks to build a foundation for complex movement and activity detection in future studies, if the measurement technology is found to be viable and reliable. Regardless, the decision to select a linear pathway does not necessarily compromise on the detection of some of the complex pedestrian dynamics either, and even single-file movement can provide insights into complex

pedestrian dynamics (Seyfried et al., 2005), as stated in Section ?? in Chapter ?? . More details regarding the selection of location will be provided in Section 1.3.1 of this chapter.

Quick Recap

Observer: BLE device in a role that allows it to only listen for BLE advertisements.

Broadcaster: BLE device in a role that allows it to only send advertisements.

Before presenting the research design and methodology, it is useful to revisit the aim of this PhD. This PhD research is aimed at assessing if a single BLE Observer is capable of detecting, identifying, and characterising the movement and activities of pedestrians in any unshared outdoor space using only the signal characteristics of BLE Broadcasters. In a nutshell, the aim can be trans-

lated as exploring the presence of recurring patterns in the received BLE signals that can be linked to particular aspects of a pedestrian's movement or activity. But how can one use the patterns in signals to identify patterns corresponding to any activity or movement when there exist several manufacturers, antenna types, and features in BLE devices?

This question highlights the need to first understand the BLE devices themselves. But again, the devices used by the pedestrians are not in the control of the researcher performing the experiments or the agency using the technology to understand the movement of the pedestrians. Therefore, due diligence is required with the choice of the Observer device, the device that is chosen by the researcher or relevant agencies. This means that the characteristics of the Observer's antenna, its deployment, and the effect of different Broadcasting devices and their configuration on the resulting patterns in the captured signals must be investigated before evaluating the usefulness of the technology for the detection and identification of pedestrians' movement and activities. Thus, this research design and methodology chapter is divided into four sections. The first is elucidating design choices to develop a platform for conducting experimental work, the second, identifying the characteristics of the selected inventory for the experiments and assessing the selected experimental location. The third, evaluating the efficacy of the BLE technology in the detection of the patterns corresponding to select pedestrian movement and activities, and the fourth, describing the analytical tool-set used for understanding the measurements.

1.2 Design and development of a BLE Observer device capable of selectively listening to BLE Advertisements

The rationale behind the selection of BLE was discussed in Chapter ?? . This research prioritises privacy preservation at each design step of the system for capturing pedestrian movement and activities. While this research does not aim to develop a fully privacy-preserving technique, it investigates off-the-shelf techniques to restrict privacy infringements. While the security aspect of privacy is critical to any system, it is outside the immediate scope of the work presented in this dissertation. Thus, the

Quick Recap

Privacy-by-Design (PbD): System's design philosophy that focuses on *minimising* data collection, hiding personal data, *separating* storage and processing of data, *aggregating* data, informing the collection and processing of data, providing *control* of the data to contributors, *enforcing* privacy policies, and *demonstrating* data collection and processing process.

system development follows the PbD approach.

This chapter details the design process and methodology for the experiments to test the research hypothesis. It covers BLE technology from a design perspective and highlights features that support the PbD philosophy. Additionally, it explores the usability of BLE for identifying and characterising pedestrian movement and activities.

1.2.1 BLE

General Guide

The term *Observer* refers to both the BLE Link Layer (LL) device role (??) and the monitoring platform built for this research. *Advertisers* and *Broadcasters* are used interchangeably for BLE devices that emit signals.

To capture BLE signals, a device can operate in *observer* or *central* roles, as presented in Section ?? in Chapter ?. To protect pedestrian privacy, the monitoring device used in this research operates in *observer* role, which prevents it from initiating connections with other devices, thereby safeguarding against spoofing from nearby devices.

Advertisers are set in a *broadcaster* role to add another layer of security against spoofing for all the experiments presented in this dissertation. But, in a real-world deployment, pedestrians may use devices that are in *peripheral* mode, which allows connections to be established with a *central* device and subsequently, data transfers. Even in such a case, the design choices for the monitoring device, based on principles presented in Section ??, mitigate privacy risks since the *observer* role is incapable of establishing connections with any BLE device regardless of its role.

The feature of *whitelisting*, as presented in Section ?? in Chapter ??, offers another advantage to the monitoring platform or the Observer. Through this *link layer* feature, a clear list of devices of participants to be monitored is defined. The Observer then, upon processing the received advertisement, checks the *Universally Unique Identifier (UUID)* of the device emanating the signals against the UUIDs in the list. In the case where the signals have emerged from a device whose UUID is not mentioned in the list, those advertisement packets are discarded, therefore *selectively listening* to advertisements from specific devices. This feature provides a robust mechanism to encourage and facilitate *participatory monitoring*. Such a feature essentially provides control to the pedestrians, where the pedestrians can choose to not carry their whitelisted devices on days when they do not want to participate, thereby adding another layer to the privacy preservation aspect.

Quick Recap

Participatory monitoring involves data collection with the contributor's awareness and consent. *Opportunistic monitoring* captures data without the contributor's knowledge.

Finally, the availability of relatively inexpensive beacons provides an opportunity for using BLE for such studies. This means that instead of asking pedestrians to use their own devices to participate in the experiments, beacons can be provided by the investigators for the purpose of studying participating pedestrians. It ensures that the captured signals do not originate from personal user devices. Further, since the range of BLE devices is constrained, the measurements only take place in relatively close proximity to the Observer. This is akin to hard-geofencing in comparison

to the geofencing defined in software in the case of GPS, as previously mentioned in Section ?? in Chapter ??.

General Guide

The word *observer* here will be used both for the BLE device role and the monitoring platform built for conducting this research which is programmed to be in the *observer* role. The words *advertiser* and *broadcaster* will be employed interchangeably for the device used for other devices that advertise or broadcast BLE signals in the vicinity.

The above reasons provide a solid grounding for BLE to be the choice of technology in the design of a platform capable of listening to BLE advertisements for studying pedestrian movement and activities. To summarise, the following are the advantages BLE offers:

1. *Observer* and *broadcaster* role that mitigate the risk of spoofing to steal personal data.
2. *Whitelisting* to facilitate *selectively listening* to BLE advertisements, thereby enabling *participatory* monitoring.
3. Provision of inexpensive beacons as an alternative to the use of personal devices and providing more control to participants over opt-in and opt-out in the study.

1.2.2 Hardware Design

1.2.2.1 Observer and Broadcaster

Once the monitoring technology is decided, the next step is to identify a portable platform equipped with the technology. One option is to design and build an embedded system that includes BLE, however, since BLE has become a ubiquitous technology in personal devices, there are several embedded systems and Single-Board Computer (SBC) on the market that are well capable of being used as a platform for this development. This reduces the time spent on planning, developing, troubleshooting, and testing circuit designs for a BLE-enabled platform. Therefore, for an Observer device, the choice of selecting an off-the-shelf device is made.

Similarly, there are several off-the-shelf beacons available that can act as an advertiser. These beacons however are pre-programmed by the manufacturer, and many, if not all, cannot be re-programmed or even re-configured by the user at a later stage. The lack of options to re-program those beacons is, however, not a limiting factor for most of the experimental cases presented in this research. One of the experimental cases where the ability to re-programme/re-configure a beacon/advertiser is useful is to evaluate the effect of the advertisement interval on the correspondence of RSSI with pedestrian movement and activities. For such experiments, off-the-shelf embedded systems or SBCs, similar to the ones used for the Observer, can be selected and programmed to be used as a beacon/advertiser. This will be presented in Section 1.3.2.5 in this chapter.

Quick Recap

Advertisement interval is the duration between subsequent advertisements by a BLE broadcaster or peripheral. These advertisement intervals are in the range of 20ms to 10.24s in a multiple of 0.625ms.

Since only two hardware devices are required to facilitate the BLE experimentation, the Observer and Advertiser, an investigation is required to identify the candidate devices to be used in the experiments, and subsequently, select optimal candidates. There are several off-the-shelf SBC products available in the market such as the RPi 4B (RaspberryPi-Foundation, 2023), BeagleBoard BeagleBone (BeagleBoard, 2023), and DFRobot LattePanda SBC (DFRobot, 2023) that can be employed as an Observer in the platform. A relative normalised price¹ comparison of a few such components is listed below in Table 1.1 (Farnell, 2023). This relative price is computed by dividing each product's price by the price of the most expensive product.

SBC	Relative Price
RPi 4B	0.44
BeagleBone	0.85
DFRobot	0.71
Polyhex Debix	0.14
Hackboard 2	1
Asus Tinker Board R2.0	0.59

TABLE 1.1: Price Comparison of Commonly Available Single-Board Computers

General Guide

BlueZ is the official Linux Bluetooth protocol stack which features command-line interfaces such as *hcitool* and *hciconfig* to use and configure BT and BLE devices (BlueZ n.d.).

The RPi 4B is selected due to its cost, availability, and community adoption. The community presence and size are especially important as it provides greater support through community projects, which reduces the time spent on trivial setup and troubleshooting processes. Furthermore, the popularity, strong community, and availability of RPi also mean that it is more likely

to garner the attention of other researchers and is easier for them to replicate the published results or extend the research. Therefore, amongst the mainstream choice of components, it is more impactful to select RPi. In particular, the RPi 4 Model B variant with 8 GB RAM with integrated BLE 5.0 support is selected as this was the latest and most powerful variant available at the time this research work was initiated. Henceforth in this dissertation, the selected variant, RPi 4 Model B, for this research will be referred to simply as RPi. Figure 1.1 illustrates the RPi board. *The prominent face visible on the figure with the electronic and IO components will be referred to as the front face hereon.* Further details about the RPi is provided in Appendix D.

A weather-proof high-density plastic enclosure was used (enclosure, n.d.) for housing the RPi. Four pipe straps are mounted on two opposite sides of the enclosure and to facilitate previous related work, a couple of magnets are also secured behind the enclosure to enable it to be mounted on ferromagnetic surfaces such as lamp posts, if required. However, these pipe straps are removed for all of the experiments that do not require deployment of the Observer on metallic objects to reduce the likelihood, if any, of these pipe straps affecting the signals due to any propagation mechanisms highlighted in Section ?? in Chapter ?. The front and back of the enclosure are shown in Figures 1.2 and 1.3. Finally, the orientation of RPi inside the

¹As of November 2023

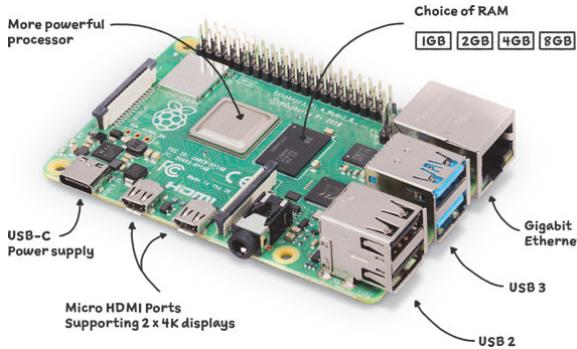


FIGURE 1.1: Raspberry Pi 4B ((RaspberryPi-Foundation, 2023))

enclosure is depicted in Figure 1.4.

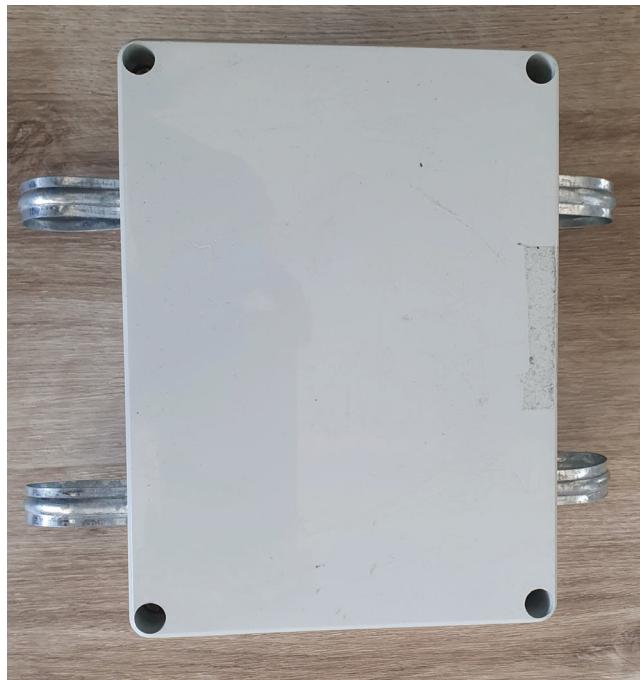


FIGURE 1.2: Front of the Enclosure for Observer

To provide power supply to the RPi, an off-the-shelf 10,000 mAh power bank ([Ansmann Powerbank n.d.](#)), as shown in Figure 1.5, is used. An RTC module, DS3231 ([DS3231 RTC Module for Raspberry Pi n.d.](#)), shown in Figure 1.6, is employed.

For the Broadcaster, a beacon is selected. Beacons are all generally in the same price range, and, as previously stated, since no programming is required on the Broadcaster side, any off-the-shelf Broadcaster is sufficient for this study. The cost benefits of the BLE beacon make it feasible to distribute them to the participants, thereby, monitoring the behaviour of pedestrians without them having to agree for their own personal devices to be used. The Media Access Controller (MAC) addresses of the delegated BLE beacons can be *whitelisted* to facilitate participatory



FIGURE 1.3: Back of the Enclosure for Observer

monitoring, thereby, despite BLE's ability to be used for opportunistic monitoring the Observer becomes incapable of being opportunistic. A RuuviTag (Ruuvi, [n.d.](#)) was selected as a Broadcaster. The rationale behind the selection of the RuuviTag was simply because it was readily available in the *tPOT* research group inventory. Figure 1.7 illustrates the Ruuvi Broadcaster that was used for this work.



FIGURE 1.4: Orientation of RPi Inside the Enclosure



FIGURE 1.5: Power Bank Housed in the Enclosure to Power the Observer (*Ansmann Powerbank n.d.*)

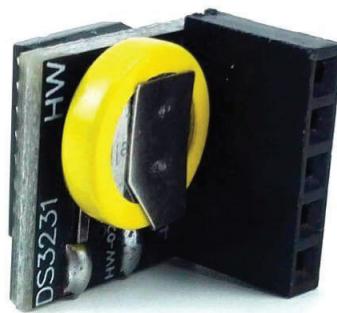


FIGURE 1.6: DS3231 RTC Module (*DS3231 RTC Module for Raspberry Pi n.d.*)



FIGURE 1.7: Ruuvi Tag (Ruuvi, n.d.)

An alternative Broadcaster was developed using the RPi solely to modify the advertisement rate for one of the experiments, as the RuuviTag does not allow alterations to the advertisement rate. The system includes a RPi connected to the same battery pack and RTC module as used in the Observer. Along with providing control over the advertisement rate, this RPi Broadcaster is also designed to transmit an advertisement sequence number in the message of the advertisement. This number corresponds to the number of the advertisement the Broadcaster is emitting.

1.2.3 Ground Truth

Ground truth is essential to validate the results of the experiments pertaining to the selected cases of pedestrian movement and activities. Those selected cases were identified to reflect a diverse range of pedestrian activities and movement, and will be presented in Section 1.3. Three different approaches were investigated to capture the ground truth. The first approach was hardware-based which was dropped due to its unreliability, and is discussed in Appendix C. The other two were software-based approaches and are presented in the next section.

1.2.4 Software

Scripts to employ BLE on the RPi to facilitate experiments were developed primarily in Python. The core script dealing with capturing and processing the BLE advertisements was written in Python and will be referred to as *main script* from here on. Since the beacons used for most of the experiments were pre-programmed by the manufacturer and non-reprogrammable devices, they were used without alteration. However, for one study, on the effect of the advertisement rate of the Advertiser on the detection and identification of pedestrian activities, programming was required to adjust the advertisement rate. Therefore, an alternative Advertiser was developed using an RPi, which will be presented in Section 1.3.2.5 in this chapter. This alternative Advertiser was programmed using NodeJS.

MEAN stack was also used for the development of a local web app or, an interactive portal, to observe the acquisition of BLE signals by the Observer in real time.

A web app was developed, that also ran on the same RPi which was used as an Observer through the *main script*. InfluxDB was used as the preferred database to store data acquired from BLE devices and ground truth devices, and will be presented in Section 1.2.4.6. Finally, both MATLAB and Python scripts were used to implement and perform analysis of the acquired data.

1.2.4.1 Main Python Script

This Python script was developed to continuously scan advertisements from nearby BLE devices without any sleep cycles to ensure no data point was missed. *BluePy* ([IanHarvey/bluepy n.d.](#)) Python library, based on the *BlueZ* ([BlueZ n.d.](#)) stack, was used for this purpose. As presented in Section ?? in Chapter ??, the *hybrid* approach is selected for capturing measurements. This decision was taken to ensure that both the RSSI and the time of measurement of RSSI are captured by the device. This enables spatio-temporal analysis of the measurements which is necessary to study the change in movement of pedestrians.

The library was modified and two features were added to it. The first code snippet is included to save received BLE signals to the InfluxDB database. The data stored includes raw timestamped RSSI, a filtered RSSI using the sliding-window SMA algorithm (will be described in Section 1.4.2), and the MAC address of the Broadcaster. The rationale behind the selection of these parameters will be presented in Section 1.2.5 in this chapter. The second modification is the addition of the simulation of the *whitelisting* feature to ensure that only the signals emerging from the Advertiser carried by the volunteer pedestrian are intercepted and stored. Signals originating from any other BLE device are discarded by the library. A flowchart depicting this is presented in Figure 1.8. This script and the modified library will be referred to as the *main script* and *library* respectively from here on in this report.

1.2.4.2 NodeJS Broadcaster Script

For the alternative Broadcaster, which was developed for the RPi, NodeJS was utilised to develop the required software. The choice for NodeJS over Python is because the library for NodeJS allows a simple way to create a Broadcaster which can be invoked with a defined advertisement rate at the time of execution. The library employed is Abandonware's Noble ([@abandonware/noble - npm n.d.](#)). The NodeJS script performs two tasks, first, it sets up the advertisement rate defined by the user and second, it sends the sequence number in the advertisement message. This sequence number is the number of the advertisement being broadcast. So, the first advertisement packet contains 1, the second packet contains 2, and the n^{th} contains n . The sequence number is included in the message to understand the number of dropped advertisement packets.

1.2.4.3 Blue Dot (Ground Truth)

BlueDot ([Blue Dot - Apps on Google Play n.d.](#)) is an Android phone application comprised of a blue button on the screen, as shown in Figure 1.9. This application allows transmitting the acknowledgement of button presses

General Guide

Pairing is the process of establishing a connection between BT and BLE devices. Two connected devices are called *paired* devices.

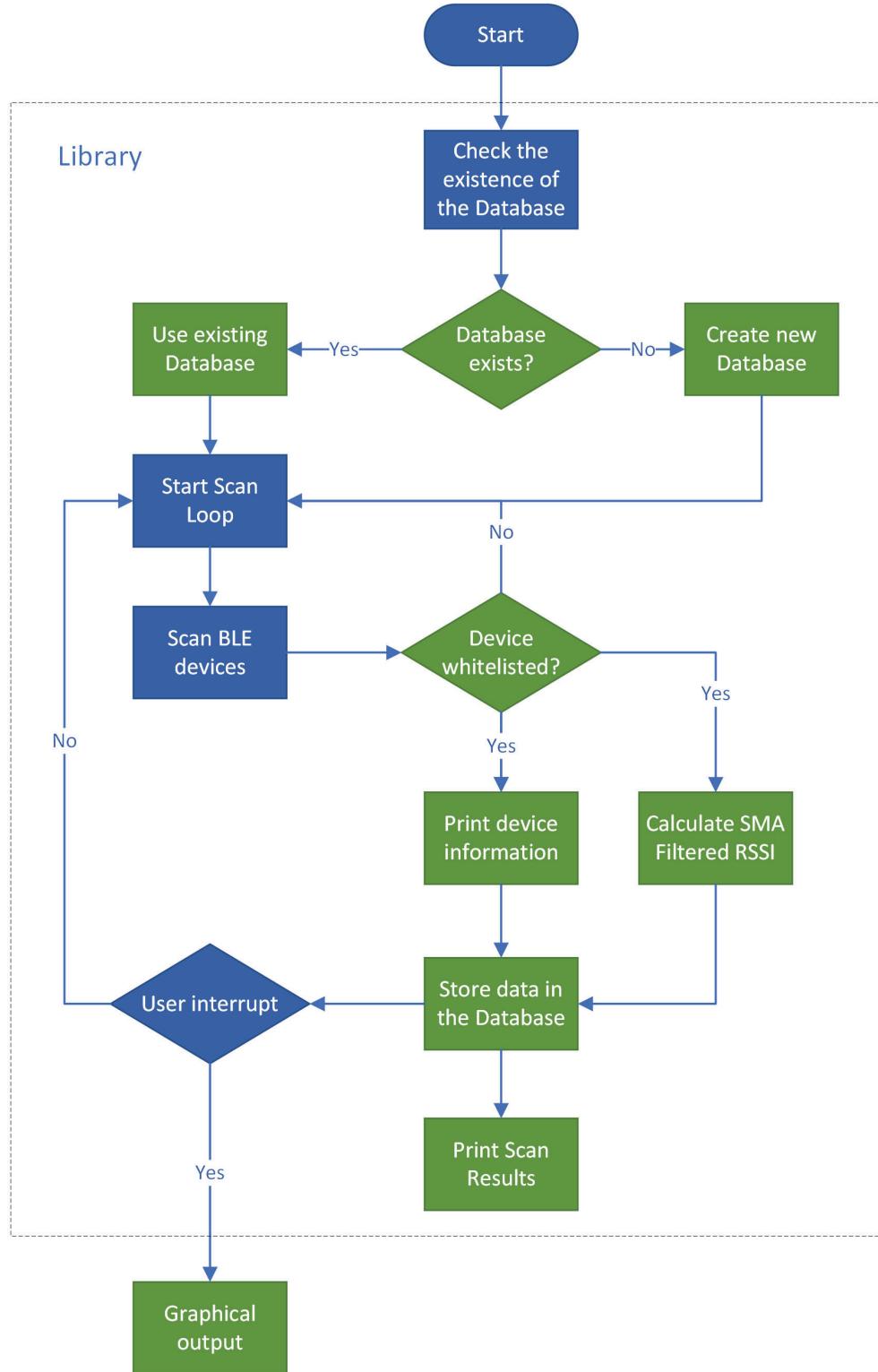


FIGURE 1.8: Flowchart for the Main Script and Library with Modification Highlighted in Green Coloured Elements

to an already paired BT-enabled device at a time. To use this application, the Android phone employed was paired up with the RPi that was used as an Observer.

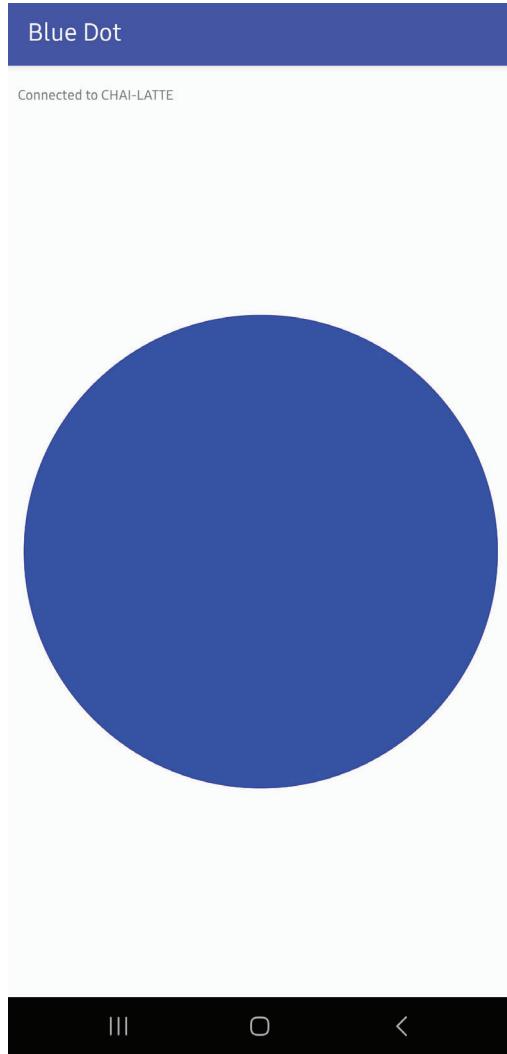


FIGURE 1.9: BlueDot Android Application

The Python script developed earlier to capture the hardware ground truth was repurposed to listen for messages over BT. The only information received from the application is that the the volunteer pedestrian has pressed the button on the application screen, the Python script contains its own list of location aliases. Upon reception of a button press message, this script picks up the first location in the list, records it in the database along with the timestamp and moves to the next location to be stored upon the subsequent button press. This happens in a round-robin fashion to ease the repetition in the experiment.

Since no crucial information about the location itself is transmitted between the two devices, this approach prevents any spoofing during transmission that may require any Privacy-Enhancing Techniques (PET) under the "private communication" subsection described in Appendix B.

The Python script described here will be referred to as the *ground truth script* henceforth. Figure 1.10 describes this process.

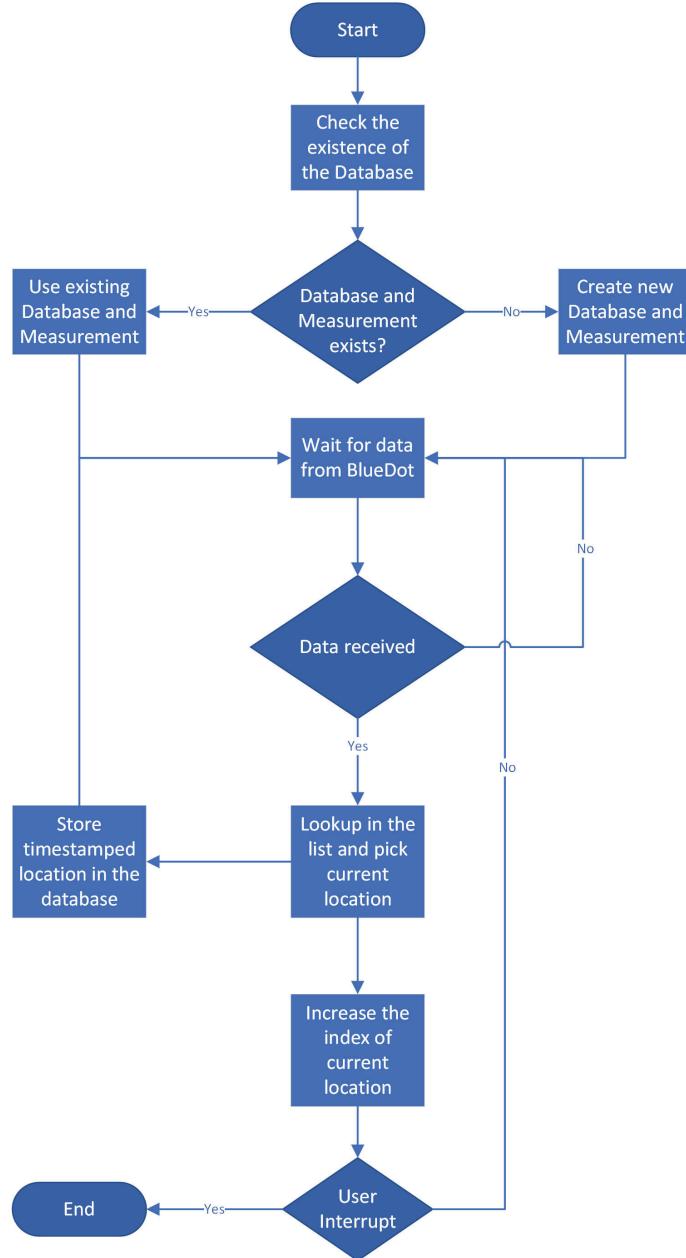


FIGURE 1.10: Flowchart for the Ground Truth Script²

1.2.4.4 GPS Logger (Ground Truth)

Another alternative to capturing ground truth data is through logging GPS locations of the volunteer pedestrians. An Android mobile application, GPS Logger ([GPS](#)

²Note that "Measurement" in the figure is logically equivalent to "Tables" in SQL in InfluxDB.

Logger - Apps on Google Play n.d.), is installed on the phone carried by the volunteer pedestrian. This application stores the geo-location of volunteers as they move on the predefined pathway for the purposes of the experiment. The geo-locations captured remain on the phone and are not transmitted to any other device. At the end of the study, the captured data is exported to the laptop for analysis. Since, the micro-level ground truth data, as captured through this method, is only required for some specialised cases, this method was used sparingly. A screen capture of the GPS Logger application is shown in Figure 1.11.



FIGURE 1.11: GPS Logger Android Application

1.2.4.5 Interactive Portal

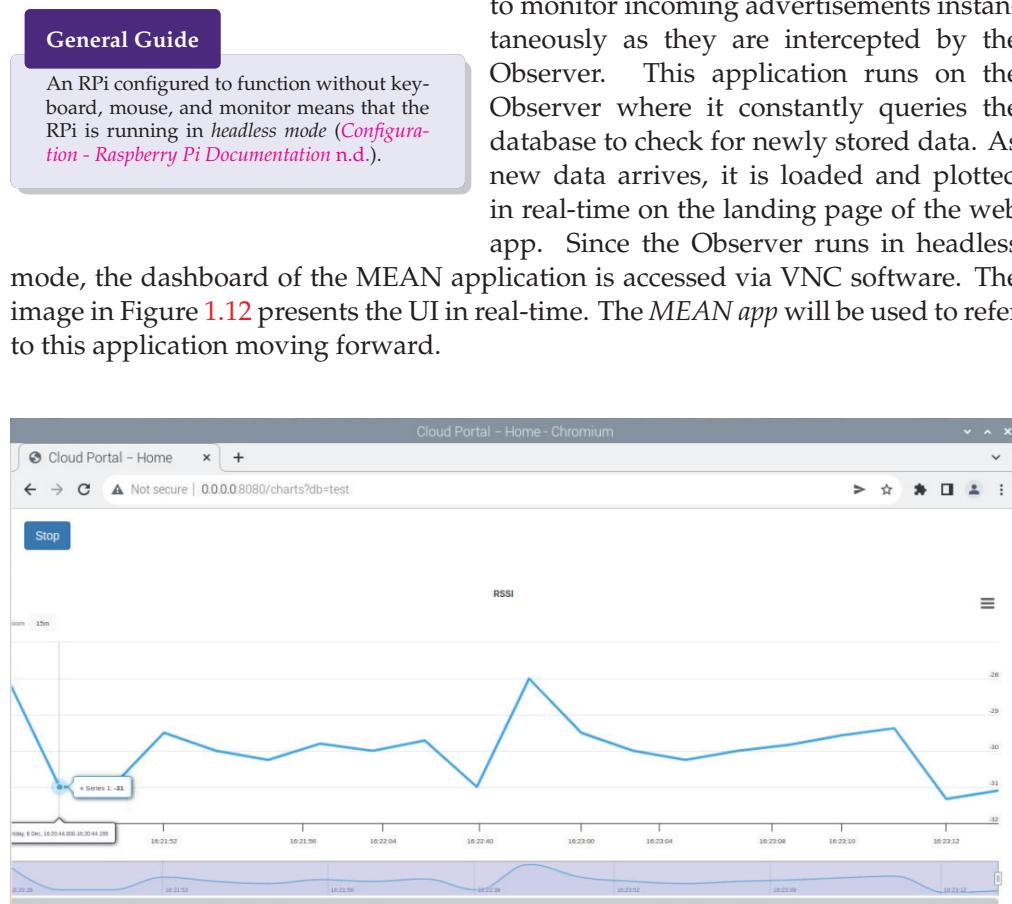


FIGURE 1.12: UI of MEAN App

1.2.4.6 Database

InfluxDB ([InfluxDB Time Series Data Platform | InfluxData n.d.](#)) is used as the database to store all the data locally on the Observer. InfluxDB is selected for two reasons, first, it boasts a feature called *ephemerality*, and second, it is a purpose-built database for time-series data. As it is a time series database, InfluxDB specialises in indexing, collecting, and storing data based on their timestamps. It enables fast querying on time ranges and ingesting high-throughput data which is ideal for the application presented in this dissertation where data is continuously generated and is centred around its time of acquisition. The former reason, *ephemerality*, implies that the database has an expiry, after which, the data is permanently deleted by the database itself. This strengthens the core principle of privacy-preservation which is a common theme throughout this PhD research. The latter reason is due to the correlation between the features of BLE Advertisements and any other time-series signal. While the literature has no evidence of BLE signals as time-series data, the signals can be treated as noisy or scrambled time-series data as explained below.

Highlights

There is an underlying default advertisement rate at which any BLE Broadcaster advertises. The signals when intercepted continually correspond to regularly sampled time-series data. However, the Observer may not always be able to reproduce the received advertisements as regularly-sampled signals. This is due to several limitations of the environment and the antenna characteristics such as packet collision, multipath collision, or even the delay in the reception of the data due to latency of the software to process the signals on the Observer. However, for all but the most aggressive advertisement rates, the short delay in reception and processing of signal on the Observer should not be a major concern. This is because the non-aggressive rates will correspond to macro level pedestrian movement dynamics, as opposed to finer granularity of movement that aggressive advertisement rates are able to capture. Therefore, for the case of aggressive rate, say 100 ms for example, even a delay of a second in reception and processing of that signal by an Observer might mean that the pedestrian has covered just over 1 metre distance as per a slow-walking pace of 4 kilometres per hour. Moreover, other signals might be intercepted by the Observer before the delayed signal, which might introduce an anomaly in the dataset. Since the non-aggressive advertisement rate corresponds to spatio-temporally dispersed signals, short latencies do not pose significant concerns. This analogy of BLE advertisements with sparsely populated, regularly sampled time-series signals also opens avenues for the application of time-series techniques to analyse the data.

1.2.5 Protocol for Data Collection

In line with the guiding principles outlined in Section ?? of Chapter ??, the following elements are identified for data collection.

1. Main Script

- **UUID:** While UUID does not fit the criteria of the guiding principles since it infringes the privacy of the pedestrian, it is not a necessary element for data analysis. It has been captured only for the sake of experimentation to be able to identify the beacon used. During real-world deployment, methods such as hash encoding, as described in Chapter ?? can be used for pseudo-anonymisation of the UUIDs and to generate a unique identifier to distinguish between different devices .
- **RSSI:** RSSI is an important measure since it is one of the main parameters utilised in the data analysis.
- **Timestamp:** Another important parameter that is used in data analysis. It is important to note that the timestamp here is not the time of the advertisement event at the Broadcaster. It is the time that is assigned to the measured advertisement at the *application* layer in the Observer, therefore, it accounts for the propagation duration and the time it takes for the advertisements to travel through the layers up to the Observer's application layer. This extra duration could be negligible or substantial based on the propagation route and the processing capabilities of the Observer.
- **SMA-filtered RSSI:** This parameter is not acquired through the advertisement but derived from the RSSIs.

- Standard Deviation (SD): Another parameter that is derived using the acquired RSSIs.
- Sequence Number: In one of the experiments, to follow in Section 1.3.2.5, sequence number, an incremental number generated by the Broadcaster for each advertisement broadcasted, is enclosed within the advertisement packet. This is performed to understand which packets are not measured by the Observer. Since this number is a system-generated sequence, it aligns with the privacy preservation philosophy presented in this dissertation.

2. NodeJS Broadcaster

- UUID: Similar to the *main script*, UUID is transmitted for the purpose of identification of the device. In a real-world deployment, many commercial devices frequently change their UUIDs. Even if a user device is not programmed to change its UUID, measures such as hash encoding of UUIDs in the Observer can preserve the privacy in the data.
- Sequence Number: It is an incremental number sent by the Broadcaster that corresponds to the number of advertisements broadcasted. This is a system-generated parameter and does not correspond to any user data.

3. MEAN Stack Dashboard

- The dashboard only utilises the data acquired by the Observer and displays only measured RSSI and timestamp. In the case where multiple Broadcasters are simultaneously visible to the Observer since the Observer can hash encode the UUIDs, only the stored hash encodings are used to differentiate between the whitelisted devices.

4. BlueDot and GPS Logger

- Both of these Android applications are required only to calibrate the deployment in a configuration. These applications are not a necessity for real-world deployment.
- On a volunteer’s device, the BlueDot application only relays the acknowledgement of the button being pressed on the volunteer’s phone to the Observer. The Observer associates the button press via Bluetooth Classic to a pseudo-location from a series of locations hard-coded in the Observer, and increments the sequence in the set of locations. The Observer attaches a timestamp to the received location associated with the button press when the information of the button press is received at the application layer.
- GPSLogger, on the other hand, captures the geolocation of the volunteer during the experimental phase, and the log files are stored on the device provided to the volunteer for the duration of the experiment. The log files contain geolocations along with timestamps on the provided phone. These files are then copied to a local computer for investigation.

5. Parameters for Data Analysis

- RSSI, SMA-filtered RSSI, and timestamps are used for most analysis. In experiments where more than one Broadcaster is employed, UUID is used to distinguish the Broadcasters.

- For ground truth obtained from BlueDot, the pseudo-location and timestamps are used.
- For ground truth obtained from GPSLogger, geolocations and timestamps are used.

1.2.6 Privacy by Design

Each element of the platform is designed to minimise privacy concerns. Emphasis has been placed on restricting any collection of non-consensual data, for instance through the use of whitelisting. Furthermore, measures have been implemented to ensure that even consensually collected data is protected against privacy threats in real-world deployment. The steps taken to ensure that privacy preservation is intrinsic to the design of the platform are discussed below. Some of these have already been discussed earlier but are revisited here in this section for further emphasis.

1.2.6.1 Hardware

1. Observer

- The device chosen as the Observer is a RPi which has an embedded chipset for BLE. This restricts certain types of cyber attacks such as eavesdropping, that could target the *L1* or the *Perception Layer* of this platform since these attacks occur in the presence of a physical separation of the sensor and the device that processes the data from the sensor.
- As an *Observer* type BLE device, the device is incapable of establishing a connection with any other BLE enabled device, and hence, reduces the need for any of the PET discussed under *private communication* and overcomes the *inventory* privacy threat, as discussed in Appendix B.

2. Broadcaster

- The Broadcaster employed is a non-personal device. Since it is a device handed out to the volunteer, there is no possibility of a privacy attack establishing a connection with this device through *fingerprinting and impersonation*.
- The Broadcaster is also selected as a *Broadcaster* type of BLE device which is inherently incapable of establishing a connection. Even if this platform is to be deployed in a real-world setting, where the devices used by the public can have *peripheral* roles, *whitelisting* will be useful to disallow interaction with the devices of users who are not on the list, and even if whitelisting is not used, the role of the monitoring platform is *Observer* which will be incapable of connecting to them.

1.2.6.2 Software

1. Main Script and Library

- The library is modified to enable whitelisting, and hence captures the data emanating only from the whitelisted Broadcasters of volunteer pedestrians. This modification was necessitated as the Python library, BluePy, employed for the experiment did not support whitelisting. This feature was hence manually written to perform as per the BT/BLE specification

(Sig, 2014), however, in the application layer. Also, the code snippet to add data to the database is written in the library instead of on the script at the user side for future-proofing for real-world deployment as the library can then be obfuscated using a tool such as PyArmor (Zhao, 2023) and stored in the read-only parts of the OS, to prevent intruders modifying or accessing the sensitive portion of the script.

- The main script is visible to anyone with access to the device but contains no sensitive code. It only *delegates* to the library all the tasks related to the interception of BLE signals, identification of whitelisted advertisements, and their storage in database.

2. Database

- The database, InfluxDB, is chosen because it is a suitable option for time-series data. Another feature that InfluxDB offers is the ability to make *ephemeral* data collections, that is the data stored can be given an expiry date after which it is automatically deleted. This means that when the platform is deployed in the real world, this feature can be used to further auto-delete the detailed data captured on the local device. Data expiry can be set such that on-board processing can be performed in chunks in quick succession, and raw data can be deleted in hours if not minutes to prevent leakage of data in case the device itself is accessed in an unauthorised manner.

3. Ground Truth Script

- This script only listens to the signals emerging from the MAC address of the device employed for ground truth storage, which is required to be paired up with the Observer beforehand. The phone application only sends a message to signal to the Observer that the button on the application is pressed and that message contains no information about the location itself. The locations for the experiments are hard-coded in a list and the script associates the timestamp of when an incoming message is received with one location from the list in a round-robin fashion and stores it in the database. Since all the experiments conducted are controlled experiments, the locations are given pseudo-names in the form of alphabets, such as "Location A", "Location B", etc.

4. Ground Truth GPS Logger

- The data from the GPS Logger application is never transmitted to any other device. It only stores the data locally which is, at the end of the experiment, copied over to a University laptop for analysis.

5. MEAN App

- The dashboard shows the chart of time vs RSSI in real-time and does not show the MAC address of the device of origin.

These decisions were carefully taken to approach the system with a PbD outlook. Admittedly, the MEAN app is a potential threat where an intruder can access the data to fetch the time and RSSI of the Broadcaster. This has the potential for a *linkages* privacy threat and a *data inference* privacy attack. However, this application is only made to ensure the experiments are running without any errors by observing the

data stream and is not a necessary feature for the platform to function and therefore, can be removed when needed.

1.3 Experimental Setup for Evaluation

The evaluation mechanism here is twofold. The first parameter to evaluate is the technology, BLE, itself. This involves identifying the specifics of the hardware used and its effect on the application including the characteristics of the antenna of the Observer and how it performs outdoors, the range at which the antenna of the device can receive usable signals, and the rate at which it can capture signals that are useful to draw assertions. The second part then, is an investigation of scenarios of pedestrian movement and activities that such a system can detect, identify, and characterise, given that the experiments are conducted using the configuration identified through the experiments from the first part. This section is hence divided into two parts to discuss the aforementioned parts by category. However, before outlining the specifics of the evaluation, it is essential to describe the location of the experiments since except for one, all the other experiments are conducted in the same outdoor space.

1.3.1 Experimental Location

As presented in Section 1.1 of this chapter, for assessing the feasibility of BLE for measuring pedestrian dynamics, a study must first be carried out in an urban canyon with a relatively simple topology. This facilitates more the technology and clearer understanding of the extent of its application with minimal supplementary influence before extending the application in complex scenarios. Another reason behind this decision is that the research presented in this dissertation is a natural progression of work carried out by Ahlam AlAnbouri, another researcher in the same research group. A key aspect of AlAnbouri's work (AlAnbouri et al., 2019), as also seen in Section ??, utilised BLE in an open space, but near a large water body (that reduces the reflection of signals) with no buildings apart from a pump house and a lighthouse, and with low footfall. The location selected for my research includes a tall vertical wall on one side and an open space on the other. The vertical wall is the side of a 5-storey tall building, mostly covered by glass windows. While the open ground used in the study is not shared by any other pedestrians or vehicles, the immediate surroundings are open to the public, and being inside the university campus, the area is busier than the location of AlAnbouri's related work (AlAnbouri et al., 2019). The choice of location in the research presented in this dissertation is hence, a natural extension of the study by AlAnbouri that is applied in a slightly more complex urban canyon. All the outdoor experiments conducted during the course of this PhD were deliberately performed in the same location to have a coherent set with same topology.

As previously mentioned, since the selected location had an open ground on one side, there were no defined paths, which allowed for marking pathways suited to the experiments. The satellite image of the experimental space is shown in Figure 1.13. A low wall adjoining a five storey building, composed of brick work and large glass windows, and also provided a suitable place to deploy the Observer. Figure 1.14 depicts the building on one of the sides of the experimental location and the adjacent

low wall where the Observer (in white enclosure) was deployed. For each experiment, the marked path was always 24 metres long, irrespective of the distance of the path from the deployed Observer. The path was also marked with five equidistant key points, *start*, *approach*, *centre*, *depart*, and *end*, each 6 metres apart. Whiteboard markers were used to label those key points, as shown in Figure 1.14. The '*path*' was a surface covered with grass and wildflowers with height spanning between 30mm to 150mm, as presented in Figure 1.14. As pointed out by Rappaport in (Rappaport, 2002) and Liao et al. in (Liao, 1990), the presence of grass may have an effect on BLE signals caused by diffraction or absorption of signals, described in Section ?? in Chapter ?. The layout is presented in Figure 1.15. The paths used for different experiments varied between the combinations of 3 metres, 5 metres, 7 metres, and 9 metres away from the deployed Observer. For one of the experiments, four points, *P*, *Q*, *R*, and *S*, were also chosen at random to merely highlight the effect of distance on the signal strength, RSSI.



FIGURE 1.13: Satellite View of the Experimental Location

The other side of the experimental location was covered by 2-3.5 *metre* high metal and wooden infrastructure at a distance of approximately 30 *metres* from the low



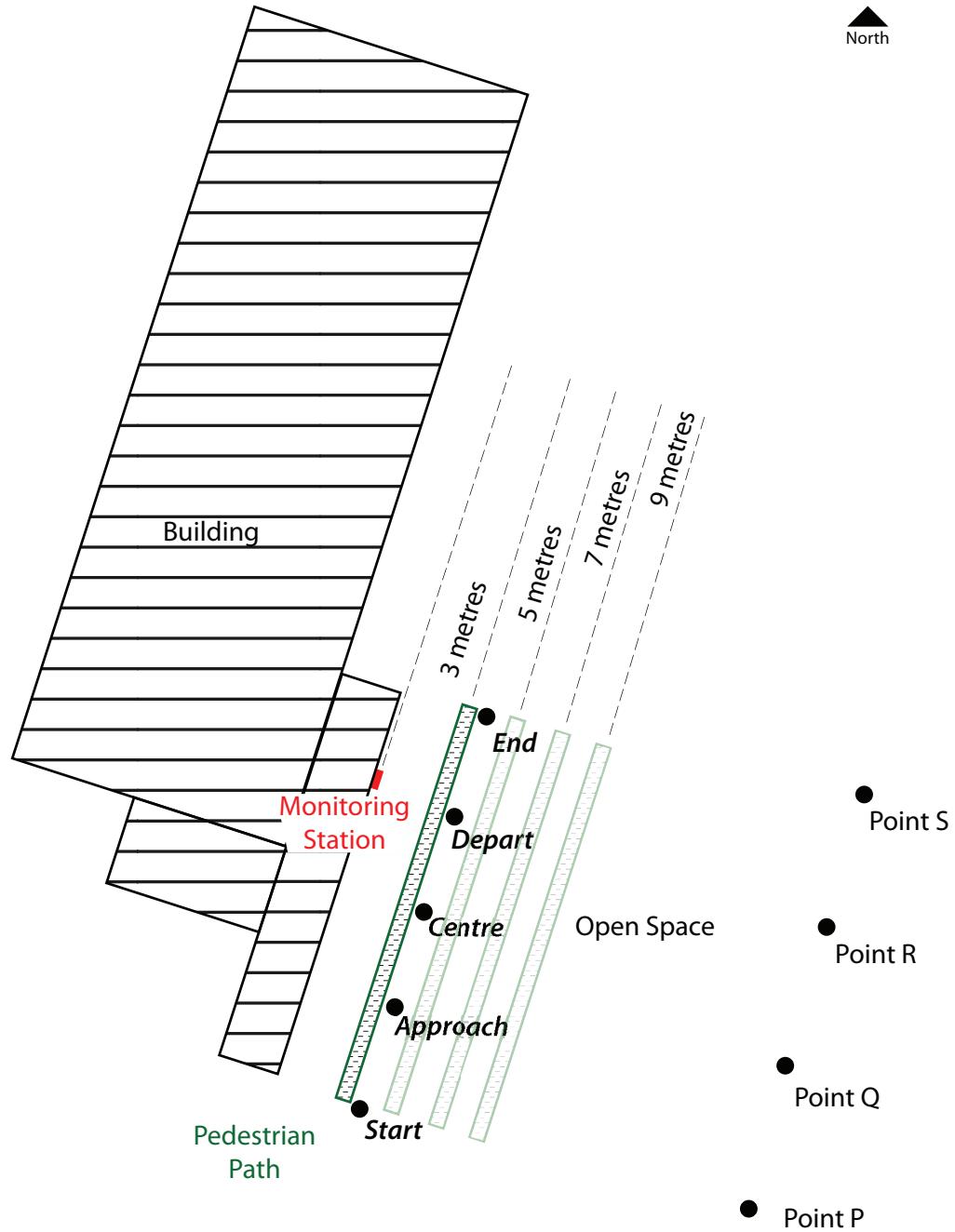


FIGURE 1.15: Experiment Location Layout

wall where the Observer was deployed. This is presented in Figure 1.16. The volunteer pedestrian, depending on the specifics of the experiment, was stationary or moving along the selected pathway with a Broadcaster in their hand. Figure 1.17 depicts a volunteer pedestrian on a representative experimental pathway.



FIGURE 1.16: Metal Infrastructure on the Far End of the Experiment Location

1.3.2 Experimental Setup for Evaluation of the Technology and its Features

1.3.2.1 Characteristics of the Antenna of the Observer

Different manufacturers of embedded systems and SBC with BLE may use different chipsets, antenna designs, and transmission characteristics such as antenna gain. In the literature review, it has been noted that devices from different manufacturers have behaved differently and therefore, those researchers have forewarned about the issues that may pertain. If the platform is to be deployed in the real world, it will be difficult to control such characteristics as the chipset and antenna on the user devices. However, it is intuitive that such devices are made to connect and/or communicate with other BLE devices and therefore, their performance must be comparable. Keeping this in mind, an investigation to understand the antenna on the RPi (that is used as an Observer) and its characteristics must be carried out. This is to ensure that there is a clear understanding of the sensitivity of the antenna on the Observer. Such sensitivity features can be employed to aid the assertions. Moreover, measurements of RSSI from the anechoic chamber represent measurements in an ideal scenario, where no external noise is influencing the signal strength. This provides a baseline set of patterns which can be used to compare data collected in the comparatively noisy outdoor environment to identify the degree of their deviations



FIGURE 1.17: Volunteer Pedestrian on Representative Experimental Pathway

from ideal conditions.

The antenna of the RPi was studied in two environments: in a noiseless Anechoic Chamber and in an outdoor space adjoining the office in the Grangegorman campus, and was conducted in three parts. The first part was studied in an anechoic chamber, whereas, the second and third studies were conducted in an outdoor environment, as discussed in the following paragraphs.

In the anechoic chamber, the Observer was kept on a rotating platform. The Broadcaster was then placed exactly 3 metres away from the Observer. The rotating platform was then rotated starting from 0° at discrete intervals of 45° until 180° along the horizontal plane of the RPi in the enclosure, as shown in Figure 1.18, where the orientation of the RPi within the enclosure is illustrated in Figure 1.4. This particular orientation of the RPi has shown optimum results in the study conducted by Fleuratoru et al in (Fleuratoru et al., 2022). For each rotation interval, the *main script* captured the BLE advertisements for 3 minutes. This experiment allowed for the understanding of the strength of the signal and the fluctuation in the signals for the entire duration of the experiment in a noiseless environment. This was then used as a basis to assess the strength and fluctuation of the signals obtained in a noisy outdoor environment. The intended methodological decision while conducting this experiment was that if the BLE signals are too weak in the outdoor setting or are too jittery to obtain meaningful patterns from the acquired signals, then the technology must be sidelined for this application.

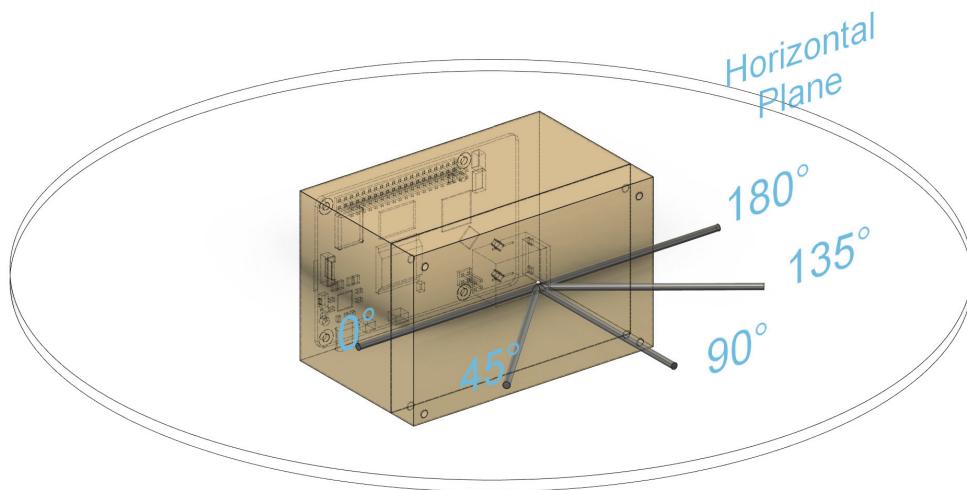


FIGURE 1.18: Depiction of the Horizontal Plane of the Observer

The experiment was then taken to the outdoor environment for conducting the remaining two parts. In the second part, the characteristics of the antenna were assessed in the same manner as they are in the anechoic chamber, that is, by assessing the directional sensitivity of the antenna towards the signals emerging from the Broadcaster at a constant distance from the Observer. The Observer was deployed at the outside wall of the building as illustrated in Figure 1.15, and the volunteer pedestrian holding a Broadcaster device was instructed to stand 3 metres away from it at

the same angles, 0° , 45° , 90° , 135° , and 180° , as in the case of the anechoic chamber. Since there was no rotating platform outdoors, in this case, the volunteer was instructed to move around the Observer. The advertisements were collected by the Observer for 3 minutes at each location. This case was further expanded and the measurements were also collected for all those angular resolutions at a distance of 5 metres, 7 metres and 9 metres between the Observer and the volunteer Broadcaster.

Secondly, two 24-metre long paths were laid out 3 metres and 5 metres away from the Observer. Five equidistant points, also illustrated in Figure 1.15 and Figure 1.19, *Start*, *Approach*, *Centre*, *Depart*, and *End*, were marked on each of the two paths where the volunteer pedestrian was directed to stand for a period of 3 minutes each. The Observer collected the advertisements at each of these locations using the *main script*. This experiment did not fully corroborate with the experiment conducted in the anechoic chamber, as in the anechoic chamber each point was on a radius exactly 3 metres away from the Observer. This sub-experiment on the linear pathway was performed, in addition to the replication of the anechoic chamber scenario in an outdoor setting, as a representation of a real-world scenario where we have a linear path. Due to the linear nature of the pathway, the distance between the Broadcaster and the Observer on the paths at 3 metres and 5 metres away from the Observer were 3 meters and 5 metres respectively only at the *centre* point of each of those pathways, which was when the Broadcaster is directly opposite the Observer. The linear nature of the pathway meant that moving away in any direction along the path increased the distance, which can be calculated using Pythagoras rule. An illustration of the increase in distance on different key points on the pathway is also depicted in Figure 1.19. For instance, on the selected 24 metre long pathway, with five equidistant key points, separated by a distance of 6 metres from the neighbouring key points, the distance between the Broadcaster and the Observer at the *approach* and *depart* point was 4.24 metres.

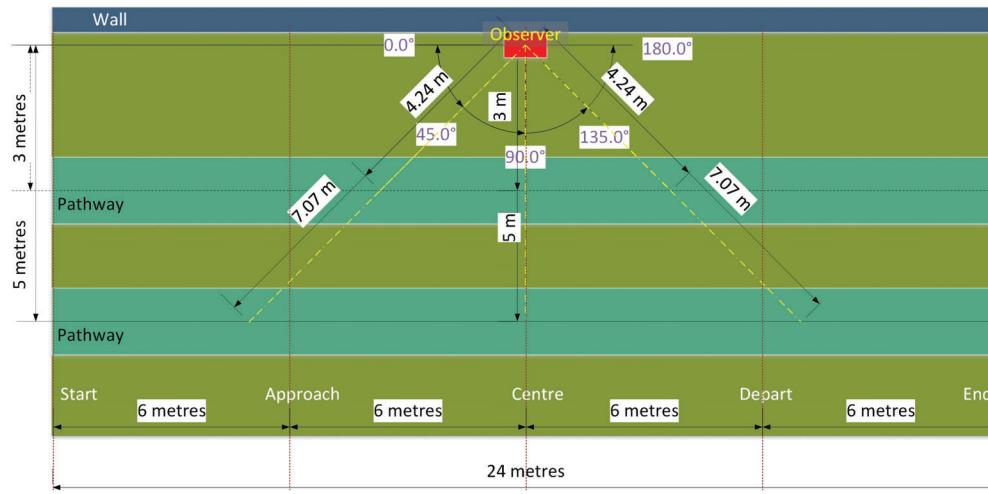


FIGURE 1.19: Experimental Layout and Effect of Pythagorean Rule

Since the distance between the Observer and the Broadcaster varied at different key points along the linear pathway, unlike the fixed distance between them in the anechoic chamber, the signals were expected to attenuate differently on each key point. That means the least distance-related attenuation must have occurred when

the Broadcaster was at the *centre* key point since it was the closest approach to the Observer along the pathway, followed by the *approach* and the *depart* key points, and finally, the *start* and the *end* key points. The attenuation is in accordance to the path loss, as described in Section ?? in Chapter ?. Therefore, the selection of BLE as a technology to measure pedestrian dynamics could be assessed by evaluating the quality of the signals emerging from the Broadcaster at the *start* and *end* key points, that is, the farthest locations on the pathway.

For both the experimental setups, advertisements are collected for 3 *minutes*, in three 1-*minute* segments. The data collected consists of the MAC address, timestamp RSSI, and SMA of RSSI, details of SMA will be covered in Section 1.1 of this chapter.

The following is the inventory for this experiment:

1. Observer
 - (a) RPi based device running *main script* which stores the data in InfluxDB.
 - (b) The device is also running MEAN stack dashboard for live monitoring.
 - (c) Battery pack and RTC connected to the RPi.
2. Broadcaster
 - (a) RuuviTag beacon with advertisement rate of 2 Hz (advertising once every 500 ms).

The following are the protocols for the experiments conducted in the anechoic chamber and in the outdoor environment:

1. Anechoic chamber experiments
 - (a) The Observer, the entire unit with RPi connected to a power bank and RTC in an enclosure, is placed on a rotating platform.
 - (b) The pipe straps are removed from the enclosure to avoid any interference with the signals.
 - (c) One of the holes for the screw to position one of the pipe straps in place is enlarged just enough to pass a CAT6 Ethernet cable through. This Ethernet cable is a standard feature and part of the anechoic chamber at the University, used to monitor and control the device remotely from outside the chamber during live experiments.
 - (d) The Broadcaster is stationed at a platform which is exactly 3 metres away from the rotating platform where the Observer is placed.
2. Experiments are outdoor locations beside an office building
 - (a) The Observer is positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
 - (b) The Observer is the entire unit used in the case of the anechoic chamber experiment except without the Ethernet cable.
 - (c) For the measurements around the Observer, the volunteer pedestrian is instructed to stand at the designated location holding the Broadcaster facing directly in line of sight of the Observer.

- (d) The orientation of the Broadcaster is with its front face facing towards the Observer, as in the case of the anechoic chamber.
- (e) The researcher, after ensuring that the volunteer is appropriately in position, launches the *main script* and stopwatch. The volunteer is instructed to be at ease by the researcher once the measurements are taken for the desired period.
- (f) The volunteer is instructed to stand at the location and the process is repeated.

The results of these experiments will be presented in Section 2.1.1 in Chapter 2. Moreover, the outcome of the anechoic chamber portion of this experiment will be used in the experiments described in Sections, 1.3.2.2 and 1.3.2.4 to evaluate reference RSSI at reference distance to calculate fading and to compare the RSSI measurements in anechoic chamber against outdoor environment respectively.

1.3.2.2 Large-Scale (LS) and Small-Scale (SS) Fading

Fading is a prominent feature in wireless communication that is prevalent in an outdoor environment. As discussed in Section ?? in Chapter ??, one of the important factors, aside from the propagation distance, behind the attenuation of BLE signals is fading. Empirically, this relationship has been presented in Equation ?? in Chapter ??.

This experiment was designed to understand the effect of the environment where the experiments were conducted through Small-Scale (SS) and LS fading. While SS fading provided insights into the frequency and intensity of the reflections signals undergo in the experiment location, LS fading shed light on shadowing due to the presence of large objects in the environment.

RSSI measurements were collected in two different scenarios. First, measurements were recorded at a fixed distance of 3 metres in an anechoic chamber for different antenna angles – 0°, 45°, 90°, 135°, 180° – relative to the Observer, as mentioned in Section 1.3.2.1. Second, RSSI measurements were taken at five key points – *Start*, *Approach*, *Centre*, *Depart*, and *End* – on a linear pathway, 3 metres from the deployed Observer.

Since, the signals at the Observer’s antenna were not received with a sensitivity that was symmetric around the face of the Observer (RPi), measurements captured at all five angles were averaged. Reference path loss, PL_{d_0} was calculated using the measurements in the anechoic chamber.

For LS fading, a path loss model was fitted to the outdoor environment data to predict the expected RSSI at each location based on the calculated reference path loss from the anechoic chamber. This predicted path loss was calculated using the expression presented in Equation ?? in Section ?? of Chapter ???. Residuals, which will be were calculated by taking the difference between the measured and fitted RSSI at each location.

To calculate SS fading, Rician Distribution (which will be explained in Section 1.4.11) was fitted to the deviations in the measured RSSI. Shape and Scale parameters were obtained from the result of the Rician distribution fitting.

The inventory and protocol for this experiment is the same as in Section 1.3.2.1. The results of this experiment will be presented in Section 2.1.2 in Chapter 2.

1.3.2.3 Optimal Deployment Distance of the Observer

Once the suitability of BLE was confirmed through the experiment presented in Section 1.3.2.1 in this chapter for measuring pedestrian dynamics, the next question then was about the placement of the Observer. In this particular experiment, the optimal deployment distance of the Observer that could produce discernible patterns in the resulting RSS values to indicate any characteristic of the pedestrian movement or could provide descriptive patterns of any pedestrian activity was tested.

To do this, a set of deployment distances was chosen and the distance between the Observer and the linear pathway was varied for each distance in that set. Since the research location offered an open space on one side, as illustrated in Figure 1.15, the task of varying the distance was achievable. Three new paths were laid out at a distance of 5-metre, 7-metre, and 9-metre, parallel to the 3-metre path seen in the previous experiment. Each path had the same five equidistant key points, from *Start* to *End*, marked as illustrated in Figure 1.20. The volunteer pedestrian was directed to stay stationary at each of those five points on all four pathways for three repetitions of 1-minute stationary pauses each for two cases: first, when the Broadcaster faced towards the Observer, that is LoS, and second, when the Broadcaster faced away from the Observer with the body of the volunteer between the two devices, that is nLoS.

It is also important to note here that occlusion in the experiment does not conform to complete occlusion as in the case of nLoS. Complete occlusion occurred only when the volunteer pedestrian was at the *centre* key point on the pathway. A partial occlusion occurred at *approach-centre* and *depart-centre* points. And, line-of-sight occurred at *start* and *end* points. However, for the sake of simplicity, all of the nLoS cases are referred to as cases of occlusion from here on. This differentiation between full occlusion and partial occlusion is illustrated in Figure 1.21.

Advertisements were collected for 3 minutes, in three 1-minute segments. The data collected consisted of MAC address, and timestamped RSSI, and SMA RSSI (1.1).

The following was the inventory for this experiment:

1. Observer
 - (a) RPi based device running *main script* which stored the data in InfluxDB.
 - (b) The device also ran a MEAN stack dashboard for live monitoring.
 - (c) Battery pack and RTC connected to the RPi.
2. Broadcaster
 - (a) RuuviTag beacon with advertisement rate of 2 Hz (advertising once every 500 ms).

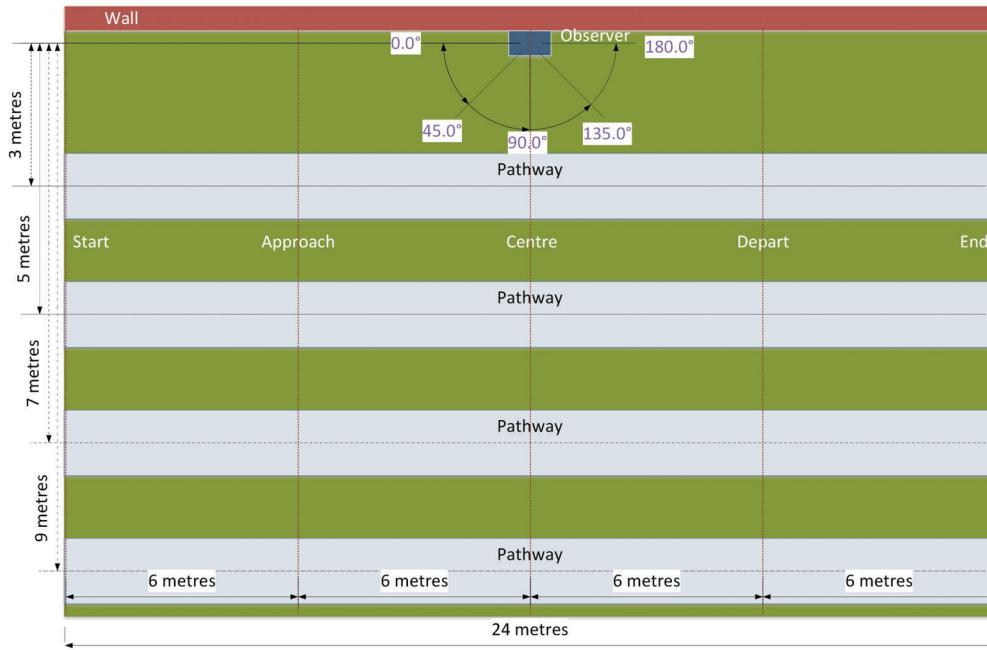


FIGURE 1.20: Experimental Layout for Assessing Optimal Deployment Distance Showcasing all Pathways

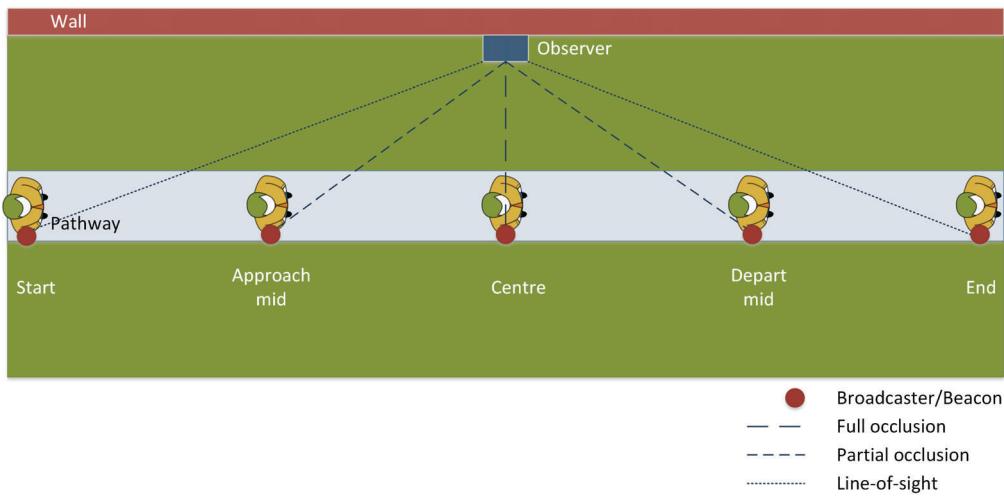


FIGURE 1.21: Depiction of Occlusion States in the Case of nLoS

The following were the protocols for the experiment conducted in the anechoic chamber and the outdoor experiment:

1. The Observer was positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
2. The Observer was the entire unit that includes a battery pack and RTC, all enclosed in the enclosure.
3. The volunteer was instructed to stay stationary at each point on all of the pathways one after the other.
4. The volunteer was instructed to hold the Broadcaster in the hand facing towards the Observer for the case of LoS. The volunteer was provided with clear instructions to keep the face of the Broadcaster uncovered, that is to not enclose the Broadcaster in their fist.
5. Similarly, the volunteer was asked to switch the hand in which the Broadcaster was held for the nLoS case.
6. After ensuring that the volunteer was appropriately in position, the researcher launched the *main script* and stopwatch. For the entire period of the experiment, the MEAN app dashboard was executing.
7. The *main script* was tweaked to collect data from different locations and pathways in different "measurements" of the database through arguments passed through the script's shell.
8. The volunteer was instructed to stand at a new location and the process was repeated.

The results of this experiments will be presented in Section 2.1.3 in Chapter 2.

General Guide

Measurements in InfluxDB are conceptually analogous to tables in SQL. The use of different measurements in the design of this research allows neat separation of the readings from one location to another and from one pathway to another, making the analysis part simpler as separation and sorting of the data is no longer required.

1.3.2.4 Detecting Self Body Occlusion between the Broadcaster and the Observer

In real-world situations, there could be many ways a BLE-enabled device can be carried. The broadcasting device may be held in a hand, worn on a body like a smartwatch, kept in a pocket like key fobs, or can be kept in a bag. This presents a challenging situation where the capabilities and performance of BLE must be tested to evaluate its usefulness in highlighting patterns in resulting RSSI even in situations where they are occluded by the body of either the pedestrian carrying the device themselves or by the bodies of other pedestrians. Or simply, to test if, through the RSSI, there is a likelihood of asserting the presence or absence of occlusion. However, to limit the scope of the work, the only occlusion considered is self body occlusion, or in other words, whether the Broadcaster is carried by the volunteer on the side facing towards the deployed Observer (LoS) or, facing away from the deployed Observer (nLoS).

The experiment was performed in the same location as the previous experiment, presented in Figure 1.15, with five key points, *start*, *approach*, *centre*, *depart*, and *end*,

as depicted. Since the experiment site was not heavily used by pedestrians, it prevents the presence of any obstruction on the advertisement signal other than body occlusion. The Observer was facing clear ground to ensure low signal contention and so that the signal undergoes minimal reflection, refraction, and diffraction. The deployment distance of the Observer was 3 metres from the selected path, and the volunteer pedestrian was instructed to traverse the path in a straight line to ensure that the RSS values were not influenced by the pausing of the volunteer pedestrian. Three sub-experiments were set in this experiment as follows:

1. *Sub-experiment 1: RSS values from a stationary pedestrian at each key-point.* At each key point, *start to end*, on the path, a volunteer pedestrian was stationed with a Broadcaster for 30-second intervals. RSS values were collected over each 30-second observation period, first, when the Broadcaster was in LoS (with no body occlusion), with the Observer and second, when the Broadcaster was in nLoS (with body occlusion) with the Observer. This experiment was used to understand variations in RSS values of the received signals. The information gathered may also be used as a calibration mechanism, aiding the identification of the point of closest approach of a pedestrian to the Observer. If the advertisement rate of a Broadcaster is known, the number of dropped signals can be identified from the total measurement period, which has greater likelihood to be impacted by distance due to the increase in multipath propagation fading, and by occlusion.
2. *Sub-experiment 2: RSS values from a stationary pedestrian at each key-point from P to S.* Four random points *P*, *Q*, *R*, and *S* were selected at a distance greater than 9 metres away from the Observer on the open ground at the experiment location. This sub-experiment was designed only to develop an understanding of the behaviour of BLE signals emerging from a distance to the Observer. While this experimental setup did not directly concern the objective of this experimental case, which was to determine the impact of occlusion on close proximity detection through RSS values of the advertisements, the comprehension of the pattern of the RSS values of the advertisements emerging from a distant Broadcaster could be beneficial in scaling this study in the future to detect far away devices, mark them as anomalies, and filter them out.
3. *Sub-experiment 3: Continuous RSSI collection while the pedestrian traverses the path in both directions on the path 3 metres away from the Observer.* In this sub-experiment, RSS values of the advertisements from the Broadcaster carried by a pedestrian while walking on the predefined path, shown in Figure 1.21, were assessed. The volunteer pedestrian was instructed to walk on the pathway starting from each end one after the other while carrying the Broadcaster in both LoS and nLoS manner. Thus, the experiment under this sub-experiment was divided into four scenarios, where each scenario was repeated five times.
 - (a) Pedestrian walking from *start* to *end* with Broadcaster in LoS of the Observer (no occlusion).
 - (b) Pedestrian walking from *end* to *start* with Broadcaster in LoS of the Observer (no occlusion).
 - (c) Pedestrian walking from *start* to *end* with Broadcaster in nLoS of the Observer (occlusion).

- (d) Pedestrian walking from *end* to *start* with Broadcaster in nLoS of the Observer (occlusion).

Following was the inventory for this experiment:

1. Observer

- (a) RPi based device running the *main script* which stores the data in InfluxDB.
- (b) The device was also running the MEAN stack dashboard for live monitoring.
- (c) Battery pack and RTC connected to the RPi.

2. Broadcaster

- (a) RuuviTag beacon with advertisement rate of 2 Hz (advertising once every 500 ms).

Following were the protocols for the experiment conducted in the outdoor experiment with a stationary pedestrian:

1. Sub-experiment 1 and Sub-experiment 2

- (a) The Observer was positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
- (b) The Observer comprised of an RPi with power bank and RTC in an enclosure.
- (c) The pipe straps were stripped off the enclosure.
- (d) The Broadcaster was held by a volunteer pedestrian who was instructed to remain stationary at each of the five key points on the pathway and at each of the four random key points far away from the Observer.
- (e) While the volunteer pedestrian remained stationary at those points, the *main script* and the *MEAN* app was launched by the researcher, simultaneously noting the time on a stopwatch to ensure data was collected for during real-world deployment duration.

2. Sub-experiment 3

- (a) The Observer was set up in the same manner as in sub-experiments 1 and 2.
- (b) The researcher launched the *main script* and the *MEAN* app, and provided vocal instruction to the volunteer to begin walking. The *main script* was stopped as soon as the volunteer reached the other end of the pathway.
- (c) Upon completion of the first case, the LoS for the experiments, the nLoS case was performed in the same manner.

The experiment conducted in this case did not require the use of ground truth as the data collection process was started with the pedestrian's walk and was stopped upon the end of each walk, and no intermediate location during the walk was scrutinised in this study. The results obtained from this experiment will be presented in Section 2.1.4 in Chapter 2.

1.3.2.5 Effect of Advertisement Interval of the Advertiser on the Capturing Capabilities of the Observer

When such a platform is deployed in the real world, the users of the pathway could be carrying different types of advertising devices. One of the parameters of BLE *Advertisers* is *advertisement rate*, which is the rate at which advertisement packets are emanated by the devices to allow other BLE devices to discover them. Or conversely, advertisement interval, which is the time difference between subsequent advertisements. It was noted in Section ?? in Chapter ?? that advertisement interval affects the measurement quality (Montanari et al., 2017). Different manufacturers and even different devices from the same manufacturers use different advertisement intervals. Therefore, understanding this is crucial when it comes to acquiring signals from devices carried by moving pedestrians. A device with a longer advertisement interval will result in fewer intercepted signals for the entire traversal of the path, which may result in no discernible patterns. On the contrary, a shorter advertisement interval may overwhelm the Observer and/or may result in more packet collisions, and therefore, result in no discernible pattern or loss of entire journey detail.

In order to assess the effect of advertisement interval, this experiment tested three different advertisement intervals, 100 ms, 500 ms, and 1000 ms. The volunteer pedestrian is directed to walk on the pathways at a distance of 3 metres from the deployed Observer. This experiment explored the pedestrian walking from both ends of each pathway, from *start* to *end* and from *end* to *start*, for both LoS and nLoS cases. Each scenario was repeated three times and provided a total of 36 individual walks. For this experiment, an alternative Broadcaster, designed with an RPi as mentioned in subsection 1.2.2.1, was used since the beacon, a RuuviTag, did not allow modifying the advertisement interval. A *GPS Logger* mobile application, as mentioned in subsection 1.2.4.4, was used to collect the ground truth. The ground truth collection was performed in this experiment as the same measurements could then also be used to identify if there was a correlation between the packets dropped and the location of the transmission of the signals.

Advertisements were collected for the entire duration of the walk by the volunteer pedestrian. Three sets of measurements are performed for each configuration of the experimental setup. Since a custom Broadcaster, as presented in Section 1.2.2.1 in this chapter, was used in this experiment, additional information in the form of a *sequence number* was also transmitted with each advertisement. This sequence number was iterated beginning at a value of 1 with the first broadcast and incremented with each subsequent broadcast of advertisements. InfluxDB was also maintained on the Broadcaster to store timestamped sequence number on the device. This allowed accurate assessment of packet drops since both the number of advertisements broadcasted and the number advertisements intercepted were known. Therefore, the Observer was also modified to process and store the sequence number of the received advertisement in addition to the timestamped RSSI, and SMA RSSI (Section 1.1).

The following was the inventory for this experiment:

1. Observer
 - (a) RPi based device running the *main script* which stores the data in InfluxDB. The *main script* was further modified to also process and store

the sequence number of the received advertisement.

- (b) The device was also running the MEAN stack dashboard for live monitoring.
- (c) Battery pack and RTC connected to the RPi.

2. Broadcaster

- (a) RPi Broadcaster since it provided the capability to adjust the advertisement interval and maintain a database to store the timestamped sequence number.
- (b) The functionality of this Broadcaster was developed using *NodeJS*.
- (c) Battery pack and RTC connected to the RPi.

3. Ground Truth

- (a) A mobile phone with a *GPS Logger* application running on it.

The following were the protocols for the experiments:

General Guide

Bash, Bourne-Again SHell is a command language interpreter for unix-type OS. Bash provides a platform, as a macro processor, that is both a command interpreter and a programming language.

1. The Observer was positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
2. The Observer was the entire unit that includes a battery pack and RTC, all enclosed in the enclosure.
3. The volunteer was instructed to walk from one end of the pathway to the other.
4. The volunteer was instructed to hold the RPi Broadcaster in their hand facing towards the Observer for the case of LoS. The volunteer was provided with clear instructions to keep the face of the Broadcaster uncovered, that is to not enclose the Broadcaster in their fist.
5. Similarly, the volunteer was asked to switch the hand in which the Broadcaster was held for the nLoS case.
6. The researcher, after ensuring that the volunteer was appropriately in position, launched the *main script* on the Observer and the *NodeJS* script on the Observer and provided an oral cue to the volunteer to start walking.
7. The researcher kept a check on the status of the script and the volunteer's progress along the path.
8. The volunteer was instructed to stop upon reaching the other end of the pathway, and the researcher stopped both the scripts, *main script* on the Observer and the *NodeJS* script on the Broadcaster.
9. Both the *main script* and the *NodeJS* scripts were tweaked to store data from different locations and pathways in different "measurements" of the database through arguments passed through the *bash* to call the script. The advertisement interval for the Broadcaster was also changed directly from the shell when calling the *NodeJS* script to launch it.

10. To ease conducting the experiments, the bash scripts were written in advance for each scenario of the experiment. The researcher only executed the appropriate bash script before performing a corresponding scenario for the experiment. The bash scripts contained all the arguments regarding advertisement interval, database, and measurement name in the database where the measurements of the RSSI were to be stored.

The results of these experiments will be presented in Section 2.1.1 in Chapter 2.

1.3.2.6 Section Summary

The experiments presented in this section were aimed at assessing the suitability of BLE to measures pedestrian dynamics. The experiment described in Section 1.3.2.1 was aimed at understanding the capability of the antenna to measure RSSI in an outdoor environment with sufficient resolution to produce discernible patterns pertaining to pedestrian movement, which were assessed in reference to the measurement performed in the anechoic chamber. This experiment was also used for identifying any quirks of the antenna that could aid or be used to detect pedestrian movement dynamics or activities.

The next experiment, presented in Section 1.3.2.2, was used to understand the extent of environmental contribution to the propagation mechanism, such as the level of LoS components at various location on the pathway and the extent of the effect of shadowing. This experiment was to understand the suitability of the selected environment for conducting the subsequent experiments.

In the experiment presented in Section 1.3.2.3, the optimal deployment distance of the Observer was studied. This was performed by testing candidate deployment distance, that is the perpendicular distance between the Observer and the linear pathways. Through this experiment, a distance that resulted in discernible patterns on the RSSI was aimed to be identified.

During pedestrian walks, a direct LoS between the Broadcaster carried by a pedestrian and the deployed Observer is not guaranteed. Therefore, in the experiment presented in Section 1.3.2.4, effect of body occlusion on BLE signal strength was examined. To limit the scope of study, only self body occlusion was scrutinised. Through this experiment, the focus was to assess whether self body occlusion can be detected and whether the resulting RSSI from the self-occluded Broadcasters produced recognisable patterns to gain insights into pedestrian activities and movement dynamics.

The final experiment conducted to assess the overall suitability of BLE before applying it to understand selected heterogeneous pedestrian activities and movement dynamics was to understand the effect of advertisement intervals on the measuring capabilities of the Observer, as presented in Section 1.3.2.5. Since different Broadcasters advertise at different rates, it was important to assess the range of advertisement rate or advertisement intervals that selected Observer was able to measure accurately.

Through the results of these experiments, that will be presented in Section 2.1 in Chapter 2, feasibility of BLE as a measuring technology, usefulness of selected

experimental location, identification of suitable deployment location and extent of advertisement rate measurement capability of the Observer, and ability of the developed system to detect and produce recognisable patterns even in the presence of self body occlusion were evaluated. Based on the outcome of the evaluation, the suitability of BLE was determined.

1.3.3 Evaluation of the Platform's Abilities to Detect, Identify and Characterise Pedestrian Movement and Activities

1.3.3.1 Pause Detection in the Movement of the Pedestrians

One of the movement characteristics of pedestrians on pathways is pauses in movement, or as presented in Section ?? in Chapter ??, *staying* at a location. The pauses can be attributed to several reasons – it could be a place of interest, some event may be occurring, perhaps the pedestrian met some acquaintance or simply paused to pet a dog, or the pedestrian could be tying their shoelaces, etc. Pauses are a common occurrence in a walk, and their repetitions may provide development agencies with significant insights. For instance, a park may have a spot that provides a scenic view of a sunset that the urban designers did not consider in their planning. If many people carrying BLE devices are pausing or stopping at that location for extended periods, the aggregated data collected would highlight that aspect, and could signify a point of interest such as the presence of an attractive or scenic location. This could then feed into redevelopment plans where benches could be set up to aid elderly and vulnerable visitors of the park to absorb the scenery. Thus, such a tool could potentially highlight the interaction of pedestrians with the local environment and/or other pedestrians without compromising the privacy of their sensitive data.

To identify pauses in the movement of the pedestrian, RSS values of advertisements acquired by the Observer were used. The experiment was conducted in the same open space as the previous experiments. Like previous experiments, since it was conducted in a location with open ground on one side, in the absence of a defined pathway, it was feasible to define a path at any chosen distance from the wall where the Observer was deployed. Therefore, a 24-metre long path is defined as 3 metres away from the Observer. Markers were laid at the starting point of the pathway, called *Start*, the end point of the pathway, *End*, and at a point 6 metres away from *Start*, called *Approach*. These markers were used to aid the volunteer pedestrian's linear movement and also worked to highlight the exact point for pausing. Figure 1.22 illustrates the location of the experiments.

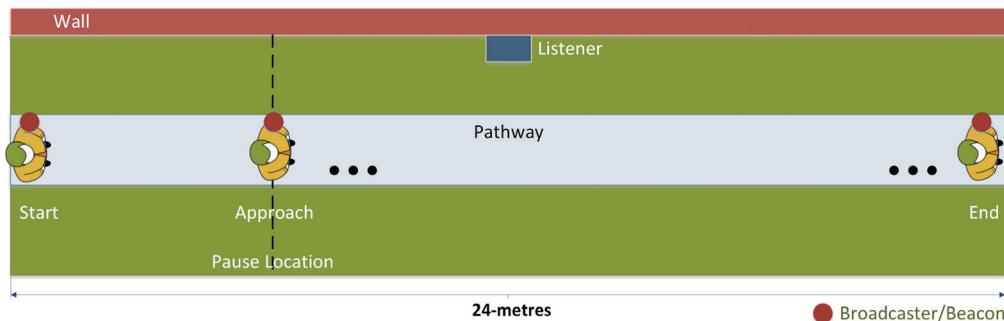


FIGURE 1.22: Experimental Location and Setup

A volunteer pedestrian was instructed to walk on the linear path from *Start* to *End*. The pedestrian set off from *Start* in the direction of *End* with a pause at *Approach*, 6 metres away from *Start*. The duration of this pause was varied between 5 seconds, 15 seconds, and 25 seconds. For each pause duration, the process was repeated four times. Hence, a total of 12 individual repetitions were considered in this experiment. The Broadcaster was held by the pedestrian facing in the LoS of the Observer throughout this experiment. The *Blue Dot* mobile application was used in this experiment to obtain the ground truth.

The following was the inventory for this experiment:

1. Observer
 - (a) RPi based device running *main script* which stores the data in InfluxDB.
 - (b) The device was also running MEAN stack dashboard for live monitoring.
 - (c) Battery pack and RTC connected to the RPi.
2. Broadcaster
 - (a) One RuuviTag beacon, advertising once every 500 ms.
3. Ground Truth
 - (a) Android phone running the BlueDot application.

The protocols for the experiment conducted are as follows:

1. The Observer was positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
2. The Observer comprised of a RPi with power bank and RTC in an enclosure.
3. The Volunteer pedestrian was provided with a Broadcaster and an Android phone with the BlueDot application to capture ground truth data.
4. The volunteer pedestrian was instructed to begin walking on vocal cue from the researcher and pause at the *approach* key point for the *pause duration* before continuing the walk.
5. The set of *pause duration* includes 5 seconds, 15 seconds, and 25 seconds.
6. The researcher initiated the *main script* and MEAN app just before providing a vocal cue to the volunteer.
7. The volunteer pedestrian was instructed to stop upon reaching the *approach* key point, press the button on the BlueDot application on the phone, and to resume their walk upon receiving vocal instruction from the researcher.
8. When the pedestrian stopped at the *approach* key point on the pathway, the researcher started a stopwatch for the duration based on the pause duration under test. After the pause at *approach* for the duration being tested, the researcher provided another vocal cue to the pedestrian to continue walking.
9. The volunteer was instructed to stop at the end of the pathway and the scripts were stopped by the researcher.
10. Once four bouts were measured, the volunteers were asked to pause for the duration based on the next candidate from the set.

The results of this experiments will be presented in Section 2.2.1 in Chapter 2.

1.3.3.2 Identifying Interactions with Other Pedestrians or the Environment

This experiment is an extension of the previous experiment. Interacting with other pedestrians is one of the common activities people engage in during their walks. Here, instead of employing a single volunteer pedestrian, two volunteer pedestrians assisted in the experiments. The purpose of the experiment is to understand if the approach used in the previous experiment, presented in Section 1.3.3.1, is useful in the identification of pauses in measurements of RSSI that are measured simultaneously from two Broadcasters.

The Observer was deployed at a distance of 3 *metres* from the pathway. Two Broadcasters were provided to two volunteer pedestrians and the advertisements were measured from both the Broadcasters simultaneously for the duration of the walk. Since both pedestrians were instructed to walk simultaneously from the same starting point, only one device to gather ground truth was used, a phone running the GPS logger application. The measured parameters on the Observer were timestamps, raw RSSI, sliding-window mean of RSSI, and UUID. A pause period of 20 *seconds* was selected for this experiment.

The experiment was performed in the same outdoor space as the previous experiments. Several cases were studied in this experiment and these cases are listed below,

1. Both pedestrians holding the Broadcaster in LoS of the Observer.
 - Both pedestrians start from the same end of the pathway, *start* point.
 - The volunteers pause at the *approach* point and continue walking.
 - The volunteers pause at the *centre* point and continue walking.
 - The volunteers pause at the *depart* point and continue walking.
2. Both pedestrians holding the Broadcaster in nLoS of the Observer.
 - Both pedestrians start from the same end of the pathway, *start* point.
 - The volunteers pause at the *approach* point and continue walking.
 - The volunteers pause at the *centre* point and continue walking.
 - The volunteers pause at the *depart* point and continue walking.
3. One pedestrians holding the Broadcaster in nLoS of the Observer and the other in LoS.
 - Both pedestrians start from the same end of the pathway, *start* point.
 - The volunteers pause at the *approach* point and continue walking.
 - The volunteers pause at the *centre* point and continue walking.
 - The volunteers pause at the *depart* point and continue walking.

Each case was repeated three times for both paths, at a deployment distance of 3 *metres* away from the Observer. Therefore, the measurements were taken for a total of 27 individual walks.

The following was the inventory for this experiment:

1. Observer
 - (a) RPi based device running *main script* which stores the data in InfluxDB.

- (b) The device was also running MEAN stack dashboard for live monitoring.
 - (c) Battery pack and RTC connected to the RPi.
2. Broadcaster
 - (a) Two RuuviTag beacons with advertising once every 1280 ms.
 3. GroundTruth
 - (a) Android phone running GPSLogger application.

The following were the protocols for the experiment:

1. The Observer was positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
2. The Observer comprised of RPi with power bank and RTC in an enclosure.
3. The pipe straps were removed from the enclosure.
4. Volunteer pedestrians were provided with a Broadcaster each.
5. One of the volunteer pedestrians was provided with an Android phone with a GPSLogger activated to capture ground truth data.
6. The volunteer pedestrians were instructed to walk together alongside each other upon a vocal cue from the researcher and pause at one of the three key points, *approach*, *centre*, or *end*, depending on the location under scrutiny.
7. The researcher executed the *main script* and *MEAN* app just before providing a vocal cue to the volunteers.
8. When the pedestrians stopped at the point of focus on the pathway, the researcher started the stopwatch. Upon pausing for 20 seconds at the location of interest, the researcher provided another vocal instruction to the pedestrians to resume walking.
9. The volunteers were instructed to stop at the end of the pathway and the scripts were stopped by the researcher.
10. Once three bouts were measured for one specific location, the volunteers were asked to pause at the next locations for the next three repetitions and so forth.
11. The same protocol was used to measure both LoS and nLoS cases.

The results of these experiments will be presented in Section 2.2.2 in Chapter 2.

1.3.3.3 Asserting the Direction of Travel

The direction of travel is yet another important movement characteristic. Knowledge of the direction of travel can potentially reveal the entry and exit choices of the pedestrian. Such information is useful not only for assessing preferred routes but also in characterising the nature of walks with the help of some contextual information. For instance, if it is known through, for example, the help of maps that one of the exits of a pathway leads to a commercial area, the direction and the time of a walk can help to characterise the walk to be likely as a leisurely walk for shopping

purposes.

The experiment was conducted in the same open ground where all the other experiments were conducted. This is presented in Figures 1.15 and 1.13. Two paths were defined for this experiment, first, 3 *metres* away from the deployed Observer and second, 5 *metres* away from the deployed Observer. Five key points, *start* to *end*, as discussed previously in the Section 1.3.1 in this chapter were marked on both the pathways. Findings from 1.3.2.1 were employed in this experiment. The topology of the experiment was depicted in Figure 1.23.

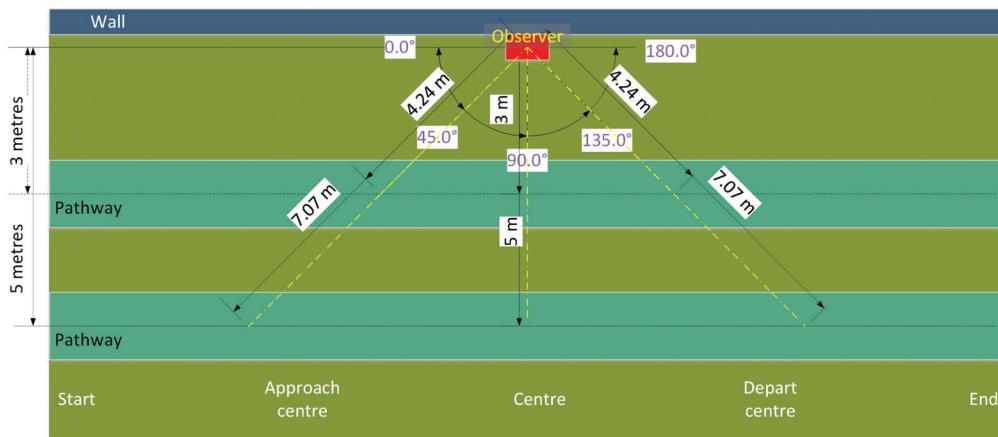


FIGURE 1.23: Topology for the Experiment

A volunteer pedestrian was instructed to walk along the linear path with a Broadcaster and a phone with the preinstalled *Blue Dot* application to establish ground truth locations as they pass over the selected marked points on the path. The volunteer pedestrian was instructed to press the button, as presented in Figure 1.9 in Section 1.2.4.3 in this chapter, once on the Blue Dot application while crossing each of those key points. The pedestrian traversed the path in both directions on the pathway, that is from *Start* to *End* and from *End* to *Start*. This sequence of actions was performed by the pedestrian at both pathways, 3 *metres* and 5 *metres* away from the deployed Observer. This experiment was repeated for two cases, first, when the Broadcaster is in LoS of the Observer, and second, when the Broadcaster is in nLoS of the Observer. A total of 24 individual runs were captured for this experiment. This approach differed from earlier work in the literature to detect the direction of travel of pedestrians, studied by AlAnbouri in (AlAnbouri et al., 2019). In AlAnbouri's work, two Observers were used and were strategically deployed to impede advertisements emerging from Broadcasters behind them. In contrast, the work carried out in this experiment utilised a single Observer, leveraging its antenna characteristics instead of strategic placement.

The structure of the experiment is described below:

- 3 *metres* deployment distance
 - LoS case
 - * *Start* point to *end* point scenario
 - Repetition 1

- Repetition 2
- Repetition 3
- * *End* point to *start* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3
- nLoS case
 - * *Start* point to *end* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3
 - * *End* point to *start* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3
- 5 *metres* deployment distance
 - LoS case
 - * *Start* point to *end* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3
 - * *End* point to *start* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3
 - nLoS case
 - * *Start* point to *end* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3
 - * *End* point to *start* point scenario
 - Repetition 1
 - Repetition 2
 - Repetition 3

The following was the inventory for this experiment:

1. Observer
 - (a) RPi based device running *main script* which stores the data in InfluxDB.
 - (b) The device was also running MEAN stack dashboard for live monitoring.
 - (c) Battery pack and RTC connected to the RPi.
2. Broadcaster
 - (a) One RuuviTag beacon, advertising once every 500 *ms*.

3. Ground Truth

- (a) Android phone running the BlueDot application.

The following were the protocols for the experiment:

1. The Observer was positioned on the side wall of the building shown in the layout in Figures 1.15 and 1.13.
2. The Observer comprised of RPi with power bank and RTC in an enclosure.
3. The pipe straps were removed from the enclosure.
4. Two pathways were marked 3 *meters* and 5 *meters* away from the deployed Observer.
5. The volunteer pedestrian was provided with a Broadcaster and an Android phone with the BlueDot application preinstalled to capture ground truth data.
6. The volunteer pedestrian was instructed to walk after receiving vocal cues from the researcher from one of the ends of the pathway to the other end of the pathway.
7. The volunteer pedestrian was also instructed to press the button on the Blue Dot application while they were crossing over each of those key points.
8. The researcher initiated the *main script* and *MEAN* app just before providing a vocal cue to the volunteer.
9. After three repetitions, the next case and point of start of the walk were selected as per the aforementioned structure of the experiment.

The results of these experiments will be presented in Section 2.2.3 in Chapter 2.

1.3.3.4 Analysing the Behaviour of Pedestrians in a University Campus for an Extended Period

The data collection for this experiment was conducted by another researcher in our research group. I contributed to the data analysis and writing a paper for the work. BLE beacons were provided to 28 participants and the RSS values were collected at 17 different locations in the Grangegorman campus at TU Dublin (TUD). Of the 28 volunteers who activated their Broadcaster devices, 21 of them were local residents, 3 were TUD postgraduate researchers, one was a TUD undergraduate student and three were TUD staff members. There were 15 males and 13 females among the volunteers. One volunteer was in the 18-24 age group, 8 were in the 25-34 age group, 14 were in the 35-44 age group, 4 were from 45-54 age group and one was in the age group 55-64. Over the study period, 126,130 BLE advertisements were collected by the 17 Observers from the Broadcasters carried by the participating volunteers. The MAC addresses of the devices were hash encoded and the RSSI data was accompanied by a record of their intentions.

Two types of Observers were employed in this part of the study, a Raspberry Pi 3-based indoor Observer, and an ESP32-based outdoor Observer. This second Observer was selected due to the high power requirements of the Raspberry Pi 3-based Observer which could not be used for more than a few hours on portable power

banks. The outdoor Observer employed ESP32 with a real-time clock and SD card breakout attached, housed in weatherproof enclosures, Figures 1.2 and 1.3, which could be magnetically attached 3-4 *metres* high on a metallic pole using a purpose-built hook and form mechanism, Figure 1.24. The ESP32s were each powered by an off-the-shelf 10000 *mA H* power bank as shown in Figure 1.5. In normal operation, preliminary experiments showed that a power bank was able to continuously power an ESP32-based Observer for over 48 *hours*. For this reason, depleted power banks were swapped with fully charged power banks every two days by the researchers. At the same time, the data recorded on their SD cards was also downloaded to a database on a laptop for analysis.

Since the indoor Observer was based on the Raspberry Pi, which featured a complete operating system, the data was stored locally on a MySQL database, whereas, on the ESP32-based outdoor Observer, data was written to an SD card in a CSV file format. The full-fledged operating system on the Raspberry Pi allowed it to be easily configured via remote connection through single-glazed windows using a phone hotspot, which meant that its database can be checked from outside the window without needing to enter the room.

The deployment of the Observer in the campus is depicted in Figure 1.25 where the yellow coloured rectangles signify the sub-experiments is in the control of TU Dublin Estates, green rectangles are controlled by either local residents or businesses, and red rectangles are controlled by Dublin City Council.



FIGURE 1.24: Attaching the Enclosure to a Pole.

One key intended output of the research is to show that by grouping the RSSI record for a particular shared hash encoded alias across the data set, it is possible to generate an aggregated record of the activity of the pedestrians. The results obtained from these experiments are discussed in Section 2.2.4 in Chapter 2.

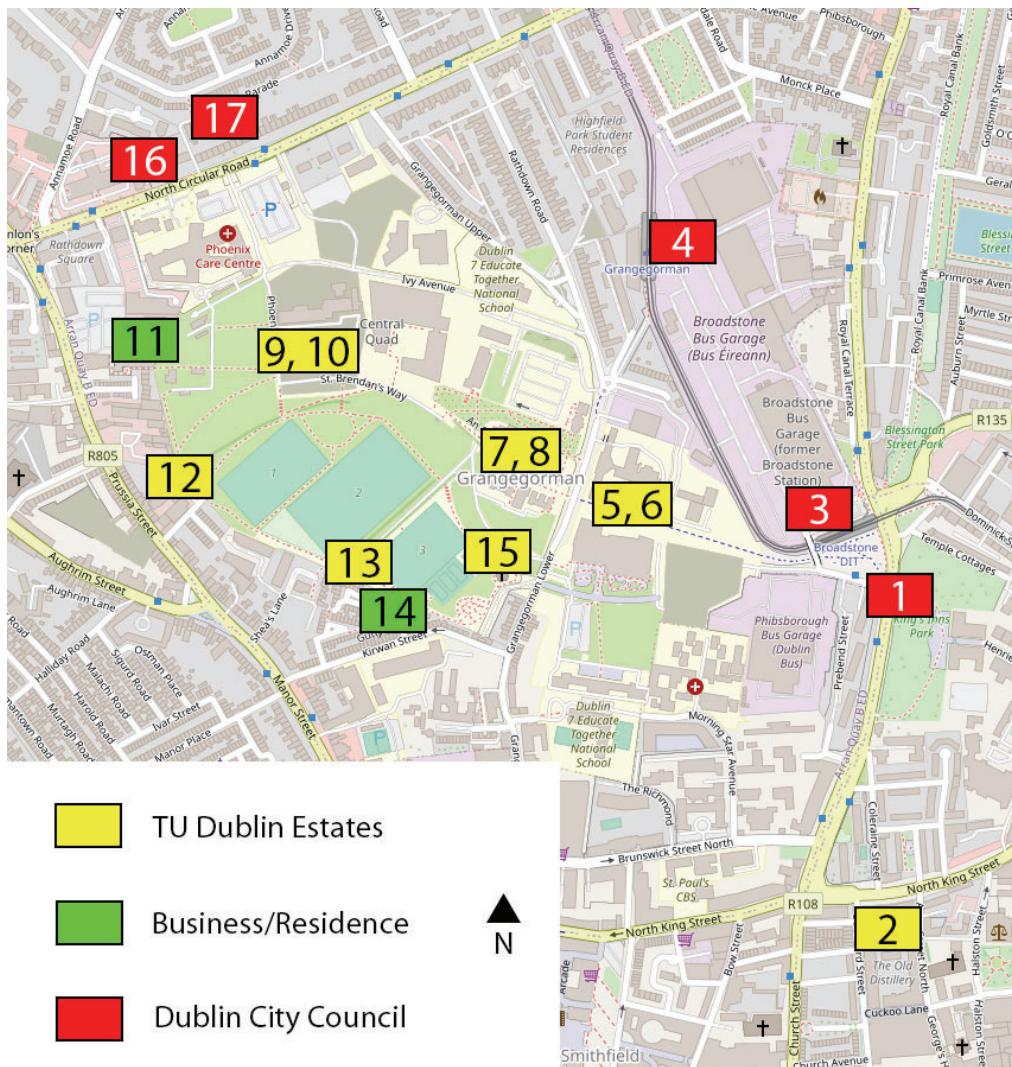


FIGURE 1.25: Deployment of Devices for the Study

1.3.3.5 Section Summary

The experiments presented in this section were aimed at testing a range of pedestrian dynamics. One of the important aspect of walking is *staying* or pausing, as identified in the literature, described in Section ?? in Chapter ?. Thus, the experiment presented in Section 1.3.3.1 was focused on identification of pause, a movement dynamic, in the movement of a pedestrian. There could be many reasons for pausing during a leisurely walk or a commute, one of them being to engage with other pedestrians. The next experiment, as presented in Section 1.3.3.2, was therefore aimed at assessing if BLE could be used to identify pauses in two Broadcasters simultaneously and subsequently, inferred as an interaction between the two pedestrians. This can be attributed to both, movement dynamic and an activity.

The next experiment was focused on the identification of a pedestrian's travel direction using a single Observer. The knowledge of a pedestrian's travel direction can provide valuable insight, especially if combined with contextual information such as time of walk and the location of experiment. For instance, if the location of the experiment is known to be a commercial region and a walk is carried out during office hours, leading into the commercial zone, there is a strong likelihood that the walk is a purposeful commute to the office could be asserted. The outcome of this experiment was therefore, aimed to assess both movement dynamic and activity.

Finally, an extended study of monitoring the behaviour of pedestrians in a campus was conducted by a research colleague, in which, my contribution towards data analytics was important in assessing the capabilities of BLE for understanding macro-level pedestrian movement dynamics, pedestrian activities, and pedestrian behaviour.

1.4 Methodological Elements of Data Analysis

In this section, the general approach taken for data analysis is discussed. The approach includes an overview of the selected statistical tools and techniques applied to the measurements obtained from experiments mentioned in previous sections and the rationale behind their selection.

1.4.1 Median

Since all of the experiments presented in this dissertation collect a series of RSS measurements, there is a need for a statistical tool to evaluate a single empirical representation or description is required for the set of measurements. A *mean* or *average* value is one of the methods for calculating a single value that can represent the trend of measurements, however, this value is susceptible to anomalous measurements. For instance, if five measurements belong in a set with an anomalous measurement, viz. -10, -12, -15, -13, and -82, the average or mean value will be -26.4. This average reading is pulled away from the general trend of the measurement set only because of one anomalous measurement. The median value, however, for the same set is -13. Therefore, *median* shows greater resilience in comparison to *average* towards anomalous measurements. Since, RSSI is susceptible to factors such as the topology of the surrounding environment, weather, etc, as presented in Section ??, there is an increased likelihood of capturing an anomalous measurement in the data collection

process. In addition to its resilience towards anomalous measurements, the function to calculate median value is a standard part of several programming languages and libraries, and its calculation is not computationally expensive. Therefore, it is selected for calculating empirical description or representation of a collected set of measurements.

In the experiments, multiple rounds of the same scenarios are repeated to ensure that the measurements are representative. This repetition helps mitigate the influence of any external factor that could cause the measurements to be unrepresentative if only collected once. Medians of each of those rounds belonging to a single scenario are taken, but it means that there is a representing value for each of those rounds and not for the overall scenario. Therefore, averaging of those medians is then performed to evaluate a representative value of the scenario. Since the calculation of the median for each repetition has already reduced the effect of the presence of any anomalous readings, the average of medians will also remain uninfluenced by any anomalous value.

1.4.2 Simple Moving Average (SMA)

Evaluation of both BLE technology and its capabilities are obtained from assessment of RSSI measurements. These measurements are susceptible to several factors in the outdoor environment, such as weather conditions and the presence of other infrastructure elements, as discussed in Section ?? in Chapter ?. These factors affect the obtained RSSI measurements, which have the potential to cause fluctuations in those measurements. One of the approaches to counter these effects is to smooth the obtained signals, and SMA is one of the statistical techniques that are employed for smoothing the data (Koledoye et al., 2018). As mentioned in Section ?? in Chapter ?, Koledoye et al. in (Koledoye et al., 2018) compared a variety of statistical techniques from smoothing the measurements taken from BLE devices and identified SMA as a reliable technique. SMA is calculated by creating subsets of the measurements of length determined by the window size and averaging each subset. The calculated value then replaces the latest acquired measurement. The equation used to calculate SMA is presented in Equation 1.1.

$$sma_{rss}i(i) = \begin{cases} \frac{1}{i} \sum_{j=0}^i rssi_j, & 1 \leq i < k, \\ \frac{1}{k} \sum_{j=i-k+1}^i rssi_j & i \geq k, \end{cases} \quad (1.1)$$

where:

- k is window size,
- j is iterator or the current value, and
- $rssi$ is the set of RSSI measurements.

This SMA is a Finite Impulse Response (FIR) (Shenoi, 2006) filter, meaning that it does not use a feedback loop with the output signal. The FIR Filters output is only based on the current and past input values which means that the filter's output only responds to a certain fixed number of samples. In the Equation 1.1, it means that the filter response will stop after k samples. This is opposed to the Infinite Impulse Response (IIR) filters in which the output of the filter depends on both current and past inputs as well as past outputs, causing the filter to continue responding to an input signal indefinitely. Since the FIRs have no feedback loop, they are more stable

and do not require the same care in handling as IIR Filters.

SMA demonstrates characteristics of a low-pass Filter, which means it allows the low frequencies to pass through without attenuation while attenuating the high-frequency components of any signal. To visually illustrates this, the frequency response of SMA with window sizes of 6, 8, 10, 12, and 14 are presented in Figure 1.26. In the figure, the main lobe at around 0 frequency on the x-axis has a magnitude approaching 1 on the y-axis, indicating that the low-frequency components can pass through. The Cutoff Frequency, the boundary between passband and stopband, is determined by the expression presented in Equation 1.3. For instance, the cutoff frequency, f_c , for window size 10 is approximately 0.1 cycles/sample. This signifies that frequencies below 0.1 cycles per sample are allowed to pass through the SMA filter with minimal attenuation, while other frequencies are attenuated. Whereas, the ripple-like side lobes indicate partial attenuation of high frequencies. The time-domain equivalent, to evaluate the attenuation in dB can be elucidated using an example. With a sampling frequency, f_s , of 2 Hz and a window size (k) of 10, the cut-off frequency, f_c , is approximately 0.1 cycles/sample. The time domain equivalent cutoff frequency can be computed by using the expression in Equation 1.4, which gives us the value 0.2 Hz. Magnitude response ($|H_f|$), illustrated on the y-axis of Figure 1.26, is then required to evaluate the attenuation in the time domain for any particular frequency. Assuming the attenuation of 0.4 normalised frequency is to be evaluated, the magnitude response for the selected frequency is approximately 0.027. The attenuation can then be calculated using the expression in Equation 1.5, which computes to -31.36 dB.

Normalised Frequency

Normalised frequency is used to express frequency relative to the sampling rate of the signal. The expression to calculate normalised frequency is shown in Equation 1.2. It is a unitless representation and in the context of this dissertation, allows consistent comparison across varying window size.

$$f_n \approx \frac{f}{f_s} \quad (1.2)$$

where:

- f is actual frequency of the sample, and
- f_s is sampling frequency of the signal

$$f_c \approx \frac{1}{k} \quad (1.3)$$

where:

- f_c is cutoff frequency, and
- k is window size

$$f_c(\text{Hz}) = f_c \times f_s \quad (1.4)$$

where:

- f_c is cutoff frequency, and
- f_s is sampling frequency of the signal

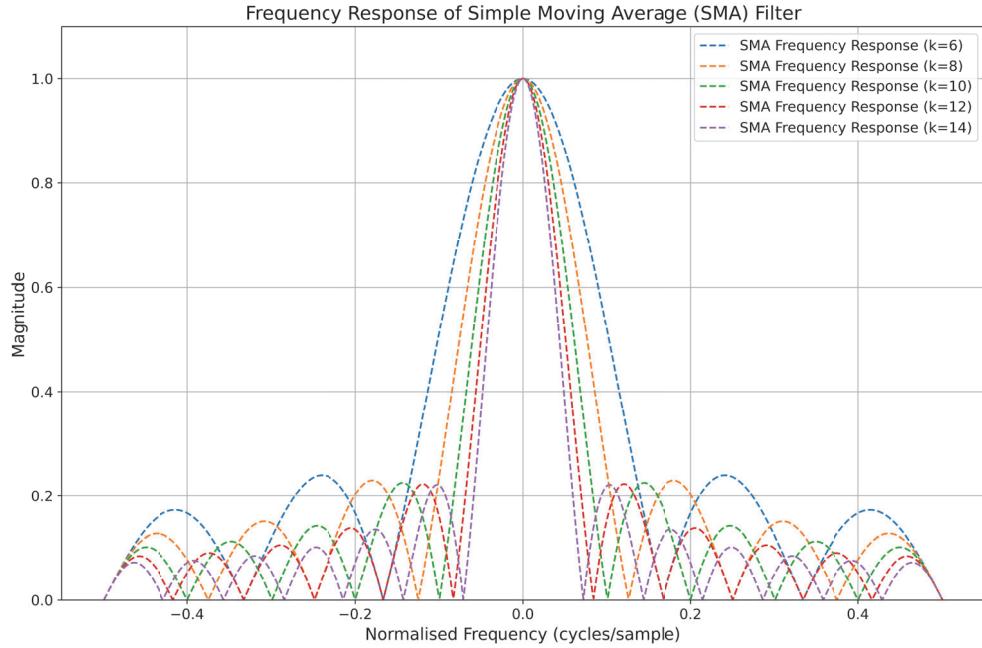


FIGURE 1.26: Frequency Response of SMA Filter for Varying Window Sizes

$$\text{Attenuation}(dB) = 20 \log_{10}(|H_f|) \quad (1.5)$$

where:

- f_c is cutoff frequency, and
- f_s is sampling frequency of the signal

The farther away from the main lobe, the greater the attenuation. Note that in Figure 1.26, the frequency response is presented over both positive and negative frequencies (from -0.5 to $+0.5$ normalised frequency units), therefore, the greater attenuation is at higher frequencies only. At certain points the frequency response dips down to zero magnitude, corresponding to the complete attenuation of those frequency components. This can be understood using the expression in Equation 1.5, where zero magnitude in place of $|H_f|$ will result in $-\infty dB$ in attenuation, indicating complete attenuation of the signal. As presented in the figure, the width of the main lobe is inversely proportional to the window size. This implies that the efficacy of a low-pass (FIR) filter increases with the increase in window size, that is more attenuation of higher frequencies albeit at higher computational cost.

The frequency response of raw RSS measurements and that of SMA (with varying window sizes) filtered sample raw RSS measurements are presented in Figure 1.27. It can be identified that high frequencies of the sample RSSI are attenuated when they are subjected to SMA. It can be seen that a larger window size results in greater attenuation of higher frequencies. While this is favourable as it provides

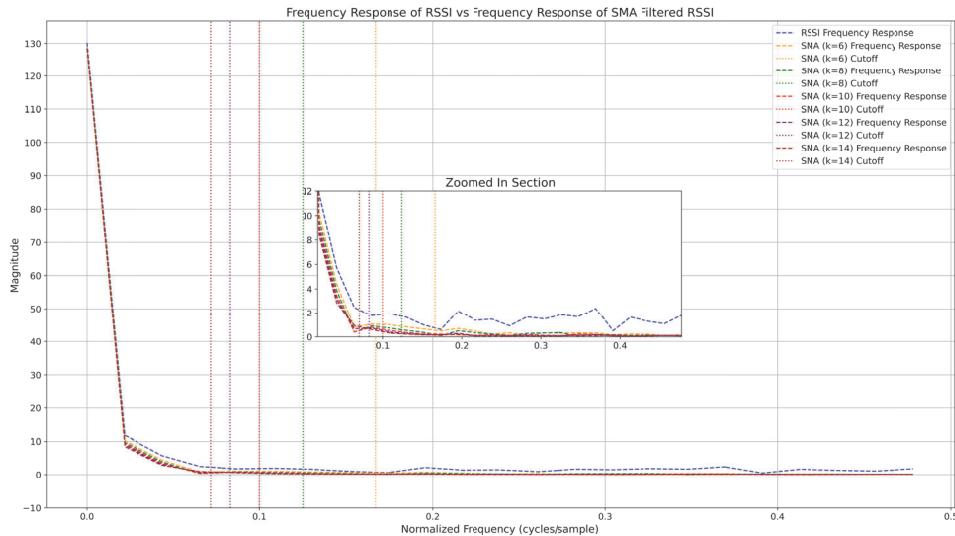


FIGURE 1.27: Frequency Response of SMA Filter and Cut-off Frequency on Sample RSSI for Varying Window Sizes

more smoothing, the larger window size also has the potential to reduce details in the actual RSSI measurement. This is depicted in Figure 1.28, where raw RSSI and SMA-filtered RSS evaluated with varying window-sizes are plotted against sample index. It can be observed in the figure that a larger window size for SMA resulted in a smoother representation of raw RSS. A window size of 10 is selected for all the experiments presented in this dissertation to balance the smoothness and details of the measurements. The window size of 10 samples was deliberately chosen to ensure the noise was significantly curtailed with the awareness that it would also reduce temporal sensitivity. This will be discussed further in Section ?? in Chapter ?. For the first nine measurements in any given experiment, the averaging is performed over the available measurements, for instance, on the fifth measurement, only the first five measurements are averaged. Note that in real-world deployment this will work differently, as a pedestrian approaching the range of measurement of the Observer will result in RSSI in the ranges of -100 dB . Once the number of collected measurements reaches beyond ten, the most recent measurement and the last nine measurements are used to calculate SMA.

SMA of the RSSI is calculated by the *main python script* at the time of acquisition of the RSSI measurement and is stored in the database along with the raw RSSI, timestamp, and UUID.

1.4.3 Standard Deviation (SD) and Variance

Both of these statistical tools, SD and Variance are employed to measure dispersion. That is the spread in the values collected in a measurement set. SD is a key descriptive value for a measurement set and since it uses the same units as the units of a sample, it is comparatively intuitive. Whereas, Variance is typically used in intermediate steps of complex calculations such as ANOVA. Variance is less intuitive in comparison as it uses squared units. The equations for SD and Variance are given in Equation 1.6 and 1.7 respectively.

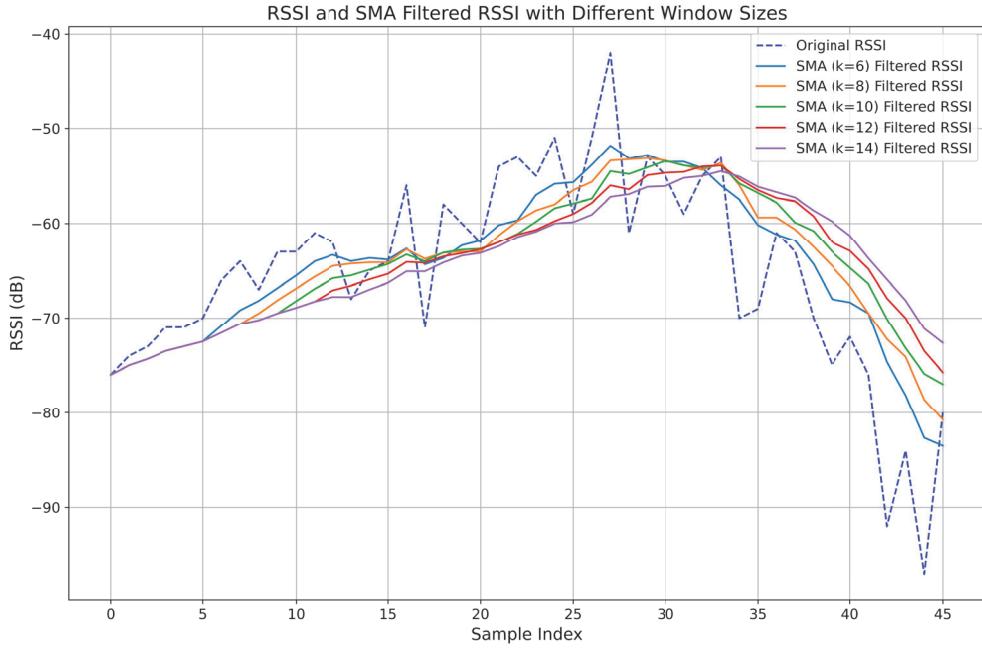


FIGURE 1.28: Raw RSS Measurements and Subjected to SMA (with Varying Window Size) Filtered RSS

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (1.6)$$

where:

- σ^2 is the population variance,
- N is the number of data points in the population,
- x_i represents each data point, and
- μ is the mean of the population

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.7)$$

where:

- s^2 is the sample variance,
- n is the number of data points in the sample,
- x_i represents each data point, and
- \bar{x} is the mean of the sample

1.4.4 Sliding Window Standard Deviation (SD)

Sliding window SD is the application of SD on a subset of measurements within a set of measurements, where the window that encompasses the subset slides over the entire measurement set. It is similar to the SMA technique discussed in subsection 1.4.2, only that SD is computed instead of computing average. This value is computed using the formula depicted in Equation 1.8.

$$\sigma_k = \sqrt{\frac{1}{w} \sum_{i=k}^{k+w-1} \left(x_i - \frac{1}{w} \sum_{j=k}^{k+w-1} x_j \right)^2} \quad (1.8)$$

where:

- σ_k is the standard deviation of the k -th window,
- k is the starting index of the current window,
- w is the window size,
- x_i represents each data point within the window, and
- $\frac{1}{w} \sum_{j=k}^{k+w-1} x_j$ is the mean of the k -th window

The rationale behind the selection of a sliding window SD is to perform a localised analysis of dispersion. Such an analysis is important in the assessment of dispersion or fluctuation in sections of an entire measurement set. This technique is also useful in detecting sudden changes or anomalies in the values contained within an entire measurement set since those sudden changes would introduce a spike in the standard deviation.

In the context of this dissertation, the measurements are timestamped RSS values. Therefore, the sliding window could be applied to such a measurement in two ways. First, *time-based window*, and second, *sample sequence-based window*. The time-based window is easier to understand, where, if a five-second window is chosen, all the measurements within the first five seconds are used, which is then shifted by one second and so forth. In the sample sequence-based window, if a five-sample size window is selected, the first five measurements are utilised in the computation regardless of how many seconds duration that corresponds to, and then the window is shifted by one measurement and so forth. While both have their use cases, in the context of the research described in this dissertation, a time-based sliding window is more applicable. To understand this, let's consider an example. Say, a pedestrian carrying a BLE device is being observed and the observed RSSI are irregularly distributed in time – due to reasons such as artefact in the environment impeding the signal travel – a sample sequence-based window may result in fewer subsets of measurements. This has the potential to hamper the use of data to infer any pedestrian activity or movement dynamics at a fine granularity. However, a time-based window will operate with however many samples available in a time-defined subset. While this may have the negative consequence of resulting in variable-sized subsets, for this study achieving finer granularity is more important. This is because a continuous walk may only last for a few seconds to a minute from a BLE Observer's point of view. The difference between the two windowing techniques is visually presented in the Section 2.2.1 in Chapter 2. It is noteworthy that the timestamps are obtained at the top of the BLE stack, which mean that the timestamps are not representative of the arrival of the signal at the hardware level, and may be affected by

any delay in the stack caused by the operating system of the Observer.

1.4.5 Median Absolute Deviation (MAD)

It has already been discussed in Section ?? in Chapter ?? that the RSSI of BLE, due to its susceptibility to external factors, fluctuates. Even if several measurements of RSSI are taken with the Observer and Broadcasters stationary, the value of collected RSSI will fluctuate within a range of values. This will be clarified in section 2.1.1, where the RSSI measurements collected for stationary devices in an anechoic chamber, devoid of any external electromagnetic interference, also showcase, however small, RSSI values bound within a range. SD and Variance are good descriptive measures of this dispersion, however, another tool is required that specifically measures the deviation of extreme values with respect to a central value and is robust against these outliers. Since SD and Variance use average or mean to quantify the spread, these techniques are susceptible to the effects of the outliers.

Median Absolute Deviation (MAD) is a robust standard statistical technique to quantify dispersion in any measurements, and a resilient solution to outliers (Wilcox, 2021). Since it uses median instead of mean, MAD offers lesser sensitivity to outliers or extreme values. Equation 1.9 describes the formula to evaluate the MAD.

$$\text{MAD} = \text{median}(|x_i - \text{median}(x)|) \quad (1.9)$$

where:

- x_i represents each data point within the data set,
- $\text{median}(\text{rss})$ is the median value of x ,
- $|x_i - \text{median}(x)|$ is the absolute deviation of each data point from the median

1.4.6 Median Deviation Analysis (MDA) (Positive and Negative Errors)

Median Deviation Analysis (MDA) is a non-standard statistical tool specifically developed for this purpose. While I did not find the use of MDA or a similar technique in the literature, the simple nature of the formulae for MDA could mean that such a technique may have been used previously under a different name. MDA is employed to calculate the dispersion of values across the median value in a set of measurements. The following are features of MDA,

- Central tendency: MDA uses median as the central value which makes it more robust, as compared to the use of mean, against outliers.
- Sensitivity towards extremes: As opposed to SD and Variance, which considers all data points in the measurement, the MDA uses only the *maximum*, *median*, and *minimum* values of a measurement set.
- Asymmetry in data distribution: MDA directly measures the difference between the extreme values from the median value.

In the context of BLE RSSI, MDA calculates the deviation of the maximum and minimum values from the median, focusing only on the extremes. In contrast, MAD,

uses all measurements in the data to determine the overall deviation from the median value. The deviation of maximum values from the median indicates the presence of occasional strong signals, whereas the deviation of minimum values from the median value indicates the presence of occasional weak signals. This helps in the assessment of the consistency of the received signals, or RSSI. As will be presented in Section 2.1.1 of Chapter 2, the outcome of MDA on the data acquired in an anechoic chamber results in smaller error bars, thereby, suggesting the consistency of received signal which can also be linked to the measure of electromagnetic noise and other influential factors in the environment affecting the use of this technology.

MDA is calculated using the equation presented in Equation 1.10. The equation results in two values, $Error_{positive}$ and $Error_{negative}$, which can then be graphed along with *median* value.

$$\begin{aligned} Error_{positive} &= \max(rssi) - \text{median}(rssi) \\ Error_{negative} &= \text{median}(rssi) - \min(rssi) \end{aligned} \quad (1.10)$$

where:

- $rssi$ is the set of measured RSSI,
- $\max(rssi)$ is the maximum value in $rssi$,
- $\min(rssi)$ is the minimum value in $rssi$, and
- $\text{median}(rssi)$ is the median value of $rssi$

1.4.7 /Analysis of Variance (ANOVA) and Tukey's Honest Significant Difference (HSD)

Null Hypothesis, Type I and II Errors

Null Hypothesis is the assumption that all group means are equal, that is, absence of any statistically significant difference in the data groups.

Type I Error: The probability of rejecting the null hypothesis when it is true, that is, false positive.

Type II Error: The probability of failing to reject the null hypothesis when it is false, that is false negative.

To detect the presence of any statistically significant differences in the measurements, the ANOVA technique is applied to the measurements. Statistically significant differences mean that the differences in the means or averages of multiple groups – reflecting that the groups compared are not the same – are unlikely to have occurred by random chance but are present due to true differences in the performance of the groups. With ANOVA however, only the presence or absence of significant differences between group means can be identified. Hence, to understand which specific

groups are different, post-hoc comparisons are required. Tukey's HSD test is one of those post-hoc analysis tools. Tukey's HSD identifies specific groups that are significantly different from each other. One of the outcomes of Tukey's is *p*-value, which is the probability of obtaining the measurements by random chance assuming the null hypothesis is true. This *p*-value is subjected to hypothesis testing where it is compared against a user-defined significance level, called α . For all the experiments presented in this dissertation where ANOVA and Tukey's HSD techniques are applied,

the value of α is chosen as 0.05. While no reliable source for this value is found, many articles state that this value is believed to offer moderate stringency and a balance between *Type I* and *Type II* errors, while, it could also be due to Fisher's influence (Hackshaw and Kirkwood, 2024; Geraghty, n.d.; Zaiontz, n.d.). Finally, CI provides a range of values, derived from the sample measurements, within which, population parameters such as means or difference of means are likely to be present with a certain level of confidence, commonly 95% (Winer, Brown, and Michels, 1991).

The outcome computed through ANOVA can be visually supported using a box plot that demonstrates the median, first and third quartile, range of the measurement, and 95% confidence interval. This will be shown in the Section 2.1.3 in Chapter 2. Both ANOVA and Tukey's HSD are standard statistical tools and are part of many standard libraries.

1.4.8 Dropped Advertisement Rate (DAR) and Dropped Advertisement Percentage (DAP)

Another technique used for analysing the measurements acquired in some of the experiments in this dissertation is the measure of dropped advertisements. This is akin to *packet loss ratio* presented by Peng et al. in (Peng et al., 2013). In Section ?? in Chapter ??, it is already discussed that BLE device emit signals containing basic information regularly. It has also been discussed in Section 1.2.5 in this chapter that the experiments presented in this dissertation rely on the acquisition of these signals, called advertisements, to measure the activity and movement-related trends of pedestrians. Not all of these advertisements however are captured by the Observer. Some Advertisements are lost in a collision or are lost to delay in the software stack of the Observer, as discussed in Section ?? in Chapter ?. Therefore, the advertisements that are not captured by the Observer are dropped advertisements.

The count of dropped advertisements can be calculated in two ways. First, by maintaining a record of Advertisements emitted by the Broadcaster on the Broadcaster itself, which however, may not be possible in off-the-shelf Broadcasters like BLE beacons. Second, which is applicable to those off-the-shelf BLE beacons, by extrapolating the total count of advertisements emitted by the Broadcaster. This is achieved by checking the datasheet of the off-the-shelf Broadcaster to identify the advertisement rate and using it to extrapolate that information. For instance, say a beacon has an advertisement rate of 2 Hz, that means that the beacon advertises two times in a second. With this information, the advertisement count can be identified over a period of measured activity of a pedestrian. Say, the measurements from a pedestrians BLE beacon are observed by the Observer for ten seconds. Since two advertisements are emitted in one second by the Broadcaster, in a measurement period of ten seconds, twenty advertisements must have been emitted by the Broadcaster. Furthermore, since the Advertisement counts are whole numbers, if the measurement period is a fraction, a *floor* operation can be applied to the calculated advertisement count to bring the count to the nearest whole number. Finally, all the observed advertisements are already measured and stored in the Observer.

The calculation of Dropped Advertisement Rate (DAR) and Dropped Advertisement Percentage (DAP) is expressed in Equations 1.11 and 1.12 respectively.

$$\text{DAR} = \frac{(\text{Count of Advertisements Broadcasted} - \text{Count of Advertisements Observed})}{\text{Count of Advertisements Broadcasted}} \quad (1.11)$$

$$\text{DAP} = \frac{\text{Count of Advertisements Broadcasted} - \text{Count of Advertisements Observed}}{\text{Count of Advertisements Broadcasted}} \times 100 \quad (1.12)$$

1.4.9 Curve Fitting

The outcome of a plot of observed RSSI during a walk of a pedestrian with respect to time highlights the fluctuations in the RSSI. As has been discussed, this is due to the sensitivity of these signals to the surrounding environment and other external influences such as weather. A relatively smooth approximation of the resulting plot can be achieved through curve fitting. This is performed by tuning a polynomial curve to follow the trend of the discrete measurements of observed RSSI. An interpolated curve can be obtained through an expression presented in Equation 1.13. The equation requires the degree or order of polynomial to superimpose a curve over the actual measurements. Based on the spread and distribution of the measurements, a higher order polynomial may be required to closely follow the actual measurements. Therefore, to evaluate the optimal degree for the measurements obtained in the experiments presented in this dissertation, the outcome of this curve fitting is subjected to a subsequent test. SSE is calculated to evaluate the sum of the square of differences between each estimation obtained from curve fitting and the actual RSSI measurement to test how deviant predicted or estimated curve is from actual measurements. The expression to calculate SSE is presented in Equation 1.14.

$$\hat{y}(x) = p_1x^n + p_2x^{n-1} + \dots + p_nx + p_{n+1} \quad (1.13)$$

where:

- $\hat{y}(x)$ is polynomial function of x , and
- x is the input RSSI to the polynomial

$$SSE = \sum (\hat{y}_i y_i)^2 \quad (1.14)$$

where:

- $\hat{y}_i(x)$ represents estimation obtained from polynomial, and
- y_i represents the original RSSI

1.4.10 Advertisement Ratios

Advertisement Ratios are used in some of the experiments where a certain percentage of all the measurements falling within any condition being tested are considered against the total number of measurements. For instance, measurements in the range of 10% of the peak or the maximum value over the total number of measurements. If $rssi_i$ is a set of RSS values, Equation 1.15 represents the peak or the maximum value in the set. Equation 1.16 expresses the identification of the RSS values that fall within the 10% range of the maximum RSS value. Finally, the ratio is taken over the total number of advertisements received, as depicted in Equation 1.17.

$$rss_{\max} = \max\{rss_i \mid i = 1, 2, \dots, n\} \quad (1.15)$$

$$rss_{10\%} = \{rss_i \mid rss_i \geq 0.9 \cdot rss_{\max}\} \quad (1.16)$$

$$\text{Ratio} = \frac{|rss_{10\%}|}{n} \quad (1.17)$$

1.4.11 Rician Distribution

The Rician distribution, (Molisch, 2012), is a probability distribution function used for describing the distribution of signals that have both LoS component and several nLoS components. This probability distribution is often used in digital signal processing and is suitable for analysis for the research carried out in this dissertation due to the likelihood of reflection, diffraction, and scattering of advertisements in the outdoor environment. This investigation allows for understanding the degree of influence such phenomenon have on 2.4 GHz Radio Frequency (RF) propagation in the environment, subsequently, enabling the evaluation of fluctuations in RSSI at various locations on selected pathways for conducting experiments, and aiding in the identification of a suitable pathway.

The distribution is defined by its Probability Density Function (PDF) for a random variable R , representing the amplitude of the received signal as expressed in Equation 1.18

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + s^2}{2\sigma^2}\right) I_0\left(\frac{rs}{\sigma^2}\right), \quad r \geq 0 \quad (1.18)$$

where:

- r is the magnitude of receiver signal,
- s is the amplitude of dominant LoS component,
- σ is SD of scattered components, and
- $I_0(\cdot)$ is the modified Bessel function of the first kind and zero order.

The Rician Factor, K , signifies the ratio of power in the LoS component to the power of scattered components. The greater the value of K , the stronger the indication of LoS signal relative to the scattered signals. The expression to calculate the Rician factor, K is given in Equation 1.19.

$$K = \frac{s}{2\sigma^2} \quad (1.19)$$

where:

- s is the amplitude of dominant LoS component, and
- σ is SD of scattered components.

There is a special case of Rician distribution, where, $K = 0$. In this case, the Rician distribution reduces to what is known as *Rayleigh Distribution*, which signifies the situation where the received signal consists purely of scattered components. The significance of other values is as follows:

$$K \geq 0 \quad (1.20)$$

Dominance of LoS component.

$$K \approx 0 \quad (1.21)$$

Dominance of multipath propagation, with no or weak LoS component.

$$K \rightarrow \infty \quad (1.22)$$

Rician distribution approaching Gaussian distribution, with little to no multipath components.