

Henriette Kolby Rohde Garder

# Obtaining situational awareness using Wi-Fi geolocation

Master's thesis in Information Security

Supervisor: Katrin Franke and Kyle Porter

June 2020



# Obtaining situational awareness using Wi-Fi geolocation

Henriette Kolby Rohde Garder

CC-BY 2020/06/01





# Abstract

The primary goal for this master's thesis was to investigate if using geolocalization techniques (based on Wi-Fi and rogue access points) may be a practical way to increase situational awareness by locating and tracking persons of interest for first response personnel. Specific subgoals of the project were to see how the collected data can be used for geolocation and how precise the results can be to locate and predict the movement of a unit.

During the project, the data collection was done by a project partner within a confined area inside a building using colleagues' mobile phones, with their knowledge and consent. The data was collected based on specific scenarios. The different scenarios were set up to see how the captured data could vary based on different realistic situations e.g if an individual is standing still, moving in different patterns or moving between different rooms.

This master's thesis presents an overview of ways to clean a data set and which of the cleaning methods were used on the data set. We also explain why exactly these cleaning methods were used and how they affected the data set. The data set was analyzed using four main geolocation methods: Angel of Arrival, Time of Flight, Fingerprinting using signal strength and Triangulation or Multilateration using signal strength. The methods were evaluated and tested with the data set to see which methods worked and the degree to which they worked. If the methods did not work with the given data set it was explained what had to be changed for the method to work.

It turned out that only one of the four geolocation methods (assuming triangulation and multilateration are grouped into the same category) could be used on the given data set. The two methods, Angel of Arrival and Time of Flight, could not be tested since the special equipment needed was not available to ensure correct data information and the data it selves did not contain enough information. The third method, fingerprinting using signal strength, did not need any extra equipment, but the lack of enough data packets due to time and other uncertainties meant that the method could not be used on the given data set. The fourth and final method, triangulation or multilateration using signal strength, was the only one that could be used on the given data set with sufficient accuracy. This method seemed to provide an accuracy of a few meters and could possibly be used to obtain increased situational awareness.



# Sammen drag

Det overordnede målet for denne masteroppgaven var å undersøke om bruk av geolokaliseringsteknikker (basert på Wi-Fi og rogue-tilgangspunkter) kan være en praktisk måte å øke situasjonforståelse ved å lokalisere og spore personer av interesse for politi, brannvesen og ambulansepersonell. Spesifikke delmål for prosjektet var å se hvordan de innsamlede dataene kunne brukes til geolokalisering og hvor presise resultatene var for å lokalisere og forutsi bevegelsen til en enhet.

I løpet av prosjektet ble datainnsamlingen gjort av en prosjektpartner innenfor et avgrenset område inne i en bygning ved bruk av kollegers mobiltelefoner, med kollegaenes kjennskap og samtykke. Dataene ble samlet inn basert på spesifikke scenarier. De forskjellige scenariene ble satt opp for å se hvordan dataene som ble fanget kan variere basert på forskjellige realistiske situasjoner, for eksempel hvis en person står stille, beveger seg i forskjellige mønstre eller beveger seg mellom forskjellige rom.

Denne masteroppgaven presenterer en oversikt over metoder å rense et datasett på og hvilke av disse metodene som ble brukt for å rense datasettet som ble brukt i prosjektet. Det forklarer også hvorfor disse rense metodene ble benyttet, og hvordan de påvirket datasettet. Datasettet ble analysert ved hjelp av fire hovedmetoder for geolokalisering: Angel of Arrival, Time of Flight, Fingerprinting using signal strength og Triangulation or Multilateration using signal strength. Metodene ble evaluert og testet på datasettet for å se hvilke metoder som fungerte og i hvilken grad de fungerte. Dersom metodene ikke fungerte med det gitte datasettet ble det forklart hva som måtte endres for at metoden skulle fungere.

Kun en av de fire geolokasjonsmetodene (forutsatt at triangulering og multilaterering er gruppert i samme kategori) kunne brukes på det gitte datasettet. De to metodene, Angel of Arrival and Time of Flight, kunne ikke testes siden spesialutstyret som trengs ikke var tilgjengelig for å sikre korrekt datainformasjon og at dataene ikke inneholdt nok informasjon. Den tredje metoden, Fingerprinting using signal strength, trengte ikke noe ekstraintstyr, men mangelen på nok datapakker på grunn av tid og andre usikkerheter gjorde at metoden ikke kunne brukes på det gitte datasettet. Den fjerde og siste metoden, Triangulation or Multilateration using signal strength, var den eneste som kunne brukes på det gitte datasettet med en viss troverdighet. Denne metoden så ut til å gi en nøyaktighet på noen få meter og kan muligens benyttes til å oppnå økt situasjonforståelse.



# Acknowledgements

I would like to thank my supervisor Prof. Katrin Franke for making this master's thesis possible and all the support. I would also like to thank my co-supervisor Kyle Porter for continues support and thorough feedback throughout the project. Furthermore I would like to thank my project partner for arranging and collecting the data used in this master's thesis. Lastly I would like to thank my family for valuable support.



# Contents

<b>Abstract</b> . . . . .	<b>iii</b>
<b>Sammendrag</b> . . . . .	<b>v</b>
<b>Acknowledgements</b> . . . . .	<b>vii</b>
<b>Contents</b> . . . . .	<b>ix</b>
<b>Figures</b> . . . . .	<b>xiii</b>
<b>Tables</b> . . . . .	<b>xvii</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Topic covered by the project . . . . .	1
1.2 Keywords . . . . .	2
1.3 Problem description and research questions . . . . .	2
1.3.1 Research questions . . . . .	2
1.4 Justification, motivation and benefits . . . . .	3
1.5 Planned contributions . . . . .	4
1.6 Report structure . . . . .	5
<b>2 Background</b> . . . . .	<b>7</b>
2.1 Technical terms . . . . .	7
2.2 Wi-Fi and Probe request . . . . .	11
2.2.1 Probe request and probe response . . . . .	15
2.3 Related work . . . . .	15
2.3.1 Crowd monitoring research . . . . .	16
2.3.2 Individual tracking research . . . . .	18
<b>3 Methodology</b> . . . . .	<b>21</b>
3.1 Possible data collection methods . . . . .	21
3.2 Possible data cleaning methods . . . . .	22
3.2.1 Unwanted observations . . . . .	23
3.2.2 Structural Errors . . . . .	24
3.2.3 Unwanted Outliers . . . . .	24
3.2.4 Missing Data . . . . .	25
3.2.5 Distance based filtering . . . . .	25
3.2.6 Time-based filtering . . . . .	26
3.2.7 Time compression . . . . .	26
3.2.8 Cycle removal . . . . .	27
3.3 Possible location methods . . . . .	27
3.3.1 Angle of Arrival . . . . .	27

3.3.2	Time of Flight . . . . .	28
3.3.3	Fingerprinting using Signal strength . . . . .	30
3.3.4	Triangulation or Multilateration using Signal strength . . . . .	31
3.4	Methodology flowchart . . . . .	43
<b>4</b>	<b>Experiment setup . . . . .</b>	<b>45</b>
4.1	Experiment location and participants . . . . .	46
4.2	Equipment . . . . .	47
4.3	Sensor locations . . . . .	47
4.4	Grid pattern . . . . .	48
4.5	Scenarios . . . . .	48
4.5.1	Preparation scenarios . . . . .	49
4.5.2	Geolocation Scenarios . . . . .	52
<b>5</b>	<b>Data Prepossessing and Analysis . . . . .</b>	<b>57</b>
5.1	Data types . . . . .	57
5.2	Data prepossessing and cleaning . . . . .	58
5.2.1	Probe type removal . . . . .	59
5.2.2	Extraction of data collected during the scenarios . . . . .	60
5.2.3	Removing data based on signal strength . . . . .	63
5.2.4	Time compression . . . . .	66
5.3	General analysis of data . . . . .	66
5.3.1	Difference in number of registered phones per scenario . . . . .	66
5.3.2	Difference in number of packet per phone . . . . .	67
5.3.3	Synchronization . . . . .	68
5.3.4	Calibration . . . . .	69
5.3.5	Time stamps . . . . .	69
5.4	Angle of Arrival - Analysis . . . . .	69
5.5	Time of Flight - Analysis . . . . .	69
5.6	Fingerprinting using signal strength - Analysis . . . . .	70
5.6.1	Fingerprints . . . . .	75
5.7	Triangulation or Multilateration using signal strength - Analysis . . . . .	79
5.7.1	Experiment location and placement of mobile units . . . . .	79
5.7.2	Calculations used by both methods . . . . .	82
5.7.3	Triangulation . . . . .	84
5.7.4	Trilateration/Multilateration . . . . .	84
5.7.5	Uncertainties: Triangulation and Multilateration . . . . .	85
<b>6</b>	<b>Triangulation and Multilateration results . . . . .</b>	<b>91</b>
6.1	Scenario 4 . . . . .	92
6.1.1	Scenario 4.1 . . . . .	93
6.1.2	Scenario 4.2 . . . . .	95
6.2	Scenario 5 . . . . .	96
6.2.1	Scenario 5.1 . . . . .	96
6.2.2	Scenario 5.2 . . . . .	98
6.3	Scenario 6 . . . . .	100
6.4	Scenario 7 . . . . .	102



6.5	Scenario 8	104
6.6	Scenario 9	106
<b>7</b>	<b>Discussion and Conclusion</b>	<b>109</b>
7.1	Limitations of the thesis	110
7.2	Future work	112
7.3	Conclusion	112
	<b>Bibliography</b>	<b>115</b>
<b>A</b>	<b>Additional localization results</b>	<b>121</b>
A.1	Scenario 4.1	122
A.2	Scenario 4.2	125
A.3	Scenario 5.1	127
A.4	Scenario 5.2	129
A.5	Scenario 6	131
A.6	Scenario 7	133
A.7	Scenario 8	135
A.8	Scenario 9	137



# Figures

2.1	Access point visualization . . . . .	12
2.2	SSID and BSSID visualization . . . . .	13
2.3	Wi-Fi passive search . . . . .	14
2.4	The difference between passive and active Wi-Fi search . . . . .	14
2.5	Initial phase of communication . . . . .	15
2.6	The scenario presented and used by Groba . . . . .	16
2.7	Sensor locations used by Chilipirea et al. . . . .	17
2.8	Sensor locations used by Schauer et al. . . . .	18
2.9	Sensor locations used by Musa et al. . . . .	19
2.10	Tracking concept presented and used by Meng-Hsuan et al. . . . .	20
3.1	Angle of Arrival visualization . . . . .	28
3.2	Time of Flight visualization . . . . .	29
3.3	Fingerprinting visualization . . . . .	30
3.4	Triangulation with two receivers . . . . .	34
3.5	Triangulation with three receivers . . . . .	35
3.6	The values that must be calculated . . . . .	36
3.7	The calculated values . . . . .	37
3.8	Triangulation with multiple receivers . . . . .	37
3.9	Trilateration . . . . .	38
3.10	Multilateration . . . . .	38
3.11	Multilateration with individual circle equations . . . . .	39
3.12	The circles intersecting at the same location . . . . .	42
3.13	Methodology flowchart . . . . .	43
4.1	Experiment environment . . . . .	46
4.2	The location of the sensors . . . . .	48
4.3	Grid pattern . . . . .	49
4.4	Preparation 1 - Fingerprinting . . . . .	50
4.5	Preparation 2 - Attenuated signal strength . . . . .	51
4.6	Preparation 3 - Standing still . . . . .	52
4.7	Scenario 4 and 5 - Moving around inside the room . . . . .	53
4.8	Scenario 6 and 7 - Moving around in the corridor . . . . .	54
4.9	Scenario 8 and 9 - Moving between the corridors and the room . . . . .	55

5.1	Initial phase of communication . . . . .	59
5.2	Scenario 1.1 - Sensor 1-4 - Strongest Signal Strength . . . . .	61
5.3	Scenario 1.1 - Crucial points for locating where the scenario begins . . . . .	61
5.4	Scenario 1.1 - Crucial points for locating where the scenario ends . . . . .	62
5.5	A typical scenario . . . . .	64
5.6	One section of a typical scenario . . . . .	64
5.7	Visualization of multipath . . . . .	65
5.8	The signal strength boundary . . . . .	66
5.9	Different number of packets per phone and scenario . . . . .	68
5.10	Fingerprint visualization . . . . .	70
5.11	Packet detection rate . . . . .	71
5.12	Packet clustering . . . . .	72
5.13	Number of data packets per phone - scenario 1.1 and scenario 1.2 . . . . .	74
5.14	Fingerprint 01 - Preparation 1.1 - Average and Strongest dBm value . . . . .	75
5.15	Data packet registration . . . . .	76
5.16	Fingerprint 02 - Preparation 1.2 - Average and Strongest dBm value . . . . .	77
5.17	Data packet registration . . . . .	77
5.18	Fingerprint 03 - Combination of 1.1 and 1.2 - Average and Strongest dBm value . . . . .	78
5.19	Scenario 3 - Standing still . . . . .	79
5.20	Data packet distribution over time . . . . .	81
5.21	Scenario 2 - Attenuated signal strength . . . . .	82
5.22	The change from dBm to meters . . . . .	83
5.23	Triangles formed by the sensors . . . . .	84
5.24	Circles formed by the sensors . . . . .	85
5.25	Triangulation and Multilateration with ideal measurements . . . . .	86
5.26	Triangulation and Multilateration with errors . . . . .	86
5.27	Phone 5 results . . . . .	88
6.1	Visualization of movement from multiple people . . . . .	91
6.2	Color code - mobile phones . . . . .	92
6.3	Data packets recorded during Scenario 4 . . . . .	92
6.4	Scenario 4.1 - Movement and location results . . . . .	94
6.5	Scenario 4.2 - Movement and location results . . . . .	95
6.6	Data packets recorded during Scenario . . . . .	96
6.7	Scenario 5.1 - Movement and location results . . . . .	97
6.8	Scenario 5.2 - Movement and location results . . . . .	99
6.9	Data packets recorded during Scenario 6 . . . . .	100
6.10	Scenario 6 - Movement and location results . . . . .	101
6.11	Data packets recorded during Scenario 7 . . . . .	102
6.12	Scenario 7 - Movement and location results . . . . .	103
6.13	Data packets recorded during Scenario 8 . . . . .	104
6.14	Scenario 8 - Movement and location results . . . . .	105
6.15	Data packets recorded during Scenario 9 . . . . .	106

6.16 Scenario 9 - Movement and location results . . . . .	107
A.1 Visualization of movement from multiple people . . . . .	121
A.2 Color code - mobile phones and connecting circles for interaction point . . . . .	122
A.3 Scenario 4.1 . . . . .	123
A.4 Scenario 4.1 - Multilateration and Triangulation . . . . .	124
A.5 Scenario 4.2 . . . . .	125
A.6 Scenario 4.2 - Multilateration and Triangulation . . . . .	126
A.7 Scenario 5.1 . . . . .	127
A.8 Scenario 5.1 - Multilateration and Triangulation . . . . .	128
A.9 Scenario 5.2 . . . . .	129
A.10 Scenario 5.2 - Multilateration and Triangulation . . . . .	130
A.11 Scenario 6 . . . . .	131
A.12 Scenario 6 - Multilateration and Triangulation . . . . .	132
A.13 Scenario 7 . . . . .	133
A.14 Scenario 7 - Multilateration . . . . .	134
A.15 Scenario 8 . . . . .	135
A.16 Scenario 8 - Multilateration and Triangulation . . . . .	136
A.17 Scenario 9 . . . . .	137
A.18 Scenario 9 - Multilateration . . . . .	138
A.19 Scenario 9 - Triangulation . . . . .	139



# Tables

2.1	Wi-Fi spesification [27–29] . . . . .	12
3.1	ToF approximately distance resolution per time unit . . . . .	29
3.2	An ideal measurement . . . . .	41
3.3	Calculated X and Y values for an ideal measurement . . . . .	42
5.1	Data packets per scenario . . . . .	63
5.2	Which phones were captured during which scenarios . . . . .	67
5.3	Amount of data packets sent and received . . . . .	80
5.4	Ideal example with added errors . . . . .	87
5.5	Calculated X and Y values . . . . .	87





# Chapter 1

## Introduction

### 1.1 Topic covered by the project

Localization of mobile phones is useful in many situations. It can help find a single person, or can track movement and direction for a single person or groups of persons. The localization of persons or a group of persons can be used to get an overview of the current situation for a geographical area of where people are, where they are moving and their interaction with others. In this masters thesis, this is defined as situation awareness.

Situational awareness is critical for first responders, law enforcement and military to get an accurate and up to date overview of the current situation when providing security in public places as this ideally provides actionable and timely intelligence, either during public events or in everyday life. Situational awareness is a tool for the decision makers at the tactical and operational levels for preventing or reducing the impact of unwanted incidents, such as fires, burglary, medical emergency, trespassing on restricted areas or disorderly behavior. Others who can benefit from situational awareness are Health authorities for pandemic control/research, transport, public service, customer service, who can improve their strategic decision making [1].

This master's thesis focuses on analyzing user data from publicly placed rogue Wi-Fi access points to see if the data can be used enhance situational awareness in a timely and actionable fashion. There is a lot of information that can be collected from a Wi-Fi access point, and the analysis of user data can potentially provide a multitude of benefits with respect to situational awareness, but this master's thesis focuses on only a subset of the data for Wi-Fi device geolocation.

This device geolocation will estimate the position of a device and if a specific individual can be tracked to see how a person has most likely moved from one place to another. As an example, this can be used if a criminal act has occurred and the police are trying to find out what direction a suspect has taken when fleeing the scene. Another example is the need to locate a person who needs help in different situations like if he is in a fire or in need of medical assistance [1].

## **1.2 Keywords**

Wi-Fi geolocation, Situational awareness, Surveillance mechanisms, Wireless access points

## **1.3 Problem description and research questions**

Fires, medical emergencies and criminal acts are problems for all communities and it is the first responders' task to assist in emergencies. The main task for law enforcement is to maintain and protect the law and order in the society. There will always be a desire to reduce the number of emergencies, criminal acts or public disorder, such that it results in a safer society.

A problem for the first responders, such as the police and emergency medical responders, is to get a good overview of the situation when there are many people gathered at a location, or if there are locations where there are no normal police activities e.g. patrolling units. In these situations, first responders depend on the public to report emergencies before they get knowledge of the situation and can respond. This reactive approach of handling the situation will not prevent the situation from happening and it may take the first responders some time to arrive and assess the scene, thus giving the situation time to escalate or giving the perpetrator of a crime an opportunity to escape. First responders would be more effective and the public would be safer if the first responders had tools that could enhance their intelligence capabilities for the purpose of situational awareness. The first responders could act and deploy firemen, ambulances or police officers to a situation before irreversible damage is done, or even prevent an impending criminal act.

In an ongoing operation, the estimation of device geolocation may give the first responders enhanced situational awareness, allowing first responders to be more proactive and prepared for different events. With geolocalisation data, the Intelligence officers may analyze the movement of an involved party before the incident occurs, during the incident and may even predict the movements after the incident. Another benefit will be that the analysis, evaluation, and the tactical decision making may be done in a safe position and not take away valuable time and efforts from the operational command of the commanding officer at the location, but to give him the timely, actionable, and supporting information needed.

### **1.3.1 Research questions**

The goal of this master's thesis is to investigate if using geolocalization techniques based on Wi-Fi and rogue access points may be a practical way to increase the situational awareness by locating and tracking persons of interests for first response personnel where this is important.

It is very important that the information produced is of good enough quality so that the results are sufficiently reliable for e.g. first responders and law enforce-

ment to use the intelligence to effectively allocate their resources. The research question for this master's thesis is therefore:

- How can the information gathered from rogue access points increase situational awareness by Wi-Fi geolocation of mobiles and what are the challenges that may occur when analyzing the collected data?

There are a some questions that need to be answered before an answer to the research question can be found.

- What type of information can be gathered by the rogue access point?
- How to handle messy and intentionally misleading data?
- How to ensure reliability of the analysis?
- Can data from several access points at the same location be correlated?
- How effective is the collection of geolocation data with respect to:
  - What type of data and how much data will be registered at an access point from each device?
  - How often will data from one device be registered at an access point?
  - Will data from one device be registered simultaneously at all access point in reach?
  - Will there be a limit to how many devices that can be detected at each access point?
  - Are the clocks of all the receivers synchronized?
- Is there any privacy (GDPR) issues handling user data?

## 1.4 Justification, motivation and benefits

The motivation behind this project is to be able to provide a method to improve the capabilities of the Norwegian first responders by increasing their situational awareness by way of device Wi-Fi geolocation. There is a desire to take a proactive approach to public safety. As an example, for the police, who are often responsible for security during public events, an increase in situational awareness via geolocation can provide tactical level intelligence of ongoing situations to detect indicators of criminal behavior.

There can never be enough situation awareness. By increasing the first responders situational awareness, they can be more effective in their response by dispatching the necessary units to a situation before the severity of the event escalates. For example by providing quick medical assistance or stop a criminal act in progress. The benefits this will provide are quicker response for the public in emergencies. In addition to increasing the security of the public, the data collected may also be used by the police for the sake of prosecution. If a criminal act were to occur, the data collected at that time might be used to potentially place a suspect at the location during the criminal incident or maybe even vindicate the suspect.

A challenge for the first responders is to optimize their use of resources (units

and equipment) in the most effective way and an increased situation awareness will assist the first responders in this assignment.

## 1.5 Planned contributions

Many similar studies have been carried out with very different goals. There has also been large variations in types of locations where different studies have collected their data. Several studies have collected their data indoors, in malls or in office buildings. These projects have mainly focused on how to use the data for marketing or how to adjust the environment, like power and light based on where people tend to stay [2].

The different studies that have collected their data outdoors have focused more on crowd analysis, such as how people move during festivals, demonstrations and other larger gatherings [3, 4]. There are a few studies that have focused on being able to locate individuals, such as the work by Musa et al. [5], Cunche [6] and Meng-Hsuan et al. [7]. The problem with most of the studies and research reports this thesis has studied is that they mainly focus on how they collected their data and results, but not so much the transition from the collected data to the actual results. It is unclear how they interpreted the data. The data they have collected is likely to have contained noise and perhaps misleading data. The method and criteria's for cleaning and handling this situations are not clear.

The overall goal for this master's thesis is to increase the situational awareness for first responders, which in turn can be used for e.g. mitigation actions or better handle ongoing situations. During the project, the data collection was done by a project partner in "controlled environments" on colleagues within a confined area inside a building, with the colleagues' knowledge and approval. This master's thesis will present an overview of some ways to clean the data set and which methods that were used on the data set in this master's thesis. This master's thesis will also explain why exactly these methods were used and how it changed the data set. The data set was analyzed for the four main geolocation methods, and the methods were evaluated and tested with the data set used in this master's thesis. This was done to see which methods work and which methods do not work and why this is the case. The masters thesis presents how the methods can be used for different scenarios such as:

- Pin-Pointing individuals
- Tracking movement of individuals
- Tracking speed and direction of person(s)
- Identifying possible cooperation of peoples (e.g. a team working together)
- Identifying any interaction between persons (e.g. identify if a criminal has been close to a victim)

## 1.6 Report structure

The rest of the report is structured in the following way:

### Background

The **Background** chapter contains a description of some technical terms that are important to better understand the report. There are also information on some related work, which includes the various goals that can be achieved using Wi-Fi location methods.

### Methodology

The **Methodology** chapter presents possible methods for handling and processing the data set. The chapter is divided into 3 sections. The first part contains possible data collection methods and the second part contains several possible data cleaning methods. The last section contains four different possible geo-location methods.

### Experiment Setup

The **Experiment Setup** chapter describes the layout and information surrounding the data collection. The test area and test participants will be explained, together with the type of equipment used and where the different sensors are located. The experiment and data collection are based on different scenarios and these will be explained in detail.

### Data analysis

The **Data Analysis** chapter describes how the data, after it was collected, was preprocessed, cleaned and the actual analysis of the data. This analysis will, in addition to a general analysis, focus on the geolocation methods that, based on the equipment available, can be implemented to varying degrees.

### Triangulation and Multilateration results

The **Triangulation and Multilateration results** chapter presents the results from the triangulation and multilateration of Scenario 4 to Scenario 9. Each of the results will be interpreted and discussed.

### Discussion and Conclusion

The **Discussion and Conclusion** chapter contains a discussion of all the results in the report which include which methods have worked and which have not worked. Other interesting findings is also discussed in this chapter. Limitations of

the report and potential for improvement will be presented and discussed. The chapter will also discuss future work and conclude the entire report.

## Chapter 2

# Background

This chapter gives background information that may be helpful for the reader during reading the rest of the report. The chapter is divided into 3 sections:

- Technical terms
- Wi-Fi and Probe request
- Related work

The Technical terms section describes important technical terms for a better understanding of the later chapters. The Wi-Fi and Probe request section describes the Wi-Fi technology used to geolocalize a mobile using Wi-Fi. The Related work section presents past research done by other researchers for similar experiments related to Wi-Fi geolocation.

### 2.1 Technical terms

This chapter contains several important technical terms that are described for a better understanding of the later chapters.

#### **dBm**

dBm (decibel-milliwatts) is a decibel unit with a fixed value equal to one milliwatt that is often used as a unit of measurement for absolute signal strength received at an antenna [8]. Signal power becomes most accurate when expressed in milliwatts (mW), but it is more convenient to use dBm. A challenge with using milliwatts is that there will be very small numbers using many decimal places to describe the value (e.g. ranging from 100 mW to 0.000001 mW) which is not practical. Using a logarithmic scale like dBm is more practical when describing the signal strength [9]. A signal strength of -40 dBm corresponds to 0.0001 mW. A typical Wi-Fi reception level at an access point is -30 dBm (0.001 mW) to -60 dBm (0.000001 mW) which makes it difficult to read the signal strength when presented in milliwatts [9].

When using dBm to represent the signal strength, it is important to acknowledge that signal receptions are normally negative numbers, which means that -20 dBm is a stronger signal than -70 dBm. Since dBm is a logarithmic unit, it does not scale in a linear fashion. As an example, a 3 dB of loss (-3 dB) means that the signal strength is halved (e.g. 0 dBm equals 1 mW, while -3 dBm equals 0,5 mW) or if one have 3 dB of gain (+3 dB) this means that the signal strength is doubled (e.g. 0 dBm equals 1 mW, while +3 dBm equals 2 mW)[9].

### **Received signal strength indication (RSSI)**

RSS (Received Signal Strength) is a value that indicates the signal strength between a transmitter and a receiver [10]. The RSS value is the actual signal strength that a receiver receives, and this value can be used to determine how far away a device is from the receiver [11].

The RSSI (Received Signal Strength Indicator) is a relative value of the received signal strength and there is no standardized relationship between RSSI and milliwatts. RSSI is a common measurement, but each mobile vendor provides their own accuracy, granularity and range of the RSSI values [12]. The mobile vendors used different scales for RSSI, e.g. 0-60 or 0-255 [9]. Since vendors use different scales to define RSSI values, RSSI does not work as an absolute measurement of distance.

### **Ground truth**

Ground Truth, in this context, is the information that was obtained on the test location e.g. the actual location of a device. Results produced from the data collected from the test location can be compared to the ground truth to see if the results are correct and how accurate they are.

In this master's thesis, ground truth will mainly refer to the actual location of different mobiles during the experiment. This master's thesis is using a video recording of the location as the ground truth. Based on the ground truth video, one can find the physical location of the mobiles and see how they moved around the room. This allows the results of the geolocation to be compared to where the mobiles actually were, and can be used to say how accurate the geolocation methods are.

### **MAC address**

A MAC address (Media Access Control Address) is a unique identifier assigned to a network interface, and is a static address [13]. All networkable devices that exist each have a unique MAC address, which means that there are millions of possible addresses. The MAC addresses cannot be changed since it is manufactured into the unit such as the Wi-Fi card or Ethernet card [13, 14]. Although, some mobile vendors claim that they are randomizing the MAC address at different intervals to counter tracking [15].



## **Multipath**

Multipath describes a situation where a radio signal uses several ways to reach the same location. This can happen as a result of the signals sent from a device being reflected by nearby objects or the signals being scattered. Multipath will appear as fluctuations in signal strength, often making devices appear to be further away from the sensors than they actually are [16].

An example of multipath is when the different radio signal paths may cause two data packets arrive to the same access point at different time and different signal strength even if they were transmitted approximately at the same time and may interfere with each other. A radio signal from a mobile phone can reach the access point directly, and at the same time the signal may also have been reflected in several nearby buildings, causing a delay in an identical signal reaching the access point.

## **Rogue access points**

A rogue access point is a wireless access point that has been placed in an environment without the knowledge or approval from the users or network administrators with the purpose to lure users to connect to the access point or to listen in on the data traffic [17].

A rogue access point can be a significant threat to a company, and are often used in data attacks such as data theft or DoS attack [17, 18]. Rogue access points will in most cases not broadcast their existence in order to stay hidden. They will often not require authentication nor use encryption, making them a major security threat [18].

There are several ways rogue access points can get a mobile to connect to them. In some instances, the mobile devices may broadcast its list of known SSIDs, and the rogue access point listens for these broadcastings. The rogue access point will then change its SSID to one of the SSIDs known by the mobile. Alternatively, the rogue access point can use a default SSID name for a network, and people who never updated their SSID names will fall prey to the rogue access point.

Rogue access points will, in this paper, refer to a wireless access point that has been placed in an environment without the knowledge or approval from the users or network administrators in the purpose to listen in on data traffic.

## **Signal strength measurement**

The Signal Strength is measured at the rogue access point and used to calculate an indication of the distance of the transmitting unit. The Rogue Access Point gives a value that indicates the measured signal strength received either in RSSI or in dBm depending on the access point's vendor's preference. In some cases, the signal strength will also be expressed in percentage [19].

RSSI and dBm are the most common ways to represent signal strength. The difference between the units of measurement is that dBm is an absolute represent-

ation of the signal strength and RSSI is a relative index determined by the vendors [10].

### **Time/Clock synchronization**

Time or Clock synchronization is the process of making sure two or more units are synchronized together with respect to timing. Time/Clock synchronization is important in many situations like fault diagnosis and recovery. It is also important for security systems, database systems and scheduled operations. Time synchronization is critical to be able to correlate events. If, for example, a security breach has occurred, it is crucial to know in which order the events happened. For such situations, time synchronization is critical, since otherwise it may appear that some events occurred before or after they actually happened, which may give a false picture of the event [20, 21].

Another situation where time synchronization is important is in situations where one has to correlate data from different devices and the combined data should be used to make important decisions. Failure to synchronize devices may result in incorrect decisions as the data gives a wrong picture of the situation [22].

Since each device has its own internal clock, units that are initially synchronized may over time end up being out of sync. Small differences in the clocks, with regard to tick rates, can cause one of the clocks to drift e.g. one second a day [20]. If it is not detected that the clocks are no longer in sync, it may have major consequences over time. There are many systems that require the devices to be synchronized and choose to use a global clock that all the devices synchronize against at regular intervals [21].

### **Triangulating**

Triangulation is a way of calculating positions of a unit by use of triangles. When a transmitter (mobile phone) sends a signal that is received by two or more receivers (Access points), the transmitter and receivers form one or more triangles. By calculating the distance from the transmitter to the receivers by converting the received signal strength, all the sides of the triangle are known if also the locations of the receivers are known. By use of trigonometry, all the angles in the triangle can be calculated, and by using this information, the transmitting unit can be pin-pointed in an coordinate system [23, 24].

When the signal strength (dBm) from the mobile to the sensors is measured, the distance from the mobile to the sensors can be calculated, and calculating one of the angles will be enough for finding the position. See Section 3.3.4 for an example of the calculation. The triangulation calculations may use the direction from where the signal arrives (Angle of Arrival), the time the signal needs to reach the sensors (Time of Flight) or the signal strength. The signals can be affected by the surrounding objects such as buildings and give incorrect results. Therefore, since the signal strength can be affected, it may be a good idea to con-

firm the location by carry out multiple independent triangulation's when locating the transmitting unit, if possible.

### **Trilateration/Multilateration**

Trilateration/Multilateration uses received signals from a transmitter to multiple access nodes to calculate an estimated location point for the transmitting unit (i.e. received signals on several access nodes can be used for pinpointing a transmitting Wi-Fi transmitter). Trilateration/Multilateration is based on the intersections between circles around the sensors. All the sensors that captured data from a mobile phone will know, among other things, the signal strength. The signal strength received by all the sensors may be converted into distance (meters). The sensor will know that the transmitting unit will be somewhere on a circle with radius equal to the calculated distance from the sensor. The intersection of circles around several sensors indicates the location of the mobile. Using two sensors, the circles will intersect in two places, which means that one get two possible locations for the mobile [24, 25].

Using three sensors, called trilateration, the location of the mobile will be where the three circles intersect. In the best case, all three circles will intersect at the same location, and one will get the exact location of the mobile. Due to interference and other possible causes of faulty measurement, it is probable that the circles do not intersect at the exact same location, but an approximation of the probable location one can still be calculated. Multilateration is when more than three sensors are used to locate the mobile, more circles can give more confidence to the result if all the sensors intersect at the same location as well as being able to perform geolocation in three dimensional space[24, 25].

The Trilateration/Multilateration calculation may use the time the signal uses to reach the sensors (Time of Flight) or the signal strength to calculate the distance between transmitter and receiver. The signals can be affected by the surrounding objects such as buildings and give incorrect results. Therefore, since the signal strength can be affected, and it may be a good idea to confirm the location by carrying out multiple independent Trilateration's/Multilateration's when locating the transmitting unit, if possible.

## **2.2 Wi-Fi and Probe request**

The Wi-Fi, an abbreviation of "Wireless Fidelity", is based on a the IEEE 802.11 standards that define the technical specifications for the wireless system [26]. IEEE 802.11 defines several Wi-Fi specifications with different aspects for Wi-Fi. Some of the main difference are listed in the Table 2.1.

The Wi-Fi system are based on two main elements, the remote units (e.g. mobile phones or PC) and Access points that are the fixed connection point to a network. The Access point will typically be connected wirelessly to several mobile units, while the mobile unit will be connected to only one Access point at the time.

Wi-Fi Genreation	IEEE Standard	Year adopted	Maximum Linkrate	Frequency
Wi-Fi 6	802.11ax	2019	600-9608 Mbit/s	2.4/5 GHz
Wi-Fi 5	802.11ac	2014	433-6933 Mbit/s	5 GHz
Wi-Fi 4	802.11n	2009	72-600 Mbit/s	2.4/5GHz
Wi-Fi 3	802.11g	2003	3-54 Mbit/s	2.4 GHz
Wi-Fi 2	802.11a	1999	1.5-54 Mbit/s	5 GHz
Wi-Fi 1	802.11b	1999	1-11 Mbit/s	2.4 GHz

Table 2.1: Wi-Fi spesification [27–29]

The access point will process the data from several mobile units and funnel this to the connected network. In the other direction, the access point will get data from the connected network and distribute this to the correct mobile unit. Figure 2.1 illustrates how an access point funnels the data.

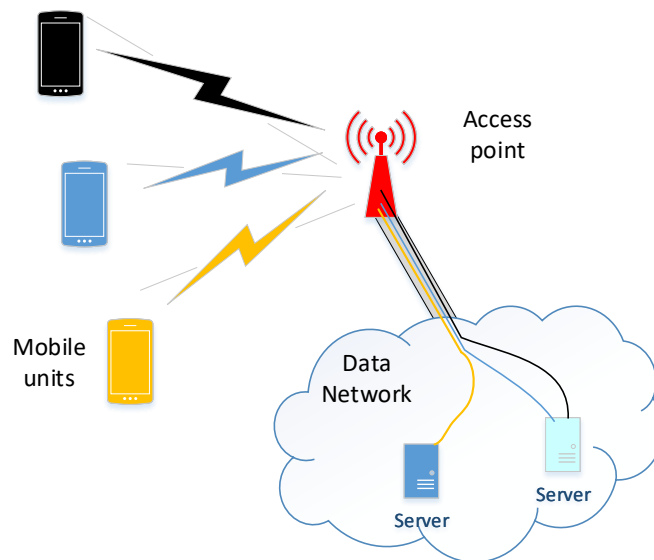
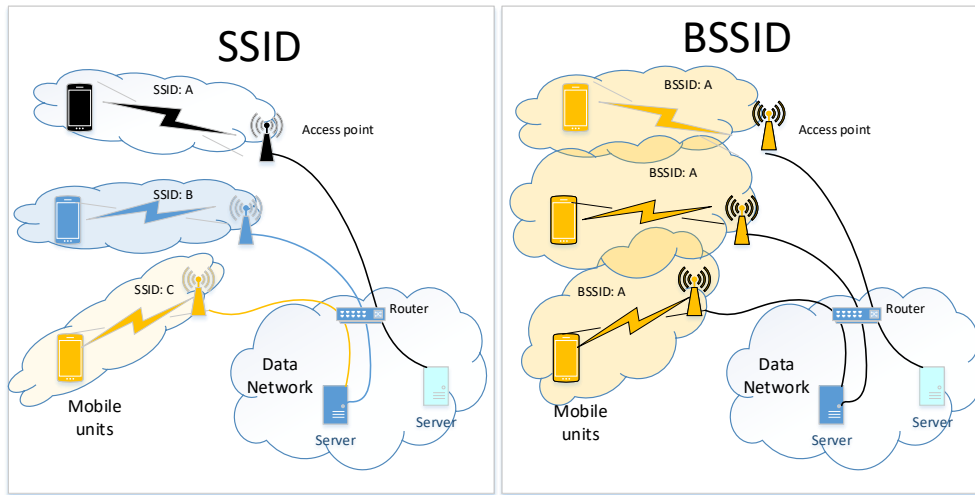


Figure 2.1: Access point visualization

The access points can be associated with service set and service set identifiers (SSIDs) that can differentiate the different wireless networks. The SSID that defines each access point has their own unique network with its own network name. Each access point will form a separate wireless network and can restrict access to the network (a mobile unit must manually affiliate to each of the access points when moving around between Wi-Fi networks). BSSID (Basic Service Set ID) can be used to cluster several access points within the same wireless network. The access points will be set up with Basic Service Sets and all access points with the same Basic Service Sets will form an associated network. When an access point

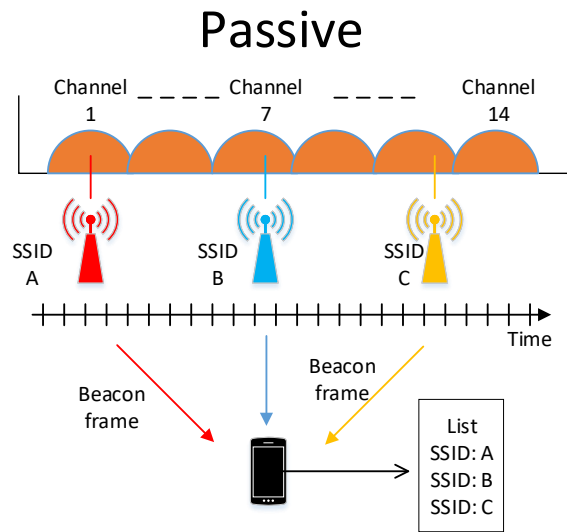
receives a data packet request from a mobile, the receiving access point and will mark each received data packet from a mobile with a Basic Service set ID number based on the access points own ID before this is sent to the server. This enables the server to respond to the correct access point for responding to the mobile initiating the request. Several access points can work together as a single network and a mobile unit can move between the different access point sectors and automatically be connected to the next access point [15, 30]. SSID and BSSID are illustrated in Figure 2.2.



**Figure 2.2:** SSID and BSSID visualization

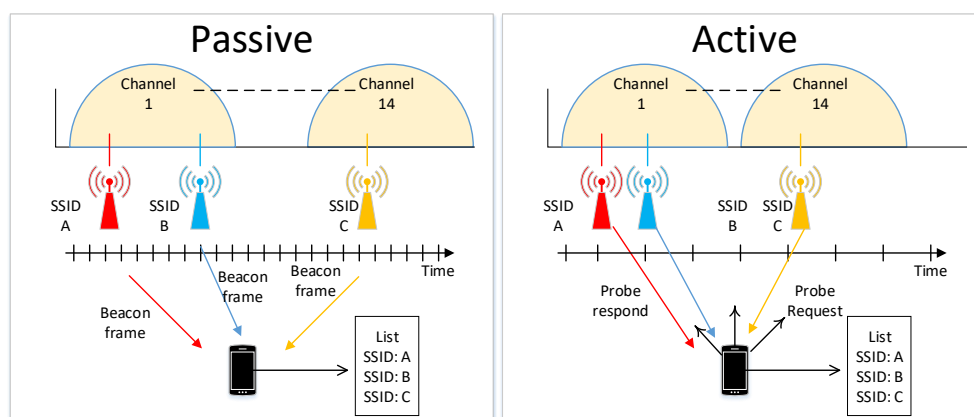
When an access point is configured, it will be programmed to work on one of the radio channels defined in the IEEE 802.11 specifications. As an example, the Wi-Fi band 2,4 GHz consist of 14 channels. When a mobile unit connects to a Wi-Fi network, it will start to search for a Wi-Fi network nearby. This search can be either in passive mode (the mobile is just listening for networks by listening for access points that are sending out network beacons to announce themselves), or in active mode (where the mobile unit is actively sending out a “Probe request” message to all access points). The access points receiving the “Probe request” message will answer with a “Probe response” message to announce themselves for the mobile unit [31].

The mobile unit will either listen in passive mode or send out a “Probe request” on all the channels available in a sequence since the access points are fixed in one of the channels and therefore only listen and transmit on the specific channel. The access points are only transmitting network beacons at regular intervals on their programmed channel, a mobile unit in passive mode must listen for a defined period of time on each of the Wi-Fi channels to discover any network beacons [15]. This may take a long time.



**Figure 2.3:** Wi-Fi passive search

For the active search in the mobile unit, the mobile unit will shift to a channel and send a “Probe request”, and all access points on that channel will immediately respond so the mobile unit will rapidly change to next channel to send out a “Probe request” and so on. When the mobile unit has gone through all the channels it will have a list of all available access points it can be connected to and present this to the user who will choose one of the networks [15]. Figure 2.4 illustrates the difference between the passive and active mode.



**Figure 2.4:** The difference between passive and active Wi-Fi search

### 2.2.1 Probe request and probe response

The signaling protocol between the mobile unit and the access point is based on a handshake system where the units are sending requests and responses to affiliate themselves with each other. The mobile unit start with sending a “Probe request” and the access point will respond with “Probe response” and the mobile unit together with the access point will initiate an affiliation process where they exchange authentication and security parameters to mutually accept each other if they are allowed to do so, all before any data transmission can start [15, 31]. Figure 2.5 illustrates the initial phase of communication.

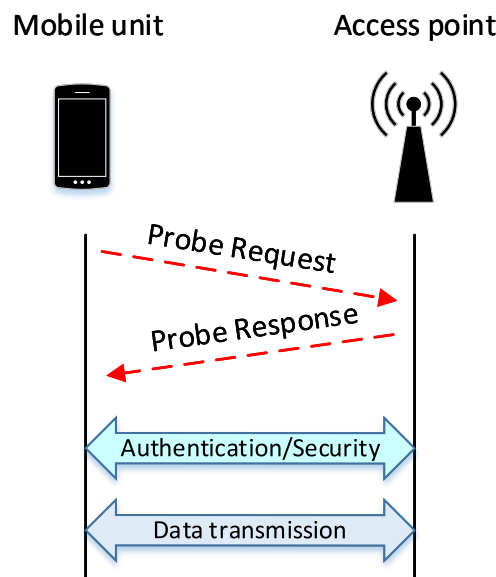


Figure 2.5: Initial phase of communication

## 2.3 Related work

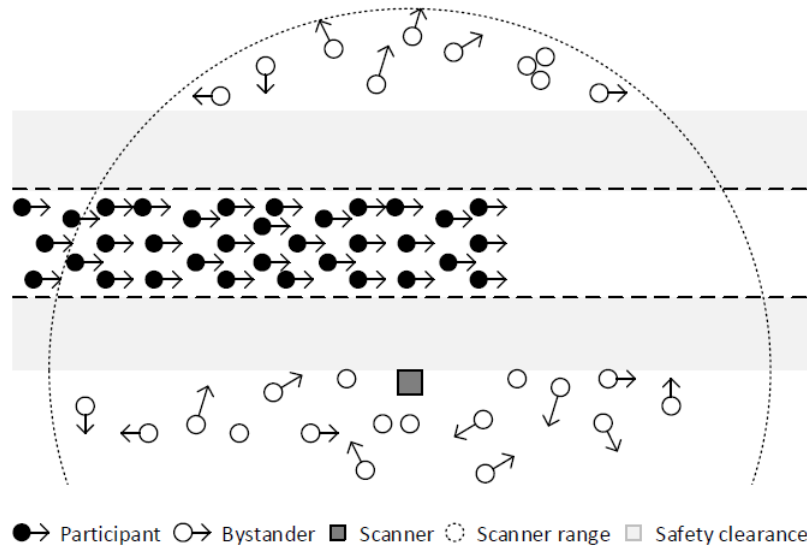
One of the most important questions regarding this master’s thesis is whether the information collected actually can be useful for geolocation in a situation awareness setting. By looking at past research done by others we can see how such information has been used for similar experiments. The collected information can be used for many different purposes, but there are some practices that are more often discussed than others. The following research can be categorized into two main categories:

- Crowd monitoring
- Tracking individuals

### 2.3.1 Crowd monitoring research

Many people have looked at crowd monitoring, both outdoors as in the work by Groba [3] and Chilipirea et al. [4] but also indoors as in the work by Schauer et al. [32]. Generally, the papers describe crowd monitoring by tracking individuals and groups that enter a scanner's range and approximates how far away from the scanner it moves before the crowd leaves the scanner's range again. By using several scanners placed in the path of the crowd, the crowd flow can be tracked.

The paper written by Groba [3] focuses on a moving crowd as part of a demonstration. Groba placed the scanners alongside the route where the demonstration would pass. When the participants were passing by Groba where able to track them.



*Illustration by Groba [3]*

**Figure 2.6:** The scenario presented and used by Groba

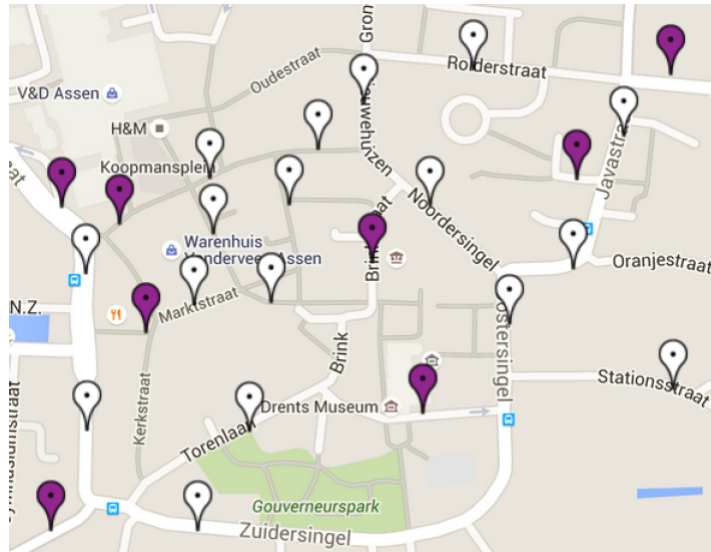
Groba [3] based the tracking of the crowd on both the Distance-based filtering based on RSSI measurement and Time-based filtering. The experiment could not pinpoint the location of a unit accurately, but could indicate approximately when it comes into the range of a scanner and approximately how far away it is to the scanner and when it exit the scanner's range. Groba [3] placed several scanners on the path of the moving crowd and by correlating the distance based filtering and the Time based filtering they could track the crowd movement through the demonstration.

Groba's experiments show that different phones signaling behavior varies significantly and because of this they where not able to find a common threshold, making it impractical to use the distance filtering based on RSSI. Another observation Groba identified was that the number of collected Wi-Fi probe requests



represented only a small fraction of the actual participants after the filtering [3].

Chilipirea et al. [4] based the tracking methods on the same techniques as Groba [3], with both the Distance-based filtering based on RSSI measurement and Time-based filtering, but Chilipirea et al. did test the tracking system in a slightly different scenario. The researchers gathered their data set during a 3-day festival in a Dutch city. Around 130 000 people attended the festival which had multiple stages around the city. Chilipirea et al. used 27 Wi-Fi scanners that they had scattered around to gather the data [4].



*Illustration by Chilipirea et al. [4]*

**Figure 2.7:** Sensor locations used by Chilipirea et al.

They emphasized problems with the data set and how to clean it. In particular, they describe a lot of noise and error in the data, and in order to clean the data they used three main techniques. The first technique uses RSSI values, where the goal was to remove data with low RSSI values. This indicates low quality and that the unit is far away from the scanner. The second technique used time frames. When a data packet is detected several times at one or several sensors within a certain time interval, the detection with the strongest RSSI is kept, the rest are removed. The last cleaning technique uses cycles, and removed detections that indicates repetitions of data set [4].

The work by Schauer et al. [32] based their research on tracking a crowd at a major airport in Germany. They used ground truth information collected at the security check provided by the German airport. They discussed the difference in pedestrian flow estimations, with respect to feasibility and quality, when looking at data collected both using Wi-Fi and Bluetooth [32].

Schauer et al. [32] conclude that one can get approximations about the movements of a crowd using both Bluetooth and Wi-Fi. They also explain that there is difference between the use of Bluetooth and Wi-Fi with respect to accuracy. The

estimations created using the data collected using Bluetooth are less accurate, with a moderate and average correlation to the collected ground truth. The estimations created using the data collected using Wi-Fi gives a better approximation when looking at crowd movement [32].



*Illustration by Schauer et al. [32]*

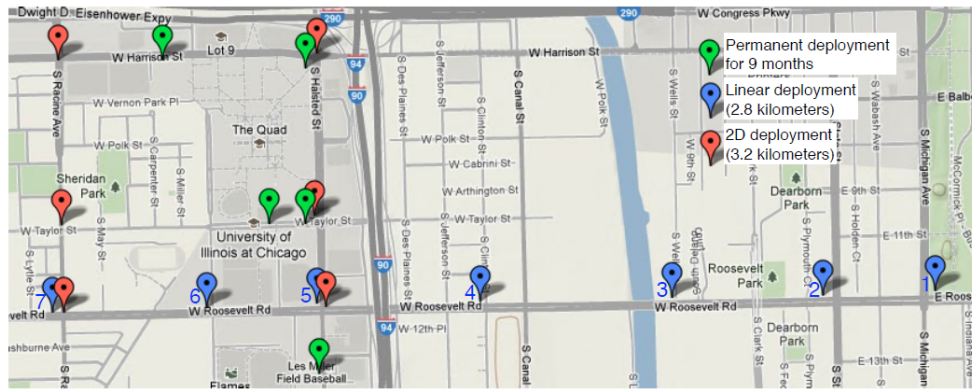
**Figure 2.8:** Sensor locations used by Schauer et al.

### 2.3.2 Individual tracking research

Individual tracking research is how the information collected from the access points or scanners can be used to locate and track an individual. There seems to be less research in this area, which may be because the scope is smaller. Locating an individual or, in most cases, a device will not be of much use to anyone except possibly by the law enforcement or individuals trying to monitor/pursue other individuals. Such invasions of privacy of individual would clearly be a misuse of this type of research. Three research reports that look into tracking of individual based on Wi-Fi information that have slightly different methods of locating devices are in the work by Musa et al. [5], Cuncu [6] and Meng-Hsuan et al. [7].

The research by Musa et al. [5], describes a system they have developed for passively tracking a device. They focus on collecting mobile detections captured by the access points, processing this data and provides a phone's possible trajectory. All registered packets from a specific mobile are used to indicate which direction the mobiles were moving in. Once they had enough registrations from enough sensors, they could indicate which route the mobiles were most likely to have used to move from one point to another [5].

Musa et al. had scattered many sensors around on their test site. They placed a few sensors slightly randomly, while the remaining sensors were placed in a fixed pattern as illustrated in Figure 2.9. As the mobiles move past one of the



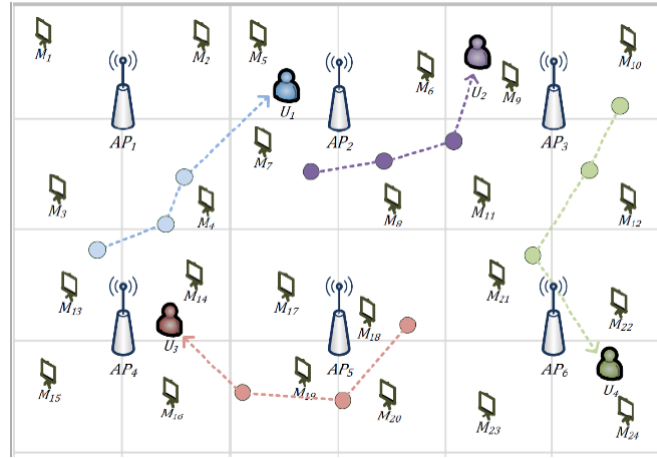
*Illustration by Musa et al. [5]*

**Figure 2.9:** Sensor locations used by Musa et al.

sensors, they are registered as close to this sensor. As the mobile moves on, it will be detected by the next sensor. These records were sent to a server where they were processed. After the mobile has been registered with several sensors, it became clearer which sensors the mobile had been near, and one could track the mobile [5].

The research by Cunche [6] was to find the MAC address to associate an individual to a device and then track a targeted individual. They presented two methods to do this based on Wi-Fi monitoring techniques. Cunche [6] concluded that as long as the individual has a wireless device and if one has enough computer resources that he or she is willing to use, they can identify the individual connected to the device. This means that it is possible to pursue a device and connect it to an individual, which allows one to pursue a specific individual (assuming the individual has the device and it has not been lent to anyone else) [6].

The research by Meng-Hsuan et al. [7] focuses on how the knowledge of how the MAC address to connect an individual to a device can be used by the law enforcement to pursue a suspected person. They present a method by which the wireless footprints of a suspect can be predicted using a passive location tracking system. This system uses passive wireless nodes and relies on these nodes to eavesdrop on the suspect's device. This means that the nodes eavesdrop on the packet sent from this device to wireless base stations [7]. The signal strength is then used to calculate the distance from the base station to the device and using multiple results from several different base stations will allow the device to be located. If they compare the location of the device at different times, they will be able to find the path of movement for this device [7].



*Illustration by Meng-Hsuan et al. [7]*

**Figure 2.10:** Tracking concept presented and used by Meng-Hsuan et al.

The different tracking systems with use of rogue access points or scanners and the type of data gathered such as RSSI and timestamps is well known and there are several papers on this topic, but most of the papers focus on tracking crowd movements or groups of people to identify or assist in handling different situations that may occur. There is a smaller number of papers that look at the tracking of individuals as presented in this chapter. This master's thesis will focus on tracking individuals to see if the information gathered by rogue access points can be used to provide situational awareness.

While most of the published papers are focused on the actual data collection of unknown mobiles moving randomly in general public area, this master's thesis will focus on data collected by a project partner in a controlled environment with known mobiles and with a defined set of movement based on scenarios to verify different aspects of geolocalization. Other papers are mostly focusing on data collection techniques and the results and not so much on the process from the data collection to the result. This paper describes the whole process including data collection, data preprocessing, data cleaning, data analyzing, geolocation calculation, analyzing and finally presenting the results.

Most papers focus on one geolocalization technique while this paper will explore different geolocalization techniques, compare these, and see how these can complement each other (e.g. triangulation versus multilaterations) to see how the methods works compared to the actual physical location of mobiles verified by a video surveillance.

## Chapter 3

# Methodology

This master's thesis methodology is made up of three main parts.

- Data collection
- Data cleaning and preprocessing
- Data analysis based on geolocation methods

Each of these sections can be performed in several different ways and this chapter presents some of these possibilities.

### 3.1 Possible data collection methods

The data could be collected by Bluetooth, Wi-Fi or Mobile phone network and correlated to get the best possible accuracy. This is often not possible or practical with respect to privacy, legal issues or other technicalities. The most flexible solution is to use a Wi-Fi based tracking system. The data can be collected by fixed or portable rogue access points or sensors.

The tracing sensors can operate as stand alone units with very little accuracy as they will only indicate what is within the sensors' field of reception. Another option could be a network of sensors that can work together to triangulate, pinpoint and track a unit. The tracking system for sensors can be used both outdoors and indoors but each of the scenarios will have their own challenges and opportunities that must be evaluated before implementation. Outdoor collections may have many other units and access points that are sending and disturbing the overall network system of sensors, or there may be multipath issues with nearby buildings, vehicles moving around, and people moving freely. On the other hand indoor tracking systems may be in a more controlled environment and the persons will move around within a confined space.

Technically, the data can be collected and stored in each access point and extracted later on for analysis and evaluations (e.g. single or multiple access points can be placed in a restricted area over time to see who has been within a location in a period of time). The data can later be correlated and used if there is a situation to evaluate. There can be several access points that are connected together

through a server to give data in real time for instant analysis and evaluation (e.g. for situations that need immediate reactions to something that may occur). There can also be a combination of these setups. If an access point is placed at a location just for collecting data for later evaluation, but also has an internal trigger function either on the received signal level raising above a certain threshold or certain whitelisted/blacklisted MAC addresses, upon triggering the access point may send an alarm and all the data to a server for instant evaluation.

The data collection for this master's thesis was collected by a project partner and the data was cleaned and anonymized before receiving the data. The data collection was performed over approximately a 3 hours time span in a controlled area consisting of a corridor and a room within a building. Six sensors were used during data collection. Two of the sensors were located on each side of the corridor and the last four were located approximately in each corner of the room. The data collected by the sensors were forwarded and stored at a central server and the data was extracted later on for analysis and evaluations. The layout of the collection location, the equipment used and placement of sensors is described in detail in Chapter 4 Experiment setup.

6 different mobiles of different models and from different vendors were used during the data collection. The data from all mobile phones are broadcast data packets called "probe requests" to search for access points at different times and at different intervals. During the data collection it varied whether the phones were in use e.g. streaming a video, or if the phones are not in use. This was determined based on specified scenarios.

The data collection was based on specified scenarios that were designed to collect data to test the property of data collected in different settings. The different scenarios were set up to see how the captured data could vary based on different situations e.g. if an individual is standing still, moving in different patterns or moving between different rooms. Each scenario described how the individual should move, so that the data collected could be used to create a ground truth and to compare other collected data against ground truth (in addition to using a camera to compare the data collected with the physical location for the individual following the pattern). The scenarios were divided into preparatory scenarios and geolocation scenarios. The preparatory scenarios were used to gather information which was further used to interpret the results of the scenarios tested against the different geolocation methods. All scenarios are described in detail with illustrations on how the person moved in the Experimental setup chapter (see Section 4.5).

### **3.2 Possible data cleaning methods**

Similar research as those described in Section 2.3 shows us that the information gathered can have many uses. Different users have varying requirements for the accuracy of the geolocation results. Some may only need to know if the units are located in the northern or southern part of a city, while others would prefer to

have the exact location. No matter what the goal is, they need the data set as clean as they can get it. They need that the data set contains as little noise as possible. They also do not want data that cannot be used or that can produce misleading results.

It is difficult to know exactly which data cleaning methods to use before looking at the data set. In this section, several possible data cleaning methods will be presented. Some of the methods are very general that can be used on most data sets, while others will only be used in specific situations (for example, some of the methods used in the studies by Groba [3] and Chilipirea et al. [4]).

Note, not all the cleaning methods were used but the cleaning methods used in relation to the data set in this masters thesis and how they changed the data set is elaborated and can be found in Section 5.2. As mentioned earlier the data set was collected by the project partner and it appear as some basic cleaning such as removing duplicates, removing data packets with structural errors and removing data packets with missing data has been performed before it was handed over to the project.

### 3.2.1 Unwanted observations

Unwanted observations can be one of the easiest to clean up and cleaning/removing them is one of the first steps in data cleaning/preprocessing. Unwanted observations often involve two groups of observations: Duplicate observations and Irrelevant observations. [33]

#### Duplicate observations

Duplicate observations are several identical observations. This type of observation usually comes in the data collection phase of a project, especially if the data is a combination of multiple data sets [33]

In the paper written by Chilipirea et al. [4], they used a technique they called "Simultaneous detection". Simultaneous detection was just one technique of a data cleaning method they called "Basic data cleaning". This method consists of three techniques, each removing a type of unwanted observations from the data set [4]. The "Simultaneous detection" technique merged data that was recorded several times at the same location at the same time. If the same registration occurs on several different receivers at the same time, the data will only be retained on the receiver with the highest RSSI score [4].

#### Irrelevant observations

Irrelevant observations can cover many different types of observations, but can be defined as data that are not needed to solve a specific problem [33]. Irrelevant data can also be data collected outside the timeframe of interest and therefore may be discarded.

The other two techniques in Chilipirea et al. [4] basic data cleaning, removed irrelevant observations and were called "misconfigured scanners" and "unknown manufactures" [4]. The technique of "misconfigured scanners" removed all recorded detections that were captured during a period when not all receivers were working properly [4]. The second technique, "unknown manufacturers", removes all data that fails to match the Organizationally unique identifier (OUI) value of a network-enabled device manufacturer. Chilipirea et al. compared the OUI with a publicly available manufacture list [4]. There were few cases where they failed to match the OUI value, so it did not remove much data from their data set [4].

The removal of irrelevant observations became important in this master's thesis and was mainly used to remove irrelevant probe types (see Section 5.2.1) and in combination with a type of Time-based filtering (see Section 3.2.6) to remove data outside the time frame of interest (see Section 5.2.2).

### **3.2.2 Structural Errors**

Structural errors are often small things that in some cases can have quite serious consequences, such as inconsistent capitalization, typos or mislabeled classes. These are often errors that can occur as a result of error in e.g. data transfer or during data collection. [33]

### **3.2.3 Unwanted Outliers**

In a data set it can occasionally be found that there are some values that differ greatly from the rest of the data and in some cases these outliers may be unwanted (e.g. measurements that theoretically should not be possible). Although these values are very different from the rest of the data set, this is not in itself a reason to remove them. Even if one has a value that is much larger or smaller than the remaining values, this can be very valuable to the results. Since these values can prove to be very valuable, an outlier should never be removed without a very good reason [33].

This method was used by Chilipirea et al. [4] which they called "Weak detection removal". The "weak detection removal" was one of two techniques in a data cleaning method they called "advanced data cleaning", that together try to smooth the way a device moves through a city. The "weak detection removal" technique removes all data associated with detection where the RSSI value is low. [4]

The removal of Unwanted Outliers were used in this master's thesis in combination with the Distance based filtering method (See Section 3.2.5) and removed data based on signal strength (see Section 5.2.3). The removed values are values that where theoretically not possible based on the size of the room used for data collection and the distance to the sensor.



### 3.2.4 Missing Data

Missing data in a data set is something that cannot be ignored and must be handled. Knowing how to handle missing data can be difficult, but two of the most common methods are either dropping data or inputting data [33].

Dropping data means that if there is a row where one of the values is missing, the entire row will be removed from the data set. In many cases, it will be a bad idea to remove the row since missing values can be important information in itself [33]. Inputting data means that if there is a row where one of the values is missing, the missing value is filled in with some value. This method can also cause one to lose information for the same reason as choosing to drop the data [33].

Neither of the methods are without drawbacks, but appear to be the ones recommended when it comes to handling Missing Data [33].

### 3.2.5 Distance based filtering

The paper by Groba [3] focused on crowd monitoring in connection with an outdoor demonstration in a major city, noting that the receivers not only collected data from those walking in the demonstration but also random passersby and people in cars or public transportation. The researchers wanted to try to remove the data from people who were not part of the demonstration, since passerby data could give an incorrect picture of the situation [3]. They came up with two methods for filtering the data, one of this is their Distance based filtering method.

They assumed that signals that were further from the receivers than the majority belonged to pedestrians and that the same was true if the signals was much closer to the receiver than the others [3]. They knew about the demonstration in advance and had set up receivers along the route where the protesters were going which is why they could base the filtration on the distance. The distance based filtering method uses the Received Signal Strength Indicator (RSSI) and the distance between the device and the receiver. Using this method, they could filter out those devices with a strong RSSI, which means that the device was near the receiver, and those with a low RSSI that would say that the device was far away from the receiver[3]. They tested with different RSSI thresholds, to try to get the most accurate result. One of the major drawbacks they discovered was that different mobiles gave different RSSI even when they were just as far away from the receivers. This means that RSSI was very dependent on what kind of model the mobile itself was, including different models from the same provider. As a result, the RSSI filtering did not work in their case, as they could not find a threshold that would work for all devices [3].

The Distance based filtering method were used in this master's thesis in combination with the removal of Unwanted Outliers (See section 3.2.3) and removed data based on signal strength (see Section 5.2.3). The removed values are values that where theoretically not possible based on the size of the room used for data collection and the distance to the sensor. The unwanted outliers were removed based on a form of distance based filtering based on what the expected signal

strength value is when located in the given test area.

### **3.2.6 Time-based filtering**

There are several ways to filter based on time, and this depends on the use of the data. If the data set is based on people and movement, the time filtering may be based on whether a person is standing still for a long time, or whether they are moving faster than what is possible or normal or other similar criterias.

Again we refer to Grobas' [3] demonstration scenario, where they used a time-based filtering method that assumed the participants will only be registered by a receiver for a short period of time as they move past it. Participants will then be registered by the next receiver along the route. This means that devices that are registered over an extended period of time by a receiver can be excluded, since they are most likely not part of the demonstration. They can also exclude devices that are not captured by multiple receivers, since those devices most likely belong to bystanders [3]. One of the problems with this method is to take in account how often a device sends Wi-Fi probes, and the difference between different devices. This may indicate that some devices are not intercepted since they do not send probe requests, or that they are only intercepted by only one receiver and are therefore excluded because they appear as bystanders and not participants [3]. Groba concluded that it was difficult to distinguish bystanders from participants and that a combination of time and distance based filtering together with different monitoring techniques would probably be the best approach.[3]

Other ways to filter based on time may be if there is data in only specific time periods that are of interest. The data collected outside the scope of the defined time period can be filtered out as it is of no value to the project. Time-based filtering was used in this master's thesis when removing data outside the timeframe of interest (see Section 5.2.2).

### **3.2.7 Time compression**

Time compression may be to merge data from identical time stamps or as in the paper by Chilipirea et al. [4], where they merged larger groups of data to get a better illustration of a path. Their paper [4] focused on crowd monitoring during a festival, where they collected data from several Wi-Fi scanners located around the festival area. With their time compression technique they try to identify which receivers were the main receivers for a path [4]. They do this by dividing the data into time intervals and then identifying the receiver with the highest RSSI score for that time period. This technique results in a lower resolution of the data set and a better illustration of a path [4].

This master's thesis used time compression (see Section 5.2.4) when it was necessary to merge data packets from the same mobile during the same second and to the same sensor.

### 3.2.8 Cycle removal

Chilipirea et al. [4] used several different techniques to remove noise from their data set. They had a technique they called "cycle removal" where they remove data where the device moves abnormally in circles, or where the device abnormally moves back and forth between two locations. They included some thresholds for what kind of movements were defined as normal movement. These thresholds were based on the number of detections of the device on the various receivers and any other receivers not included in the circle [4].

## 3.3 Possible location methods

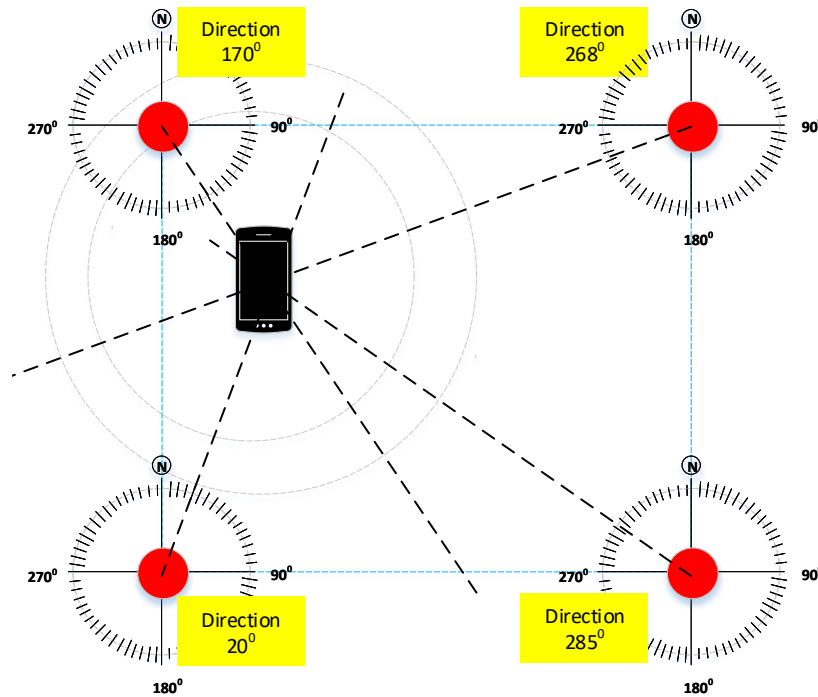
The main goal of this master's thesis is to look at the possibilities for increase situational awareness using geolocation. There are several different methods for locating a device based on Wi-Fi signals sent from the device. This master's thesis will focus on 4 different geolocation methods and whether these methods can be used. These methods are:

- Angle of Arrival
- Time of Flight
- Fingerprinting using Signal Strength
- Triangulation or Multilateration using Signal Strength

### 3.3.1 Angle of Arrival

The Angle of arrival (AoA) based location technique uses antenna arrays, which are a group of antennas that together works as a single antenna [34]. The Antenna array is located on the receiver side of the communication and can be used to determine the angle of the incoming data packet [11]. This is done by comparing the time difference and signal phase difference for when and how the different antenna received the signal. The antenna that received the signal first will be closest to the device transmitting the signal, and the direction of the received signal can be calculated, thus determining on witch side of the receiver the device is located [11]. By correlating this result with other access points it is possible to determine the position of the unit by triangulation. Figure 3.1 illustrates how Angle of Arrival works.

One of the advantages of this method is that it can find the position of the device using only two receivers, but it only applies in a 2D environment otherwise it needs at least three receivers like most other methods. This method gives quite good results as long as the distance between the device and the receiver is small. The accuracy will decrease as the distance between the device and the receiver increases since as the distance increases, small errors in the angle calculation can cause major errors in the location of the device [11].



**Figure 3.1:** Angle of Arrival visualization

Other advantages with the Angle of Arrival method are:

- It is not dependent on signal strength (damped by obstacles etc.)
- It can detect and discard multipath signals
- It can be very accurate

A major disadvantage of AoA is that the method requires more complicated and specialized hardware than most other location methods. Special equipment must be used with respect to finding the direction of radio signals and compared to the other methods, the Angle of Arrival method requires much more careful calibration [11, 35]. All access points must be calibrated to the same reference to a high degree of accuracy.

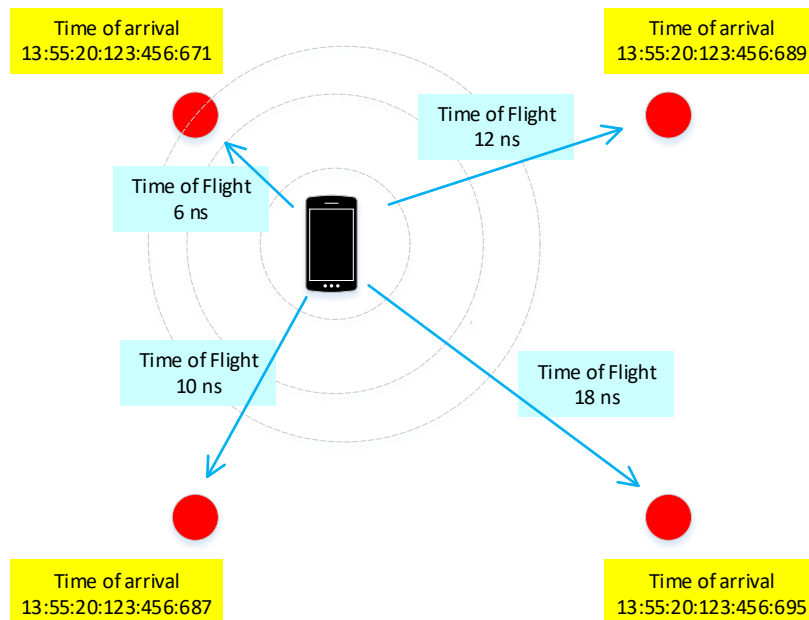
### 3.3.2 Time of Flight

The Time of Flight (ToF) based method, which can also be called Time of Arrival (ToA), calculates the distance between the device and a receiver using the signal propagation time. To calculate the physical distance between the device and the receiver, the speed of light is multiplied by the ToF value [11, 36]. The radio wave propagation speed in free space is equal to the speed of light (proximately 300 000 km/s). Using this information, the estimated distance is illustrated in table 3.1.

Time accuracy	Distance per time unit
Second (s)	300 000 km
Millisecond (ms)	300 000 m
Micro second ( $\mu$ s)	300 m
Nano second (ns)	0,3 m

**Table 3.1:** ToF approximately distance resolution per time unit

As with most other methods, this method will need at least three receivers to locate the device. The ToF method requires a strict synchronization between the devices and the receivers and in many cases it also requires timestamps to be sent with the signal so that they can use this to calculate the distance, but it varies slightly on the underlying communication [11].



**Figure 3.2:** Time of Flight visualization

The ToF will give a more accurate distance, and then by triangulation we can find the location of the unit in relation to the access points. If the exact time is known for when the data packet is sent from the unit, the distance can be calculated directly and used for triangulation. Alternatively, if the time in the sending unit is not known, the time difference (delta) between the reception from several access points can be used for the triangulation calculation.

Advantage of using ToF:

- It is not dependent on signal strength
- It can detect and discard multipath signals

- It can be very accurate

As with all other methods, the ToF localization method has some disadvantages. The first is that this method relies heavily on clock synchronization and can be severely affected by clock synchronization issues. Some attempts have been made to remove the need for clock synchronization as by Youssef et al. [37] using mathematical approaches, so it is possible that clock synchronization may become less of a problem in the future. The ToF method may also struggle with noise in the data that may result in incorrect results or problems due to multipath channel effect and data collection errors [11, 38]. Other disadvantages of the Time of Flight method are:

- All equipment must be synchronized down to nano second accuracy (Applies to an accuracy of 0.3m)
- Must use special equipment with respect to access points and clock device
- Must use atomic clock accuracy down to nano second on each access point (Applies to an accuracy of 0.3m)
- High cost equipment
- Not practical in most cases

### 3.3.3 Fingerprinting using Signal strength

The fingerprinting localization technique is a method that creates a ground truth. This ground truth is called a fingerprint which all new measured data can be compared to.

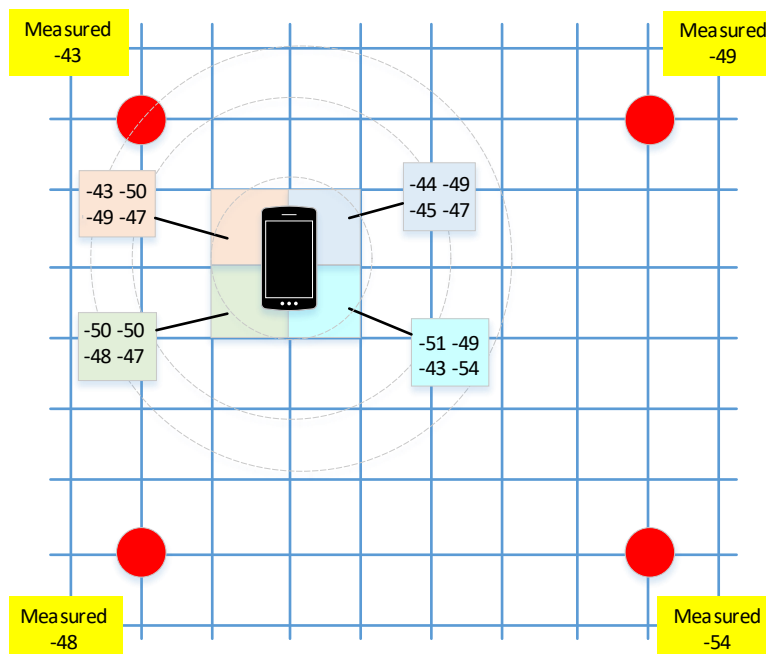


Figure 3.3: Fingerprinting visualization

The method is based on the data collection on different locations where the signal strength from a transmitting unit is measured and stored together with the location information, as illustrated in figure 3.3. The fingerprint is created for the area of interest and then allows one to compare later data collected with the ground truth to get an indication of physical location. This method requires more preparation, since the ground truth must be prepared in advance [11].

The fingerprints are usually signal strength measurements (RSSI or dBm) that are stored in a database. When the system is deployed, real-time measurements will be compared to the stored fingerprints. The result of this comparison will be able to estimate the location of the device [11]. When a sufficiently good fingerprint is obtained, there are several different algorithms that can be used to compare the real-time data with the fingerprints to find a match. Training and comparison is performed by standard machine learning classifiers, which include algorithms such as k-Nearest Neighbors (kNN), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) [11].

Fingerprinting using the signal strength method has a great advantage:

- It is easy to use and a program can compare the measured received values from the sensors with a table and select the values closest to the measured values and read out the location area.

The method has, like all other localization methods, some disadvantages. The biggest disadvantage is that fingerprints can easily change. If something changes in the environment, for example larger objects that are added or removed, this will affect the established fingerprint and a new fingerprinting should be performed. This is particularly true if collecting data indoors, since as little as moving or adding furniture can change the fingerprint [11, 36]. Other disadvantages of using Fingerprinting are:

- It can be time consuming to create the fingerprint. The gathering of Fingerprint data requires the phone (or phones) to be placed at each location for a given time period to collect enough data to establish a statistical and valid interpretation of the reception from that location.
- If the Fingerprinting is valid for different types of mobile brands and types, the fingerprinting must be done for each type.
- The different mobile units (even the same brand and type) can have individual variations of transmitting power that can affect the location estimations.

### 3.3.4 Triangulation or Multilateration using Signal strength

The signal strength method is based on the measured signal strength at the receivers and using Triangulation/Multilateration to calculate the location of the mobile. We refer to this category of geolocation as Triangulation or Multilateration using Signal Strength. The transmitting unit (mobile phones) will transmit data packets with different content and length at different intervals, and this will

be received and measured at the rogue access points. The rogue access point gives a value that indicates the measured signal strength either in RSSI or in dBm depending on the access point vendor's preference. The higher the signal strength, the closer the device is to the receiver. To be able to accurately locate a device based on signal strength, one needs at least three receivers [11].

There are two main methods for calculating the location of a transmitting unit, Triangulation and Multilateration. There is some basic input that is needed to perform the Triangulation or Multilateration calculation such as the distance from the transmitter to the receiver, or for triangulation this could also be the angle of arrival.

The distance from the transmitter to the receivers is most commonly used and can be calculated by use of timing (Time of Flight described in Section 3.3.1) or by measuring the signal strength as discussed in this section. This distance is used by both the Triangulation and Multilateration methods.

Angle of Arrival, as described in Section 3.3.1, measures the angle from where the signal arrives to the sensor. This technique can be used in the Triangulation method. The Angle of Arrival will find where the measured angles from different receivers intersect each other. The transmitter is located at this intersection and thus forming a triangle between the devices (the receivers and the transmitter). The calculation of the location of the transmitter in a coordinate system can be performed by simple trigonometry.

### Calculating the distance using signal strength

Calculating the distance from the mobile to the sensor is based on a calculation of how much a transmitted radio signal is reduced in power over a distance through air. The known parameters that are used to find the distance is the signal strength out from the mobile phone (PT), and the received signal strength (PR) measured at the sensor. When the received signal strength is subtracted from the output signal strength from the mobile, the result will be the signal reduction due to the distance the signal has traveled through air. The calculation to find the signal reduction from the mobile to the sensor in air:

$$FSPL_M = PT - PR$$

- $FSPL_M$  = Signal reduction in air by measurement (Free Space Path Loss Measured)
- PT = Signal strength out from mobile phone (P is for Power and T is for Transmitter)
- PR = Signal strength measured at sensor (P is for Power and R is for Receiver)

The calculated signal loss ( $FSPL_M$ ) value is used to find the distance between the mobile and the sensor. An equation for calculating the reduction in signal strength over a distance is called the Free Space Path Loss transmission formula that was presented by Friis [39] (Friis transmission equation). Friis based the



equation on the ratio between the received power ( $P_r$ ) and the transmitted power ( $P_t$ ). The relation between these is based on elements between the transmitter and receiver. The elements are the antennas and the antennas characteristics for receiving and sending signals (Aperture) which receives antennas Aperture ( $A_r$ ) and the transmitting antennas Aperture ( $A_t$ ), the distance ( $d$ ) and the signals wave length ( $\lambda$ ).

$$\frac{P_r}{P_t} = \left( \frac{A_r A_t}{d^2 \lambda^2} \right)$$

The Friis transmission equation is used to derive a free-space path loss formula where the Aperture was converted to the antennas directivity and the spread of radio waves in a sphere. In an omnidirectional antenna without directivity, the directivity element is equal to 1 and can be disregarded. The free-space loss is normally defined in logarithmic scale (dBm) and the formula for free-space loss is:

$$FSPL(dB) = 10 \log_{10} \left( \left( \frac{4\pi d f}{c} \right)^2 \right)$$

- $d$  = distance between the transmitter and the receiver
- $f$  = Frequency for the radio transmission (note  $\lambda = c/f$  )
- $c$  = Speed of light in free space
- $\log_{10}$  = Logarithmic calculation in base 10

Using this equation with the Free Space Path Loss calculated earlier ( $FSPL_M$ ) it is possible to find the distance if we use a known frequency. When the  $FSPL_M$  is used in the Free Space Formula from Friis [39] and solved based on the distance, the formula appears as follows:

$$d = 10^{((FSPL_M - 20 \log(f) - 20 \log(4\pi/c))/20)}$$

- $d$  = Distance in meters
- $f$  = frequency in MHz
- $c$  = Speed of light
- $FSPL_M$  = Signal reduction in air by measurement

As an example:

If we assume the phone is transmitting with  $PT = 20$  dBm (100 mW) and the receiver is measuring the incoming signal to be  $PR = -40$  dBm (0,0001 mW), then the signal reduction will be:  $FSPL_M = 20 - (-40) = 60$  dBm reduction. This reduction is used in the distance formula together with the value for frequency  $f = 4224$  MHz. The distance will be:

$$d = 10^{((FSPL_M - 20 \log(f) - 20 \log(4\pi/c))/20)} = 10^{((60 - 20 \log(4224) - (-27,55))/20)} = 5,7m$$

This value will be used in the Triangulation and Multilateration equation for Geolocation. Note, the RSSI/dBm geolocation method may be subject to interference and signal variations that can affect the accuracy of localization that must be considered when utilizing the localizations.

The advantage of using Friis transmission equation:

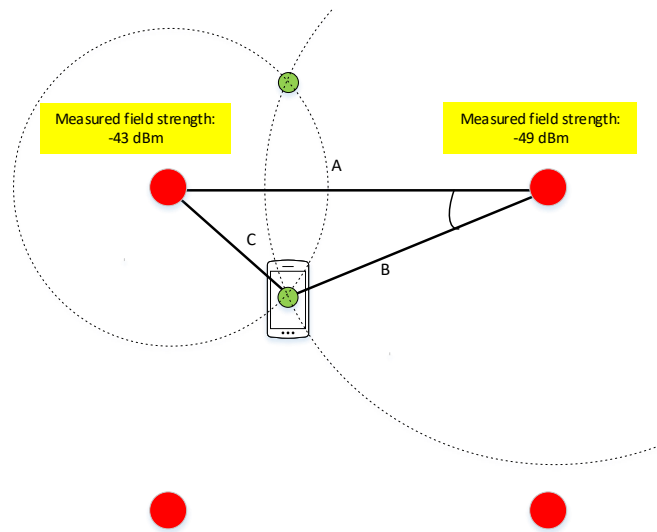
- It is easy to calculate if one know the received signal strength and basic parameters of Access point and transmitting unit.

Disadvantages of using Friis transmission equation:

- The transmitting unit may be of different mobile brand and types that may have different transmission characteristics and this may be unknown.
- Each individual phone can (also) have individual deviation to the transmission characteristics that may affect the result of the calculation.

### Triangulation

Triangulation is a way of calculating positions of a transmitter by use of triangles. When a transmitter (mobile phone) is sending a signal that is received by two or more receivers (Access points), the transmitter and receivers form one or more triangles.

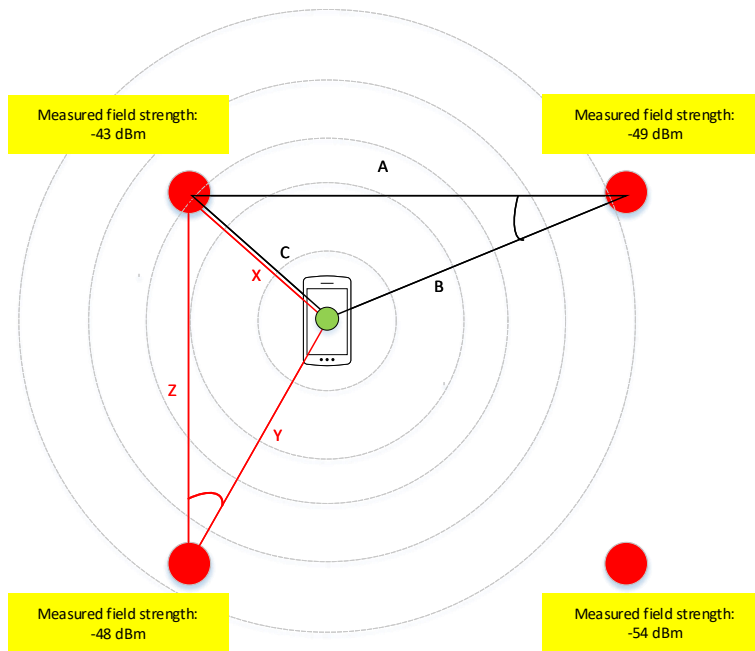


**Figure 3.4:** Triangulation with two receivers

When receiving a data packet from the transmitter, the receivers will calculate the distance based on the signal strength and this will form a circle around the receiver for where the transmitter can be located anywhere. Another receiver will also receive the same signal and calculate the distance and form a circle for where the transmitter can be located. The intersection of these circles will be the location

of the transmitter and form the triangle with the receivers as illustrated in Figure 3.4.

If the circles are overlapping, then there will be two intersections (green dots in the figure) and the transmitter could be in either of these places. To find out which of these intersections is correct, it can be decided by either excluding one of the intersections since this is outside the area of interest (e.g. outside the test area) or by adding the information from a third receiver and the intersection of that circle as well.



**Figure 3.5:** Triangulation with three receivers

By use of trigonometry, all the angles in the triangle can be calculated, and by using this information, the transmitting unit can be pin-pointed in a coordinate system that could be an X/Y coordinate system or converted to a Longitude and Latitude coordinate system.

*Example of triangle calculation by Trigonometry:*

The triangulation is based on calculating the distance to a mobile unit from two or more fixed locations (i.e. where sensors are located). An example is a mobile sending a data packet that is received at sensors 1 and 2 as illustrated in Figure 3.6. We define an X/Y coordinate system where the line between the two sensors is the X axis and we refer to the X coordinate of the mobile phone simply as X. A line perpendicular to the X axis acts as our Y axis, where the distance from the X-axis to the mobile phone is refers to the Y coordinate of the mobile phone.

Sensor 1 receives the signal with -46 dBm converted this gives  $\Rightarrow 8,6$  m

Sensor 2 receives the signal with -42 dBm converted this gives  $\Rightarrow 7,6$  m

This forms a triangle of vertices of A,B and C with the lengths of the edges a,b and c.

- $a = 7,6$  m Distance from transmitter to sensor 2
- $b = 11,7$  m Distance between sensors
- $c = 8,6$  m Distance from transmitter to sensor 1

To be able to calculate the x and y coordinates of the phone, the angle at one of the sensors is calculated.

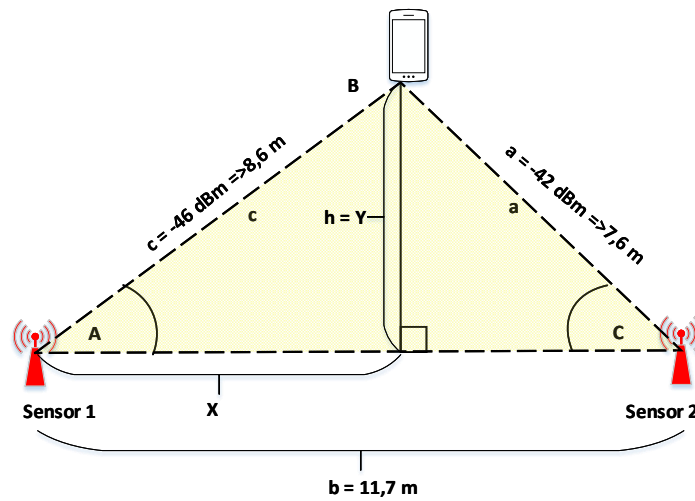


Figure 3.6: The values that must be calculated

To calculate the angle A one uses the Cosine formula:  $a^2 = b^2 + c^2 - 2bc \cos(A)$ . Converting this to find the Cosine value for (A) gives:

$$\cos(A) = (b^2 + c^2 - a^2) / 2bc$$

In this case this is  $(11,7^2 + 8,6^2 - 7,6^2) / 2 * 11,7 * 8,6 = 0,760$  which is 40,5 degrees. Using this Cosine value to calculate the x coordinate of the phone, the trigonometric formula for a right triangle using the Cosine value:

$$\cos(A) = \frac{X}{c} = \frac{\text{adjacent}}{\text{hypotenuse}}$$

and rewrite this as:

$$\cos(A) * c = X$$

In this case this will give:  $X = 0,760 * 8,6 = 6,58$  m

To calculate the Y value, the Pythagorean Theorem is used:  $c^2 = X^2 + Y^2$  and rewrite this to:

$$Y^2 = C^2 - X^2$$

This will give in this example:  $Y = \text{Root}(8,62 - 6,582) = 5,53\text{m}$

The X,Y coordinate value is then:  $X= 6,58\text{ m}$  and  $Y= 5,53\text{ m}$  (this could be converted to a longitudinal and latitudinal coordinate system)

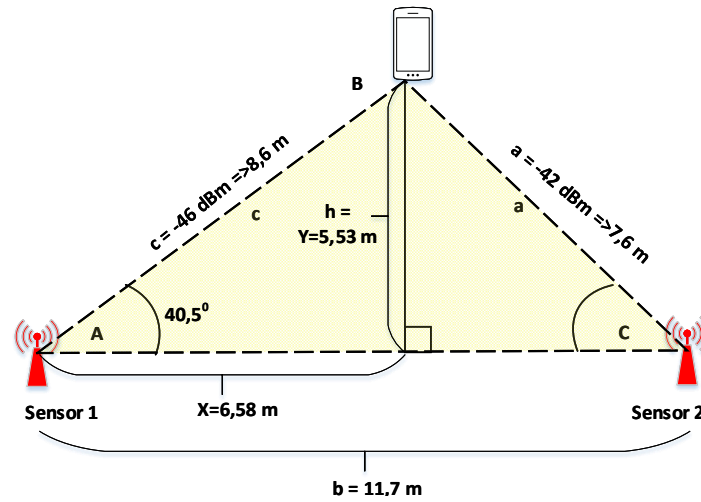


Figure 3.7: The calculated values

When other sensors are in use, this is calculated in the same manner but please note, all results are converted to the same coordinate system i.e. result from sensor 1 – sensor 2 and sensor 1 – sensor 3 and sensor 2 – sensor 4 and sensor 3 – sensor 4 is converted so they use the same X axis and Y axis as illustrated in Figure 3.8.

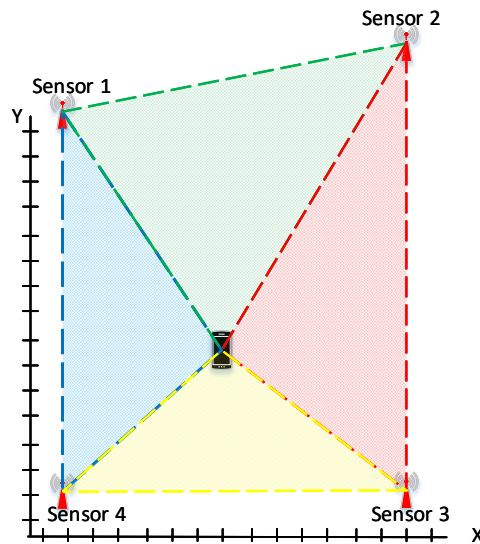


Figure 3.8: Triangulation with multiple receivers

### Trilateration/Multilateration

Trilateration uses measurements from three sensors to calculate the position of a transmitter, while Multilateration uses measurements from more than three sensors to calculate the position of a transmitter.

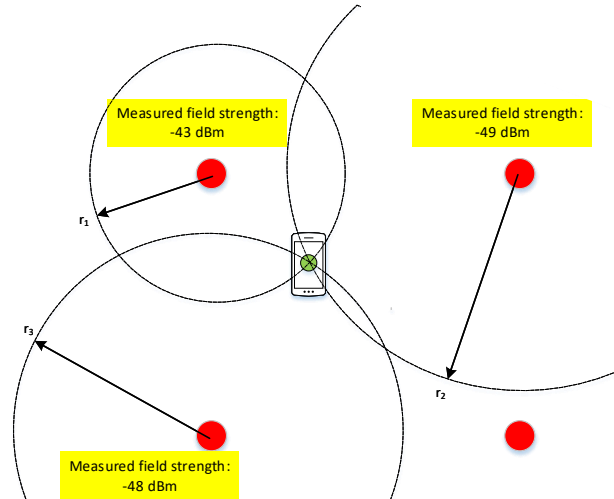


Figure 3.9: Trilateration

Multilateration is the term that will be used in this document as a generic term for Trilateration or Multilateration since it is six sensors in use even though sometimes only three of them is used in that specific calculation.

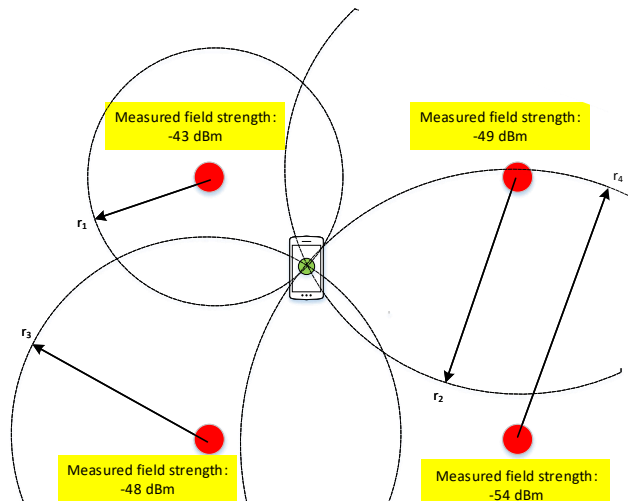
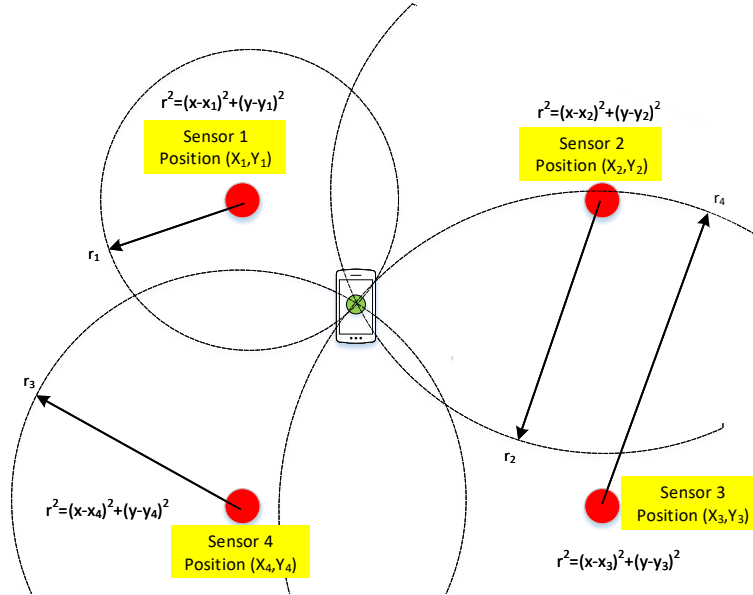


Figure 3.10: Multilateration

Multilateration uses the distance to the transmitter, and the sensor will know that the transmitting unit will be somewhere on a circle with radius equal to the

calculated distance from the sensor, just like the triangulation method is using. The difference is that the calculation is not based on trigonometry like the Triangulating method but uses the circles to calculate the localization using algebra. Multilateration uses the coordinates of the sensors and the radius of the circles to estimate the location of the transmitter. Multilateration calculations can use the information from several sensors in the same calculation.

The intersection of circles around several sensors indicates the location of the mobile. The well-known equation for a circle in an X/Y coordinate system is:  $r^2 = x^2 + y^2$  and the coordinates for x and y will be the location of the mobile. When the coordinates is adjusted for each of the sensors the equation will be:  $r_s^2 = (x - x_s)^2 + (y - y_s)^2$  where “s” is the number of the sensor as illustrated in Figure 3.11.



**Figure 3.11:** Multilateration with individual circle equations

Each of the circles will have their own equation for the estimated distance to the mobile, a radius of their own circle [40]:

Sensor 1:  $r_1^2 = (x - x_1)^2 + (y - y_1)^2$

Sensor 2:  $r_2^2 = (x - x_2)^2 + (y - y_2)^2$

Sensor 3:  $r_3^2 = (x - x_3)^2 + (y - y_3)^2$

Sensor 4:  $r_4^2 = (x - x_4)^2 + (y - y_4)^2$

Since all circles intersect at (x,y) where the mobile is located, the (x,y) value is common for all circles. To identify the location of the mobile, the equations for all circles give us a system of equations needed to solve with respect to x and y. By expanding the squares we get the squared unknowns ( $X^2$  and  $Y^2$ ) which are inconvenient. By subtracting the expressions from pairs of radii, we can eliminate the squared unknowns and the result are:

$$\text{Equation 1: } R_1^2 - R_2^2 = 2(-x_1 + x_2)x + 2(-y_1 + y_2)y + x_1^2 - x_2^2 + y_1^2 - y_2^2$$

$$\text{Equation 2: } R_2^2 - R_3^2 = 2(-x_2 + x_3)x + 2(-y_2 + y_3)y + x_2^2 - x_3^2 + y_2^2 - y_3^2$$

$$\text{Equation 3: } R_3^2 - R_4^2 = 2(-x_3 + x_4)x + 2(-y_3 + y_4)y + x_3^2 - x_4^2 + y_3^2 - y_4^2$$

The new expressions above is linear equations, with all variables but x and y are constants. By rewriting the equations to get x and y on one side of the equal sign the expressions are:

$$\text{Equation 1: } 2(-x_1 + x_2)x + 2(-y_1 + y_2)y = r_1^2 - r_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2$$

$$\text{Equation 2: } 2(-x_2 + x_3)x + 2(-y_2 + y_3)y = r_2^2 - r_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2$$

$$\text{Equation 3: } 2(-x_3 + x_4)x + 2(-y_3 + y_4)y = r_3^2 - r_4^2 - x_3^2 + x_4^2 - y_3^2 + y_4^2$$

By rewriting the equations and substitute the coefficients and the right hand side of the equations with single letters, we can see that we have a system of equations we can solve. The coefficients are substituted as follows:

$$2(-x_1 + x_2) = A$$

$$2(-y_1 + y_2) = B$$

$$R_1^2 - R_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2 = C$$

$$2(-x_2 + x_3) = D$$

$$2(-y_2 + y_3) = E$$

$$R_2^2 - R_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2 = F$$

$$2(-x_3 + x_4) = G$$

$$2(-y_3 + y_4) = H$$

$$R_3^2 - R_4^2 - x_3^2 + x_4^2 - y_3^2 + y_4^2 = I$$

This gives us the following equations:

$$\text{Equation 1: } Ax + By = C$$

$$\text{Equation 2: } Dx + Ey = F$$

$$\text{Equation 3: } Gx + Hy = I$$

When solving this for trilateration (only sensors, 1, 2, and 3) the equation 1 and equation 2 is used and solved with respect to x and y gives the trilateration solution for the coordinates of the mobile. Solving using linear algebra (only sensors, 1, 2, and 3) gives us:

$$x = \frac{CE - FB}{EA - BD}$$

$$y = \frac{CD - AF}{BD - AE}$$

When substituting the letters with the actual expression again gives the following solution:



$$x = \frac{((r_1^2 - r_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2)(2(-y_2 + y_3)) - ((r_2^2 - r_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2)(2(-y_1 + y_2)))}{(2(-y_2 + y_3))(2(-x_1 + x_2)) - (2(-y_1 + y_2))(2(-x_2 + x_3))}$$

$$y = \frac{((r_1^2 - r_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2)(2(-x_2 + x_3)) - ((r_2^2 - r_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2))(2(-x_1 + x_2)))}{(2(-y_1 + y_2))(2(-x_2 + x_3)) - (2(-x_1 + x_2))(2(-y_2 + y_3))}$$

When solving this for multilateration (all four sensors) using the equation 1, equation 2 and equation 3 and solved with respect to x and y gives the multilateration solution for the coordinates of the mobile. Solving using linear algebra again gives us:

$$x = \frac{C}{A} - \frac{B(GF - DI)}{A(GE - DH)}$$

$$y = \frac{C}{B} - \frac{A(HF - EI)}{B(DH - EG)}$$

When solving this with respect to all four sensors the equation will be:

$$x = \frac{r_1^2 - r_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2}{2(-x_1 + x_2)} - \frac{2(-y_1 + y_2)((-x_3 + x_4)(r_2^2 - r_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2) - 2(-x_2 + x_3)(r_3^2 - r_4^2 - x_3^2 + x_4^2 - y_3^2 + y_4^2))}{2(-x_1 + x_2)(2(-x_3 + x_4)2(-y_2 + y_3) - 2(-x_2 + x_3)2(-y_3 + y_4))}$$

$$y = \frac{r_1^2 - r_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2}{2(-y_1 + y_2)} - \frac{2(-x_1 + x_2)(2(-y_3 + y_4)(r_2^2 - r_3^2 - x_2^2 + x_3^2 - y_2^2 + y_3^2) - 2(-y_2 + y_3)(r_3^2 - r_4^2 - x_3^2 + x_4^2 - y_3^2 + y_4^2))}{2(-y_1 + y_2)(2(-x_2 + x_3)2(-y_3 + y_4) - 2(-y_2 + y_3)2(-x_3 + x_4))}$$

If we are using, in this example, the parameters in Table 3.2 for an ideal measurement of the positions and radius of the sensors in a X/Y coordinate system.

Sensor	Location for x(m)	Location for y(m)	Measured dbm	Calculated distance "r"
Sensor 1	0	12,5	-31,59 dbm	2,3m
Sensor 2	11,7	15	-41,39 dbm	9,7m
Sensor 3	0	0	-51,69 dbm	16,48m
Sensor 4	11,7	0	-46,12 dbm	13,59m

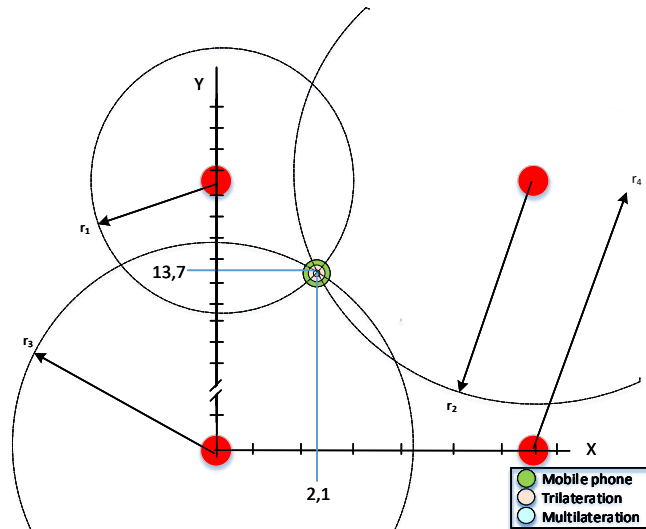
**Table 3.2:** An ideal measurement

Substituting the variables the variables of our final x and y equations above with the entries from Table 3.2 will give the location of the mobile, and since this is an ideal case, all the circles intersect at one single point. Table 3.3 shows the results from the calculation with Trilateration and Multilateration.

	X value	Y value	Comments
Sensors (1-2-3)	2,1	13,7	Trilateration
Sensors (2-3-4)	2,1	13,7	Trilateration
Sensors (3-4-1)	2,1	13,7	Trilateration
Sensors (4-1-2)	2,1	13,7	Trilateration
Sensors (1-2-3-4)	2,1	13,7	Multilateration

**Table 3.3:** Calculated X and Y values for an ideal measurement

This is then correct according to the circles intersecting at the same location as illustrated in 3.12. Using various trilaterations is a good way to verify the result. In real life, there is never an ideal situation and there may be some deviation in the measurement. This will be highlighted in Section 5.7.5.



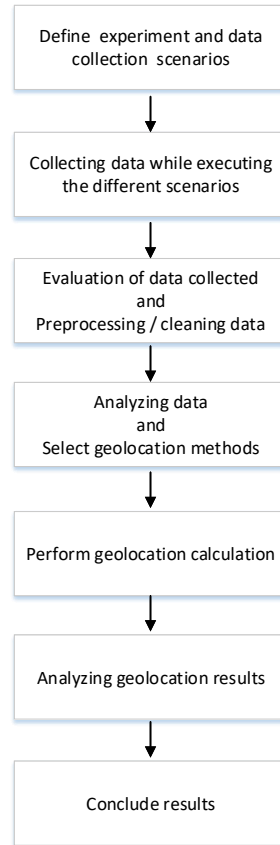
**Figure 3.12:** The circles intersecting at the same location

In this chapter, four of the methods used for localization has been described. All methods considered in this chapter can provide the desired knowledge, which is the location of a device. All methods will require information from at least three receivers to be combined to locate the device. The result from one receiver, will only give indication of how far away the device is located. This indication is usually within in a radius around the receiver. When combined with several sensors and triangulation, the intersection of the distances from each sensor to the device is where the device is located. The methods also have drawbacks that need to be considered. These drawbacks can, to a varying degree, make the methods less useful for this master's thesis. This is especially true of the Angle of Arrival method that requires special hardware that this project does not have access to. The Angle of Arrival method can already be excluded from possible methods since the data for this master's thesis was collected using multiple RaspberryPi devices without

any equipment to indicate the direction, but it is nonetheless valuable to note the existence of the method.

### 3.4 Methodology flowchart

This Master's Thesis is following the sequence illustrated in Figure 3.13.



**Figure 3.13:** Methodology flowchart

First phase defining the experiment setup and scenarios for data collection to ensure all aspects of required information is handled. Second phase is to perform the data collection by executing the different scenarios according to the defined experiment setup and the defined scenarios. Third phase consist of two functions that is closely related, this is to evaluate the collected data with respect to data integrity and relevance to the scenarios and cleaning the data by removing irrelevant data and extract data collected during the different scenarios. Forth phase is to analyze the data with respect to accuracy and usefulness and evaluating various parameters that can influence the selection of geolocation methods. And finally select the geolocation methods to be used on the data collected. Fifth phase is

performing the geolocation calculation selected during data analyses. Sixth phase is analyzing the result of the geolocation calculation and compare this with the actual location of the mobiles verified by video surveillance. Seventh and final phase is to conclude the result from the whole experiment and specially the result of the geolocation based on data collected and analyzed.

## Chapter 4

# Experiment setup

The first part of the project was to collect data based on data packets sent from the phones. The data collection for these type of projects can be very time consuming based on the goal of the project, even taking months [5]. In other situations, the actual collection of the data can take a very short time [4, 41, 42]. It depends on the requirement for the data, if the data should contain different types of data packets and the quality and quantity of the data. These factors will influence the accuracy of the geolocation result. Collecting data can therefore be one of the tasks that take the longest time during the project, which can be a challenge if the project has a strict time limit.

This master's thesis is a project that was limited in time but the data collection itself is a small part of the project. The project partner provided the data that was used in the project. The data was gathered by the project partner of the master project over approximately a 3 hours time span within a building. A controlled environment was set up, and the data was collected based on specific scenarios described in chapter 4.5.

The different scenarios were set up to see how the captured data could vary based on different situations e.g. if an individual is standing still, moving in different patterns or moving between different rooms. This gave the opportunity to use more time on analyzing the data and to evaluate different possibilities for geolocation without having to spend too much time on the data collection itself. The scenarios were set up with descriptions of how the individual should move, so that the data collected could be used to create a ground truth and to compare other collected data against ground truth (in addition to using a camera to compare the data collected with the physical location for the individual following the pattern).

One of the goals of the project was to see how the collected data can be used for geolocation and how precise the results can be to locate and predict the movement of the unit. The different scenarios consist of two different movement patterns, a predetermined pattern and a random pattern. The difference between the data from these patterns may indicate how precise the results can be. Are they precise enough to tell the difference between a predetermined pattern and a random

pattern? Is it possible to identify which data set belongs to which pattern when given both data sets? In addition to evaluating the accuracy of the result, the different scenarios were also carried out with different smartphones of different models and from different vendors. This data may indicate any difference between different types of mobile phone models.

Different mobile phones broadcast data packets called "probe requests" to search for access points at different times and at different intervals. This can indicate if there is a difference in the analyzed data due to which mobile phone is used. The different scenarios also vary between whether the phone is in use, e.g. streaming a video, or if the phones are not in use e.g. is in a pocket.

## 4.1 Experiment location and participants

The data collection was conducted indoors in a controlled environment. The experiment area consisted of a large room with a corridor on the side as illustrated in Figure 4.1.

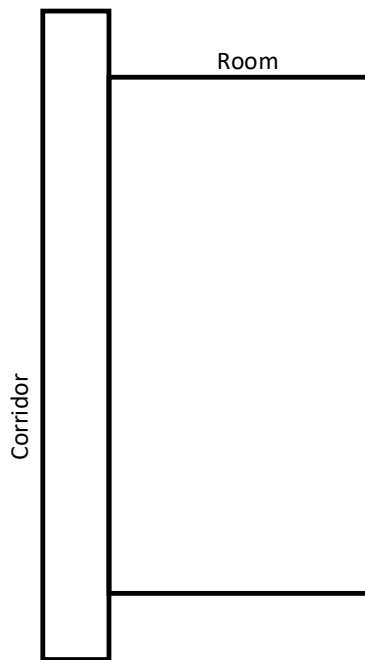


Figure 4.1: Experiment environment

The data packets were collected from phones belonging to the project partner. They both approved and participated in the data gathering experiment. The information in the data packets was anonymized and the participants' MAC addresses were whitelisted so that only the approved people's data was collected and not information of others accidentally passing by.

## 4.2 Equipment

The sensor equipment used in the experiment consist of six sensors based on 2 different types of sensors and 2 different types of antennas.

- Raspberry Pi 4
- Raspberry Pi 3
- Alpha antenna AWUS1900
- Alpha antenna AWUS036H

The mobile phones used in the experiment were 6 individual mobiles which consisted of 5 different types of phones.

- Samsung Galaxy S10 (2 phones)
- Samsung Galaxy S4
- Samsung Galaxy S8 Notes
- iPhone 6
- iPhone XS

There were 3 different types of PCs used in the experiment.

- HP Laptop
- Lenovo Thinkpad
- Apple MB Pro

In addition to the equipment mentioned above, a video camera was used to document the movement during the experiment.

## 4.3 Sensor locations

The experiment consisted of 6 sensors. Four of the sensors used the 2.4GHz and 5GHz band, while the remaining two sensors only used the 2.4GHz band. The four 2.4GHz/5GHz sensors were located in each corner of the room (numbered 1,2,3 and 4), illustrated by a 4-pointed star in figure 4.2. The two 2.4GHz sensors were located at each end of the corridor, as illustrated by a 5-pointed star in figure 4.2.

The sensors 1-4 consisted of a Raspberry Pi 4 with an Alpha antenna AWUS1900. Sensor 5 consisted of a Raspberry Pi 4 and an Alpha antenna AWUS036H. Sensor 6, consisted of a Raspberry Pi 3 and an Alpha antenna AWUS036H. The sensors were placed next to the wall and the relations between the sensors on each side of the room follow the wall of the room. The sensors were placed in a similar orientation and height, with the antenna up and placed on top of a chair or desk. The angle of the antenna was not exactly 90 degrees. The location of the sensors were used during the triangulation of the phones.

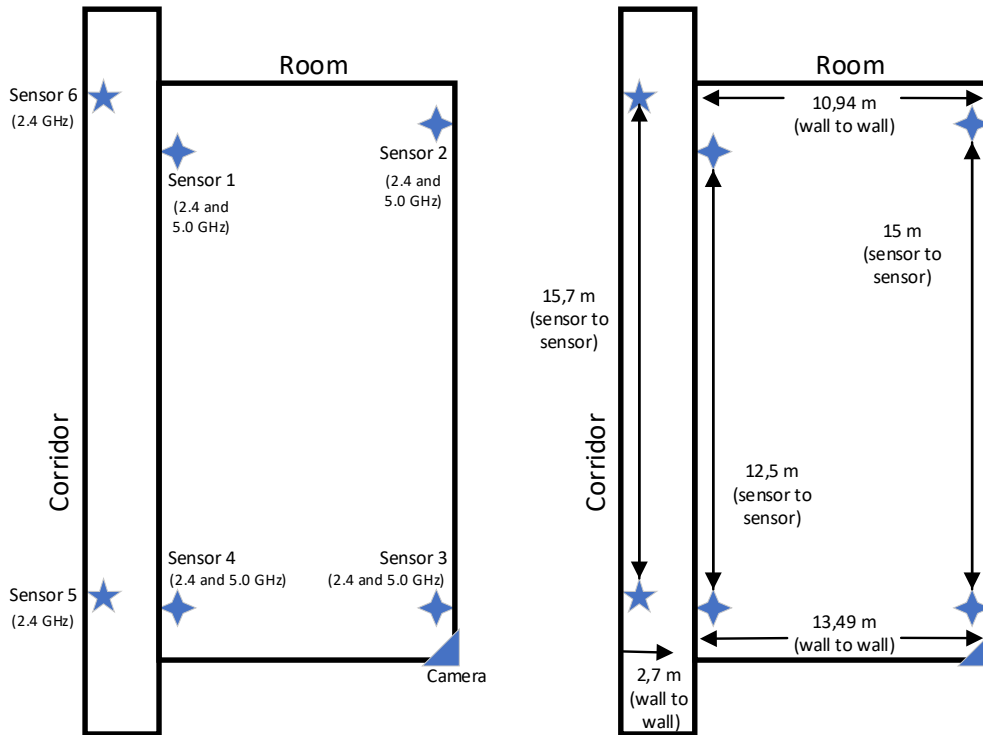


Figure 4.2: The location of the sensors

## 4.4 Grid pattern

To better describe and illustrate the different scenarios, the test area (the corridor and the room) was divided into a grid pattern. This was done to better identify and describe how the individuals would move in the different scenarios. Using a grid pattern made it easier to visualize for the test participants, which also made it easier to retry the experiment if it was proved to be necessary and if time allowed. The grid pattern is illustrated in figure 4.3.

## 4.5 Scenarios

The scenarios are divided into two groups. The first group consists of the preparatory scenarios, and this was used to gather information that was used to interpret the results of the second group of scenarios. The prepared scenarios included the preparation for the fingerprinting using signal strength method and scenarios that could give an indication of the accuracy of the different geolocation methods.

The second group consists of the scenarios to be used as tests for the remaining geolocation methods. Several of these scenarios had two different parts, where in the first part the mobiles will move in a given pattern, and in the second part the



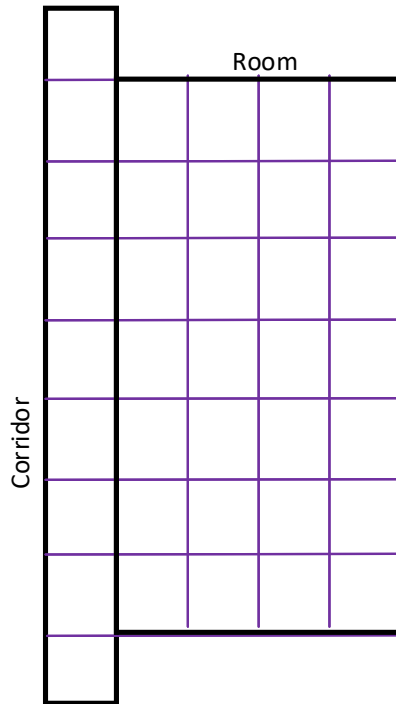


Figure 4.3: Grid pattern

mobiles will move in a random pattern.

#### 4.5.1 Preparation scenarios

##### Scenario 1 - Fingerprinting preparation

The first scenario prepares for fingerprinting geolocation (see Section 3.3.3). Using the data for fingerprinting is one of the 3 possible geolocation methods that was evaluated in this project based on the available equipment and collected data. In order to use the fingerprinting method, the rooms was mapped out. The mapping process was carried out by collecting the logged signal strength (in dBm) from the received data packets from the devices at the physical location (fingerprinting) while moving in a given pattern. These logged signal strengths measurements were used as ground truth, and other scenarios were compared against the fingerprint for estimating the localization. The video recorded by the camera was used to confirm the actual physical location of the device.

One disadvantage with this scenario is that the human body may affect (reduce) the signal strength received by the sensors based on how the mobiles are held (e.g. if this is in a pocket). The scenario was therefore carried out twice. The first time with the phones away from the body (i.e. above the head) and the second time with the phone in the pocket. This resulted in three possible fingerprints: One fingerprint where the phones are above the head during the experiment to exclude

signal weakening due to body tissue, one fingerprint with the phones in a pocket and the last fingerprint was the combinations of the two first fingerprints.

The data set for Scenario 1 consists of two individual data sets. The data set for the first fingerprint, where the phones are above the head, is defined as Scenario 1.1. The data set for the second fingerprint, where the phones are in pockets, is defined as Scenario 1.2.

The Figure 4.4 illustrates how Scenario 1.1 and Scenario 1.2 were performed. The goal was to map the room based on the grid pattern. A person was carrying the different mobile phones and moved through the room in a zig zag pattern by moving perpendicular to the room from one side of the room to the other side, then move a little further inward into the room before moving perpendicular to the room again from the one side of the room to the other side and so on as illustrated in Figure 4.4.

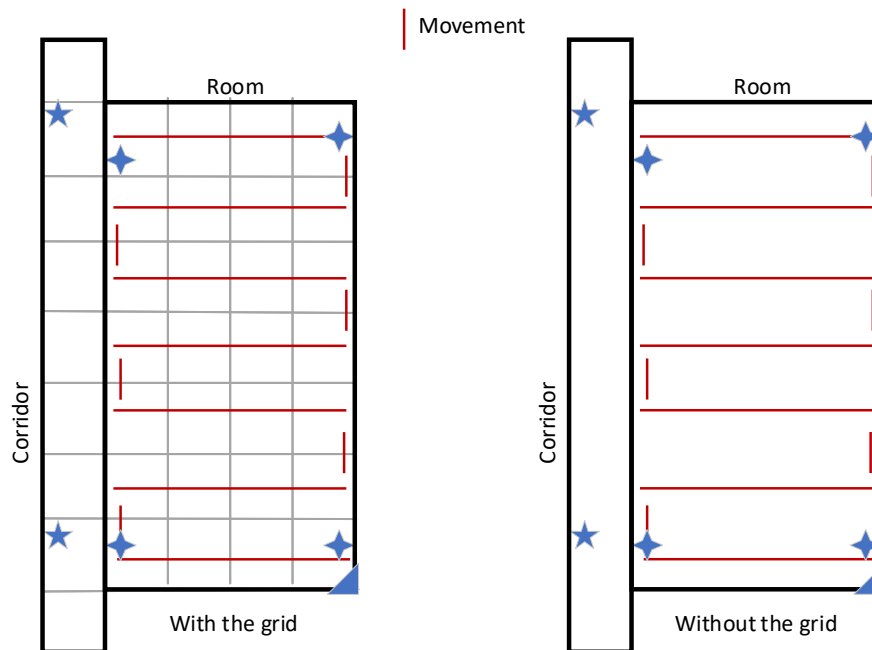


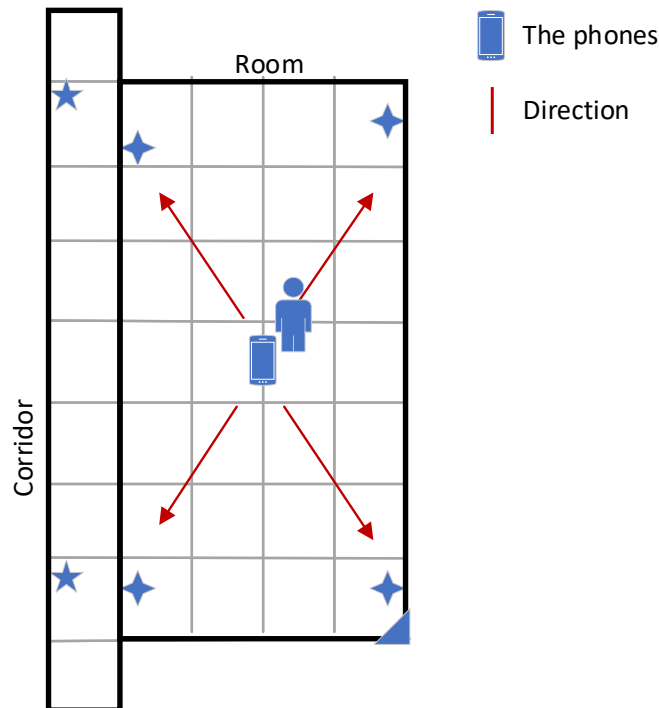
Figure 4.4: Preparation 1 - Fingerprinting

### Scenario 2 - Attenuated signal strength

The goal of Scenario 2, which were also a preparatory scenario, was to look at how much the human body affects and reduces the signal strength received by the sensors. The human body, as with other masses, will interfere and reduce the signal strength. This in turn will affect the accuracy and uncertainty of the localization method.

Figure 4.5 illustrates how Scenario 2 was performed. The mobiles were not associated with an access point but was in an active search mode for available

access points. This scenario was performed by a person standing in the middle of the room with the phones on a table. One by one, each of the mobile phones was lifted and activated by turning on the person's mobile screen and the phone was placed closer to the center of the person's body. While all the mobiles were lifted up one by one, the body was placed between the mobile and one of the sensors. Then the person was placed between the phones and the next sensor while repeating the sequence of lifting each phone. This was done for all the sensors. The data received by all the sensors simultaneously can be used to evaluate how much the body will attenuate the signal strength. This data can be used to evaluate how accurate the results are. By looking at the results of the sensors located in the corridor, the attenuation of the signal strength due to the walls between the room and the corridors can also be evaluated.



**Figure 4.5:** Preparation 2 - Attenuated signal strength

### Scenario 3 - Standing still

The goal of Scenario 3 was to evaluate how the calculated localization of the mobile phone varies over time (due to fluctuations of the received signal strength) even though the mobile phones are in sleep mode. This scenario may indicate how credible the results from an actual similar deployment of sensors will be.

Figure 4.6 illustrates how Scenario 3 was performed. The mobile phones were not associated with an access point but was in an active search mode for avail-

able access points. In this scenario, the phones were laying still for a few minutes on different locations in the room. There was six different locations selected as indicated in figure 4.6.

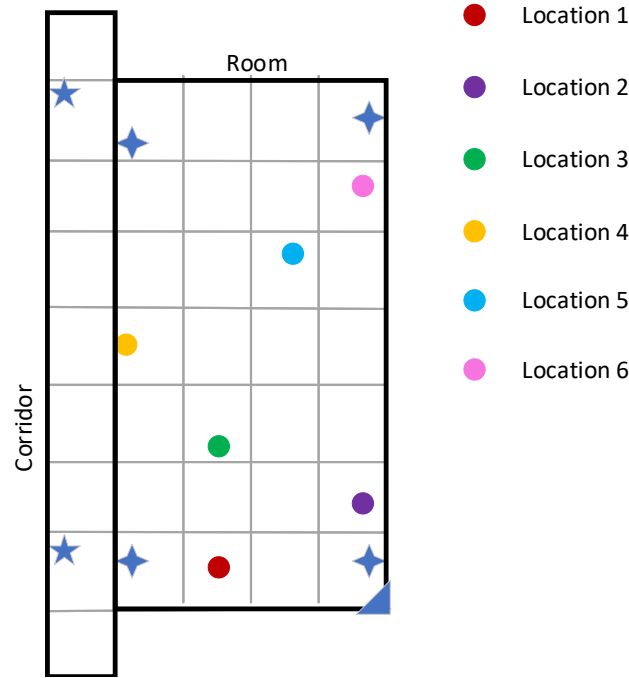


Figure 4.6: Preparation 3 - Standing still

#### 4.5.2 Geolocation Scenarios

##### Scenario 4 - Moving around inside the room while the phones are not in use (only searching for access points).

The goal for Scenario 4 was to evaluate whether, using different location methods, a device can be located accurately enough to locate the person and if possible find out how the person has moved. This scenario was used to evaluate the possibility to track the movement of a device and whether the unit can be pin pointed accurately enough to a location to be able to distinguish between a fixed pattern and a random pattern. This is indicative of how accurate the results are.

Figure 4.7 illustrates how Scenario 1 was performed. The mobiles were not connected to an access point but was in an active search mode for available access points. The scenario was performed by a person carrying the different mobile phones and moving around in the room. This scenario was divided into two phases, where in the first phase, the person moved in a set pattern and in the second phase the person moved in a randomly chosen pattern.

The data set for Scenario 4 consists of two individual data sets, one data set for each phase defined as Scenario 4.1 and Scenario 4.2. The first data set, whose

walking pattern was collected in Scenario 4.1, is the phase where the person moved in a set pattern. The second data set, whose walking pattern was collected in Scenario 4.2, is the phase where the person moved in a random pattern.

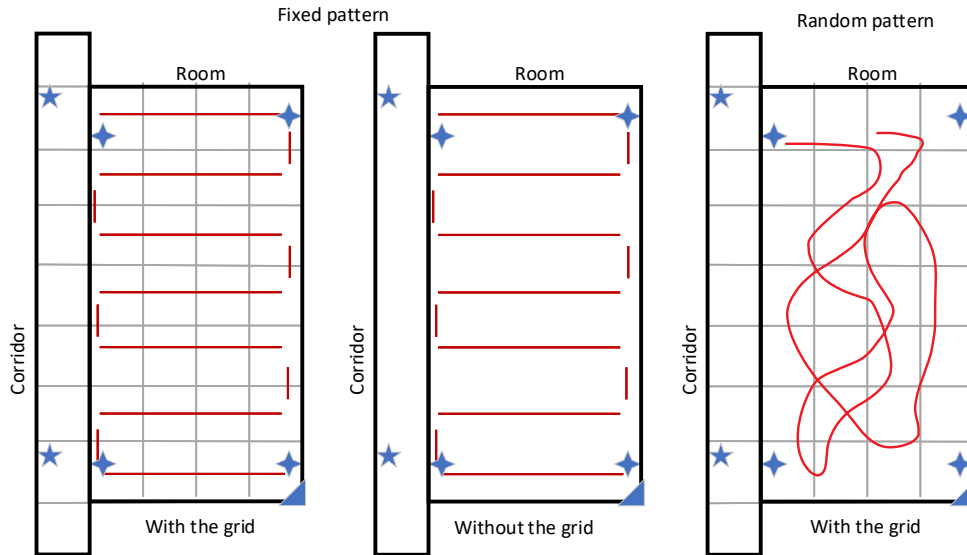


Figure 4.7: Scenario 4 and 5 - Moving around inside the room

#### Scenario 5 - Moving around inside the room while the phones are in use

The goal for Scenario 5 was the same as Scenario 4, but in this scenario the mobile phones were connected to an access point and streaming data e.g. streaming a YouTube video. The data from these scenarios was evaluated to see if the movement of a device can be tracked and pinpoint the location enough to distinguish between a fixed pattern and a random pattern. It was also evaluated how accurate the results are.

The main purpose of this scenario was to differentiate itself from Scenario 4 by sending more data packets from the phones that will be detected by the sensors. The data was evaluated and analyzed to see if the increase of data packets collected during Scenario 5 can give a better accuracy when determining the movement of the phones. The goal was to increase the amount of data collected by collecting data more frequently to calculate the location of the phones more accurately than in Scenario 4.

Figure 4.7 illustrates how Scenario 5 was performed. The scenario was conducted by a person carrying the different mobile phones and moving around the room. This scenario was divided into two phases, where the person first moved in a set pattern and in the second phase the person moved in a randomly chosen pattern.

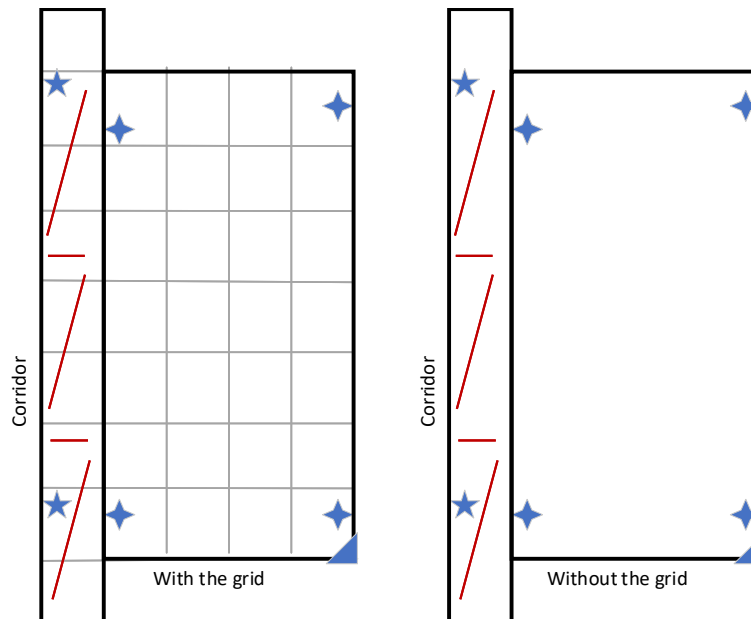
The data set for Scenario 5 consists of two individual data sets, one data set

for each phase defined as Scenario 5.1 and Scenario 5.2. The first data set, whose walking pattern was collected in Scenario 5.1, is the phase where the person moved in a set pattern. The second data set, whose walking pattern was collected in Scenario 5.2, is the phase where the person moved in a random pattern.

#### **Scenario 6 - Moving around in the corridor while the phone is not in use (only searching for access points)**

The goal for Scenario 6 was to analyze the data to see if, by using different location methods, the unit can be located accurately and to evaluate how the person has moved while restricted to a corridor. The scenario was based on Scenario 4, but applied to the corridor and not the room. During the analysis of the data it was evaluated how the wall separating the sensors was affected the accuracy of the triangulation.

Figure 4.8 illustrates how Scenario 6 was performed. The phones was not associated with an access point but was in an active search mode for available access points. The person with the mobile phones was moving around in the corridors, following a set pattern.



**Figure 4.8:** Scenario 6 and 7 - Moving around in the corridor

#### **Scenario 7 - Moving around in the corridor while the phones are in use**

The goal for Scenario 7 was to see if using different location methods can locate a device accurately enough to find the person while restricted to a corridor, and if possible, find out how the person has moved when the mobile phones are in use

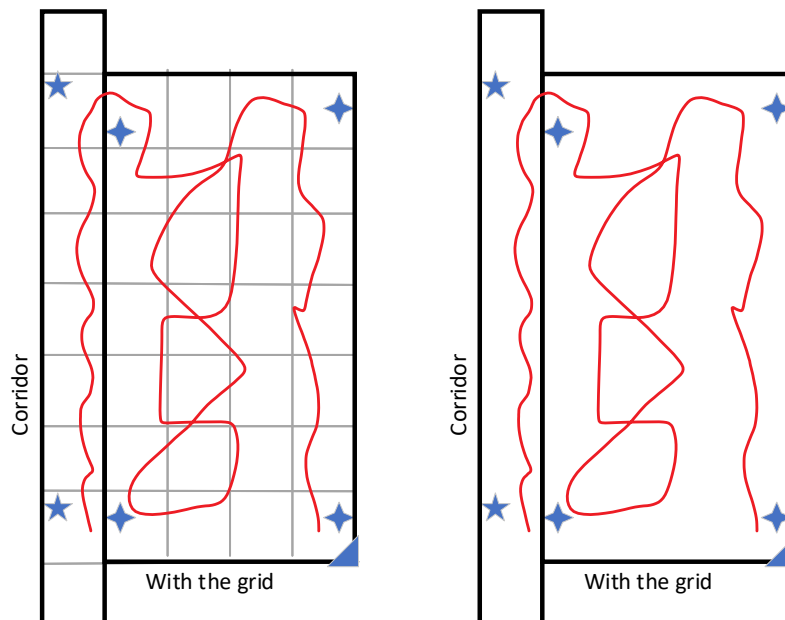
and are sending more data than when in search mode. The data from this scenario was evaluated with respect to the difference between using the mobile phone (for example streaming a YouTube video) and not actively using the mobile phone. The execution of the scenario is like Scenario 6, where the only difference is that during this scenario, the mobile phone is actively used e.g. streaming a YouTube video.

Figure 4.8 also illustrates how Scenario 7 was performed. The person with the mobile phone moved around in the corridor, following a set pattern.

#### Scenario 8 - Moving between the corridor and the room while the phones are not in use

The goal for Scenario 8 was to analyze the data to see if a person can be located correctly based on their location in the corridor or the room. The wall and the person's torso can both influence the signal strength. The focus of this scenario was to see if the device could be successfully located during the transition period between the corridor and the room.

Figure 4.9 illustrates how Scenario 8 was performed. In this scenario, the person with phones will start in one corridor and move around. After moving around in the corridor, the person moved from the corridor into the room itself. There, the person will move around in a random pattern. During this scenario, the mobile phone will not be actively in use (only searching for access points).



**Figure 4.9:** Scenario 8 and 9 - Moving between the corridors and the room

**Scenario 9 - Moving between the corridors and the room while the phones are in use**

Scenario 9 was the same as Scenario 8, but in this scenario the mobiles were in active use for example streaming YouTube video. The goal was to see if by analyzing the data a person could be correctly located in the corridor or the room.

The walls and the person can both influence the signal strength. The focus of this scenario was to see if the device could be successfully located during the transition period between the corridor and the room. In this scenario where the mobile phones are in active use, it was more frequent data packets to pinpoint the device.

Figure 4.9 illustrates how Scenario 9 was performed. In this scenario, the person with the phones started in the corridor and moved around. After moving around in the corridor, the person moved from the corridor into the room itself. There, the person moved around in a random pattern, before the person will exit to the corridor again and move around in there.



## Chapter 5

# Data Preprocessing and Analysis

This chapter contains the preparation of the data set and the analysis of which of the geolocation methods discussed in Section 3.3 that works on the given data. The chapter is divided into 7 sections:

- Data types
- Data preprocessing and cleaning
- General analysis of data
- Angle of Arrival - Analysis
- Time of Flight - Analysis
- Fingerprinting using signal strength - Analysis
- Triangulation or Multilateration using signal strength - Analysis

The first section, Data types, contains some information about what type of data was in the data set. The second section, Data preprocessing and cleaning, describes how the data set was preprocessed and cleaned before data analysis began. The third section, general analysis of data, contains discoveries in the data set that are not linked to specific geolocation methods. The last four sections analyze the four different geolocation methods and whether they can be used to locate a unit with the given data set.

### 5.1 Data types

There is a lot of different information that can be extracted from each data packet, but for this data analysis, only a few fields will be extracted for the analysis itself. The fields that will be used further in the project are the time stamp (unixtime), the sensor ID (sensorid), the hashed MAC address (machash), the signal strength (power) and the type of data probe used (prodtype).

**unixtime:** Unixtime is the date and time associated with the packet. This field contains the time information inserted by the sensor for when the sensor received the data frame. This time information is used to pair data packets with respect to time during triangulation. The time stamps are given down to milliseconds, but

as mentioned in the general data analysis (Section 5.3.3), there are uncertainties about the synchronization between the different sensors. Therefore the analysis will only be based on the time stamps down to seconds.

**sensorId:** The sensorId field identifies which sensor that received the data packet. The field contains an ID in the range 1 to 6, which is the ID of the sensor used in the project.

**machash:** machash is a hashed version of the MAC address of the mobile that sent the data packet. A hashed version of the MAC address was used to anonymize the users.

**power:** The power field is the signal strength measured by the sensor. The signal strength is given in dBm (see Section 2.1). The information in the power field is used for the geolocation methods based on the signal strength. The signal strength indicates how far from a sensor the mobile is located.

**probetape:** The probe type field contains the subtype value defined in the 802.11 frame. The subtype uses has only two possible values. Value 4 (i.e. Probe request) and value 5 (i.e. Probe Response).

## 5.2 Data preprocessing and cleaning

The data collected for use in the project was collected by the project partner over a period of 2 hours and 21 minutes. The data was collected based on the different scenarios as defined in Section 4. In the data set received from the project partner it was stated that the data collection was conducted with 6 different mobile phones and 6 access points, but the data analysis discovered data packets from 7 different mobile phones and 6 access points. It is possible that the 7th mobile phone was not a mobile phone at all and rather a laptop or other device that has been mistakenly white listed.

The total amount of data packets collected and received by the project was 26 108 data packets. The data packets were of two categories:

- Category 1: Subtype 4 in the 802.11 frame (i.e. Probe request), contained 9328 data packets
- Category 2: Subtype 5 in the 802.11 frame (i.e. Probe response), contained 16780 data packets

During the planning phase of the master project it was expected that the data contained a fair amount of noise that needed to be filtered out. It was also expected that the data could contain unintentionally misleading data which could cause some problems. The project was eventually carried out in a controlled area with a small number of units inside a building. The devices used were white listed and only packets from the white listed devices that participated in the experiment

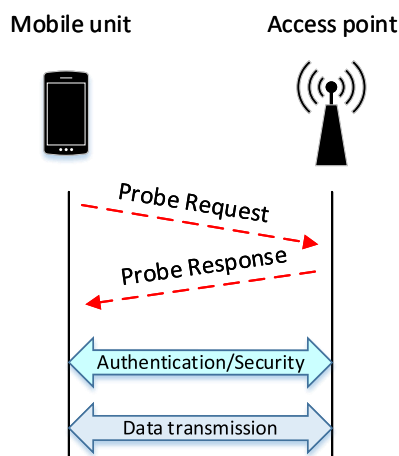
were captured. This greatly reduced the amount of data that needed to be reviewed which also affected the amount of noise that needed to be filtered out. Since the data was collected in a controlled area with a limited number of known devices, there was no risk that someone was deliberately trying to modify the data. Manipulation of data is one element to consider if this experiment was to be conducted on a larger scale outside a controlled area where all data from all passersby were collected.

The preliminary data cleaning was performed by the project partner to anonymizing the data and to remove excess data information. There were no structural errors or missing data in the data set, and it is possible that more data cleaning was performed by the project partner during the collection and before the data was given to the master's thesis project. Although the data at first initial review may appear to have been cleaned, there were still some additional preprocessing of the data set needed to be done to prepare the data for further analysis. The data set was therefore further cleaned and prepared based on 4 methods:

- Probe type removal
- Extraction of data collected during the scenarios
- Removing data based on signal strength
- Time compression

### 5.2.1 Probe type removal

When analyzing the data packets with respect to geolocation, a large group of unwanted observations were discovered. During an initial phase of communication between the mobile phones and the sensors, there will be an exchange of data to verify if it is possible to establish a connection. In an active search phase from the mobile phones, the basic data exchange is when the mobile sends a Probe request to the access point and the access point responds by sending a Probe response. This initial phase of communication is illustrated in Figure 5.1.



**Figure 5.1:** Initial phase of communication

The data set received contained only Probe requests and Probe responses. For our analysis, only data packets sent from the mobile phones are of interest. That is, the probe response packets do not give any information for performing triangulation. All packets with probe type 5 were therefore filtered out.

The data set received consisted of 26 108 data packets in total, where this count includes the data collected before, between and after the scenarios. Of the 26 108 data packets, only 9 328 were of probe type 4 (probe request), which is relevant data for this experiment since this is coming from the phone and can be used for geolocation. Since probe type 5 (Probe response) is not used, 16 780 packets (64.3% of the packets) is discarded and only 35.7% of the original data set was used for further analysis.

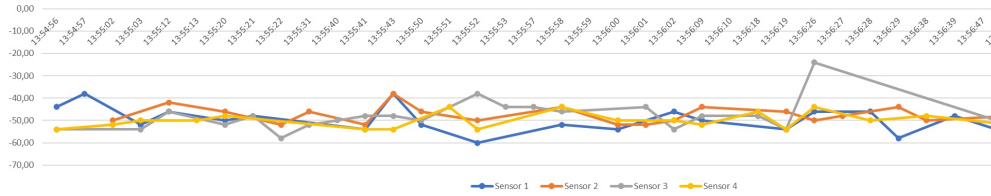
### **5.2.2 Extraction of data collected during the scenarios**

Within the 2 hours and 21 minutes of data there were several scenarios that were executed (as described in Section 4) and the collected data set consists of all the scenarios in addition to a lot of data that has been collected before, between and after the different scenarios. After removing the probe responses, the next step in cleaning the data was filtering out the scenarios into individual data sets. This was done to remove excess data that was not part of the scenarios and to simplify the work with the individual scenarios. The filtering was based on a time-based filtering method. The scenarios were filtered based on start times and end times stated in the ground truth video. A challenge was to define when the scenario started with respect to the data traffic when the ground truth video defined the start and end times only down to the minute and not down to the seconds which gave uncertainties to when the scenario started with respect to the data collected that was according to seconds and milliseconds. To correlate the different data packets from different sensors, it would have been preferable to work with milliseconds, but uncertainties about the synchronization between the sensors prevented this. To read more about the synchronization see Section 5.3.3.

Because the ground truth video only provides time down to minutes and not seconds, it was not clear exactly when the scenarios started or ended. If a scenario was stated as start at 13.05 and end at 13.06, but only lasted for 1 minute and 15 seconds, all the data from 13.05.00 to 13.06.59 still had to be included in the scenario data set since it was uncertain whether the scenario was started anywhere between 13.05.00 and 13.05.59. This also led to some temporary overlap between some of the scenarios.

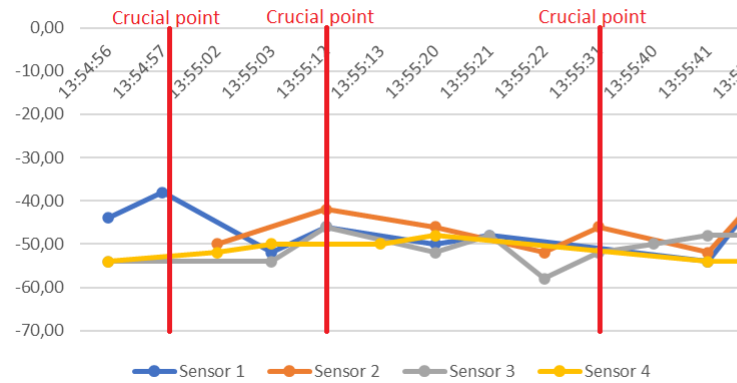
The rest of the data had to be filtered out after trying to find start times based on the changes in RSSI strength and how the RSSI was expected to change according to the scenario. As an example, in Scenario 1.1 (after all remaining data preprocessing was completed), it starts at 13.55 according to the ground truth video, and finished 13.56. The ground truth video clip for scenario 1.1 lasts for about 1 min and 23 seconds. The person holding the phones started approximately 2 seconds into the video clip, spending approximately 1 min and 17 min on

the scheduled route. The figure 5.2 indicates the dBm strength measured at the different sensors during the scenario.



**Figure 5.2:** Scenario 1.1 - Sensor 1-4 - Strongest Signal Strength

The data indicates that the mobile phones stayed close to sensor 1 just before scenario 1.1 starts, then they began to move away from sensor 1. This is consistent with the ground truth video. In Figure 5.3, the first red line marks approximately where scenario 1.1 starts. Little information was recorded by sensor 1 around this time so it is not quite clear when the scenario starts, but we can see that the signal strength drops sharply between the registration at 13:54:57 and 13:55:03.

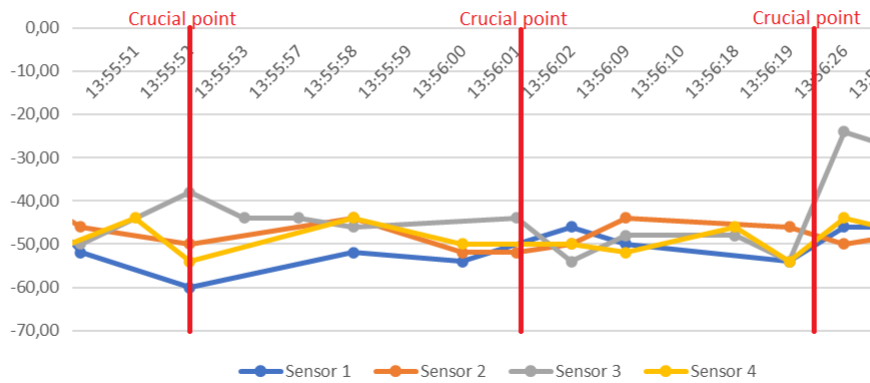


**Figure 5.3:** Scenario 1.1 - Crucial points for locating where the scenario begins

The mobiles spend about 10-12 seconds crossing the room in the ground video. After about 10-12 seconds in the data set there is an increase of signal strength at sensor 2 indicating that the mobiles are approaching sensor 2. This point is marked in Figure 5.3 with the second red line. This seems to match the ground truth video and what was expected. Again, there have been few registrations with sensor 2 in the seconds around this point. The next point that fits well with the ground truth video and data set is around 13.55.31 (marked by the last red line in figure 5.3). The mobile phones have crossed the room 3 times at this point and are between sensors 2 and 3. According to the data set, the mobile phones are now further away from sensor 2 than before, but still closest to sensor 2 which matches well with the ground truth video.

Analyzing the data in more details uncovered that it became more difficult to find places that continue to fit in well with the ground truth video. This is mainly

due to the registrations made by sensor 4. Approximately around 13:55:52 the mobiles according to the video should be close to sensor 3, but very close to sensor 4. This does not match the data, which indicates that the mobiles are very close to sensor 3 and not near sensor 4. This is marked by the first red line in Figure 5.4. This may be due to some unfiltered noise.



**Figure 5.4:** Scenario 1.1 - Crucial points for locating where the scenario ends

The next spot that fits with the ground truth video is around the time 13:56:01, as marked by the second red line in figure 5.4. The mobiles should be between sensor 2 and 3 around this time, but closest to sensor 3. This seems to match the data set. The last red line in figure 5.4 marks the expected end of the route to the mobiles. Here, the mobile phones should be located closest to sensor 3. A problem here is that there are no registrations of data with sensor 3 around this time which makes it difficult to confirm that this matches the data set. According to the ground truth video, the mobiles stay near sensor 3 for a little while after finishing the route before they start moving back to the starting position. This may be consistent with the data set at time 13.56.26, where we can also see that the signal strength decreases when the phones are moving away before the next registration.

Approximately 1 min and 17-20 seconds after the person started walking, the data set indicated that the mobiles are close to sensor 3, which is expected based on the ground truth video. There was some uncertainty due to lack of data, which is a problem that is reflected in almost all scenarios. If there is little data, the diagrams will be greatly affected if any of the registrations are incorrect. There was nowhere else in the scenario 1.1 data set where the data set fits with the scenario start as well as from around 13.55.02, giving confidence that the scenario starts at this point in time.

The filtering that was done to find the start second for a scenario mainly applied to Scenarios 1.1 and 1.2 which were to be used to create a fingerprint for the fingerprint geolocation method (see Section 5.6 for the analysis). The fingerprint is based on the data that is captured over a given period of time, and the measured signal strength with respect to all the sensors corresponds to the device's physical

location within the room's grid drawn in Section 5.6. If the start time is shifted, the data will be placed in the wrong cell and when new data is compared to the fingerprint the device will be placed at the wrong location. Since there are some uncertainties around the start time and it is difficult to find an exact start time by looking at the data set (due to limited data packets), a possible fingerprint will have a certain amount of uncertainties associated with it.

Prior to extracting the scenario data sets from the overarching data set, the data set consisted of 9328 data packets of probe type 4 (i.e. Probe request), while after the filtering the data set was left with 4588 data packets distributed across 12 scenarios/sub-scenarios. The distribution of data packets between the scenarios is illustrated in Table 5.1. After all the scenarios were filtered out and the remaining data was removed, the scenarios were further divided based on each machash to see the difference between the mobiles and analyze every single mobile data transmission.

Scenario	Total data packets
Scenario 1.1	276
Scenario 1.2	266
Scenario 2	1037
Scenario 3	320
Scenario 4.1	423
Scenario 4.2	670
Scenario 5.1	262
Scenario 5.2	454
Scenario 6	217
Scenario 7	49
Scenario 8	84
Scenario 9	530
<b>Total for scenarios</b>	<b>4588</b>
<b>Outside scenarios (not in use)</b>	<b>4740</b>

Table 5.1: Data packets per scenario

### 5.2.3 Removing data based on signal strength

The data packets will, when received at the access points, be measured according to the signal strength of the incoming signal. The signal strength is measured in dBm. The transmitting unit (mobile phones) will transmit data packets with different content and length at different time intervals and this will be received and measured at the access points. During the analysis of the data, the signal strength appears to be very random and sporadic in many cases. As an example, within 1 second the values will in some cases range from -20 to -80 for the same mobile to

the same sensor. This applies both when there are only a few registrations during the same second, and when there are many (10+) registrations during the same second. In some cases, the signal strength within a given second will appear as having groupings, for example there are 15 registrations where 5 are in the range -20 to -30, 5 in the range -50 to -60 and the last 5 in the range -80. A typical scenario measurement is shown in figure 5.5 and a section of the same scenario is shown in figure 5.6.

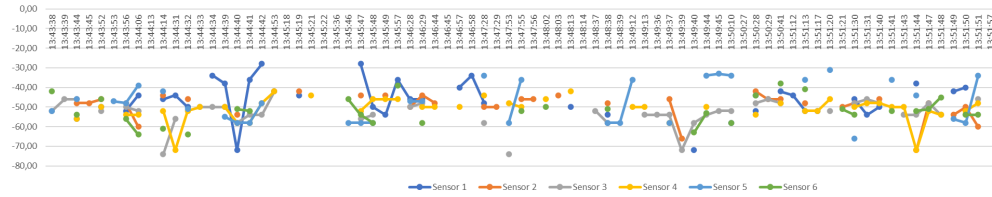


Figure 5.5: A typical scenario

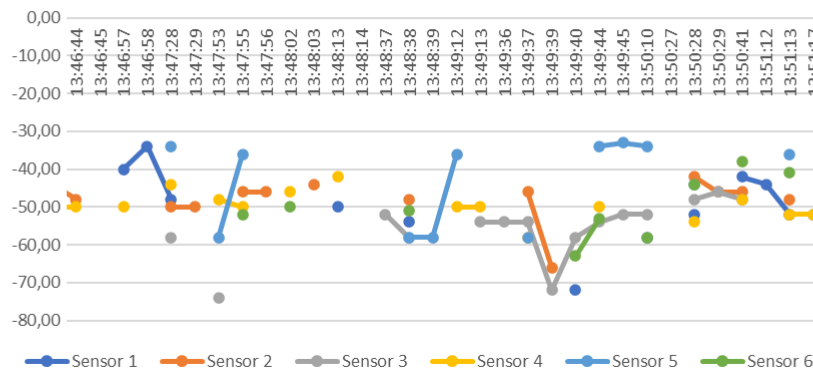
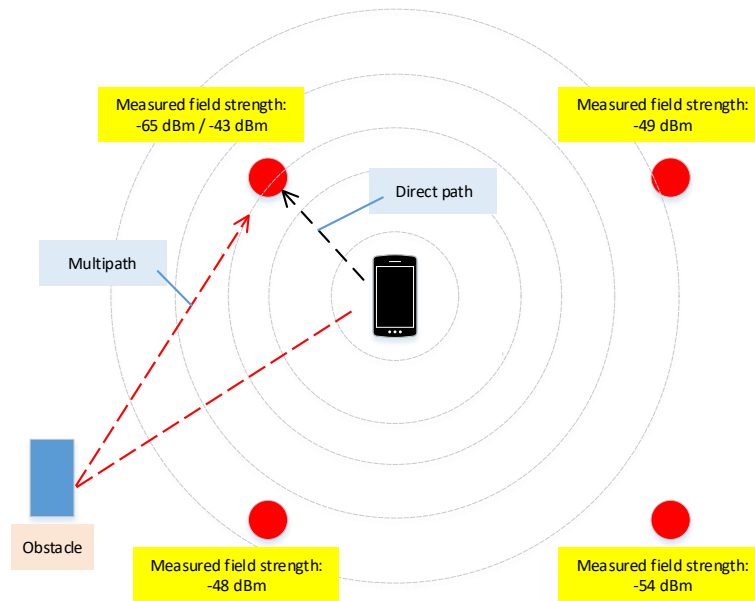


Figure 5.6: One section of a typical scenario

Many of the lowest values e.g. -70 and -80 may be due to multipath. This will affect the access points measurement of received signals, which can make the phones appear further away from the sensors than what they really are. Figure 5.7 illustrates how a packet can take multiple paths from the phone to the sensor. The figure illustrates the direct path from the mobile to the sensor and a path where the signal is reflected by an object e.g. a wall. If the packet sent from the phone uses the direct path to the sensor, it would then have a signal strength of -43 dBm. The phone will then appear to be quite close to the sensor, which is correct. If the packet sent from the phone uses the other way to the sensor, in that case it would come with a signal strength of -65 dBm. The phone will then appear to be quite far away from the sensor which is incorrect.

The values, which are not theoretically possible based on the size of the room and the distance to the sensor, will be seen as a type of unwanted outliers and should therefore be removed. Retaining these values will only cause the phones





**Figure 5.7:** Visualization of multipath

to be misplaced relative to the sensors. The unwanted outliers will be removed based on a form of distance based filtering based on what the expected signal strength value is when located in the given test area.

### Calculation of unwanted outliers

To evaluate the expected values of the received signal within the test area, a calculation of expected received signal in free space was calculated in a path loss transmission formula presented by Friis [39] (Friis transmission equation) and explained in Section 3.3.4.

When calculating the signal strength reduction through air within the test area, the longest path for the radio signal will be the furthest point a mobile can be away from any access point within the test area. This was calculated to be approximately 23 meters.

When calculating for the 4.2 Ghz band, the middle frequency 4224 Mhz (channel 7) is used and the path loss calculation indicates a loss of 72,2 dBm. When calculating for the 5 Ghz band, the middle frequency 5200 Mhz (channel 40) is used and the path loss calculation indicates a loss of 74 dBm. An indicated signal loss due to body attenuation (when phones are in pocket) is estimated to 3 dBm [43].

We estimate the output power for smart phones from Wi-Fi to be 20 dBm [44, 45]. This gives an estimated reception on the access point at the level -55,2 dBm for the 4.2 band and -57 dBm for the 5 band. When taking a 10% uncertainty into consideration, the values below -63 dBm are regarded as sporadic and not useful

and therefore is discarded. Figure 5.8 illustrates the removed data.

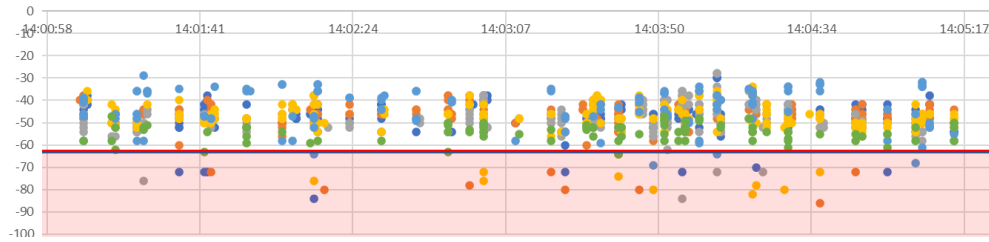


Figure 5.8: The signal strength boundary

#### 5.2.4 Time compression

During the analysis of the data, a form of time compression became necessary. In most phases of the data analysis, the focus was on the seconds, as a result of the scenarios ending up being quite short and uncertainties around milliseconds. Since there were often multiple data packet from the same mobile connected to the same second and the same sensor, these were merged.

All the packets from the same mobile, the same second and the same sensor were combined and the group was represented by a value. This value was calculated based on the average of the values or based on the strongest dBm value. In most cases, two versions were created, one based on the average and one based on the strongest dBm value.

For the fingerprint geolocalization method (see Section 3.3.3 and 5.6), where even more time compression was performed, several seconds were combined to represent a larger group of values. The time compression for the fingerprints was performed in the same way as for all values associated with one second, but this time several seconds were merged. A fingerprint was created based on average dbm calculations and a version based on calculations of the strongest dBm value.

### 5.3 General analysis of data

This section is divided into five sections. The first part discusses general discoveries related to the data set that do not apply to a specific geolocation method. The four remaining parts are based on the four geolocation methods and which findings are decisive for whether the geolocation method works or does not work on the given data set.

#### 5.3.1 Difference in number of registered phones per scenario

After a quick look at the different scenarios, it was noticed that not all the phones were registered in every scenario despite the fact that every phone was used in every scenario. In scenario 1.1, four of the mobiles were registered which means

that they sent probe requests during the time scenario was performed. The mobiles that were registered were those with IDs 1, 2, 4, 5. Already in scenario 1.2, a change had occurred, and the mobiles with IDs 1, 3, 5 and 6 were the only ones to send out probe requests. Scenario 3, is the first of the scenarios to capture probe requests from all 7 different devices. This may be because scenario 3 was performed over a longer period of time. Figure 5.2 shows which mobiles were captured during which scenarios.

Scenario	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5	Phone 6	Phone 7
Scenario 1.1	X	X		X	X		
Scenario 1.2	X		X		X	X	
Scenario 2	X		X	X	X	X	X
Scenario 3	X	X	X	X	X	X	X
Scenario 4.1	X		X		X	X	X
Scenario 4.2	X	X	X	X	X	X	X
Scenario 5.1	X		X	X		X	X
Scenario 5.2	X	X	X	X		X	X
Scenario 6	X		X	X		X	X
Scenario 7	X			X			X
Scenario 8	X	X		X			X
Scenario 9	X	X	X	X			X

**Table 5.2:** Which phones were captured during which scenarios

We can see that mobile 1 is the only mobile that is registered in all scenarios and that mobile 5 abruptly stops sending data after scenario 4.2.

### 5.3.2 Difference in number of packet per phone

The data shows that there is a big difference in the number of probe requests sent from each mobile. Some mobiles sent many probe request during the short period a scenario is conducted, while others sent very few.

The ratio of how many probe requests the different mobiles send out seems (in most cases) to remain the same between all scenarios. During all the scenarios where mobile 1 and mobile 5 are registered, we can see that these two mobiles send a lot more probe requests than all the other mobiles. It also became quite clear that the mobiles sent out fewer probe requests when the mobile had the screen off. Figure 5.9 illustrates the difference between the number of probe request each single mobile sends per scenario

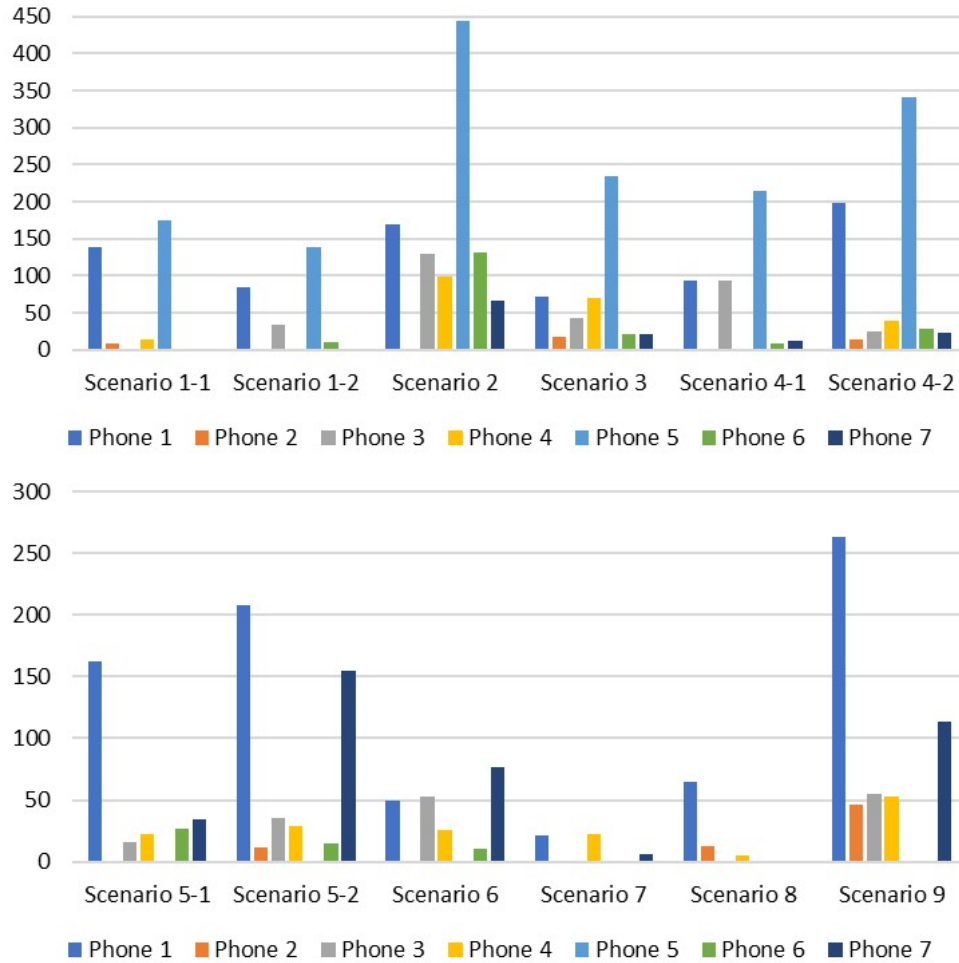


Figure 5.9: Different number of packets per phone and scenario

### 5.3.3 Synchronization

Before data collection began, all of the sensors were synchronized with a ntp server on the Web, but it was uncertain whether they were synchronized down to milliseconds. The sensors were listening to the device in promiscuous mode using a python script with scapy [46], but discarded all packets that were not probe requests or probe response. All six sensors collected the entire time simultaneously, but channel hopping and possible time offsets are possible errors. The analyzes indicate that the data is not collected simultaneously at all sensors at the same time (i.e. the sensors are not listening and therefor not detecting all data packets but only the data packets addressed to the specific sensor), but this may be because the scan through the channels happens at random intervals which are not in sync. It is therefore possible that one sensor hears/sees something that a different sensor misses because they are on different channels at the time of the event.

#### 5.3.4 Calibration

The sensors are not calibrated towards each other with respect to dBm measurement. No calibration equipment was used to compare the antenna or radios. The Wi-Fi-adapters are consumer grade, and there could very well be slight differences between them.

#### 5.3.5 Time stamps

In the video the start time for each scenario is given in hour and minute. In some cases, the analysis requires that the start time of the scenarios is given in seconds. It is uncertain exactly when the scenario started, since the start time could be anywhere during the minute. The time stamps in the data packets should be when the events happened and not when they were written to the database, but this is not confirmed.

### 5.4 Angle of Arrival - Analysis

One of the geolocation methods evaluated was Angle of Arrival described in Section 3.3.1. In the beginning, it was obvious that it would not be possible to try this method with the equipment available. The access points used for gathering data did not have any direction finding function or necessary equipment to do so. The data set given did not have any information regarding direction so there was no way to test the Angle of Arrival method. In order for Angle of Arrival testing to have been executed, all the sensors would have needed several antennas and the necessary direction-finding features.

### 5.5 Time of Flight - Analysis

Time of Flight is one of the geolocation methods described in Section 3.3.2. It was stated early on that this method, as with Angle of Arrival (3.3.1), could not be applied to the available data set. The analysis of the data indicates that the data is not collected simultaneously at all sensors at the same time. If so, then the Time of Flight method will not be possible since we do not have the same data packets to measure time delay against the different sensors.

Although the data indicates that the data packets are not being collected simultaneously, the sensors were synchronized with an ntp server and all the sensors were in promiscuous mode. As discussed in Section 5.3.3, the uncertainty of this can be attributed to the fact that the scan through the channels is at random intervals which are not in sync and that it is very likely that one sensor captures something that a different sensor misses because they are on different channels at the time of the event. Although, the reason why the sensors do not detect all data packets is not because the sensors are not in promiscuous mode or are not syn-

chronized, we still (in most cases) do not have the same data packet from several of the sensors so we cannot measure time delay .

The biggest problem with the Time of Flight method compared to this data set is the timestamps. The times go down to milliseconds, and even if there was no uncertainty about the synchronization of milliseconds, this would still not be enough to use the Time of Flight method. Milliseconds which gives a distance resolution of approximately 300 000 meters. This means that one will be able to place a unit within a 300,000 m radius. This resolution has no practical use in locating units within a few meters or hundred's of meters and therefore, the Time of Flight method cannot be used with the available data.

To be able to use the Time of Flight method with an accuracy high enough to be used in the test location, one needs to have nanoseconds. Nanoseconds gives an accuracy around 0,3m.

## 5.6 Fingerprinting using signal strength - Analysis

The method of Fingerprinting used signal strength described in Section 3.3.3. The goal was for the fingerprints to become a ground truth database that could be used to compare the next registrations against, but the fingerprints were created only for the room itself and not the corridor.

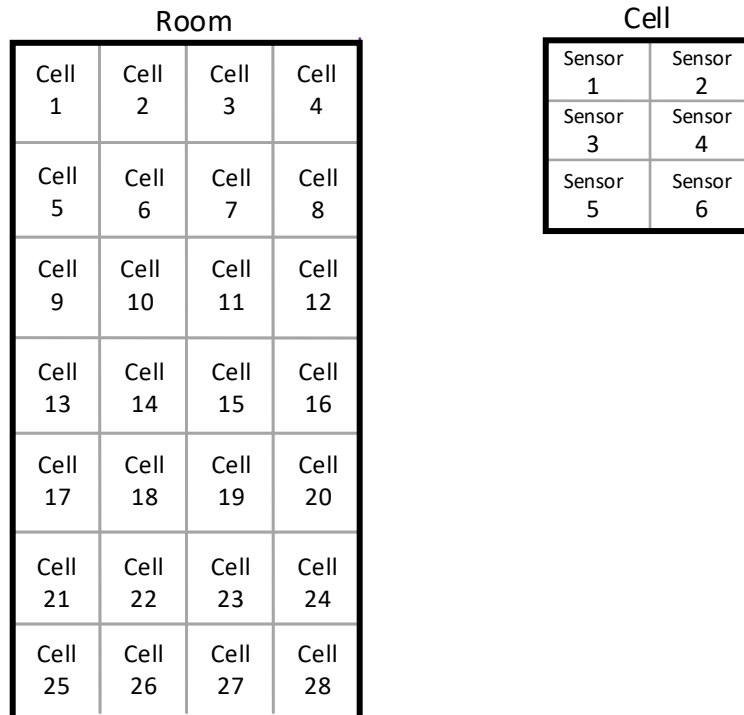


Figure 5.10: Fingerprint visualization

The ground truth database consist of several defined areas in the room, each

location with 6 values representing the signal strength from that location to each sensor. Figure 5.10 visualizes what a fingerprint will look like. Each location, defined as a cell, is illustrated in the figure. Each cell consist of the 6 values that represent the signal strength.

Based on the data, it appears that the sensors are detecting the data packets at different intervals. As this example shown in table below, it is shown that within the same time interval (1 second) some sensors are detecting a data packet and some are not. If the sensors were receiving the same data packets, it was expected that if the sensors are not synchronized in ms then at least the data packets would have the same time delay in all data packets (e.g. if data packet 1 has a time difference 100 ms between two sensors, then the next data packet should also have the same time difference, 100 ms) but they do not have this. An example of this is given in table 5.11.

Time	MS	Phone 1					
		Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6
		dBm	dBm	dBm	dBm	dBm	dBm
13:55:20	274			-52			
13:55:20	363					-48	
13:55:20	411					-51	
13:55:20	418						-50
13:55:20	438						-50
13:55:20	453					-46	
13:55:20	523			-46			
13:55:20	544			-78			
13:55:20	557			-50			
13:55:20	589	-50					
13:55:20	590				-46		
13:55:20	610				-44		
13:55:20	631	-44					
13:55:20	642		-40				
13:55:20	685		-40				
13:55:20	694	-42					
13:55:20	702				-46		
13:55:20	722				-44		
13:55:20	729		-44				
13:55:20	742				-76		
13:55:20	746		-42				
13:55:20	768				-46		
13:55:20	782		-46				
13:55:20	890				-48		

**Figure 5.11:** Packet detection rate

The data indicates that it is not possible to see which data packet is received by one sensor only or if received by other sensors also, making it difficult to create good enough cross-references. The sensors are synchronized and all the sensors listen in promiscuous mode, but the scan through the channels are at random intervals which are not in sync. This can explain the behavior illustrated in 5.11, as it is then possible that one sensor captures something that a different sensor misses because they are on different channels at the time of the event.

The data analysis also shows that the transmission from the phones are not very frequent but are clustered in groups, and the next cluster for the same phone could be more than 10 seconds later. As an example, Figure 5.12 shows how the

data packets are spread out over time for mobile 1 for all sensors in scenario 1.1.

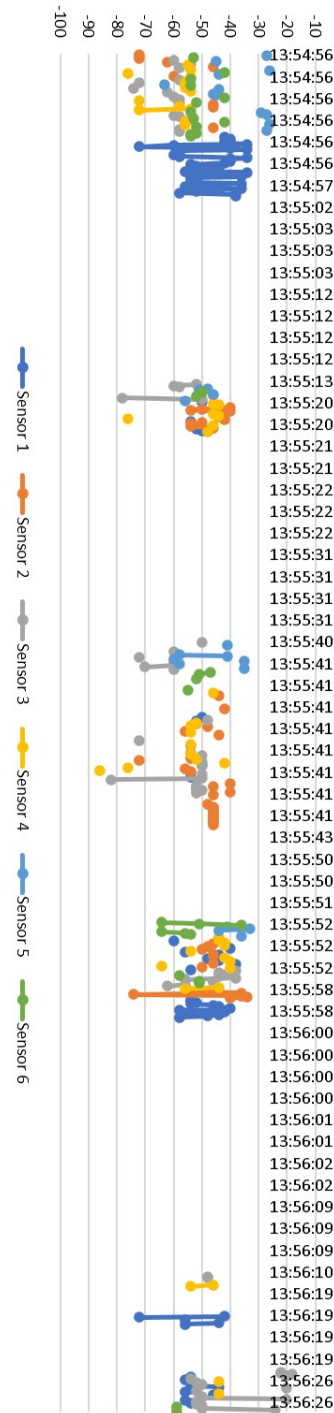


Figure 5.12: Packet clustering



The original data set contained a lot of excess data from before and in between the scenarios that have been filtered out, so only data recorded in the period indicated by the videos is used. One drawback when it comes to being able to establish a ground truth database is that the video only indicates the minute the scenario starts and not the second, which creates uncertainties when the scenario starts e.g. 14:55:03 or 14:55:23. The different scenarios ended up being very short and they lasted around 1-2 minutes which required the accuracy of the second. The fingerprints were made based on Scenarios 1.1 and Scenario 1.2. Based on the data and visualization and as well as the length of the scenario (the number of seconds they use) it seems that scenario 1.1 starts right after the minute has started approximate 13:55:02 as discussed in Section 5.2.2 and scenario 1.2 starts approximate 13:57:11.

Another issue that arose was the amount of data available to create the fingerprint. The room was to be divided into 4x7 cells as illustrated in 5.10. The person with the mobile phones crossed the room 7 times which was therefore set to the number of rows in the fingerprint. The person carrying the mobile phones spent about 10-12 seconds crossing the room from left to right, which resulted in very little data being collected. Since the row was divided into 4 cells, it only took 2.5 - 3 seconds for each cell. Since there were some uncertainties in whether the sensors were synchronized to milliseconds the data could only be distributed on the cells based on seconds. Therefore, one cell would contain data from 3 seconds which is the approximate time the mobiles are located within this location, assuming the person spent 12 seconds walking the distance.

This led to another issue as well as uncertainties about synchronization, clustering and detection rate. It is still not certain that the start times are correct and since there are few seconds in each cell, the mobiles could be placed in the wrong cell if the time is 1-3 seconds wrong. Therefore, it is not to be expected that the results for the fingerprints are accurate. The method could still be tested a little further to see if there were any other major problems or uncertainties that needed to be addressed before any eventual new data collection.

The goal was to get a unique reference point in each location that other measurement could be evaluated against. This meant that the reference point must be able to represent all the data collected during the approximately 3 seconds. How to determine this unique signal strength is considered by looking at different methods. This included looking at the use of the average, median, mode, strongest dbm value and/or weakest dbm value of the collected data in the given time period at that location.

- **Average:** The advantage of using averaging is that it takes into account all the values that were recorded.
- **Median:** The median can represent the group well if the values are quite similar, but it can give a value that wrongly represents the rest of the group if there is a large spread e.g. if some values are around -20, some values are around -40 and some around -60.
- **Mode:** The advantage of using the mode is that it can give us the value that

is repeated most times in the group of values. The disadvantage is that it gives no answer if none of the values in the group are equal. During the analysis of the data set, it was discovered that many of these groups consist of 1-3 values where all values are different.

- **Strongest dBm value:** The advantage of using Strongest dBm value is that this value is most likely to represent the group well, since the value will represent the most direct route from the mobile to the sensor.
- **Weakest dBm value:** It will not be a good idea to use the weakest dBm value, since this value is most likely to be due to multipath, and is therefore more likely to produce incorrect results.

The project ended up using the two following methods: average and strongest dbm value. The fingerprint based on the average values was calculated in 2 rounds. In the first round, all packets from the same mobile, in the same second to the same sensor, were combined using averages. In the next phase, all the packets belonging to the time period of a given cell were combined and the average for each sensor was calculated. The fingerprint based on the strongest dBm value was calculated using the same method, where the only difference is that instead of calculating the average, only the strongest dBm value is used.

Since there is a difference between mobile vendors, one should preferably have a fingerprint for each vendor, which in this case would have been for iPhone and Samsung, but it was quickly apparent that the mobiles did not send enough data to make a fingerprint for each. It varied greatly how many registrations there were per mobile. Some of the mobiles sent a fair amount of data packets, while other mobiles sent no more than 2-3 data packets throughout the scenario. The only mobiles that could have their own fingerprints were mobile 1 and 5, since they generally sent a lot more data than all the other mobiles. Figure 5.13 illustrates the amount of data per mobile for scenario 1.1 and scenario 1.2.

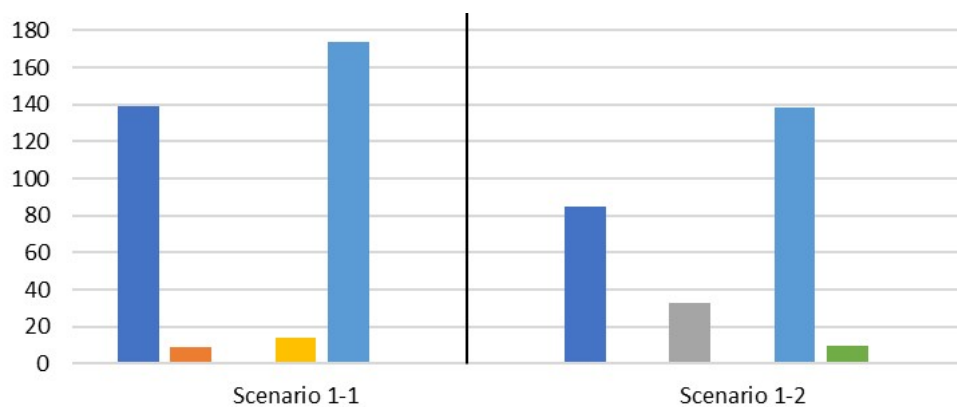


Figure 5.13: Number of data packets per phone - scenario 1.1 and scenario 1.2

### 5.6.1 Fingerprints

It was decided that two fingerprints should be created for all the mobiles for the scenarios. First, a fingerprint was created based on the average of all values and then a fingerprint based on the strongest dBm value.

#### Scenario 1.1 Fingerprint

The first set of fingerprints generated was for scenario 1.1. Figure 5.14 illustrates the two fingerprints. As described earlier in this section (Section 5.6) and illustrated in Figure 5.10, a fingerprint consists of several cells, each cell consisting of the 6 values representing the signal strength received by the 6 different sensors. The first number is the signal strength received by sensor 1, the second number is the signal strength received by sensor 2, etc. to the last number which is the signal strength received by sensor 6.

Average value				Strongest dBm value			
-50 -47	— —	— —	-46 -42	-48 -46	— —	— —	-46 -42
-48 -50	— —	— —	-46 -48	-46 -46	— —	— —	-46 -46
— —	— —	— —	-41 -55	— —	— —	— —	-33 -55
— —	— —	-45 -47	— —	— —	— —	-42 -40	— —
— —	— —	-50 -45	— —	— —	— —	-46 -44	— —
— —	— —	-41 -50	— —	— —	— —	-35 -50	— —
— —	— -44	— —	— —	— —	— -40	— —	— —
— —	-50 —	— —	— —	— —	-48 —	— —	— —
— —	— -50	— —	— —	— —	— -49	— —	— —
— —	-43 -37	— —	— —	— —	-36 -28	— —	— —
-49 —	-48 -50	— —	— —	-48 —	-48 -42	— —	— —
-36 —	-41 -51	— —	— —	-30 —	-41 -47	— —	— —
-47 -46	— —	-45 -38	-51 -50	-38 -42	— —	-40 -34	-48 -46
-44 -43	-44 —	-43 -44	-43 -48	-38 -40	-44 —	-38 -44	-42 -48
-38 -53	— —	— -51	— -54	-30 -51	— —	— -51	— -54
-44 -49	— —	-50 -44	— —	-42 -48	— —	-50 -44	— —
-49 -50	— —	-48 -47	— —	-46 -50	— —	-48 -44	— —
— —	— —	-43 -53	— —	— —	— —	-36 -53	— —
— —	-46 -46	— —	— —	— —	-42 -46	— —	— —
— —	-50 -48	— —	— —	— —	-48 -42	— —	— —
— —	-34 —	— —	— —	— —	-34 —	— —	— —

**Figure 5.14:** Fingerprint 01 - Preparation 1.1 - Average and Strongest dBm value

As shown in Table 5.14, the fingerprints have many empty cells. This is because no data packets were recorded during that time. In several of the cells there are missing values for some sensors, which is due to the fact that this sensor did not receive any packets during that time. Scenario 1.1 only captured probe requests

from 4 different mobiles, and it was mobile 1 and mobile 5 that generated the most data. The fingerprints are clearly weighted by the values of mobile 1 and 5.

The data analysis also indicates that there are few differences between the values in the two fingerprints. This indicates that fingerprints are not greatly affected by whether they are made based on average or strongest dBm value. The fingerprint based on scenario 1.1, do not contain enough data to create a good enough fingerprint without empty cells. This is because there is a large leap in time between receiving data. The person carrying the phones spent about 10-12 seconds crossing the room (from sensor 1 to sensor 2), while the sensors received data approximately every 6 to 10 seconds (if looking at the combinations of the phones).

Looking at the time stamps in Figure 5.15, one can see that there are big jumps between when data from some of the phone were captured e.g. the leap from 13:55:03 to 13:55:12 and the leap from 13:55:31 to 13:55:40. If the room was divided into 4x7 cells, there would be approximately 2 cells with no information between each cell that could actually be used to compare triangulation results against. This gives an incomplete fingerprints.

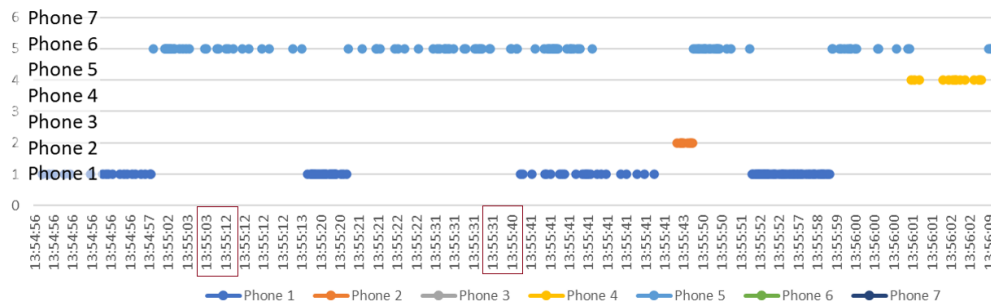


Figure 5.15: Data packet registration

### Scenario 1.2 Fingerprint

The second set of fingerprints is created using scenario 1.2. In scenario 1.2, four mobiles were registered. Mobiles 1 and 5 were both registered under scenario 1.2, and are still the mobiles that send the most data packets. One thing to note is that the person carrying the mobiles walked on average a bit faster under scenario 1.2, but not enough to lower each cell from 3 to 2 seconds. If the milliseconds had been accurate enough, the seconds could have been divided and given a more accurate fingerprint, which also applies to scenario 1.1.

A lot more data was collected during scenario 1.2 than scenario 1.1. This means that the fingerprints created based on scenario 1.2 have far fewer empty cells. By looking at the amount of data captured during this scenario illustrated in Figure 5.17, one can see that the sensors received data at smaller intervals between each received packet.

Average value				Strongest dBm value			
-37 -46	-46 -50	— —	-49 —	-36 -46	-46 -42	— —	-48 —
-45 -53	-53 -53	-52 —	-56 -55	-42 -50	-50 -48	-52 —	-56 -52
— -47	-46 —	— -58	-52 -46	— -47	-34 —	— -58	-49 -44
— —	-54 —	— —	-56 -47	— —	-54 —	— —	-56 -44
— —	-54 -48	— —	-54 -48	— —	-50 -46	— —	-54 -46
— —	-38 -50	— —	-50 -47	— —	-32 -50	— —	-50 -47
-50 -43	— —	— —	-50 -46	-50 -42	— —	— —	-50 -44
-54 -54	— —	— —	-52 —	-54 -54	— —	— —	-52 —
-43 —	— —	— —	— —	-35 —	— —	— —	— —
— —	-50 -54	-48 -44	— —	— —	-50 -50	-40 -42	— —
— —	-56 —	-52 -50	— —	— —	-56 —	-48 -46	— —
— —	— —	-48 -56	— —	— —	— —	-33 -55	— —
-54 -43	-49 -46	— —	— —	-50 -36	-48 -42	— —	— —
-56 —	— —	-48 —	— —	-52 —	— —	-46 —	— —
— -58	-41 —	— -58	— —	— -58	-33 —	— -58	— —
-52 —	-48 -48	— —	— —	-52 —	-44 -44	— —	— —
— —	-46 -47	— —	— —	— —	-46 -46	— —	— —
— -57	-49 -50	— —	— —	— -57	-36 -48	— —	— —
-53 -47	— —	— —	-51 -42	-50 -44	— —	— —	-48 -38
-43 -53	— —	— —	-49 -48	-38 -44	— —	— —	-46 -46
-45 -53	— —	— —	-39 -51	-30 -50	— —	— —	-33 -50

Figure 5.16: Fingerprint 02 - Preparation 1.2 - Average and Strongest dBm value

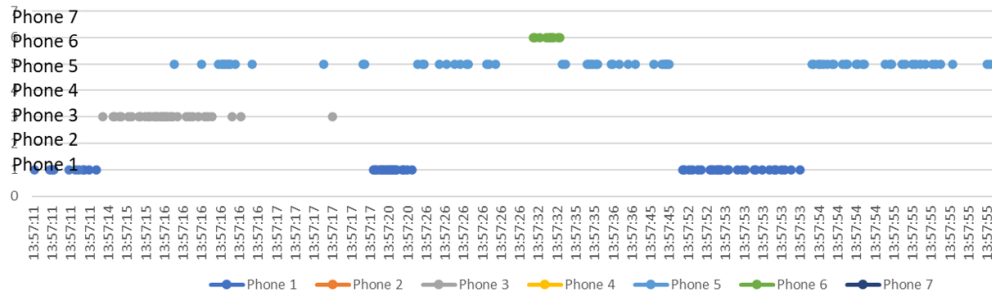


Figure 5.17: Data packet registration

### Combined Fingerprint

By looking at the various fingerprints created for scenarios 1.1 and 1.2, and by combining them, the fingerprints will fill many of each other's empty cells. This can give a better fingerprint and when the fingerprints are combined it is shown that only a few cells do not have information attached to them. These last fingerprints should be able to be used to some degree, but there are still some uncertainties associated with them due to uncertainties around start time and limited data.

Average value				Strongest dBm value			
-43 -46	-46 -50	— —	-47 -42	-36 -46	-46 -42	— —	-46 -42
-46 -51	-53 -53	-52 —	-51 -51	-42 -46	-50 -48	-52 —	-46 -46
— -47	-56 —	— -58	-46 -50	— -47	-34 —	— -58	-33 -44
— —	-54 —	-45 -47	-56 -47	— —	-54 —	-42 -40	-56 -44
— —	-54 -48	-50 -45	-54 -48	— —	-50 -46	-46 -44	-54 -46
— —	-38 -50	-41 -50	-50 -47	— —	-32 -50	-35 -50	-50 -47
-50 -43	— -44	— —	-50 -46	-50 -42	— -40	— —	-50 -44
-54 -54	-50 —	— —	-52 —	-54 -54	-48 —	— —	-52 —
-43 —	— -50	— —	— —	-35 —	— -49	— —	— —
-49 —	-46 -45	-48 -44	— —	— —	-36 -28	-40 -42	— —
-36 —	-52 -50	-52 -50	— —	-48 —	-48 -42	-48 -46	— —
— —	-41 -51	-48 -56	— —	-30 —	-41 -47	-33 -55	— —
-50 -44	-49 -46	-45 -38	-51 -50	-38 -36	-48 -42	-40 -34	-48 -46
-50 -43	-44 —	-45 -44	-43 -48	-38 -40	-44 —	-38 -44	-42 -48
-38 -55	-41 —	-41 -54	— -54	-30 -51	-33 —	— -51	— -54
-48 -49	-48 -48	-50 -44	— —	-42 -48	-44 -44	-50 -44	— —
-49 -50	-46 -47	-48 -47	— —	-46 -50	-46 -46	-48 -44	— —
— -57	-49 -50	-43 -53	— —	— -57	-36 -48	-36 -53	— —
-53 -47	-46 -46	— —	-51 -42	-50 -44	-42 -46	— —	-48 -38
-43 -53	-50 -48	— —	-49 -48	-38 -44	-48 -42	— —	-46 -46
-45 -53	-34 —	— —	-39 -51	-30 -50	-34 —	— —	-33 -50

**Figure 5.18:** Fingerprint 03 - Combination of 1.1 and 1.2 - Average and Strongest dBm value

If one divide the room differently, for example only dividing it into 4 cells, one will (if one ignores the uncertainties around start time) get a usable fingerprint since all the cells can be covered.

The fingerprinting method was a strong candidate for geolocation, although it had some drawbacks. There are many situations where the method does not work very well. This applies if the fingerprint is to be created for a larger area, e.g. a city, since it will be very time consuming to create the fingerprint. This also applies to indoor areas, where small changes such as moving a closet may require a new fingerprint. The fingerprint method can work very well in many situations, but this requires a very good fingerprint and well-done preparation. The method will work better if the goal is crowd analysis and the focus is not on the individual since then the exact positioning for the individual is not as important.

The method seems to work but not with the data set available in the project due to small amount of data and uncertainties about the start time. For the method to have worked for this project, the data collection for Scenario 1.1 and 1.2 should have been collected in a different way. The person should have been standing still for an extended period of time in each cell to collect more data. As of now, a small error in the data set will change the value of an entire cell. If the data collection

method changes, the method should in theory work.

## 5.7 Triangulation or Multilateration using signal strength - Analysis

The Triangulation and Multilateration methods using signal strength, as described in Section 3.3.4, are based on the measured dBm from each data packet sent from the mobile unit and received at the sensors. The data packets from one mobile unit was received at several sensors at the same time and these were coordinated to find an average value that can be used for triangulation and Multilateration. Both triangulation and multilateration methods were applied to the data set for the same scenarios.

### 5.7.1 Experiment location and placement of mobile units

To test the triangulation and Multilateration methods, scenario 2 and scenario 3 were used. These scenarios were chosen because in both of these scenarios, the mobile phones are at a fixed location in the room. It therefore makes it possible to see how accurate the results are.

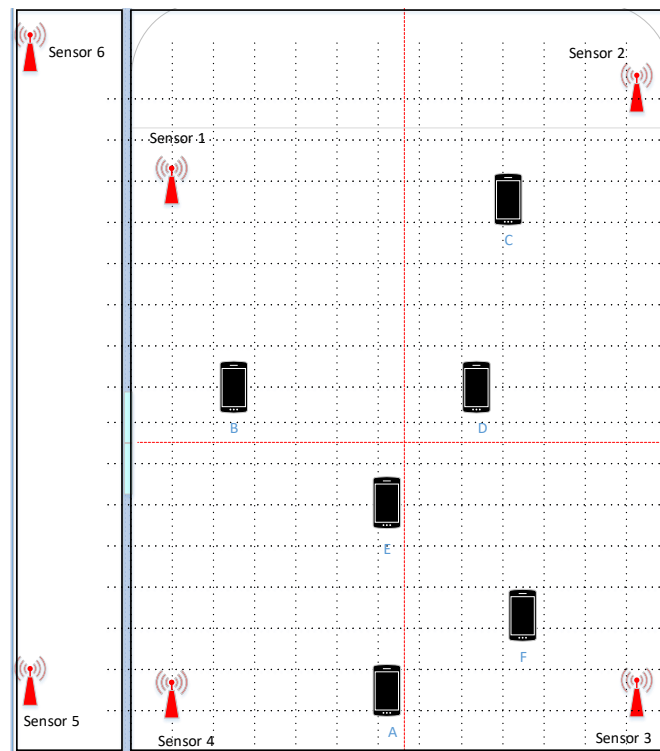


Figure 5.19: Scenario 3 - Standing still

The two scenarios were used for slightly different purposes. Scenario 3 was

used to test triangulation, while scenario 2 was used to analyze and adjust for unknown output power on the individual phones (described in Section 5.7.3).

In scenario 3, the mobile units were placed on six different locations in the room. The phones were in screen saver mode (screen turned off automatically) and was sporadically activated by “opening” the phones to look at the display at some intervals. The location of the mysterious number 7 phone was unknown, and the result of this trial could help determine where this phone was located, possibly if it was not a mobile phone at all and rather a laptop or other device that has been mistakenly white listed. The location of the six known phones are estimated and drawn in figure 5.19 based on the ground truth video.

Scenario 3 started at 14:09 and lasted for 1 minute and 50 seconds. In that period, 320 data packets were collected from the mobile phones. Table 5.3 shows the amount of data packets that was sent and received with respect to the phones and sensors in scenario 3.

Sensor Phone	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6	Total packets
Phone 1	19	9	15	15	6	0	64
Phone 2	3	4	5	3	0	3	18
Phone 3	0	8	5	4	5	7	29
Phone 4	1	5	4	8	2	0	20
Phone 5	29	31	33	30	18	24	165
Phone 6	6	6	3	3	3	1	22
Phone 7	0	0	1	0	0	1	2
Total pr phone	58	63	66	63	34	36	320

**Table 5.3:** Amount of data packets sent and received

The mobile phones were sending data packets at sporadic intervals. The diagram in Figure 5.20 show the data packets distributed over time, and how the phones are sending a cluster of Probe Request data packets (note, each point represents several data packets).

As seen in Figure 5.20, the point indicated at 14:09:51 contain many measurements of received data packets from phone 5. Each of the sensors have received several data packets within the same 2 seconds with different signal strength values. The figure shows that some sensors are receiving more data packets than other sensors, and that the sensors are not synchronized in time (down to milliseconds). It is assumed that some of the data packets are received simultaneously by several sensors but timing wise it is not possible to identify which packets are received by several sensors.



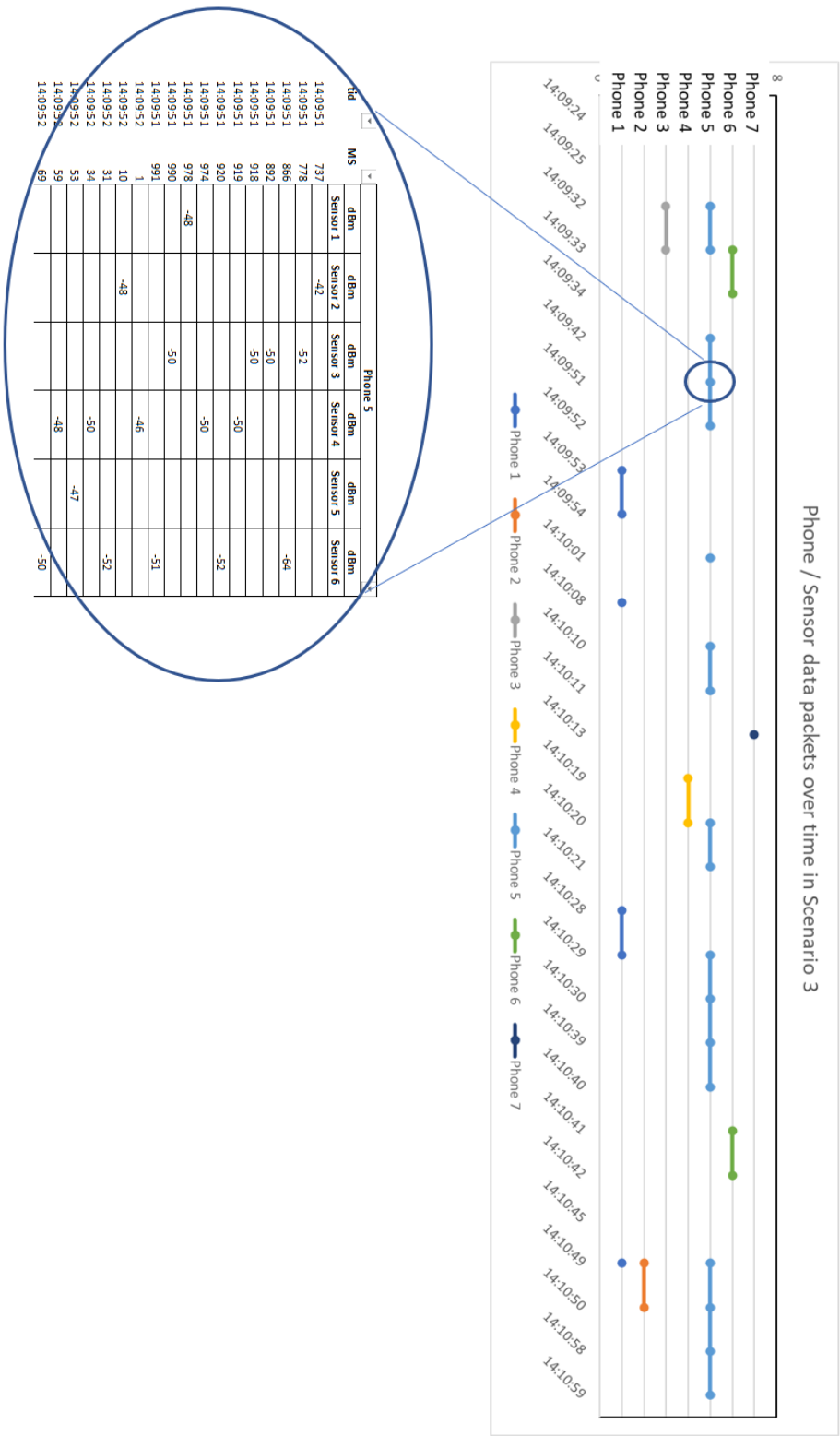
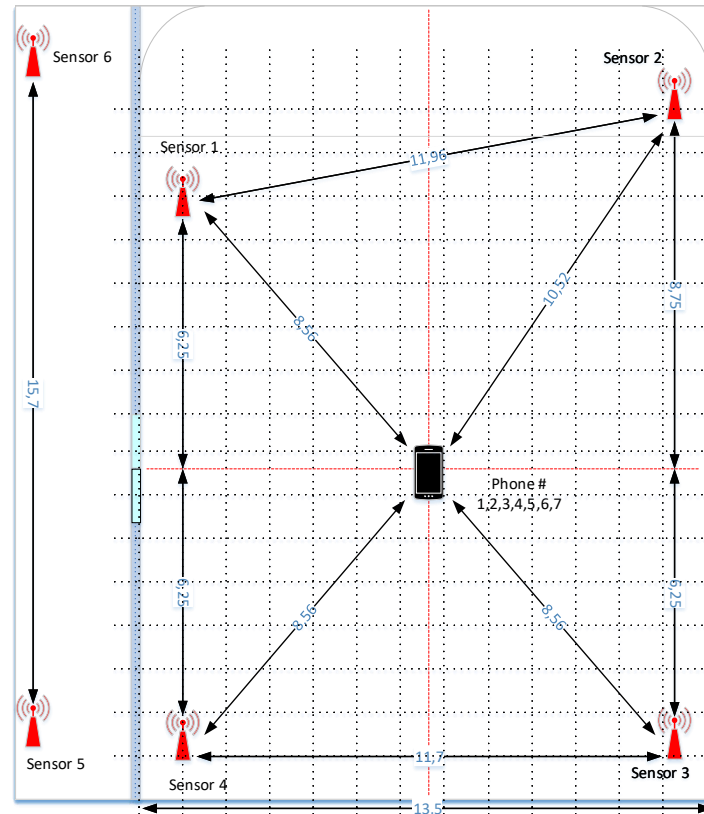


Figure 5.20: Data packet distribution over time

### 5.7.2 Calculations used by both methods

The signal strength is measured in dBm and is typically in the range -30 dBm to -55 dBm. We estimate the output signal strength from a phone for the Wi-Fi part to be between 13 dBm and 20 dBm. To calculate the distance between the transmitter (mobile phone) and the receiver (sensor) the reduction in signal strength with respect to distance has to be calculated. To evaluate the distance from the mobile phone to the sensor, the reduction in signal strength in free space is calculated by use of a path loss transmission formula presented by Friis [39] (Friis transmission equation) and explained in Section 3.3.4.

When calculating the estimated path loss within the test area, the output value of the mobile phone was first set to 20 dBm but later adjusted to different levels for the different phones based on estimated values when physical location was known in the static scenario 2 (Figure 5.21). This was to adjust for unknown output power on the individual phones.



**Figure 5.21:** Scenario 2 - Attenuated signal strength

The distance (in meters) from each phone to each sensor was then calculated for all data packets transmitted by the phones and received by all the sensors. Table 5.22 shows the change from the dBm values to the distance in meters, compensated for the individual adjustment for mobile output value and sensor re-

ceiver sensitivity value. The first table shows the original values, while the second table is the same values converted to meters.

		Phone 5					
		dBm	dBm	dBm	dBm	dBm	dBm
Time	MS	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6
14:09:51	737		-42				
14:09:51	778			-52			
14:09:51	892			-50			
14:09:51	918			-50			
14:09:51	919				-50		
14:09:51	920						-52
14:09:51	974				-50		
14:09:51	978	-48					
14:09:51	990			-50			
14:09:51	991						-51
14:09:52	1				-46		
14:09:52	10		-48				
14:09:52	31						-52
14:09:52	34				-50		
14:09:52	53					-47	
14:09:52	59				-48		
14:09:52	69						-50
		Phone 5 (meter)					
		meter	meter	meter	meter	meter	meter
Time	MS	Sen. 1	Sen. 2	Sen. 3	Sen. 4	Sen. 5	Sen. 6
14:09:51	737		4,5				
14:09:51	778			8,4			
14:09:51	892			6,7			
14:09:51	918			6,7			
14:09:51	919				10,1		
14:09:51	920						8,7
14:09:51	974				10,1		
14:09:51	978	7,5					
14:09:51	990			6,7			
14:09:51	991						7,8
14:09:52	1				6,4		
14:09:52	10		8,9				
14:09:52	31						8,7
14:09:52	34				10,1		
14:09:52	53					9,8	
14:09:52	59				8,0		
14:09:52	69						6,9

Figure 5.22: The change from dBm to meters

The sensors were not calibrated towards each other and did have different antennas and there was some local variations in the dBm measurement. Based on the test result from scenario 2, an individual constant was entered for each sensor to adjust for variations in the signal reception measurement as a coefficient. The individual constant was adjusted during the static scenario 2 by estimating and correlating the measurement for all reception of data packets from the different phones from the same physical location and received at the different sensors.

### 5.7.3 Triangulation

By using trigonometry for the triangulation based on the calculated distances, the triangulation result was placed in the same coordinate system as the room was divided into, and a plot for the location based on the triangulation was performed.

The triangulation was based on data from each of the sensors and the triangulation was done by sensor pairs. Since the geolocation was confined within a defined area, the calculated locations that was outside the defined area was discarded. The sensor pairs together with the phones (based on the dBm measurement) formed a triangle and the corresponding X/Y coordinates was calculated e.g. sensor 4 and sensor 1 formed a triangle based on the measured signal strength received at sensor 4 and 1 from phone E. At the same time sensor 4 and sensor 3 formed another triangle to the same phone based on the measured signal strength at sensor 4 and 3, and so on. This is illustrated in figure 5.23.

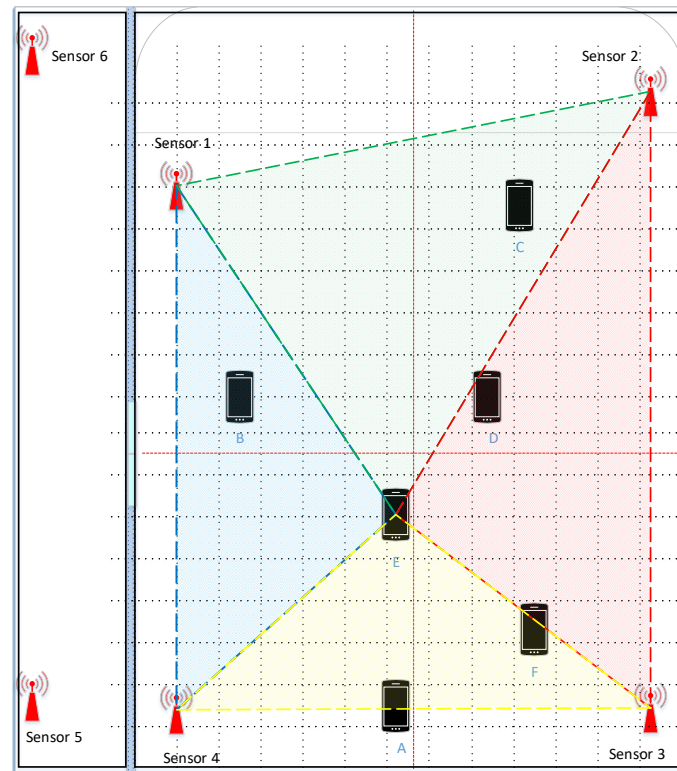


Figure 5.23: Triangles formed by the sensors

### 5.7.4 Trilateration/Multilateration

Multilateration uses measurements from more than two sensors to calculate the position of a transmitter and takes into consideration the position of each sensor used. The position of the sensors was coordinated with the rooms coordinate sys-

tem in the X/Y coordinate system. By using more than two sensors in the calculation, there will be only a single location identified for the localization of the transmitter and no extra evaluation of valid localization was needed.

The Trilateration was calculated for all combinations of three sensors within the room to locate the mobile. E.g. sensor (1-2-3), (2-3-4), (3-4-1) and (4-2-3) was used to form four locations based on the same data packet, in addition, the multilateration with all four sensors (1-2-3-4) was performed to see how the localization was affected by errors in the measurements.

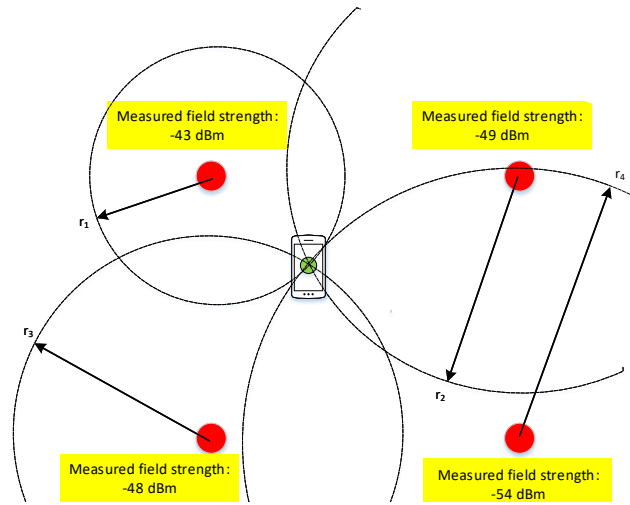


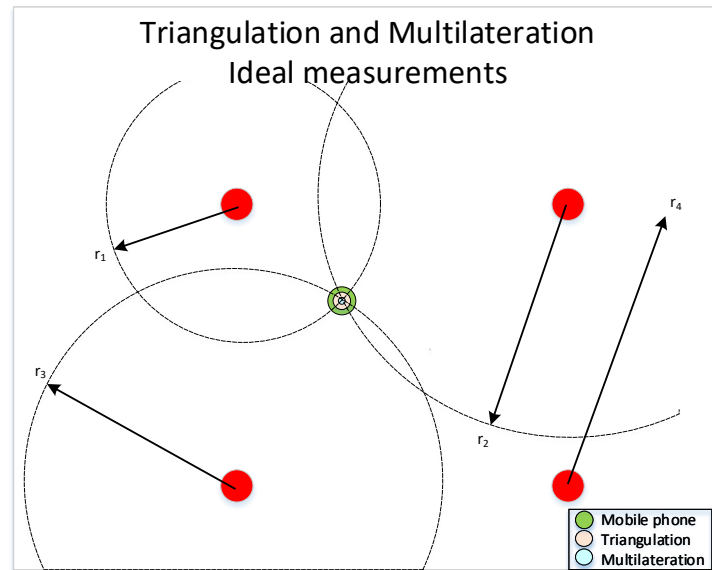
Figure 5.24: Circles formed by the sensors

### 5.7.5 Uncertainties: Triangulation and Multilateration

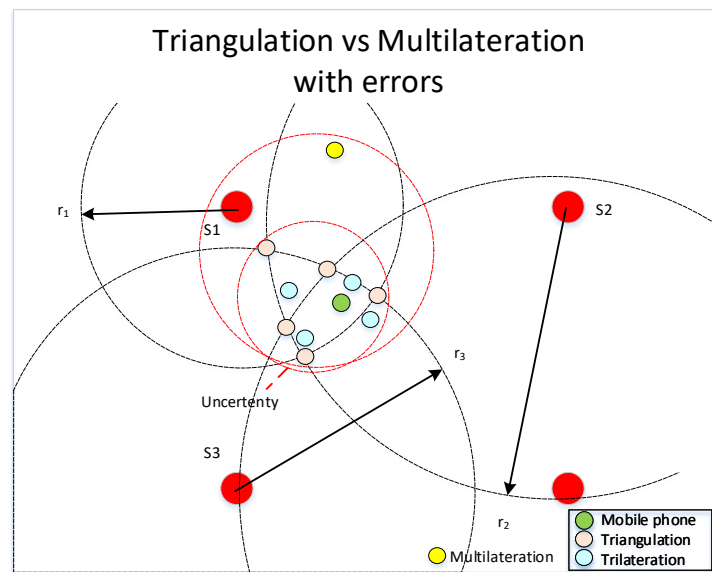
There are always uncertainties in the measurement of the signal strength and therefore the calculation of distances. This gives some uncertainties in the localization of the mobile.

The deviation in measurements can be due to signals being affected by the surrounding objects such as buildings and interference and this can give incorrect results. Therefore, since the signal strength can be affected, it may be a good idea to confirm the location by carrying out multiple independent Trilateration's/Multilateration's when locating the transmitting unit, if possible.

When calculating the positions using Triangulation or Multilateration with ideal numbers where the circles intersect each other at the same location the result will be the same both for Triangulation and Multilateration. When errors are introduced in the measurement there will be some deviation in the intersection of the circles and this cause some localization errors.



**Figure 5.25:** Triangulation and Multilateration with ideal measurements



**Figure 5.26:** Triangulation and Multilateration with errors

We see that using Triangulation on the different combinations of sensors gives six different locations, based on each of the intersection of circles. If we use Trilateration on each on the three combinations of sensors, the localization gives four different locations which are centered around the mobile, but if we use the Multi-

lateration of all four sensors, the error is aggregated by all of the sensors and the localization of the unit is far from the real position. As an example we use the ideal example presented in Section 3.3.4, but put in some errors in the measurements.

Sensor	Location for x(m)	Location for y(m)	Measured dbm	Calculated distance "r"
Sensor 1	0	12,5	-37,41 dbm	4,5m
Sensor 2	11,7	15	-41,39 dbm	9,7m
Sensor 3	0	0	-52,01 dbm	17,1m
Sensor 4	11,7	0	-45,99 dbm	13,4m

**Table 5.4:** Ideal example with added errors

The calculation of these parameters will give the location of the mobile, but since there are some errors in the measurements, there will be different intersections for all circles, which gives several different results. The calculation with Trilateration and Multilateration gives the following result:

	X value	Y value	Tri/Multi-lateration
Sensors (1-2-3)	2,6	14,1	Trilateration
Sensors (2-3-4)	1,0	14,1	Trilateration
Sensors (3-4-1)	1,0	12,62	Trilateration
Sensors (4-1-2)	2,9	12,6	Trilateration
Sensors (1-2-3-4)	2,6	21,5	Multilateration

**Table 5.5:** Calculated X and Y values

The result is not conclusive for where the mobile is located, but the average of the Trilaterations of all the sensors combinations gives a more accurate location than each of them by themselves and better than the Multilateration of all sensors together.

Another uncertainty is that it appeared that the sensors were not synchronized in time (different sensors registered the reception of the same data packet at different times with respect to milliseconds). All data that was measured during the same second at the different sensors was regarded as the same data packet and if there were several data packets at the same second to one sensor, this was averaged to get a single signal strength measuring to be used during geolocation. The average data was plotted in to the room layout. The analysis of the data discovered that the signal received at the sensors have a large spread in values even though the phones are physically standing still at one location. The average of all the data gives an indication of the location of the phones within an area.

Figure 5.27 illustrates the calculated position of mobile 5 based on the data. The four colored circles show the calculated positions based on pairs of sensors. The red circle represents where the majority of the individual positions were before the average for each sensor combination was combined. The average posi-

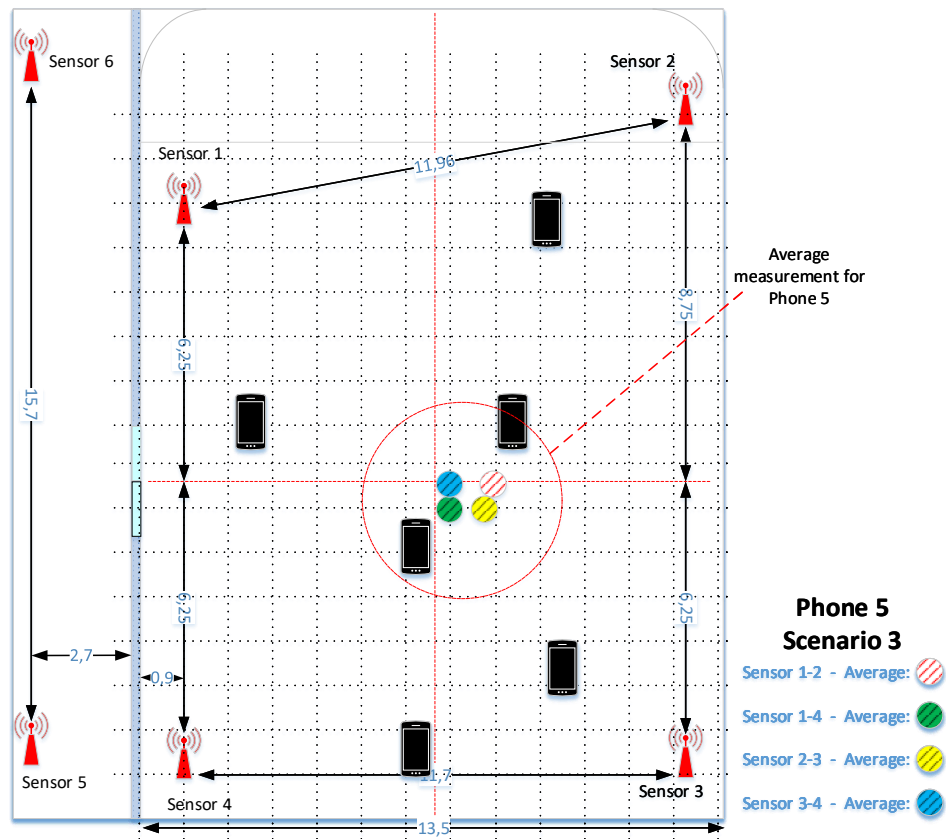


Figure 5.27: Phone 5 results

tions of the sensor combinations are centered in a small area, which means that they provide approximately the same positioning for mobile 5. Although all the locations are gathered in the same area, it is difficult to say which mobile is mobile 5 since the positioning is located between two of the mobiles. The location of the mobiles in the picture is not necessarily accurate, since they are based on the ground truth video. If the red circle is used as template, we can assume that mobile 5 is the mobile that is completely inside the circle.

This locating method can be used, but the accuracy is highly dependent on the quantity of data received and the quality of the signal strength measurement. If the received signal is distorted due to obstacles or multipath, the accuracy can be affected quite a lot. The accuracy can be evaluated by running a fast fingerprinting scenario to establish an overall feeling of the general disrupting elements in the area and use this to help evaluate the data and its accuracy.

Depending on the purpose of the localization and the requirement for the objective, it may have good enough accuracy. As an example, if the geolocation experiment is performed outside in a large area, the general signal propagation



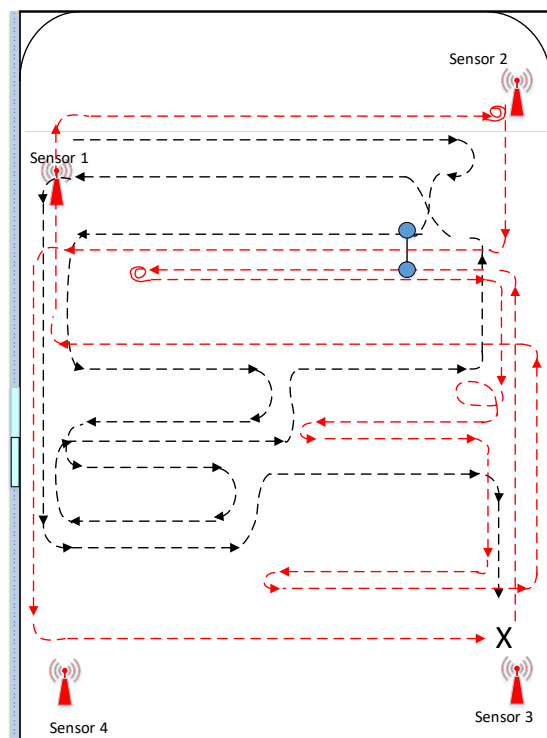
may be spread over longer distance so a higher degree of granulation of the measured signal strength can be more accurate in addition to the sensors spread over a larger area so the triangulation or multilateration are more effective. Another example is if the localization is within a building, and the room is divided into 4 segments and the requirement is only to know which of the segments the mobile is in, this method will give a good enough result as long as the mobiles are not in the boundary between the segments. This method was further tested in chapter 6, where it was tested on several scenarios to see if it could produce the desired results: Is it possible to see how the mobiles have moved? Is there enough data and accuracy to distinguish between a fixed and a random pattern?



## Chapter 6

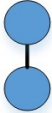
# Triangulation and Multilateration results

Each of the scenarios is described below and each of the scenarios shows the room plan with the movement of the mobiles indicated with dotted lines. The dotted lines may have two different colors indicating two different individuals walking around with different mobiles at the same time as illustrated in Figure 6.1. The two connected blue circles indicate an interaction between the persons carrying the mobile phones, for example, exchanging the mobiles or merging the mobiles.



**Figure 6.1:** Visualization of movement from multiple people

Please Note: the timing for the movement in the room is noted in the scenarios, and this is estimated by studying the video so this is approximate timing. The triangulation and multilateration calculations are pinpointing the mobiles at various locations depending on the data transmission from the mobiles. The geolocations of the different mobiles are indicated with different colors.

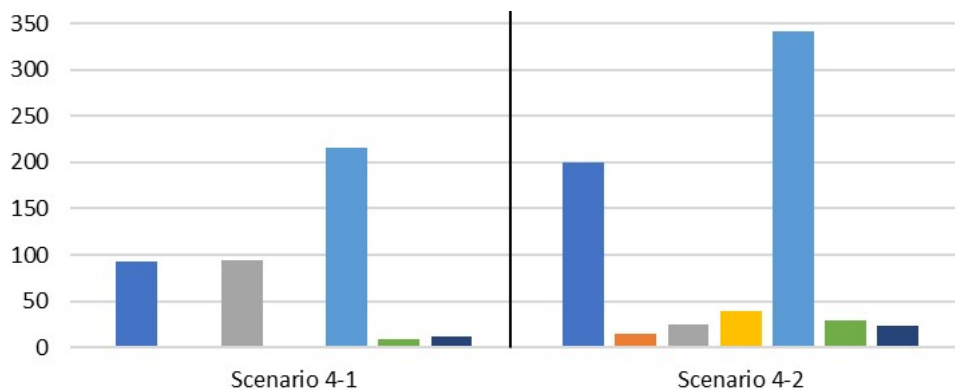
Multilateration	Triangulation	Interaction point
Mobile 1 ●	Mobile 1 ▲	
Mobile 2 ●	Mobile 2 ▲	
Mobile 3 ●	Mobile 3 ▲	
Mobile 4 ●	Mobile 4 ▲	
Mobile 5 ●	Mobile 5 ▲	
Mobile 6 ●	Mobile 6 ▲	
Mobile 7 ●	Mobile 7 ▲	

**Figure 6.2:** Color code - mobile phones

The geolocation of the mobiles is based on both triangulation and multilateration calculations and results from both the localization techniques are displayed in the scenario drawings. The timing located close to the localization is when the data packets was registered to compare with the actual timing. Please note, due to the number of locations, only a few localization's are marked with time stamp to give an indication of accuracy, the complete listing is shown in the appendix.

## 6.1 Scenario 4

The goal for Scenario 4 was to evaluate whether a mobile can be located accurately enough and if possible find out how the mobile person has moved when it is not connected to an access point but in a search mode.



**Figure 6.3:** Data packets recorded during Scenario 4

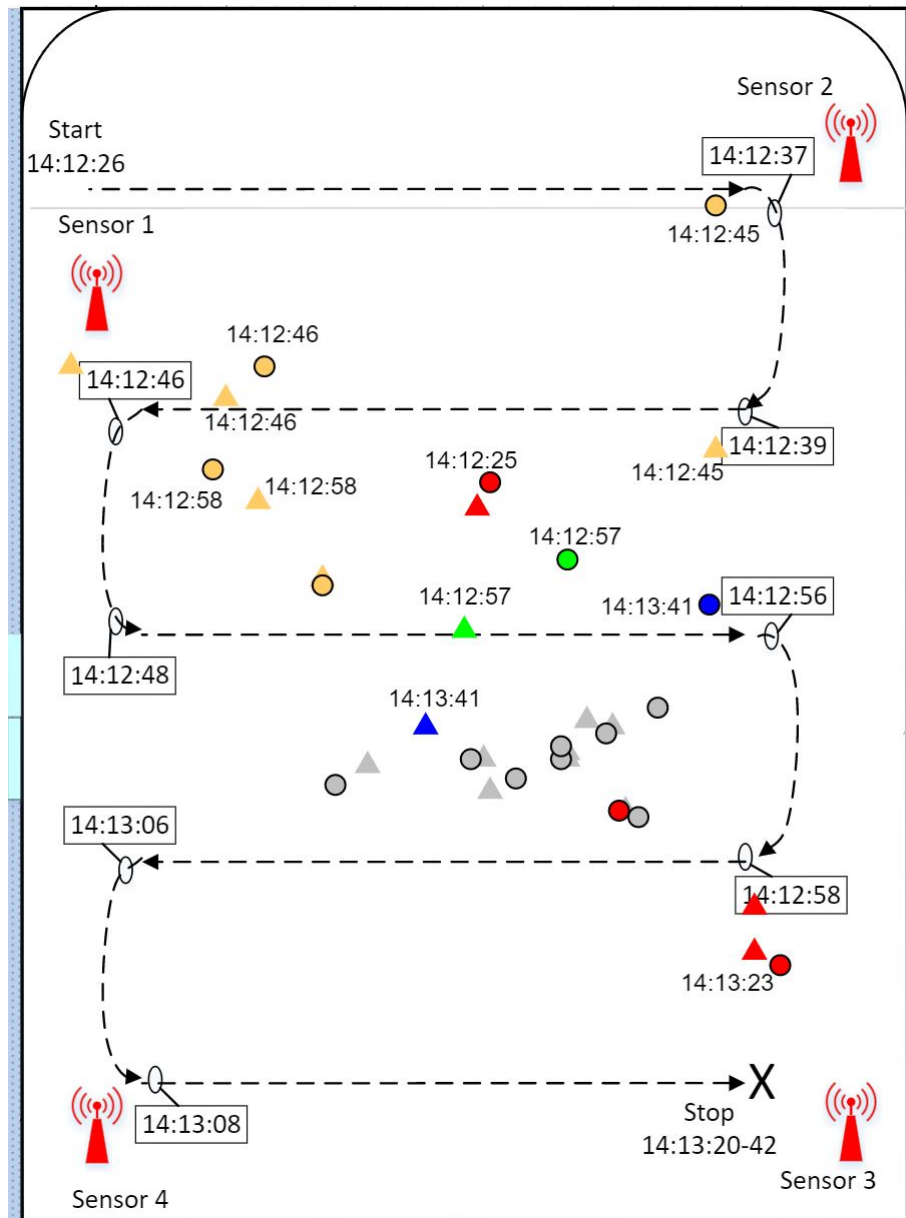
Another goal was to see whether the unit can be pin pointed accurately enough to be able to distinguish between a fixed pattern and a random pattern. This scenario is divided into two phases where in the first phase the mobile is moving in a fixed pattern, while in the second phase, the mobile is moving in a random pattern. The number of data packets recorded during Scenario 4.1 and 4.2 is illustrated in Figure 6.3.

### 6.1.1 Scenario 4.1

In Scenario 4.1 the person moved in a set pattern, and in Figure 6.4 it is indicated the movement and direction of the phones by the dotted line and the arrows estimated by studying the video. The time label at each turn indicates the approximate time when the mobile phones are in that position. The dots represent multilateration, and the triangles represent triangulation and they mark the estimated position of the mobile unit when using multilateration or triangulation. The timing located near the dots are the exact time when the sensors received the data packages used for that triangulation or multilateration. When analyzing the position compared to the actual physical location, there seem to be some deviations in the accuracy, but it gives a clear indication of where the phones are generally in the room and where it moves.

The start of the scenario was estimated to be 14:12:26. It was noticed that when the Triangulation and Multilateration deviate a lot from each other, their individual accuracies are less than when the result are close together (e.g. Mobile 3 has a Triangulation at 14:12:45 that is far away from the multilateration localization at 14:12:45, this indicates low accuracy, but compared with Mobile 3 triangulation localization at 14:12:46 that is close to Multilateration localization at 14:12:46 and compared with the actual location, this is far better accuracy). This was seen through all the scenarios.

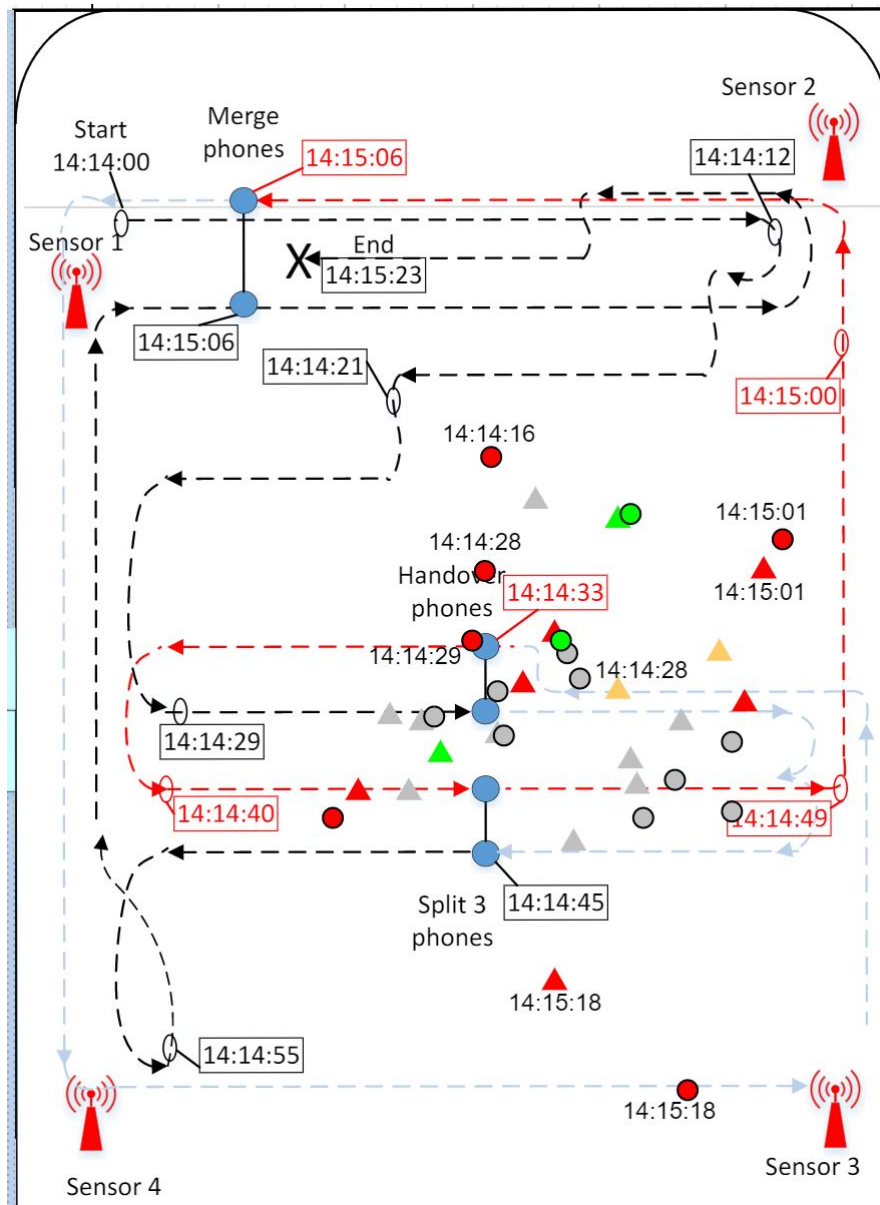
The analysis indicates that when the phones moves across the room, the triangulation follows the movement very broadly. The scenario ends approximately between 14:13:20 and 14:13:42 (the phones did linger at sensor 3 for a while). The triangulation shows that the phones have moved over to the other side of the room at the end of the scenario which indicates a direction of movement, but not following the exact path of the movement. Another observation is that the phones seem to be following each other as a group even though it seems to be a couple of meters between them (in reality they were all within half a meter apart), it could be determined that they are moving together in the same direction and speed. In a small room, this does not give enough granularity to be very accurate, but in a larger space e.g. a park or in an open area, this could give valuable information.



**Figure 6.4:** Scenario 4.1 - Movement and location results

### 6.1.2 Scenario 4.2

The Scenario 4.2 the mobiles are moving in a random pattern. The scenario started at 14:14 and lasted for 1 minute and 25 seconds. There are two persons moving around in the room, one following the black dotted line and the other following the red dotted line. At some points in the room, the persons swapped phones. Note: the light blue lines are persons moving without the mobiles.



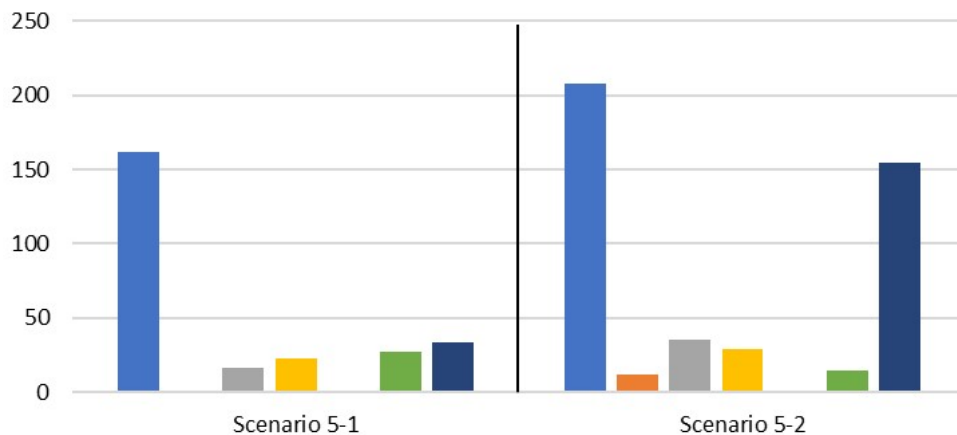
**Figure 6.5:** Scenario 4.2 - Movement and location results

As indicated the Figure 6.5, the phones are pinpointed at several locations

along the path of the movement of the persons. As seen in the figure, some of the phones are more active than others, and the sensors tend to measure signal strength that tend to place the mobiles through geolocation towards the center of the room, but the mobiles were detected in the general area and by using the information from several mobiles a general movement can be indicated. If data was collected over a longer period it is assumed that the general geolocation would be more accurate due to more data collected and evaluated.

## 6.2 Scenario 5

The goal for Scenario 5 was to evaluate whether a mobile can be located accurately enough and if possible find out how the mobile has moved when the mobile is connected to an access point and streaming data. This scenario is divided into two phases where in the first phase the mobile is moving in a fixed pattern, while in the second phase, the mobile is moving in a random pattern. The movement pattern for scenario 5 is very similar to scenario 4 but the difference between Scenario 5 and Scenario 4, is that in this scenario the mobile phones were streaming data. The data was analyzed to see if the increase of data packets collected during Scenario 5 can give a better accuracy when determining the movement of the phones. The number of data packets recorded during Scenario 5.1 and 5.2 is illustrated in Figure 6.6.



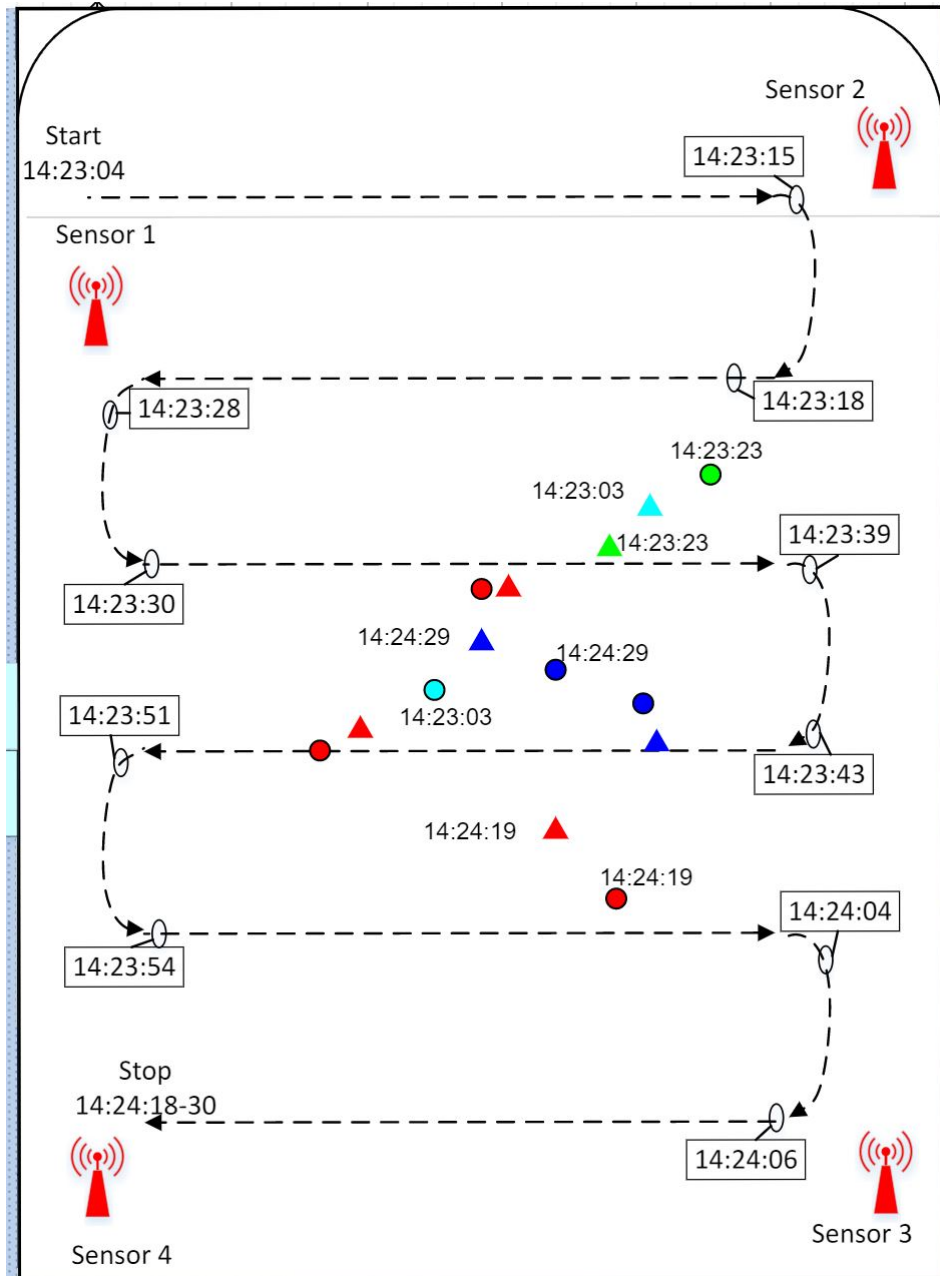
**Figure 6.6:** Data packets recorded during Scenario

### 6.2.1 Scenario 5.1

In Scenario 5.1 the person moved in a set pattern, and in Figure 6.7 the movement and direction of the phones is indicated by the dotted line and the arrows estimated by studying the video. The dots mark the estimated position of the mobile unit when using geolocation techniques. Just like in 4.1, when analyzing the triangulation-based position compared to the actual physical location, there seem



to be some deviations in the accuracy, but gives a clear indication of where the phones are generally in the room and where it moves.



**Figure 6.7:** Scenario 5.1 - Movement and location results

The start of the scenario was estimated to be 14:23:04. The actual data collected was minimum at the start of the scenario. It was a single data transmission burst at 14:23:04 but then no good data transmission to calculate triangulation or

multilateration before 14:23:23. The phones had been moving through the room and the triangulation shows the phones to be at the correct side of the room.

Just like in 4.1, the analysis indicates that when the phones move across the room, the triangulation is following the movement very broadly. The scenario ends at approximately between 14:24:18 and 14:24:30 (the phones did linger at sensor 4 for a while). The triangulation shows that the phones has moved over to the other side of the room at the end of the scenario which indicates a direction of movement, but not following the exact path of the movement. Just like in 4.1, an observation is that the phones seem to be following each other as a group even though it seems to be a couple of meters between them (just as in 4.1, in reality they were all within half a meter apart), it could be determined that they are moving together in the same direction and speed.

In a larger environment, this behavior could be an indication that they may be in some way connected. This gives the same conclusion as in Scenario 4.1. When in a small room, this does not provide enough granulation to be very accurate, but in a larger space e.g. a park or in an open area, this could give valuable information.

### **6.2.2 Scenario 5.2**

In Scenario 5.2 the mobiles are moving in a random pattern. The scenario started at 14:25 and lasted for 1 minute and 57 seconds. There are two people moving around in the room, one following the black dotted line and the other following the red dotted line. At some points in the room, the persons swapped phones.

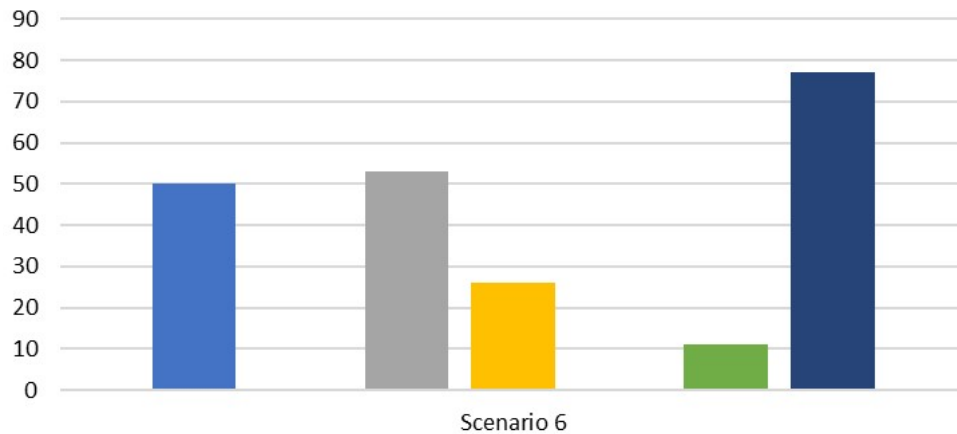
As indicated in Figure 6.8, the phones are pinpointed at several locations along the path of the movement of the persons. As seen in the figure, and just like other scenarios some of the phones are more active than others. Here the pinpointing of the mobiles seems to be following the path of the actual mobile movements. The geolocation calculation seems to locate the mobiles towards the right of the room. The actual physical movement was fast compared to how often the mobiles are sending the probe request frame so the accuracy of tracking the movement in great details are limited, and as for the other scenarios, the tracking will be of more general direction and to a certain degree, the speed it moves.



**Figure 6.8:** Scenario 5.2 - Movement and location results

### 6.3 Scenario 6

The goal for Scenario 6 was to analyze the data to see if the mobile can be located accurately and to evaluate how the mobile has moved while restricted to a corridor. In this scenario the person moved in a set pattern. The number of data packets recorded during Scenario 6 is illustrated Figure 6.9.



**Figure 6.9:** Data packets recorded during Scenario 6

The scenario started at 14:35 and lasted for 58 seconds. The data analysis gave no clear multilateration geolocation due to lack of data measured at three or more sensors that did not give a conclusive answer. The Triangulation and Multilateration method was used to identify the location of the mobile as shown in Figure 6.10.

In Figure 6.10, it is indicated the movement and direction of the phones by the dotted line. There is no video tracking the movement of the phones in the hallway so it is not possible to verify the location of mobile compare to actual physical location. The triangulation and Multilateration did indicate some locations in the hallway where the mobiles most probably was at the time it sent the data packet.

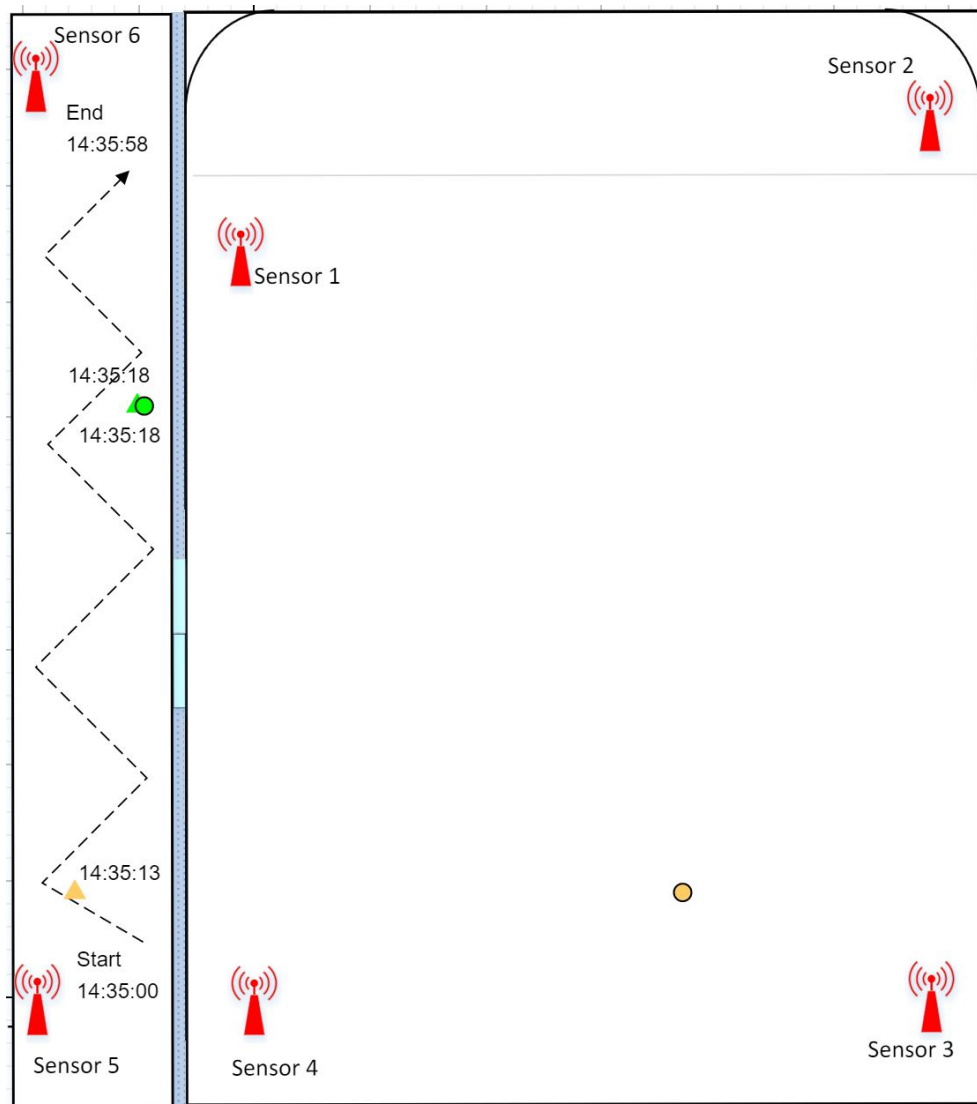
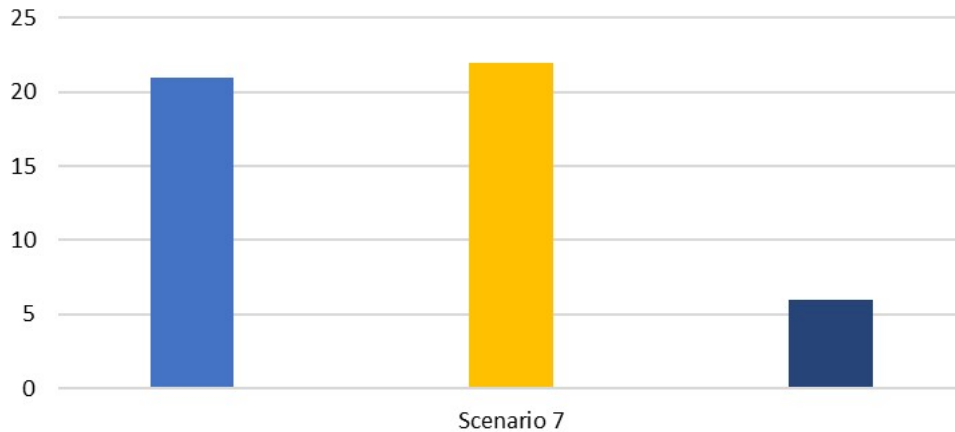


Figure 6.10: Scenario 6 - Movement and location results

## 6.4 Scenario 7

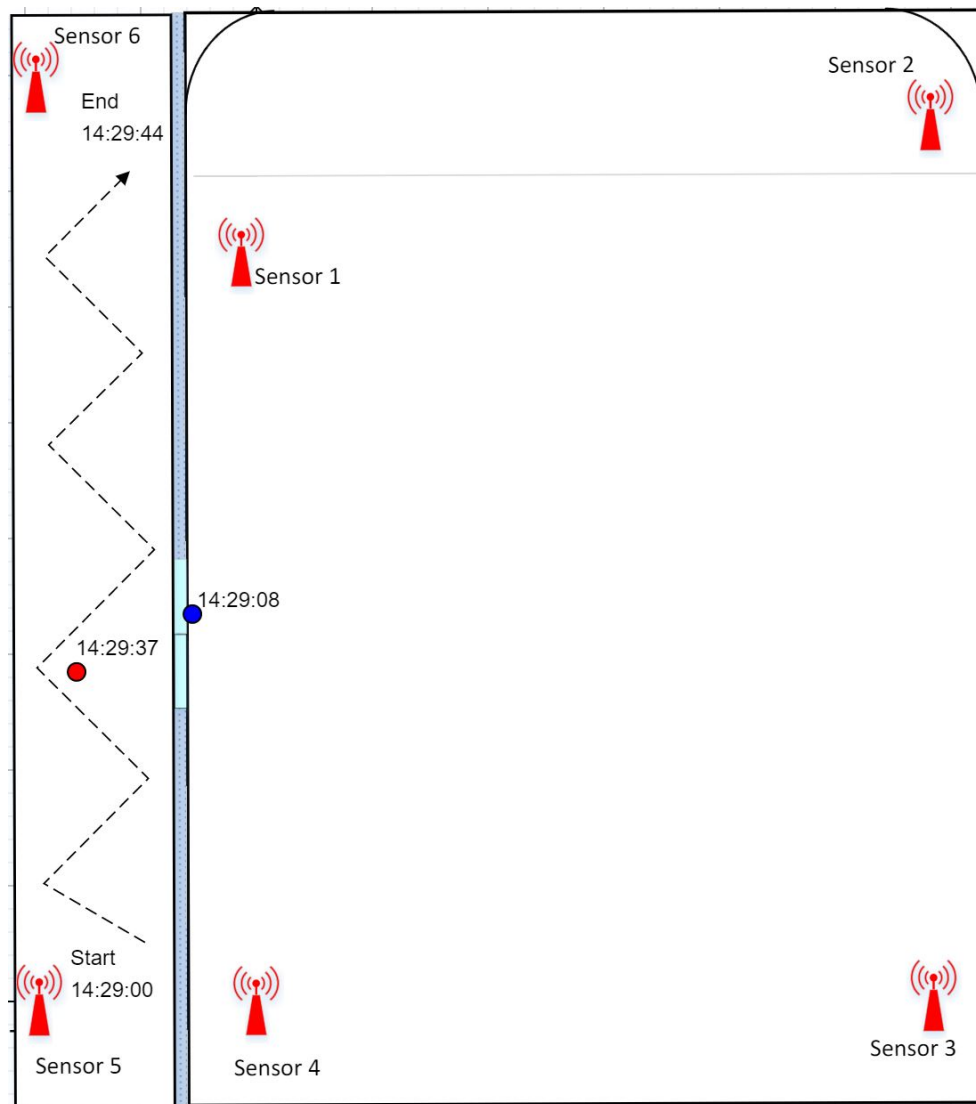
The goal for Scenario 7 was to analyze the data to see if the mobile can be located accurately and to evaluate how the person has moved while restricted to a corridor. The difference from Scenario 6, is that in this scenario the mobile phones are streaming data. In this scenario the person also moved in a set pattern. The number of data packets recorded during Scenario 7 is illustrated in Figure 6.11.



**Figure 6.11:** Data packets recorded during Scenario 7

This scenario was movement in the hallway according to a fixed pattern. The scenario started at 14:29 and lasted for 44 seconds. The data analysis gave no clear multilateration geolocation due to lack of data measured at three or more sensors that did not give a conclusive answer. The Multilateration method was used to identify the location of the mobile as shown in Figure 6.12.

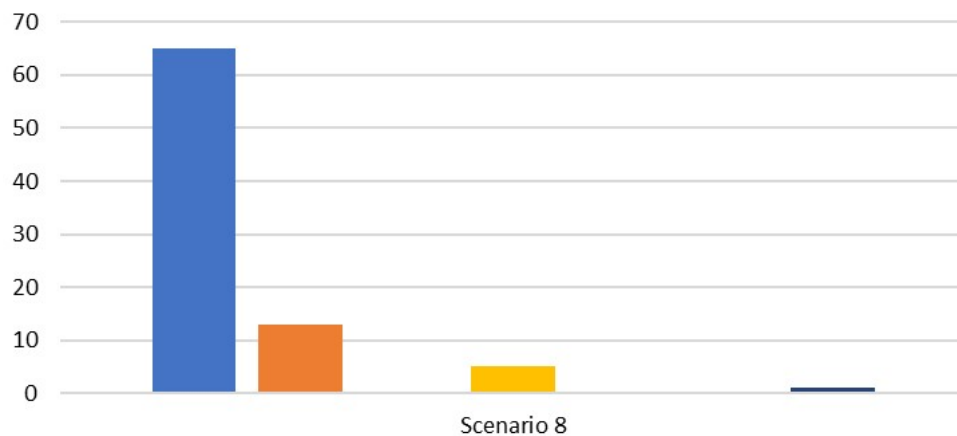
In Figure 6.12, it is indicated the movement and direction of the phones by the dotted line. There is no video tracking the movement of the phones in the hallway so it is not possible to verify the location of mobile compare to actual physical location. The Multilateration did indicate a location in the hallway where the mobiles most probably was at the time it sent the data packet.



**Figure 6.12:** Scenario 7 - Movement and location results

## 6.5 Scenario 8

The goal for Scenario 8 was to analyze the data to see if a mobile can be located correctly based on their location in the corridor or the room. The wall and the person's torso can both influence the signal strength. The focus of this scenario was to see if the device could be successfully located during the transition period between the corridor and the room. The number of data packets recorded during Scenario 8 is illustrated in Figure 6.13.



**Figure 6.13:** Data packets recorded during Scenario 8

This scenario was movement in the hallway and in the room in a random pattern as illustrated in Figure 6.14. The scenario started at 14:37:10 and lasted for 1 minute and 43 seconds. There are two people moving around in the room, one following the black dotted line and the other following the red dotted line. At some points in the room, the persons swapped phones.

The triangulation and Multilateration did indicate a location in the hallway where the mobiles most probably were at the time it sent the data packet. In the room the sensors tend to measure signal strength that tend to place the mobiles through geolocation towards the center of the room, but the mobiles were detected in the general area and by using the information from several mobiles a general movement can be indicated. If data was collected over a longer period it is assumed that the general geolocation would be more accurate due to more data collected and evaluated.



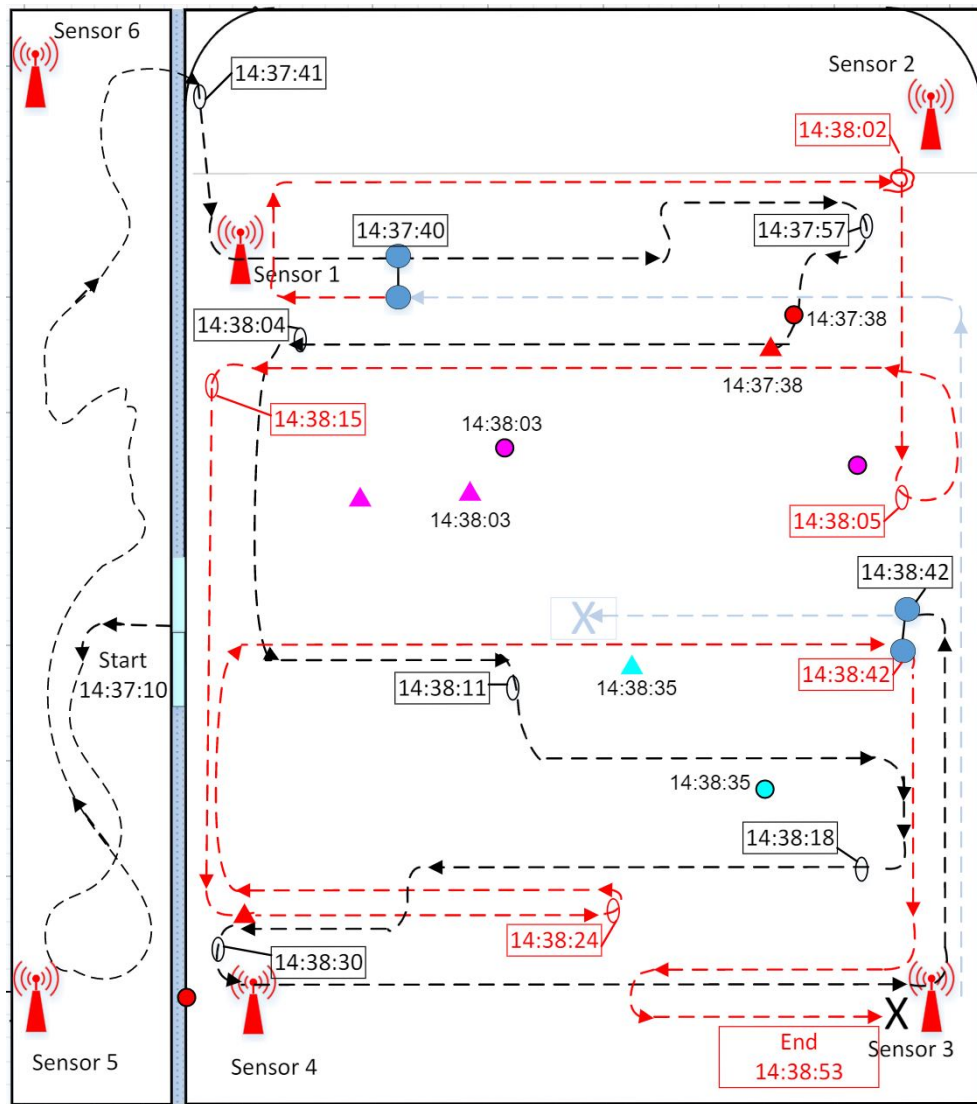
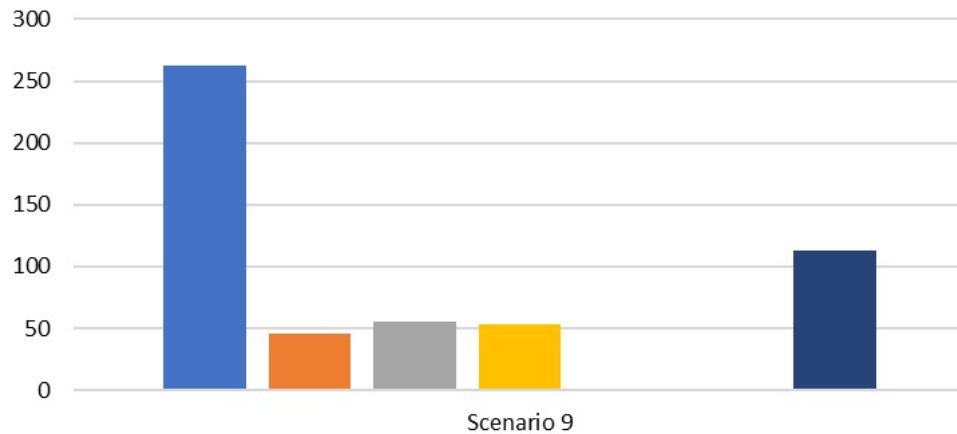


Figure 6.14: Scenario 8 - Movement and location results

## 6.6 Scenario 9

Scenario 9 was the same as Scenario 8, but in this scenario the mobiles were in active use streaming data. The goal is to see if by analyzing the data a mobile can be correctly located in the corridor or the room. The number of data packets recorded during Scenario 9 as illustrated in Figure 6.15.



**Figure 6.15:** Data packets recorded during Scenario 9

This scenario was movement in the hallway and in the room in a random pattern as illustrated in Figure 6.16. The scenario started at 14:30:45 and lasted for 1 minute and 56 seconds, but the phones did linger at sensor 3 for a while at the end. There are two people moving around in the room, one following the black dotted line and the other following the red dotted line. At some points in the room, the persons swapped phones.

As indicated in the figure, the phones are pinpointed at several locations along the path of the movement of the persons. As seen in the figure, and just like other scenarios some of the phones are more active than others. Here the pinpointing of the mobiles seems to be following the path of the actual mobile movements. The actual physical movement was fast compared to how often the mobiles are sending the probe request frame so the accuracy of tracking the movement in great details are limited, and as for the other scenarios, the tracking will be of more general direction and to a certain degree, the speed it moves.

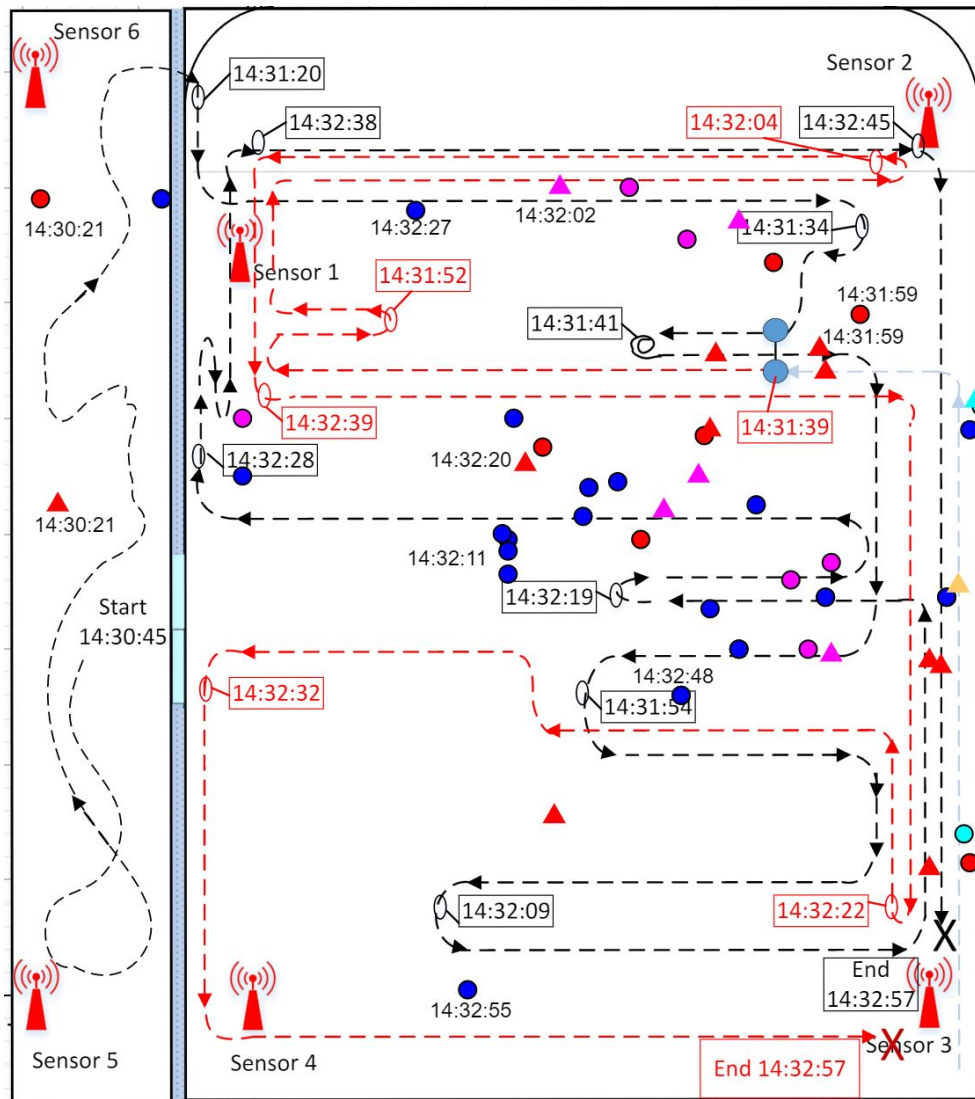


Figure 6.16: Scenario 9 - Movement and location results



## Chapter 7

# Discussion and Conclusion

The first two methods tested, Angle of Arrival (5.4) and Time of Flight (5.4) could not be used to locate the devices in the data set. The Angle of Arrival method would require some special equipment like sensors with special antenna configuration and the necessary direction-finding features, and this was unavailable for the project. The second method, Time of Flight, also required some special equipment for ensuring time synchronization down to nanoseconds that was not available. The collected data did not contain timestamps that had the necessary resolution of nanoseconds. Timestamps in the data set were stated in milliseconds giving a resolution of 300,000 meters which is not accurate enough since the entire test range was less than 25 meters in diameters. Nanoseconds could provide an accuracy of around 0.3 meters and it would have been necessary to accurately identify the location of the device in the room used for the data collection.

It does not mean that the Angle of Arrival or Time of Flight methods are not good in theory; on the contrary, it is believed to be able to produce good results especially in combination with Multilateration. Unfortunately, this could not be properly tested with the available data set and therefore it was not possible to confirm how well the methods worked.

The third method, Fingerprinting using signal strength (5.6), does not need any special equipment and therefore appeared to work well. The goal was for the fingerprint to become ground truth data that could be used to compare the next measurement against, but this turned out not to work with the available data. The obstacle for fingerprinting was the limited amount of data and uncertainties around start times. There was too little data to make good fingerprints, and the fingerprints ended up with location cells without any data. Combining multiple fingerprints could result in a stronger fingerprint with information in all location cells, but this combined fingerprint is unlikely to be of any use. One of the reasons for this is that both fingerprints have uncertainty around start time and it will amplify the uncertainty when combined since it can result in one cell containing the result of measurement belonging to two cells that are both based on incorrect data. There were uncertainties around the start times of the fingerprinting and this had to be estimated and since the data collection in each cell was limited to

few seconds, the mobiles could be registered in the wrong cell if the start time is 1-3 seconds wrong which could displace the entire fingerprint. It is therefore not to be expected that the results for the fingerprints are accurate.

The method seems to work but not with the data set available in the project due to small amount of data and uncertainties about the start time. For the method to have worked for this project, the data collection for Scenario 1.1 and 1.2 should have been collected in a different way. The person should have been standing still for an extended period of time in each cell to collect more data. As of now, a small error in the data set will change the value of an entire cell. If this changes, the method should in theory work.

The last method was Triangulation and Multilateration using signal strength (5.7), and it was the only method that gave results with the given data set. The accuracy is highly dependent on the quantity of data received and the quality of the signal strength measurement. If the received signal is distorted due to obstacles or multipath, the accuracy can be affected quite a lot. Both the Multilateration and Triangulation was sensitive to the quality of the signal strength measurement since both methods was using the same measurement as input data. When the measured data was accurate the two methods gave exact the same answer, but when the input data was not accurate, the result from the Triangulation and Multilateration deviated from each other and this could be used to give an indication of accuracy. The accuracy can also be evaluated by running a fast fingerprinting scenario to establish an overall feeling of the general disrupting elements in the area and use this to help evaluate the data and its accuracy.

Depending on the purpose of the localization and the requirement for the mission, it may have good enough accuracy. It will be difficult to identify any interaction between persons (e.g. identify if a criminal has been close to a victim), as the results indicate that one can pin-point a person within 5 meters. Based on the situation, and how many other mobiles are in the same area at the same time, it can be difficult to identify interactions. This is because one will not be able to identify if the person is next to each other or if they are on either side of a street. If there is no one else around, one will be able to indicate with greater certainty an interaction between persons.

It may be possible to identify possible cooperation of peoples (e.g. a team working together), but it may require collecting data over a longer period. If a group of mobiles move from approximately the same place, at the same time and in the same direction over an extended period, this may indicate that they are a group. If one look at the data over a longer period of time one will be able to detect if the same mobiles meet several times.

## **7.1 Limitations of the thesis**

The master's thesis can potentially be improved in several different way, and a large part of the improvements are related to the data set and how the data set was collected. Unfortunately, these improvement potentials were discovered during

the data analysis after the collection finished, and it was not possible to recollect the data at this time. The data analysis became a bit more difficult and there were some uncertainties that had to be addressed. Since there were some uncertainties, the accuracy of all the results would be more uncertain than they needed to be, but the results could still answer a good deal of the questions this master's thesis started with.

The main challenges were the amount of data collected since the devices were fairly quiet especially when the screen was off and not in use. The data analysis discovered that even when the screen was on, the phones generated very few probe requests. It might have been a better option to look at pcap files from the sensors themselves during the experiment when the devices were continuously streaming data from, for example, YouTube. In that case, there would be a lot more data to work with, and not only limited to probe request packets.

More data is especially important for the fingerprint method to be able to achieve the desired accuracy. The entire data collection related to the fingerprint should have been re-done with new test parameters. The room should have been divided into cells beforehand and the mobile phones should then have been located at each cell for a longer period so that more data could be captured for each cell. Because of how the data was collected and the limited transmission of data from the phones, the fingerprinting ended up with many empty cells, which meant that the fingerprints could not be used. The cells with data, consisted of very little data and every small unfiltered noise caused the ground truth in the given cell to be wrong or have high degree of uncertainty.

One of the other challenges was related to timing and synchronization. All of the sensors were synchronized against a ntp server on the internet before data collection was started, but it might have been a better alternative to synchronize them against a local time source. There were some uncertainties around whether the sensors were synchronized down to milliseconds, or only down to seconds. To pair the different data packets sent out (e.g. to identify the same packet at two different sensors) was difficult when they appear to arrive at different times and it was not possible to identify the packet. The start times on the various scenarios was identified down to the minute, and this caused some challenges to identify the start in the data set where each second was important. The start time should have been stated down to seconds. It was expected that the ground truth video would have a built-in time e.g. at the bottom of one corner of the video, which it did not have. Therefore, an approximate start and end time was added to the video before it was delivered to the master project. One possible improvement could be to check in advance if the camera would include time stamps on the video and as an added security use a virtual clapperboard, for example a wireless access point that broadcasted an SSID like "The time is xxxx, start of Scenario 1 ", to be picked up by all the sensors. The start times were particularly important for the fingerprint method, but it was also used for the triangulation when comparing to the ground truth.

Another issue to consider is the test area itself. The data collection took place

within a room, which was not very large in consideration of signal propagation and triangulation. Since the room was limited in space and the person carrying the mobiles spent very short time on the planned route, it is difficult to give good accuracy, since small faults can cause major impact and can cause the Triangulation/Multilateration to place the mobile far away from its actual location.

## **7.2 Future work**

There are several different changes discovered during this master's thesis study that is suggested for further work, and much of it is related to data collection. Initially, it would be better to do a data collection on a large test site to better perceive how well the different methods can identify groups and track movement in an open area.

New tests should be outside in a large area to see if the general signal propagation may be measured to a higher degree of granularity and to see if this can be more accurate. In addition, spread the sensors over a larger area to verify if the Triangulation or Multilateration will be more effective.

Furthermore, it is recommended to carry out the data collection over a longer period to collect more data to work with. The process of data collection should also be planned in more detail before it is carried out, based on the experience from this experiment. This includes what information that should be written down for each trial e.g. which mobile was located where in Scenario 3. In order to properly test the fingerprint design method, the space should be divided into the cells in advance and ensure that the mobiles stay in each cell for an extended period during the data collection. This is to ensure that the data collected is enough to get a clean and usable fingerprints. In future work, the use of pcap files and data related to streaming should be considered to see if it can enhance the data analysis. If possible, get access to the necessary equipment to test the Angel of Arrival and Time of Flight methods to see if they could produce better results.

## **7.3 Conclusion**

The majority of the published papers are mostly focusing on data collection techniques and the results, and not so much on the process from the data collection to the result. This paper describes the whole process including data collection, data preprocessing, data cleaning, data analyzing, geolocation calculation, analyzing and finally presenting the results.

While most papers focus on one type of geolocalization technique this paper was exploring different geolocalization techniques, compared these and explored to see how these can complement each other (e.g. Triangulation verses Multilaterations) and to see how the methods works compared to the actual physical location of mobiles verified by a video surveillance.



During analyzing the data through this master's thesis study, it has shown that there are various types of intelligence that can be gathered by using rogue access points and that these can be used for increasing the situational awareness, though there are some degrees of error in the data collected that has to be taken into account.

The main research question was how can the information gathered from rogue access points increase situational awareness by Wi-Fi geolocation of mobiles and what are the challenges that may occur when analyzing the collected data? The analyses of the geolocation showed that only Triangulation and Multilateration using signal strength could be used on the given data set. However, it did not provide good enough accuracy, without any uncertainties, to identify any interaction between persons (e.g. identify if a criminal has been close to a victim). The results show that one can pin-point individuals within approximately 5 meters, which in most cases will not be enough to be able to identify any interaction between persons. The accuracy may be good enough to give increased situation awareness by suggesting the connection between people (e.g. a team working together). For example, one would not be able to identify two people sitting in the same vehicle, but one could identify that they are both near a crime scene at the same time. If the test area is larger, one can track movement of individuals, their speed and direction. It could be possible to identify units that are following each other over larger area. For example, if two or more phones are moving in the same direction at the same time, but it could be difficult to see if they are close together or if they are a few meters apart. If one collect data over a longer period of time one can see phones that are often close to each other, possibly if there is a group of phones that are often isolated together.

Another research question was what type of information can be gathered by the rogue access point and how effective is the collection of geolocation data with respect to how much data will be registered at an access point from each device. There is a good deal of different data being recorded at an access point, but the most important data fields for triangulation will be the signal strength and time stamps. The analysis of the data indicates that there are very few data packets sent from each mobile, which resulted in very little data to work with. There was also quite a large variation from mobile to mobile on how often a probe request was sent out.

Another research question was can data from several access points at the same location be correlated and how to ensure reliability of the analysis? The data collected by several access points were time stamped so the collected data could be correlated to a certain degree and used in the geolocation, but unsynchronized channel hopping on the sensors resulted in data packets from one device not being recorded simultaneously at all access point in reach. It looked like all receivers were synchronized at least down to seconds, but there were uncertainties around whether they were synchronized down to milliseconds that gave some doubt to which data packet was the same data packet received at the different sensors.

It turned out that only one of the four geolocation methods could be used on

the given data set. The two methods, Angel of Arrival and Time of Flight, could not be tested since the special equipment needed was not available to ensure correct data information and the data itself did not contain enough information. The third method, Fingerprinting using signal strength, did not need any extra equipment, but the lack of enough data packets and other uncertainties meant that the method could not be used on the given data set. The fourth and final method, Triangulation and Multilateration using signal strength, was the only one that could be used on the given data set. This method seemed to provide an accuracy of roughly 5 meters and could be used to obtain increased situational awareness.

# Bibliography

- [1] D. Omand, 'Securing the state: National security and secret intelligence', English, *Prism : a Journal of the Center for Complex Operations*, vol. 4, no. 3, pp. 14–27, Sep. 2013. [Online]. Available: <https://search.proquest.com/docview/1441703466>.
- [2] E. Vattapparamban, B. S. Çiftler, İ. Güvenç, K. Akkaya and A. Kadri, 'Indoor occupancy tracking in smart buildings using passive sniffing of probe requests', in *2016 IEEE International Conference on Communications Workshops (ICC)*, 2016, pp. 38–44. DOI: 10.1109/ICCW.2016.7503761.
- [3] C. Groba, 'Demonstrations and people-counting based on wifi probe requests', in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019, pp. 596–600. DOI: 10.1109/WF-IoT.2019.8767208.
- [4] C. Chilipirea, A.-C. Petre, C. Dobre and M. van Steen, 'Presumably simple: Monitoring crowds using wifi', [Accessed 15-October-2019], Jun. 2016. [Online]. Available: [https://www.researchgate.net/publication/303843715\\_Presumably\\_Simple\\_Monitoring\\_Crowds\\_Using\\_WiFi](https://www.researchgate.net/publication/303843715_Presumably_Simple_Monitoring_Crowds_Using_WiFi).
- [5] A. B. M. Musa and J. Eriksson, 'Tracking unmodified smartphones using wi-fi monitors', in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '12, Association for Computing Machinery, 2012, pp. 281–294, ISBN: 9781450311694. DOI: 10.1145/2426656.2426685. [Online]. Available: <https://doi.org/10.1145/2426656.2426685>.
- [6] M. Cunche, 'I know your MAC Address: Targeted tracking of individual using Wi-Fi', in *International Symposium on Research in Grey-Hat Hacking - GreHack*, [Accessed 19-September-2019], Nov. 2013. [Online]. Available: <https://hal.inria.fr/hal-00858324>.
- [7] M. Tsai, J. Luo, M. Yang and N. Lo, 'Location tracking and forensic analysis of criminal suspects' footprints', in *2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT)*, 2019, pp. 210–214. DOI: 10.1109/INF0CT.2019.8710862.
- [8] R. Johnsen and K. Hofstad, 'Dbm', *Store Norske Leksikon*, May 2019, [Accessed 26-April-2020]. [Online]. Available: <https://snl.no/dBm>.

- [9] Metageek, *Understanding wifi signal strength*, [Accessed 26-April-2020]. [Online]. Available: <https://www.metageek.com/training/resources/wifi-signal-strength-basics.html>.
- [10] Metageek, *Understanding rssi*, [Accessed 8-December-2019]. [Online]. Available: <https://www.metageek.com/training/resources/understanding-rssi.html>.
- [11] F. Zafari, A. Gkelias and K. K. Leung, 'A survey of indoor localization systems and technologies', *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019. DOI: 10.1109/COMST.2019.2911558. [Online]. Available: <https://arxiv.org/pdf/1709.01015.pdf>.
- [12] G. Lui, T. Gallagher, B. Li, A. G. Dempster and C. Rizos, 'Differences in rssi readings made by different wi-fi chipsets: A limitation of wlan localization', in *2011 International Conference on Localization and GNSS (ICL-GNSS)*, 2011, pp. 53–57. DOI: 10.1109/ICL-GNSS.2011.5955283.
- [13] A. Cole, *All you need to know about mac address*, [Accessed 28-April-2020], Jan. 2020. [Online]. Available: <https://www.cleverfiles.com/howto/what-is-mac-address.html>.
- [14] TechTerms, *Mac address*, [Accessed 28-April-2020]. [Online]. Available: <https://techterms.com/definition/macaddress>.
- [15] J. Freudiger, 'How talkative is your mobile device? an experimental study of wi-fi probe requests', in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, ser. WiSec '15, Association for Computing Machinery, 2015, ISBN: 9781450336239. DOI: 10.1145/2766498.2766517. [Online]. Available: <https://doi.org/10.1145/2766498.2766517>.
- [16] R. Acharya, 'Chapter 7 - errors and error corrections', *Understanding Satellite Navigation*, pp. 243–279, Sep. 2014. DOI: <https://doi.org/10.1016/B978-0-12-799949-4.00007-5>.
- [17] A. Piltzecker, 'Chapter 7 - microsoft vista: Wireless world', *Microsoft Vista for IT Security Professionals*, pp. 345–397, Sep. 2007. DOI: <https://doi.org/10.1016/B978-159749139-6/50011-X>.
- [18] Juniper Networks, *Understanding rogue access points*, [Accessed 28-April-2020], Dec. 2016. [Online]. Available: [https://www.juniper.net/documentation/en\\_US/junos-space-apps/network-director3.1/topics/concept/wireless-rogue-ap.html](https://www.juniper.net/documentation/en_US/junos-space-apps/network-director3.1/topics/concept/wireless-rogue-ap.html).
- [19] Random solutions, *Best dbm values for wifi*, [Accessed 14-Januar-2020], Feb. 2016. [Online]. Available: <https://support.randomsolutions.nl/827069-Best-dBm-Values-for-Wifi>.
- [20] B. Sundararaman, U. Buy and A. D.Kshemkalyani, 'Clock synchronization for wireless sensor networks: A survey', vol. 3, no. 3, pp. 281–323, 2005. DOI: 10.1016/j.adhoc.2005.01.002.

- [21] Qun Li and D. Rus, 'Global clock synchronization in sensor networks', *IEEE Transactions on Computers*, vol. 55, no. 2, pp. 214–226, 2006. DOI: 10.1109/TC.2006.25.
- [22] Y. Wu, Q. Chaudhari and E. Serpedin, 'Clock synchronization of wireless sensor networks', *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, 2011. DOI: 10.1109/MSP.2010.938757.
- [23] M. L. Shafer, 'Triangulation', *The Surveying Handbook*, pp. 296–339, 1987. DOI: [https://doi.org/10.1007/978-1-4757-1188-2\\_10](https://doi.org/10.1007/978-1-4757-1188-2_10).
- [24] Circuit Cellar, *Triangulation, trilateration, or multilateration?*, [Accessed 13-May-2020], May 2014. [Online]. Available: <https://circuitcellar.com/resources/ee-tips/triangulation-trilateration-or-multilateration-ee-tip-125/>.
- [25] B. N. Sturges and F. T. Carey, 'Trilateration', *The Surveying Handbook*, pp. 340–389, 1987. DOI: [https://doi.org/10.1007/978-1-4757-1188-2\\_11](https://doi.org/10.1007/978-1-4757-1188-2_11).
- [26] D. Pogue, *What wi-fi stands for—and other wireless questions answered*, [Accessed 2-May-2020], May 2012. [Online]. Available: <https://www.scientificamerican.com/article/pogue-what-wifi-stands-for-other-wireless-questions-answered/>.
- [27] J. Kastrenakes, *Wi-fi now has version numbers, and wi-fi 6 comes out next year*, [Accessed 02-May-2020], Oct. 2018. [Online]. Available: <https://www.theverge.com/2018/10/3/17926212/wifi-6-version-numbers-announced>.
- [28] B. Mitchell, *802.11 standards explained: 802.11ax, 802.11ac, 802.11b/g/n, 802.11a*, [Accessed 02-May-2020], Apr. 2020. [Online]. Available: <https://www.lifewire.com/wireless-standards-802-11a-802-11b-g-n-and-802-11ac-816553>.
- [29] Wi-Fi Alliance, *Generational wi-fi user guide*, [Accessed 02-May-2020], Oct. 2018. [Online]. Available: [https://www.wi-fi.org/download.php?file=/sites/default/files/private/Generational\\_Wi-Fi\\_User\\_Guide\\_20181003.pdf](https://www.wi-fi.org/download.php?file=/sites/default/files/private/Generational_Wi-Fi_User_Guide_20181003.pdf).
- [30] Juniper Networks, *Understanding the network terms ssid, bssid, and essid*, [Accessed 18-May-2020], Oct. 2018. [Online]. Available: [https://www.juniper.net/documentation/en\\_US/junos-space-apps/network-director3.7/topics/concept/wireless-ssid-bssid-ssid.html](https://www.juniper.net/documentation/en_US/junos-space-apps/network-director3.7/topics/concept/wireless-ssid-bssid-ssid.html).
- [31] L. Oliveira, D. Schneider, J. De Souza and W. Shen, 'Mobile device detection through wifi probe request analysis', *IEEE Access*, pp. 98 579–98 588, 2019. DOI: 10.1109/ACCESS.2019.2925406.

- [32] L. Schauer, M. Werner and P. Marcus, 'Estimating crowd densities and pedestrian flows using wi-fi and bluetooth', in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MOBIQUITOUS '14, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014, pp. 171–177, ISBN: 9781631900396. DOI: 10.4108/icst.mobiquitous.2014.257870. [Online]. Available: <https://doi.org/10.4108/icst.mobiquitous.2014.257870>.
- [33] Elite Data Science, *Data cleaning*, [Accessed 4-February-2020]. [Online]. Available: <https://elitedatascience.com/data-cleaning>.
- [34] Tutorialspoint, *Antenna theory - antenna arrays*, [Accessed 12-December-2019]. [Online]. Available: [https://www.tutorialspoint.com/antenna\\_theory/antenna\\_theory\\_arrays.htm](https://www.tutorialspoint.com/antenna_theory/antenna_theory_arrays.htm).
- [35] S. Kumar, E. Hamed, D. Katabi and L. Erran Li, 'Lte radio analytics made easy and accessible', in *Proceedings of the 6th Annual Workshop on Wireless of the Students, by the Students, for the Students*, ser. S3 '14, Association for Computing Machinery, 2014, pp. 29–30, ISBN: 9781450330732. DOI: 10.1145/2645884.2645891. [Online]. Available: <https://doi.org/10.1145/2645884.2645891>.
- [36] A. Marcaletti, M. Rea, D. Giustiniano, V. Lenders and A. Fakhreddine, 'Filtering noisy 802.11 time-of-flight ranging measurements', in *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '14, Association for Computing Machinery, 2014, pp. 13–20, ISBN: 9781450332798. DOI: 10.1145/2674005.2674998. [Online]. Available: <https://doi.org/10.1145/2674005.2674998>.
- [37] M. Youssef, A. Youssef, C. Rieger, U. Shankar and A. Agrawala, 'Pinpoint: An asynchronous time-based location determination system', in *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services*, ser. MobiSys'06, Association for Computing Machinery, 2006, pp. 165–176, ISBN: 1595931953. DOI: 10.1145/1134680.1134698. [Online]. Available: <https://doi.org/10.1145/1134680.1134698>.
- [38] S. Lanzisera, D. Zats and K. S. J. Pister, 'Radio frequency time-of-flight distance measurement for low-cost wireless sensor localization', *IEEE Sensors Journal*, vol. 11, no. 3, pp. 837–845, 2011. DOI: 10.1109/JSEN.2010.2072496.
- [39] H. T. Friis, 'A note on a simple transmission formula', *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, 1946. DOI: 10.1109/JRPROC.1946.234568.
- [40] 101computing, *Cell phone trilateration algorithm*, [Accessed 23-May-2020], Mar. 2019. [Online]. Available: <https://www.101computing.net/cell-phone-trilateration-algorithm/>.

- [41] A. Farshad, M. K. Marina and F. Garcia, 'Urban wifi characterization via mobile crowdsensing', in *2014 IEEE Network Operations and Management Symposium (NOMS)*, 2014, pp. 1–9. DOI: 10.1109/NOMS.2014.6838233.
- [42] A. Basalamah, 'Crowd mobility analysis using wifi sniffers', *International Journal of Advanced Computer Science and Applications*, vol. 7, Dec. 2016. DOI: 10.14569/IJACSA.2016.071249.
- [43] Accolade Wireless, *Why wifi is complicated: Wifi signal issues*, [Accessed 04-May-2020], Jun. 2016. [Online]. Available: <http://www.accoladewireless.com/wlan-wifi-signal-issues/>.
- [44] Sporton International, *Fcc sar test report*, [Accessed 04-May-2020], Jun. 2015. [Online]. Available: <https://fccid.io/ZNFX150/RF-Exposure-Info/RF-Exposure-Test-Report-2662404>.
- [45] Shenzhen BALUN Technology Co, *Fcc rf test report*, [Accessed 04-May-2020], Apr. 2015. [Online]. Available: <https://fccid.io/R9C-1206/Test-Report/Test-Report-for-WiFi-2615516>.
- [46] P Biondi and the Scapy community, *Scapy*, [Accessed 05-January-2019]. [Online]. Available: <https://scapy.net/>.

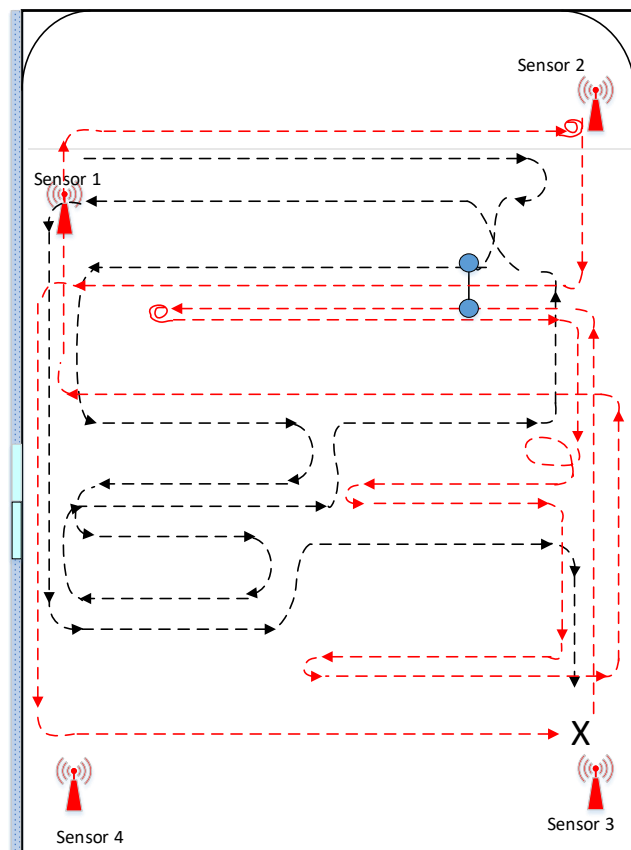




## Appendix A

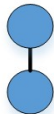
### Additional localization results

Each of the scenarios shows the room plan with the movement of the mobiles indicated with dotted lines. The dotted lines may have two different colors indicating two different individuals walking around with different mobiles at the same time as illustrated in Figure A.1. The scenarios are shown in a figure with the X/Y coordinate system.



**Figure A.1:** Visualization of movement from multiple people

The geolocations of the different mobiles are indicated with different colors, and the two connected blue circles indicate an interaction between the persons carrying the mobile phones, for example, exchanging the mobiles or merging the mobiles.

Multilateration		Triangulation		Interaction point
Mobile 1	●	Mobile 1	▲	
Mobile 2	●	Mobile 2	▲	
Mobile 3	●	Mobile 3	▲	
Mobile 4	●	Mobile 4	▲	
Mobile 5	●	Mobile 5	▲	
Mobile 6	●	Mobile 6	▲	
Mobile 7	●	Mobile 7	▲	

**Figure A.2:** Color code - mobile phones and connecting circles for interaction point

## A.1 Scenario 4.1

In Scenario 4-1 the mobiles are moving in a fixed pattern within the room. Figure A.3 illustrates the movement and calculated locations. Table A.4 shows the calculated Triangulation and Multilateration result.

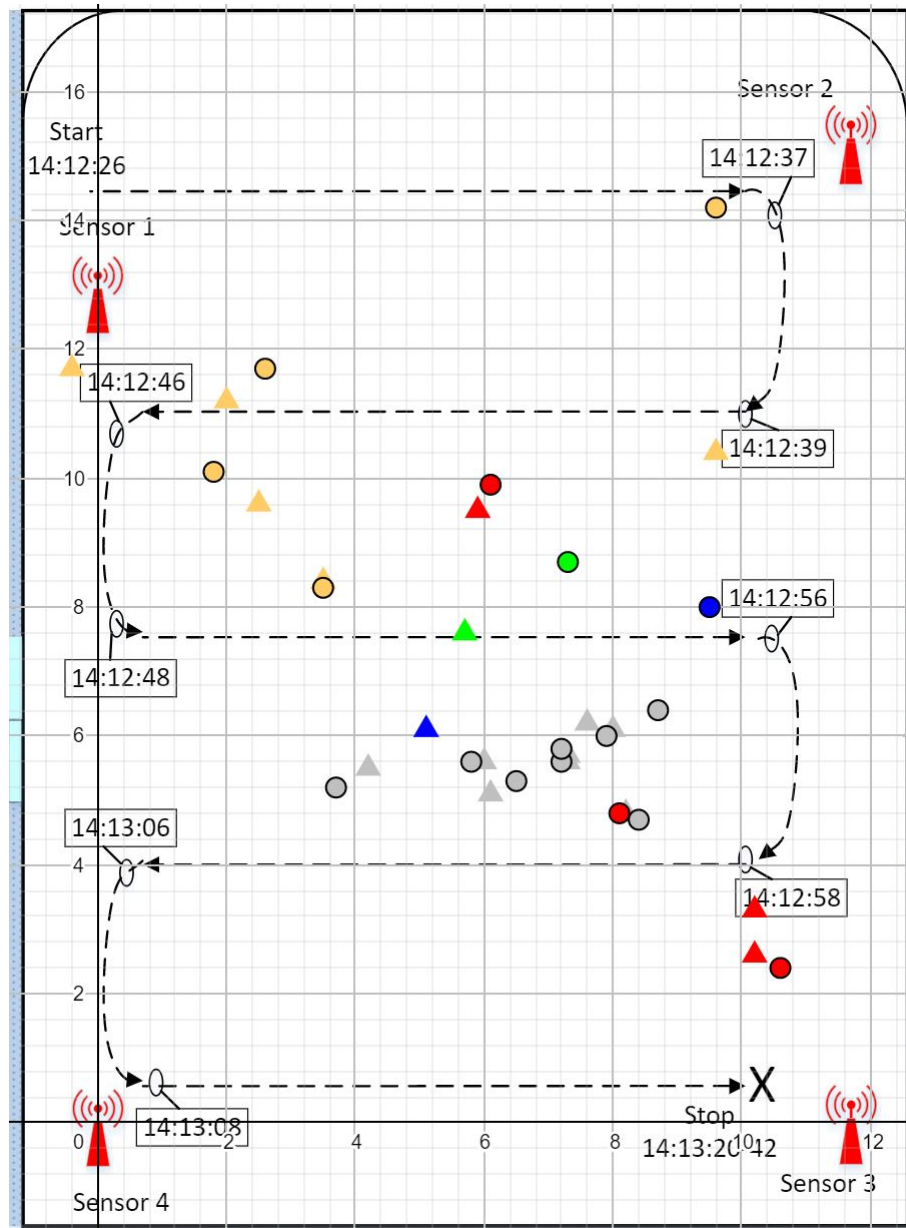


Figure A.3: Scenario 4.1

Multilateration														
Time	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:12:25	6,1	9,9							8,4	4,7				
14:12:34									7,2	5,6				
14:12:43					-3,3	6,6			20,4	7,3				
14:12:44	16,8	17,0							20,4	7,3				
14:12:45	16,8	17,0			9,6	14,2								
14:12:46					2,6	11,7								
14:12:53									7,9	6,0				
14:12:57					3,5	8,3					7,3	8,7		
14:12:58					1,8	10,1								
14:13:03									7,2	5,8				
14:13:06	8,1	4,8												
14:13:13									12,9	5,4				
14:13:22									6,5	5,3				
14:13:23	10,6	2,4												
14:13:32									5,8	5,6				
14:13:41													9,5	8,0

Triangulation														
Time	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:12:25		9,5							8,2	4,8				
14:12:34									7,3	5,6				
14:12:43					-1,7	6,6			18,8	15,0				
14:12:44									18,8	15,0				
14:12:45					9,6	10,4								
14:12:46					2,0	11,2								
14:12:47					-0,4	11,7								
14:12:53									8,0	6,1				
14:12:57					3,5	8,4					5,7	7,6		
14:12:58					2,5	9,6								
14:13:03									7,3	5,7				
14:13:06	10,2	3,3												
14:13:13									12,7	6,1				
14:13:22									6,0	5,6				
14:13:23	10,2	2,6												
14:13:32									6,1	5,1				
14:13:41													5,1	6,1

Figure A.4: Scenario 4.1 - Multilateration and Triangulation

## A.2 Scenario 4.2

In Scenario 4-2 the mobiles are moving in a random pattern within the room. Figure A.5 illustrates the movement and calculated locations. Table A.6 shows the calculated Triangulation and Multilateration result.

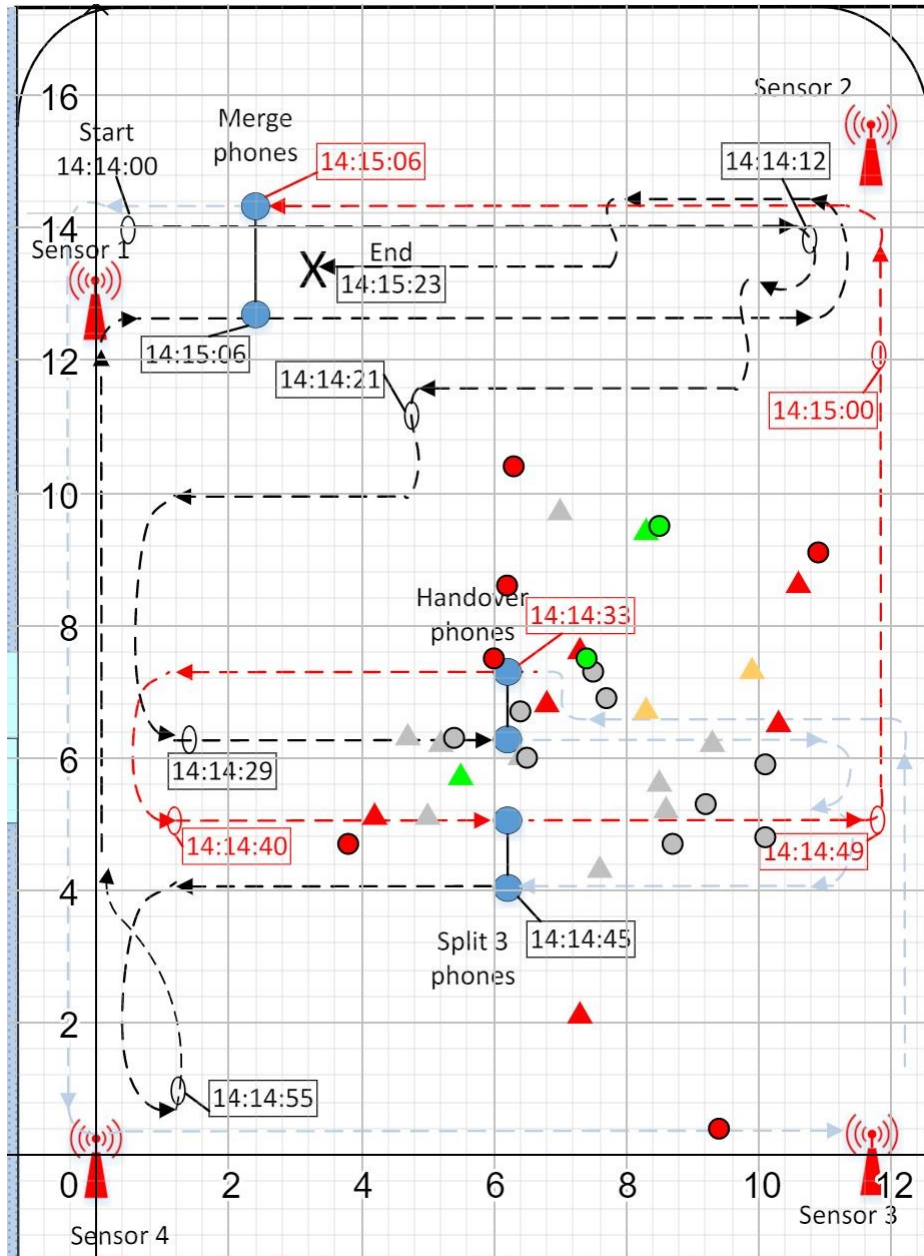


Figure A.5: Scenario 4.2

Multilateration																												
Time	Phone 1		Phone 2		Phone 3		X	Y	X	Y	Phone 4		X	Y	Phone 5		X	Y	Phone 6		X	Y	Phone 7		X	Y		
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y		
14:14:05																												
14:14:10																												
14:14:16	6,3	10,4																										
14:14:19																												
14:14:28	6,2	8,6																										
14:14:29	6,0	7,5																										
14:14:39																												
14:14:48																												
14:14:58																												
14:15:01	10,9	9,1																										
14:15:07																												
14:15:08																												
14:15:11	3,8	4,7																										
14:15:14																												
14:15:17																												
14:15:18	9,4	0,4																										

Triangulation																												
Time	Phone 1		Phone 2		Phone 3		X	Y	X	Y	Phone 4		X	Y	Phone 5		X	Y	Phone 6		X	Y	Phone 7		X	Y		
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y		
14:14:05																												
14:14:10																												
14:14:16	10,3	6,5																										
14:14:19																												
14:14:27																												
14:14:28	7,3	7,6																										
14:14:29	6,8	6,8																										
14:14:39																												
14:14:48																												
14:14:58																												
14:15:01	10,6	8,6																										
14:15:07																												
14:15:08																												
14:15:11	4,2	5,1																										
14:15:14																												
14:15:17																												
14:15:18	7,3	2,1																										

Figure A.6: Scenario 4.2 - Multilateration and Triangulation

### A.3 Scenario 5.1

In Scenario 5-1 the mobiles are moving in a fixed pattern within the room. Figure A.7 illustrates the movement and calculated locations. Table A.8 shows the calculated Triangulation and Multilateration result.

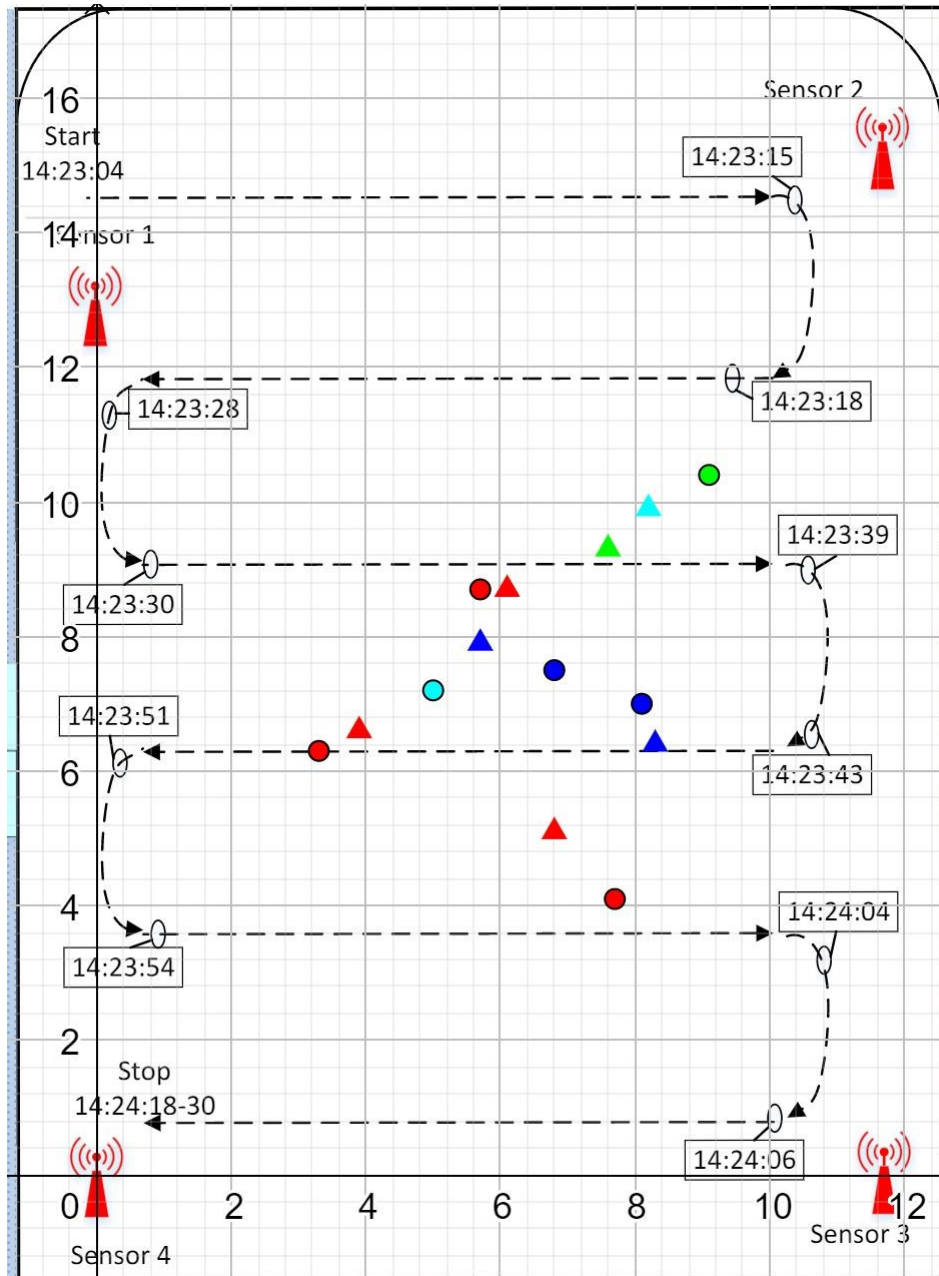


Figure A.7: Scenario 5.1

Multilateration												
Time	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:23:03							5,0	7,2				
14:23:23											9,1	10,4
14:23:29												
14:23:53	5,7	8,7										
14:24:07												
14:24:15	3,3	6,3										
14:24:19	7,7	4,1										
14:24:24	9,3	1,6										
14:24:28	6,4	5,6										

Triangulation												
Time	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:23:03							8,2	9,9				
14:23:23											7,6	9,3
14:23:29												
14:23:53	6,1	8,7										
14:24:07												
14:24:15	3,9	6,6										
14:24:19	6,8	5,1										
14:24:24	7,2	3,3										
14:24:28	6,7	5,3										

Figure A.8: Scenario 5.1 - Multilateration and Triangulation



## A.4 Scenario 5.2

In Scenario 5-2 the mobiles are moving in a random pattern within the room. Figure A.9 illustrates the movement and calculated locations. Table A.10 shows the calculated Triangulation and Multilateration result.

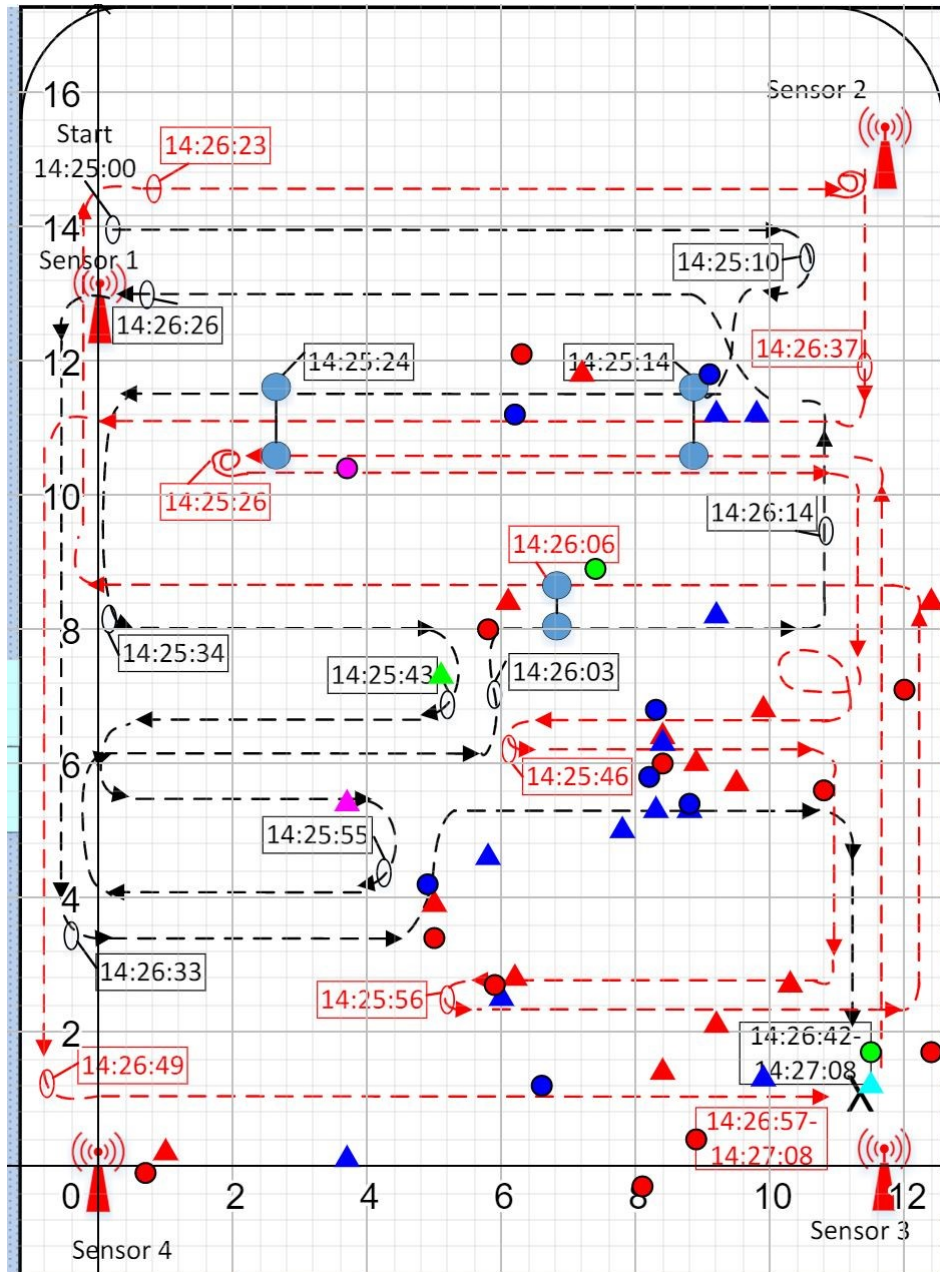


Figure A.9: Scenario 5.2

Multilateration														
Time	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:25:19														
14:25:25	5,8	8,0									7,4	8,9	9,1	11,8
14:25:31	13,3	7,2												
14:25:49														
14:25:52	8,9	0,4												
14:25:53	8,1	-0,3												
14:25:56			3,7	10,4										
14:26:01														
14:26:08	12,4	1,7												
14:26:24	12,0	7,1												
14:26:25	10,8	5,6												
14:26:28	0,7	-0,1												
14:26:29														
14:26:39	6,3	12,1											8,3	6,8
14:26:45	8,4	6,0												
14:26:49														
14:26:58							5,2	-3,9					6,6	1,2
14:27:01							11,5	1,7					21,2	1,3
14:27:03	5,0	3,4												
14:27:04	5,9	2,7												

Triangulation														
Time	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:25:19														
14:25:25	6,1	8,4									5,1	7,3	9,2	11,2
14:25:31	12,4	8,4												
14:25:47														
14:25:49														
14:25:52	9,2	2,1												
14:25:53	8,4	1,4												
14:25:54														
14:25:56			3,7	5,4										
14:26:01														
14:26:08	10,3	2,7												
14:26:24	8,9	6,0												
14:26:25	9,9	6,8												
14:26:28	1,0	0,2												
14:26:29														
14:26:39	7,2	11,8												
14:26:45	8,4	6,4												
14:26:49														
14:26:55	9,5	5,7											8,4	6,3
14:26:57														
14:26:58														
14:27:01														
14:27:03	5,0	3,9					10,7	-3,9						
14:27:04	6,2	2,8					7,9	-3,3						
							11,5	1,2					6,0	2,5
									</					

Figure A.10: Scenario 5.2 - Multilateration and Triangulation

## A.5 Scenario 6

In Scenario 6 the mobiles are moving in a fixed pattern within the corridor. Figure A.11 illustrates the movement and calculated locations. Table A.12 shows the calculated Triangulation and Multilateration result.

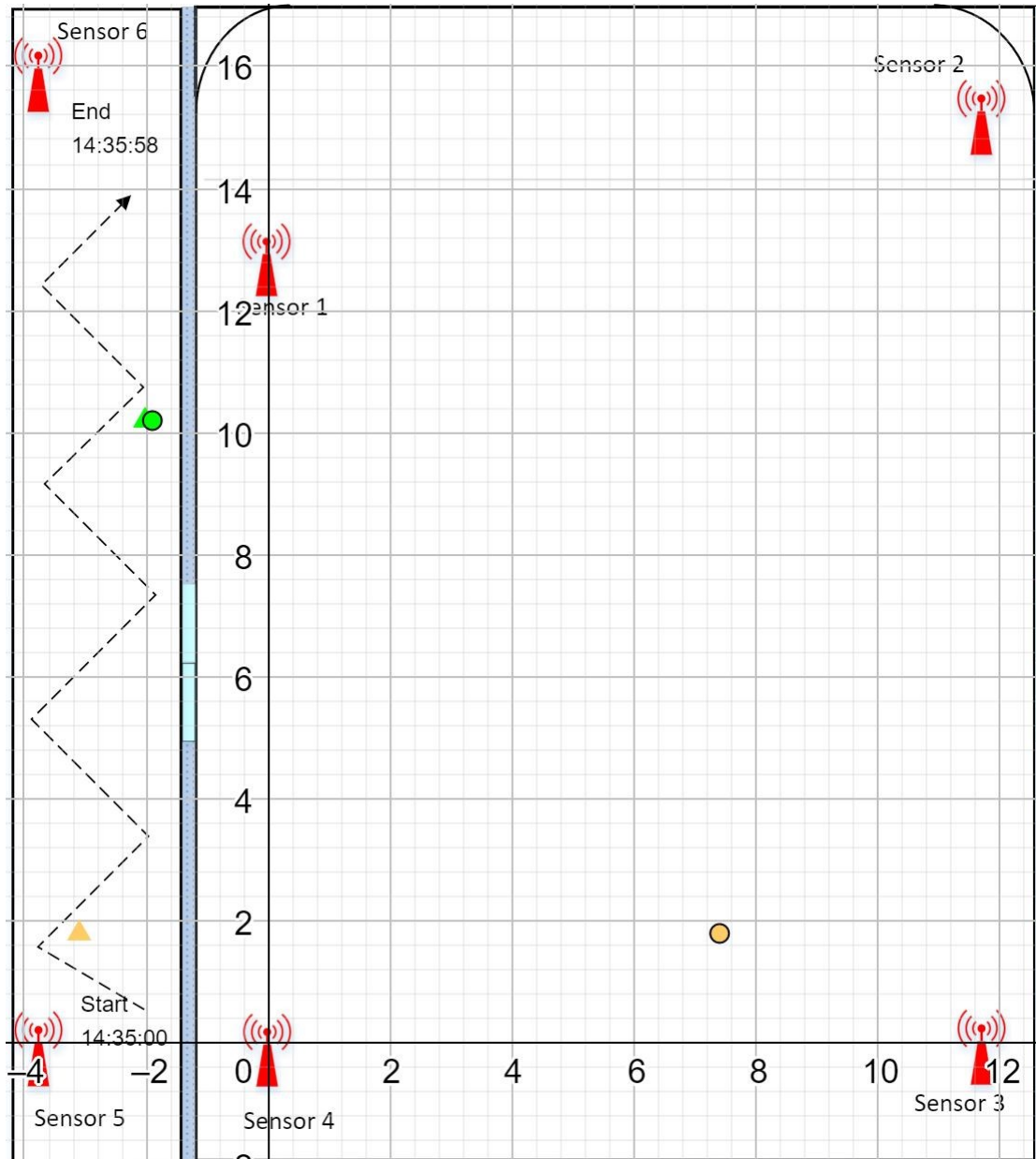


Figure A.11: Scenario 6

**Figure A.12: Scenario 6 - Multilateration and Triangulation**

## A.6 Scenario 7

In Scenario 7 the mobiles are moving in a fixed pattern within the corridor. Figure A.13 illustrates the movement and calculated locations. Table A.14 shows the calculated Multilateration result.

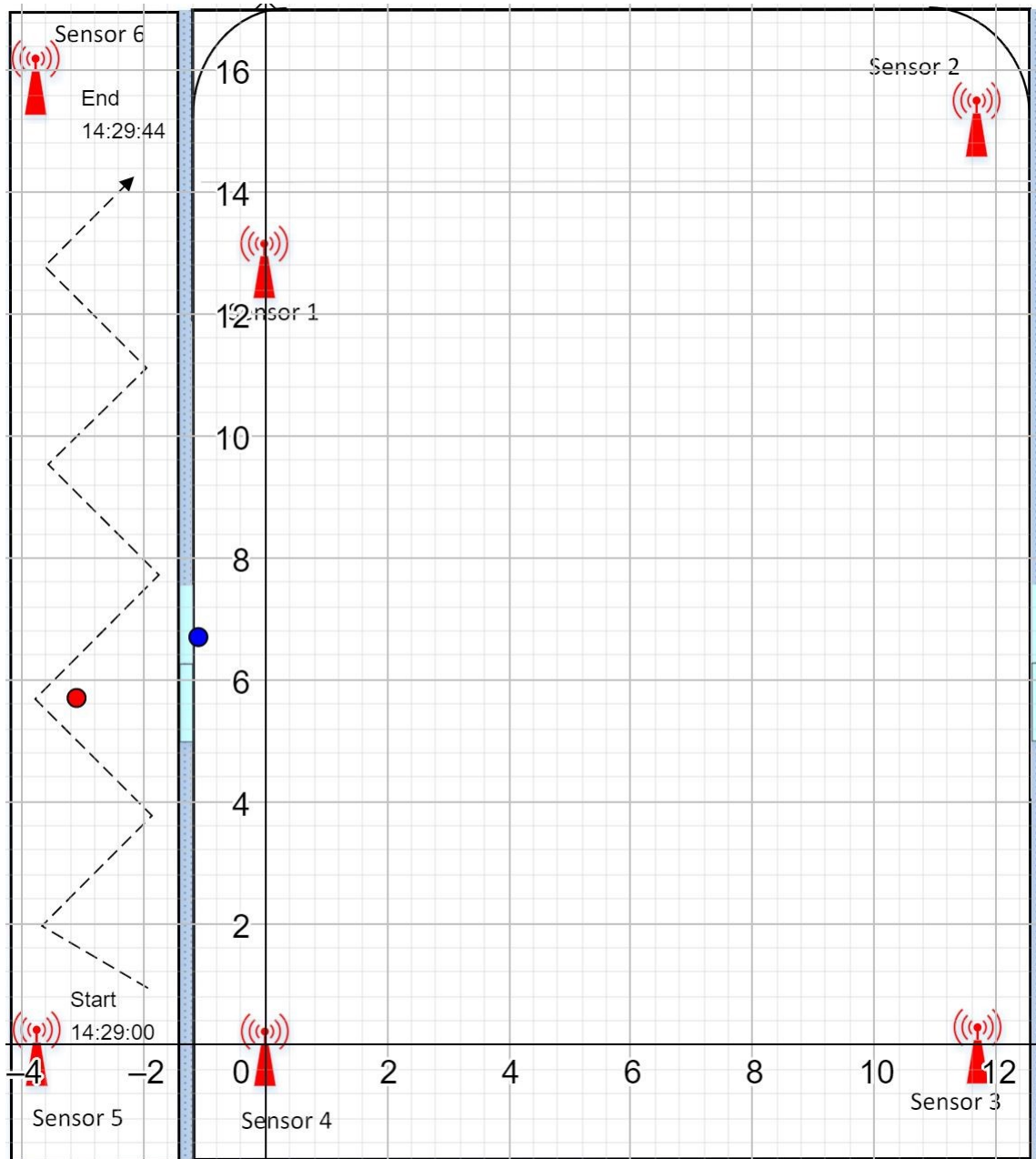


Figure A.13: Scenario 7

Multilateration														
	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
Time	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:29:08														
14:29:37	-3,1	5,7											-1,1	6,7

Figure A.14: Scenario 7 - Multilateration

## A.7 Scenario 8

In Scenario 8 the mobiles are moving in a random pattern within the corridor and the room. Figure A.15 illustrates the movement and calculated locations. Table A.16 shows the calculated Triangulation and Multilateration result.

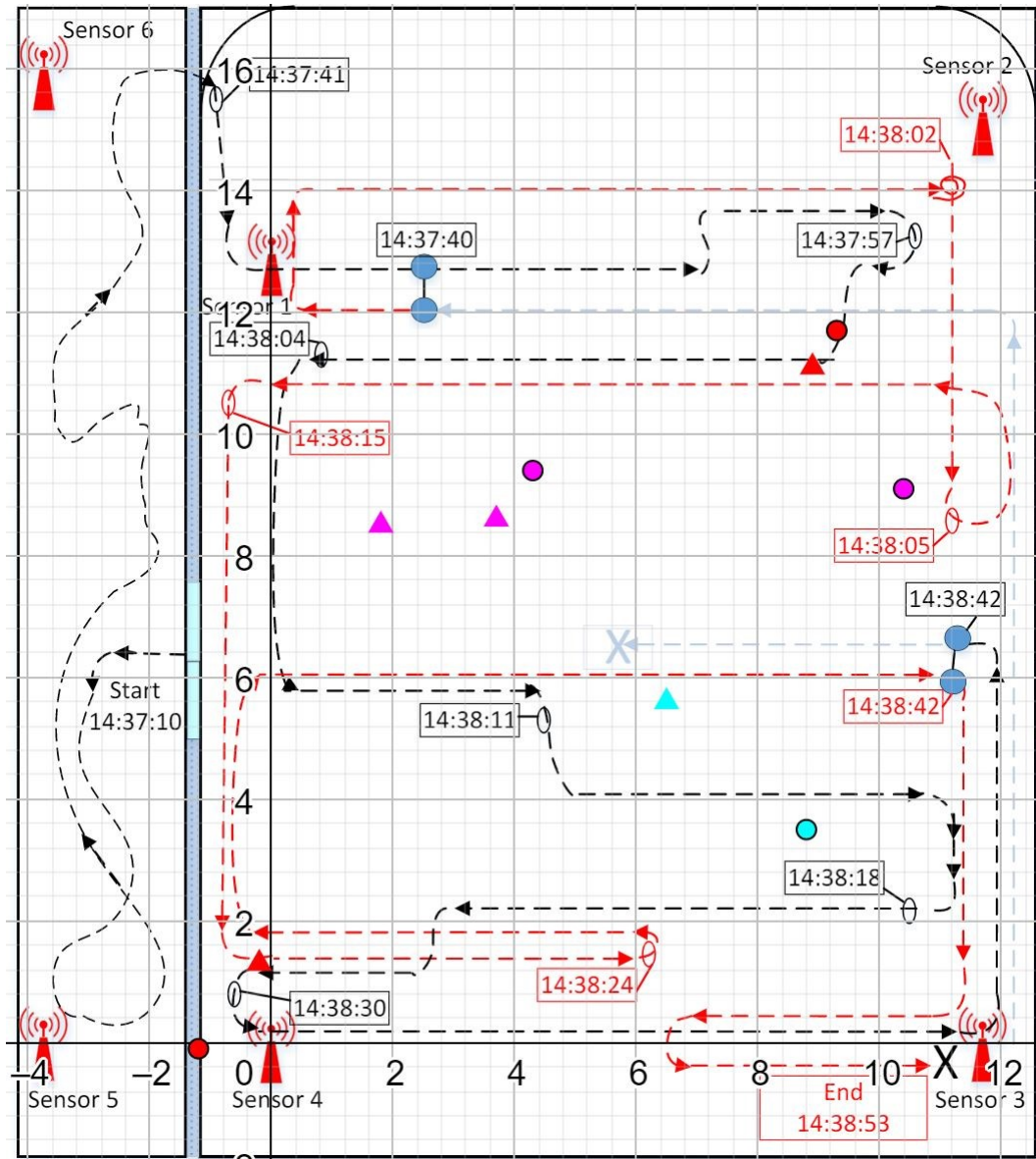


Figure A.15: Scenario 8

Multilateration										
	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5	
Time	X	Y	X	Y	X	Y	X	Y	X	Y
14:37:19	-1,2	-0,1								
14:37:58	9,3	11,7								
14:38:03			4,3	9,4						
14:38:35							8,8	3,5		
Triangulation										
	Phone 1		Phone 2		Phone 3		Phone 4		Phone 5	
Time	X	Y	X	Y	X	Y	X	Y	X	Y
14:37:19	-0,2	1,3								
14:37:58	8,9	11,1								
14:38:03			3,7	8,6						
14:38:35							6,5	5,6		

Figure A.16: Scenario 8 - Multilateration and Triangulation



## A.8 Scenario 9

In Scenario 9 the mobiles are moving in a random pattern within the corridor and the room. Figure A.17 illustrates the movement and calculated locations. Table A.18 shows the calculated Multilateration result and Table A.19 shows the calculated Triangulation results.

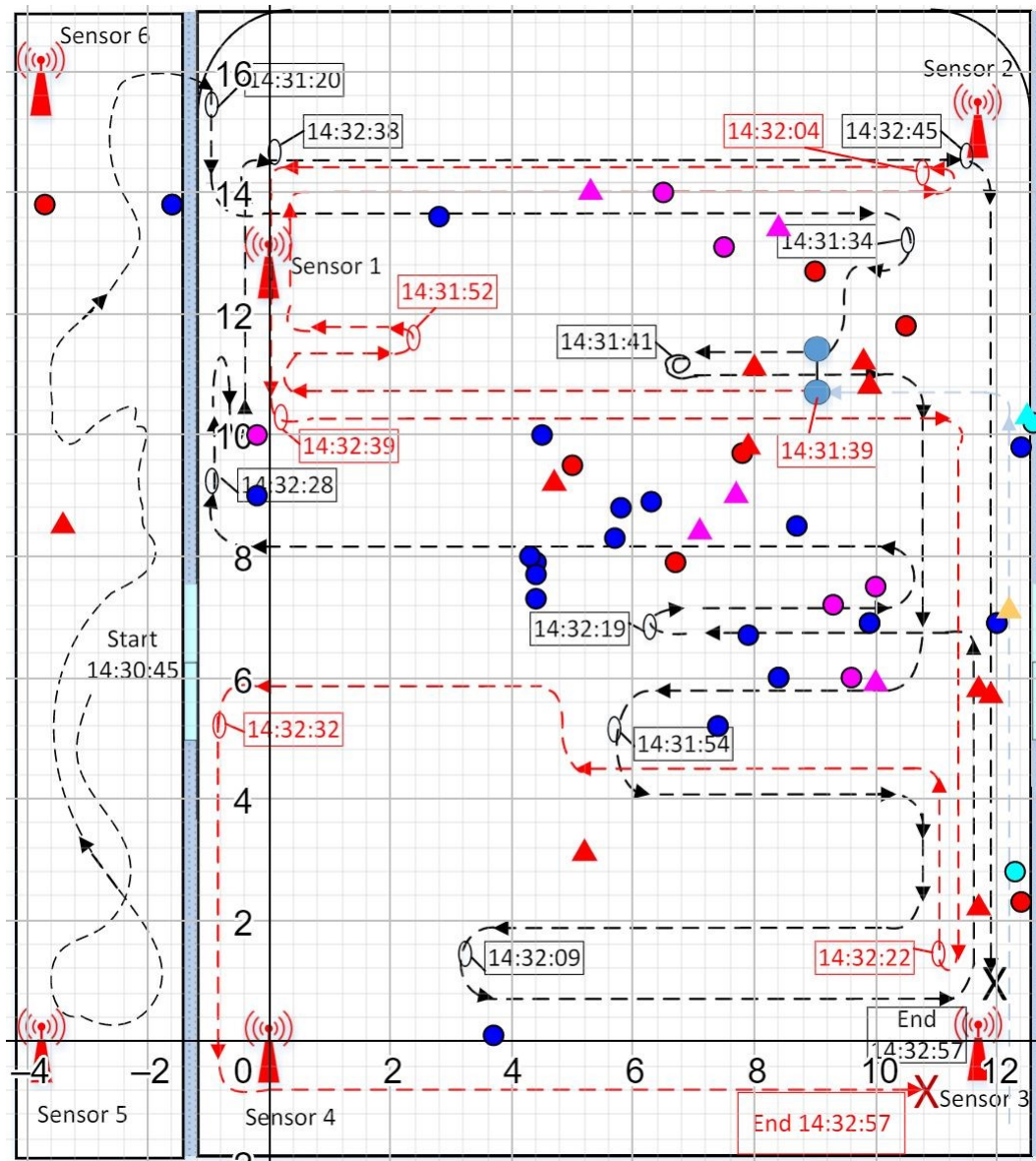


Figure A.17: Scenario 9

		Multilateration													
		Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
Time		X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:30:21		-3,7	13,8												
14:30:22															
14:30:56				-0,2	10,0										
14:31:14		6,7	7,9												
14:31:26		7,8	9,7												
14:31:45															
14:31:59		10,5	11,8											8,7	8,5
14:32:00		9,0	12,7												
14:32:02				6,5	14,0										
14:32:06								12,6	10,2						
14:32:08		15,2	1,7												
14:32:09		15,0	1,8												
14:32:11															
14:32:12														4,4	7,3
14:32:14		12,4	2,3											6,3	8,9
14:32:20		5,0	9,5												
14:32:27															
14:32:33				9,6	6,0									2,8	13,6
14:32:34				9,3	7,2									12,4	9,8
14:32:44								18,1	2,8						
14:32:45						14,4	7,1	12,3	2,8					12,0	6,9
14:32:46		12,8	-2,4											12,0	6,9
14:32:47		12,8	-2,4											8,4	6,0
14:32:48														8,4	6,0

Figure A.18: Scenario 9 - Multilateration

		Triangulation													
		Phone 1		Phone 2		Phone 3		Phone 4		Phone 5		Phone 6		Phone 7	
Time		X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14:30:21	-3,4		8,5											4,5	10,0
14:30:22														4,3	8,0
14:30:56				8,4	13,4										
14:31:14	9,9		10,8												
14:31:26	7,9		9,8												
14:31:45														7,9	6,7
14:31:55														3,7	0,1
14:31:59	9,8		11,2												
14:32:00	8,0		11,1												
14:32:02				5,3	14,0										
14:32:06								12,5	10,3						
14:32:08	11,9		5,7												
14:32:09	11,7		5,8											4,4	7,7
14:32:11														5,7	8,3
14:32:12															
14:32:14	11,7		2,2												
14:32:20	4,7		9,2												
14:32:33	5,2		3,1	7,1	8,4										
14:32:34	5,2		3,1	7,7	9,0									13,5	8,2
14:32:45						12,2	7,1								
14:32:46	12,5		-2,4			12,2	7,1							9,9	6,9
14:32:47	12,5		-2,4											9,9	6,9
14:32:48														7,4	5,2
14:32:59	15,3		4,0	10,0	5,9									7,4	5,2

Figure A.19: Scenario 9 - Triangulation

