AN ELECTROMAGNETIC SPECTRUM AWARE INDOOR POSITIONING SYSTEM

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2015

By
Myrna Margarita Rodríguez Frías
School of Computer Science

Contents

List of Tables

List of Figures

Abstract

AN ELECTROMAGNETIC SPECTRUM AWARE INDOOR
POSITIONING SYSTEM
Myrna Margarita Rodríguez Frías
A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy, 2015

The principal objectives of this research are: to investigate the performance of different fingerprint-based WiFi Indoor Positioning Systems (IPS), analyse historical long-term data signals, detection of signal change points and outliers; then present an enhanced method that generates temporal based fingerprints.

The proposed method consists of analysing signal strength profiles over time and detecting points at which the profile behaviour changes. This methodology can be used to dynamically adjust the fingerprint based on environmental factors, and with this select the relevant Wireless Access Points (WAPs) to be used for fingerprinting. The use of an Exponentially Weighted Moving Average (EWMA) Control Chart is investigated for this purpose.

A long-term analysis of the WiFi scenery is presented and used as a test-bed for evaluation of state-of-the-art fingerprinting techniques. Data was collected and analysed over a period of 18 months, with over 840 different WAPs detected in over 77,000 observations covering 47 different locations of varying characteristics.

A fully functional IPS has been developed and the design and implementation is described in this thesis. The system allows the scanning and recording of WiFi signals in order to define the generation of temporal fingerprints that can create radio-maps, which then allow indoor positioning to occur. This thesis presents the theory behind the concept and develops the technology to create a testable implementation. Experiments and their evaluation are also included.

Based on the timestamp experiments the proposed system shows there is still room level accuracy, with a reduction in radio-map size.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

This research was sponsored by the National Council of Science and Technology (CONACyT), México under grant 213616/308710. The author wishes to thank CONACyT for the support, without which this PhD would not be possible. The author further wishes to thank Dr. Martin Turner and Dr. Tim Morris for their support and guidance through out the research.

Thanks to Sam Proctor and Mía Proctor for their encouragement and love. Finally, thanks to my parents: Alicia Frías Pizano and Marco A. Rodríguez Alvarado, my sisters: Karla, Alicia and Andrómeda, and my brother Marco, for always being there for me.

Acronyms

BSSID Basic Service Set Identifier. 22, 44

CC Country Code. 22

EM Electromagnetic. 20

EWMA Exponentially Weighted Moving Average. 10

FM Frequency Modulation. 20

GPS Global Positioning System. 11, 17

HT High Throughput. 22

IMU Inertial Measurement Unit. 35

IPS Indoor Positioning System. 11, 16, 18

LCL Lower Control Limit. 36

LOS Line-Of-Sight. 20

MAC Media Access Control. 22, 63

MACD Moving Average Convergence Divergence. 32

RCS Research Computing Services. 11, 72

RFID Radio-frequency identification. 11, 20

RSSI Received Signal Strength Indicator. 22

SSID Service Set IDentifier. 22, 64

SVM Support Vector Machine. 38

UCL Upper Control Limit. 36

UWB Ultra Wide Band. 20

WAP Wireless Access Point. 11, 22, 64, 68, 73, 76, 88

WiFi Wireless Fidelity. 11

WLAN Wireless Local Area Network. 20

WPS WiFi Positioning System. 12

List of Symbols

a Counter of consecutive times an WAP is absent/undetected, used in the EWMA Fingerprint. 118

δ Acquisition probability of a WAP based on its RSSI, used in Jun-Sung Fingerprint. 111

i Index that identifies the current observation. 131

k_0 Number of observations required for initialisation. Is is used for State Estimate, Score and Control Chart. It is a constant value pre-defined by system administrator. 123, 127, 131

L Number of standard deviations to be considered in the EWMA control chart. 133

λ_x Parameter used in the EWMA algorithm for computing the state estimate. 123

l Alphanumeric string recorded in the database as a name for identifying locations. 97

N Total number of WAP within a fingerprint. 97

n Index that identifies every WAP within a fingerprint. 97

$\hat{p}$ Estimated value for the WAP presence computed with an EWMA with λ_p. 127

p Counter of consecutive times an WAP is present/detected, used in the EWMA Fingerprint. 118

Q Intermediate value used to compute the repeatedly updated mean, used in the Default Fingerprint. 103

s Score used for ranking WAPs, it is computed based on signal strength, standard deviation and WAP appearances. 118, 126

uc Binary parameter that determines when a signal strength is within acceptable boundaries within the control chart. 118

$\hat{x}_i$ The state estimate per WAP, at iteration i given measurement x_i. 121

$\hat{x}$ Expected signal strength per WAP. 97–99, 102, 104

x_i Current RSSI measurement per WAP from ongoing observation. 121

Chapter 1

Introduction

1.1 Hypothesis

Constantly analysing signals from Wireless Access Points (WAPs) in an indoor environment can lead to identification of the most valuable or trustworthy WAPs to be used in radio-maps for fingerprinting techniques.

1.2 Testing the Hypothesis

For testing the hypothesis, existing state-of-the-art fingerprinting techniques are evaluated and compared against the proposed Exponentially Weighted Moving Average (EWMA) fingerprinting technique. The evaluation is carried out using real-life data collected over a period of one year.

The updating mechanisms are evaluated simulating a real-time data stream of observations feeding into the system. A radio-map is maintained and updated for each technique and this is used for simulating real-time positioning, where an incoming "query-observation" is used for finding the best-matched fingerprint on each radio-map. Root Mean Square Error (RMSE) and Euclidian distance are computed and used as comparison mechanisms. Accuracy plots and error plots are generated for comparison and evaluation, and are discussed on the evaluation chapter. Some expected results are:

- A high accuracy can be achieved by using only the top most trusted WAPs on the fingerprint.

- A smaller number of well selected WAPs can achieve the same accuracy as systems employing all WAPs.

Measurements were also recorded in every office within the Research Computing Services (RCS), a suite of offices at The University of Manchester. A set of measurements was recorded at the start of the data collection period, and a second set of measurements was recorded at the end of this period. This is used to evaluate positioning performance over an extended time period for each fingerprinting technique.

1.3 Motivation

There is an increasing demand for applications providing location based services, [Kaa03, JW08, GKT10, Mad14, RK05] , and technologies such as Global Positioning System (GPS) [VDB76] are now widely used for outdoor positioning. The success of GPS combined with the convenience of GPS technology generates an impetus to develop a technology providing a similar service for indoor environments.

Accurate indoor positioning, where GPS is unavailable, has been significantly researched. Approaches have considered using signals across the electromagnetic spectrum (Wireless Fidelity (WiFi) [VSK$^+$10], Bluetooth [Oks14, LKJ$^+$14], visible light [MT11], infrared) and other sources of data such as audible sound [HHZ$^+$14], ultrasound [FCC10, WJH97] and inertial measurements [LCC$^+$10].

This research considers the addition of temporal information available within WiFi technology that can be exploited over an existing wireless network infrastructure. Using only the passive information broadcast by Wireless Access Points (WAPs) in a locality, we show how temporal information can be incorporated and show how this improves WiFi-based localisation algorithms. In the past, some attempts have been made to analyse system performance over extended periods of time by mapping and monitoring any changes to the wireless network.

The interest in Indoor Positioning Systems (IPSs) is showing a significant increase. There is a large number of applications currently using IPSs, this coupled with the large number of potential applications has attracted the attention of many researchers around the world. A common thread in all this research is to find both a faster and more accurate positioning system. There are two main approaches being developed to achieve this goal. The first is the use of specialised hardware such as Radio-frequency identification (RFID), lasers or wireless sensors networks. These technologies have proven to be accurate. However the need for specialised hardware results in an expensive system when deployed in large and public areas. The second approach, is an opportunistic approach, where the positioning techniques make use of existing infrastructure such as cellular and WiFi networks. These pervasive communication technologies can present the perfect set of characteristics for indoor positioning when GPS performs poorly.

WiFi Positioning Systems (WPSs) are already popular and are currently used commercially by companies such as Skyhook, Ekahau, Ubisense and Google. The two principal models used in WPS are the propagation-based model and fingerprint-based model. The latter has grown in popularity because it provides a more tractable solution to the fact that indoor WiFi signal propagation is affected by obstacles such as walls, furniture and people.

The Fingerprinting method consists of collecting and storing the signal strength of all the detected WAPs in a physical location. Then creating a WiFi signature that uniquely identifies that particular location. The earliest work successfully implementing fingerprinting using the signal strength from 802.11 beacons was presented by [BP00] in early 2000 and later by [HRF$^+$04]. The Fingerprint at a location is in essence calculated by finding and storing a representative value for each WAP found. This value is then used to match new measurements from unknown or new locations.

However, Fingerprint-based positioning is not without challenges, some such challenges are listed below:

- Large effort is required to build and maintain the Fingerprint database.
- Updating the Fingerprints due to WiFi infrastructure changes such as removal or addition of WAPs.
- Updating the database to reflect changes in physical environments such as building renovations and/or refurbishment.
- The difficult task of detecting when the physical layout of an environment has changed.
- Large data storage requirements, as potentially all measurements need to be stored.

Clearly, methods are needed to effectively compress and/or reduce the amount of stored data. Also better methods to detect environment change are needed. These two needs are closely related and form the basis of this research. The result is a system that can better learn and adapt to the WiFi environment.

The widespread adoption of WiFi technology in consumer electronics such as smart phones and laptops presents a convergence in technology. This convergence can be utilised to construct an enhanced IPS. The system designed, created and maintained allows for indoor environment radio-maps to be built that can be used to provide an increased level of spatial awareness.

1.4 Contribution Overview

This thesis introduces an improvement to indoor positioning systems, that extends current works in the concept of WiFi fingerprinting.

The concept presented in this document goes further by incorporating the use of control charts into the fingerprint definition. There are four key contributions of this work:

1. New method of evaluating the contribution of WAPs to fingerprinting. Consisting of the following:

    i. First time implementation of configurable EWMA for fingerprint-based Positioning Systems

    ii. Principle of WAP discrimination based on changes and outliers

    iii. Scoring function for ranking WAPs

2. Development of a unique and highly valuable dataset of observations

3. Database schema

4. Data capture scripts

1.5 Research Scope

It is argued that it is possible to achieve positioning by combining readings from different bands of the electromagnetic spectrum. The scope of this study addresses the use of only a specific band of the Radio Frequency, more specifically the one used by WiFi technology, to create new and more accurate dynamic mappings of indoor environments.

1.5.1 Research Objectives

- This project aims to create a novel fingerprint definition for fingerprint-based IPS to improve positioning performance.
- Testing and evaluating the proposed fingerprint definition using a dataset comprising of WiFi data captured over a period greater than 1 year.

1.5.2 Reasearch Achievements

The following publications are original contributions of the research:

- A unique and highly valuable dataset of observations [RFTM14]. Publicly available (see `http://dx.doi.org/10.5281/zenodo.12913`)
- The database schema and it Enhanced Entity-Relationship (EER) Model [RFTM15]. Publicly available (see `http://dx.doi.org/10.5281/zenodo.13793`) is the SQL script for generating the database.
- Scripts for WiFi data collection [RFMT15a]. Publicly available (see `http://dx.doi.org/10.5281/zenodo.18130`)
- Functional fingerprint based WiFi Positioning System (WPSv1) [RFMT15b] that collects data, generates fingerprints using state of the art fingerprinting techniques and the improved EWMA Fingerprint. Publicly available (see `http://dx.doi.org/10.5281/zenodo.19737`)

The remainder of this thesis is organised as follows:
Chapter 2, Background - A technical background, discussing work related to the localisation problem.
Chapter 3, System Architecture - Definition of the system designed in this research, data sources used, the hardware used to acquire the signals and the process of data acquisition.
Chapter 4, Methodology - The methods and techniques that have been employed and the implementation detail.
Chapter 5, Data Analysis - The data set used as a test-bed is presented and its descriptive statistics are explained

Chapter 6, Fingerprinting and Positioning - Describes the techniques for enhancing the fingerprinting process. The radio-map creation and update is covered, as well as the WAPs selection techniques.
Chapter 7, Evaluation and Results - Evaluation of the proposed fingerprinting techniques on the data set is presented.
Chapter 8, Conclusions and Further Work - Presents the main conclusions of the work and discusses further work.

Chapter 2

Background

This chapter presents the concepts required to comprehend the ideas that will be elaborated on throughout the rest of this thesis. The main issues around IPSs are described; including applications, challenges, technologies and techniques.

WiFi technology is then presented as a suitable, opportunistic, low-cost and pervasive solution for an IPS. In order to highlight and illustrate the pervasiveness of WiFi technology, a single measurement was made at a shopping centre (Derby, UK), and this observation resulted in a list of 67 different uniquely identifiable Wireless Access Points (WAPs). This single example is presented just to illustrate the large amount of hotspots available at any time in many private and public places. A more exhaustive analysis of a large data set covering various locations is described in further chapters.

An important concept covered in this chapter is the “Fingerprinting” approach; this technique is explored to highlight its advantages for use within an IPS. Comparisons with other existing approaches show that fingerprinting is a technique that can be made to be less vulnerable to environmental changes and therefore more robust in use.

2.1 Indoor Positioning

To introduce the concept of IPS consider the following scenario. You are arriving at a building where a very important meeting is taking place. It is in a building were you have never been before and unforeseen events make you arrive with little time to spare; so you are at the entrance of a building that is difficult to navigate and all you have is the office number and absolutely no idea of which way to go. This is a scenario that can be solved by using an IPS based on a mobile app.

The mobile device would detect signals from surrounding wireless networks to identify the user's current location, with knowledge of the user's location, the mobile device would compute the route to the destination, displaying graphical guidance and/or spoken instructions of the route to follow in order to reach the sought-after office room. This would provided a similar functionality for indoor environments as that provided outdoors by GPS devices.

Indoor Positioning is a promising, exciting and exponentially increasing area of research. The large number of potential applications makes it an appealing challenge to undertake as well as being able to be integrated within many academic and commercial opportunities. Indoor Positioning considered here involves the physical location of someone (user/client) or something (asset, electronic device, lost object, robot) in areas where technologies such as Global Positioning System GPS perform poorly or are not available. Many daily activities could benefit from the implementation of Indoor Positioning, and some potential applications include:

Mobility / Transportation/ Navigation (e.g. robotics, unmanned vehicle tracking, semi-unmanned vehicles, and navigation in universities, museums, airports, train stations, hospitals, shopping centres, etc...)

Education (e.g. handheld guides in museums, navigation in universities, image-based reconstruction of museums, academic performance)

Health (e.g. navigation in hospitals, locating staff/patients within a hospital, health services, providing help for elderly and visual impaired people)

Security (e.g. surveillance, probation tracking, access control)

Personal (e.g. location based reminders, lost property recovering, video-games, augmented reality, videoconferencing)

Industrial (e.g. process automation, inventory management and asset tracking)

Social (e.g. enhancing social networks, locating friends)

Commercial (e.g. locating staff within the office, personalised advertising, shopping centre navigation, automatic check-outs)

Private companies [urla] [urlc] [urld] [urlh] [urlg] [urlf] [urli] are actively working on indoor positioning and aim to cover one or several of these previously mentioned application areas. The In-Location Alliance (ILA) [urle], established in August 2012 is an attempt to achieve standardisation. The main ILA aim is to create a technology independent indoor localisation using WiFi and Bluetooth technologies, and is formed by well-known companies including Nokia®, Samsung® and Sony®.

2.1.1 Indoor Positioning Challenges

Achieving an accurate IPS is not an easy task and some of the concepts that should be addressed in solving indoor positioning are the following:

Nonlinearity. The construction materials, thicknesses of walls, pipes and electromagnetic sources, all make the environment highly variable and unpredictable. Because of this, the development of an Indoor Positioning System that fits any indoor environment is a challenge.

Multi-path. Refers to the different paths followed by an electromagnetic signal before reaching a receiver. This could generate interference, fading of the original source and generating signal duplication.

Reflection, Refraction and Absorption. Reflection and refraction are causes of multipath. Reflection is presented when the signal impacts a body or surface and the signal is propagated back. Refraction is presented when propagated waves change in direction at the interface between two mediums due to the change in propagation velocity. Absorption refers to the signal being reduced in signal strength; partially or completely, by the medium and/or the surfaces impacted. These phenomena affect all types of sound, light, and EM based systems. The uncertainty about the materials of the impacted surfaces affects the propagation, hence systems based on mathematical modelled propagation may work poorly indoors.

Change of indoor scenery. Changes in indoor environments include: moving objects, building renovations, refurbishing, or people walking around. All these changes affect temporarily or permanently the propagation of the surrounding signals.

All these situations should be taken into consideration when designing an IPS. Furthermore, it is required to handle wisely the tradeoffs regarding Accuracy, Scalability, Cost, Coverage, Complexity, Time response and Adaptability.

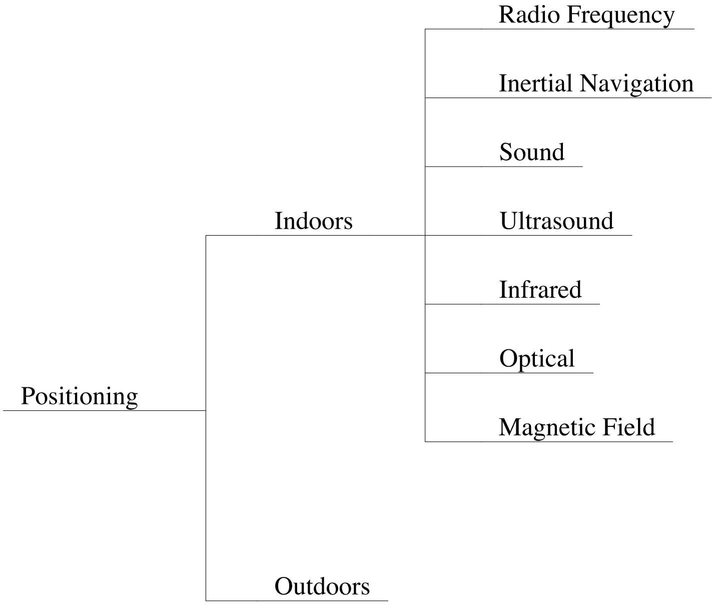


Figure 2.1: Indoor Positioning Technologies.

2.2 Indoor Positioning Technologies

Technologies that have been used for indoor positioning include: Motion (Inertial Navigation), Sound, Ultrasound, Optical, Infrared, Magnetic field and Radio-Frequency. Figure 2.1 presents a diagram listing these indoor positioning technologies. A brief description of non-radio-frequency-based technologies is covered here but please refer to [LDBL07], [MT11], [ZFI13] and [Gos13] for more detailed reviews of these technologies.

Motion based Positioning Systems are also known as inertial navigation systems, and they rely on the motion of the target device to estimate the position. Data gathered by sensors such as accelerometers and gyroscopes are used to compute the speed and direction of the target device. A common downside of this system is that small drifts can add up, leading to a large absolute error in the positioning so it is often used along with other positioning systems with fixed beacons in the environment to increase accuracy and reduce the number and size of errors.

Infrared Positioning Systems are based on infrared sensors (transmitters/receivers), that emit a unique infrared identification/code, installed at known locations. The user's position can be determined by the proximity principle, where the closest receiver is considered the position of the user's location. A downside to using Infrared technology is that dedicated and expensive hardware is required in order to achieve high accuracy and it is highly sensitive to the Line-Of-Sight (LOS) error, which means that if there are obstacles between the transmitter and the receiver the signal may not be detected. A representative example system using infrared is the Active Badge [WHFaG92], that achieves an accuracy of 7cm, and a 5m range.

Ultrasound based Positioning Systems consist of a transmitter installed on the target device that emits ultrasonic waves at a particular frequency. The waves are detected by receivers fixed at known locations within the room so that the target position is estimated based on the Time-Of-Arrival (TOA) attribute. Some limitations using ultrasound for indoor positioning include; dedicated hardware is required and the speed of sound can be affected by temperature and humidity. Active Bat [WJH97] is an example that can achieve 3D positioning with an accuracy of 10cm in the three dimensions with a 50m coverage. Another ultrasound based system is Cricket [Pri05] achieving up to 2cm accuracy within a 10m range of coverage.

Optical Positioning Systems use optic or vision based techniques in order to achieve positioning, some of these techniques include: QR codes, landmark images, light intensity, image matching techniques, rotational laser, or fluorescent light. Optical positioning systems require dedicated hardware, are affected by changes in illumination and are limited to LOS. A detailed survey about optical indoor positioning systems is presented in [MT11].

Radio Frequency based systems use electromagnetic radio waves for positioning, exploiting the large coverage offered within this range of the Electromagnetic Electromagnetic (EM) spectrum. Some advantages of a radio frequency based positioning system include that it can achieve longer range than optical, ultrasound and infrared based systems, and it is not limited by line-of-sight restrictions.

Positioning based on radio frequency can be subdivided into Wireless Local Area Network (WLAN), RFID, Bluetooth, ZigBee, Ultra Wide Band (UWB) and Frequency Modulation (FM). Figure 2.2 presents a classification of the Radio Frequency based technologies for Positioning Systems.

Figure 2.2 highlights the system which this research will focus on. WLAN, more specifically WiFi technology, will be thoroughly explained in the following section.

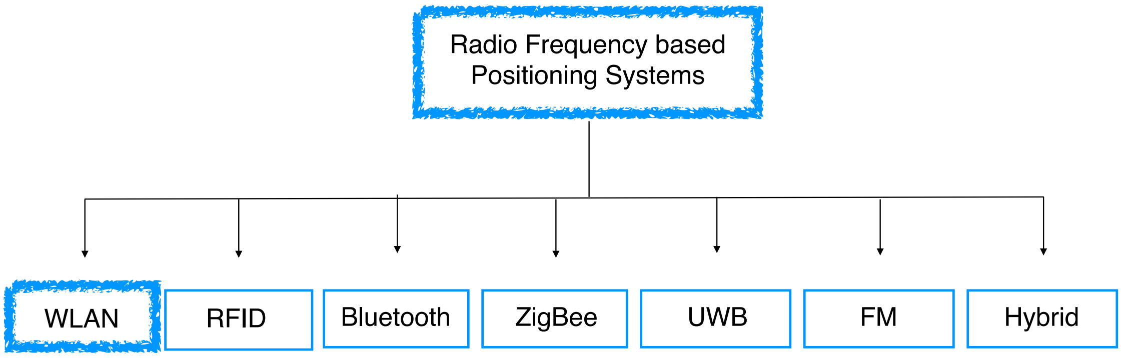


Figure 2.2: Positioning Systems Taxonomy.

2.2.1 WiFi Technology

The term WiFi stands for Wireless Fidelity. WiFi is a wireless networking technology based on electromagnetic radio waves for data transmission. The WiFi technology enables electronic devices such as computers, smart-phones, printers, digital audio players and video game consoles to interconnect amongst each other and to have untethered Internet connectivity. Many emerging electronic devices by default include a Network Interface Card (NIC) for wireless connectivity.

WiFi can now be considered a ubiquitous technology as it is used extensively in offices, universities, private residences and public places such as: shopping centres, museums, airports, train stations or libraries.

WiFi follows standards set by the Institute of Electrical and Electronics Engineers (IEEE). The standards currently in use are the IEEE 802.11 b/g/n, and a new standard, the 802.11ac is under development. The purpose of the Wireless Ethernet Compatibility Alliance (WECA), also known as WiFi Alliance, is to test and to certify the compliance of electronic devices implementing the IEEE 802.11 standards.

The frequencies established for communication under the IEEE 802.11 a/b/g/n standards for WiFi are 2.4GHz and 5GHz, and these frequencies are within the industrial, scientific and medical (ISM) radio bands.

The Basic Service Set (BSS) for WiFi networking includes a Wireless Access Point (WAP) and at least one associated station or device equipped with a NIC. In accordance with the WiFi Alliance, the radio signal emitted by WAP devices covers a range of about 200m (656 feet) outdoors, in a wide open area with no obstacles. However, it has been known for successful transmissions to up to 305 meters (1,000 feet) and even up to 1.6 kilometres (1 mile) under optimal conditions. Nevertheless, the covered distance is drastically reduced indoors. Blockage by objects such as walls, furniture and people cause signal attenuation or signal loss and typically an indoor signal travels within a range of 20 to 46 meters (65 to 150 feet).

Some of the parameters used in WiFi wireless communication are:

- The Service Set IDentifier (SSID) or WAP/network name, which consists of a 32 byte long string that identifies the network.

- The Basic Service Set Identifier (BSSID) also called the Media Access Control (MAC) address. This is a serial number 48 bits long often shown as a combination of hexadecimal characters that is stored in hardware and acts as the NIC's unique identifier. It can be changed but its modification is not an easy or common task.

- The Received Signal Strength Indicator (RSSI) is the strength of the signal measured in dB and its range varies based on the electronics, but the empirical data varies in a range from -10dB to -100dB.

- The Channel is the number of sub-divisions of the ISM range of the spectrum used by the WAP at the particular moment the data was transmitted.

- High Throughput (HT) is a binary parameter indicating the standard employed. If the HT value is equal to one, it means that the standard used by the WAP is the IEEE 802.11n.

- Country Code (CC) indicates the country where the WAP is registered.

- The Security information string reveals the security settings employed by the WAP and can vary between: None, Wired Equivalent Privacy (WEP), Wireless Protected Access (WPA) and the second version of Wireless Protected Access (WPA2). The function of these encryption algorithms is to protect the content transmitted within data frames. WPA2 is the most secure out of the three current standards.

All these parameters are broadcasted by the WAPs and are received by WiFi enabled devices.

WiFi technology has become a popular research subject for indoor positioning ([BP00], [BBP00], [KKJ$^+$04], [GJ04],[AKS04] ,[PKL07], [YAUS03], [SL14], [XSC$^+$04]) and outdoor positioning. Some advantages of using WiFi for indoor positioning are:

- Existing infrastructure can be used.
- The space around is full of electromagnetic radio-waves emitted by WAPs.
- WiFi technology is now ubiquitous in many locations.
- There is no need for specialised or expensive hardware to be added.
- It is growing in popularity.
- It is based on regulated IEEE standards.
- Signals can be detected using smartphones or light-weight sensors.
- WiFi signal transmitters also called WiFi Access Points (WAPs) broadcast their physical address which can be used as a unique identifier.
- WiFi technology presents an opportunistic approach for positioning systems

Some disadvantages of a WiFi based positioning system include signal attenuation, multi-path effects and it can suffer from low accuracy.

2.3 Indoor Positioning Techniques

Locating techniques can be classified as follows: Time-based, Angle-based and Power-based. Figure 2.3 presents the classification diagram for existing locating techniques.

The work in this thesis has focused on Power-based techniques and specifically in the RSSI-based technique. The RSSI-based technique has the advantages that can be easily deployable and implemented without requiring specialised hardware. This minimal hardware requirements are because modern smart phones, tablets and laptops are equipped with a NIC able to report the RSSI of an incoming packet. another advantage is that it does not require the knowledge of the WAP position, or time synchronisation. Furthermore, it is less susceptible to signal reflection complexities presented within crowded indoor environments unlike the angle-based and time-based methods. Angle-based and time-based techniques are suitable for outdoors environments with direct LOS. Angle-based techniques present problems in presence of multi-path, noise, and interference [PLM02]. LOS is not always possible in indoor environments, where a target device to be located is most of the time surrounded by objects, resulting in signal reception from multiple angles. Measurements from multiple angles can limit the

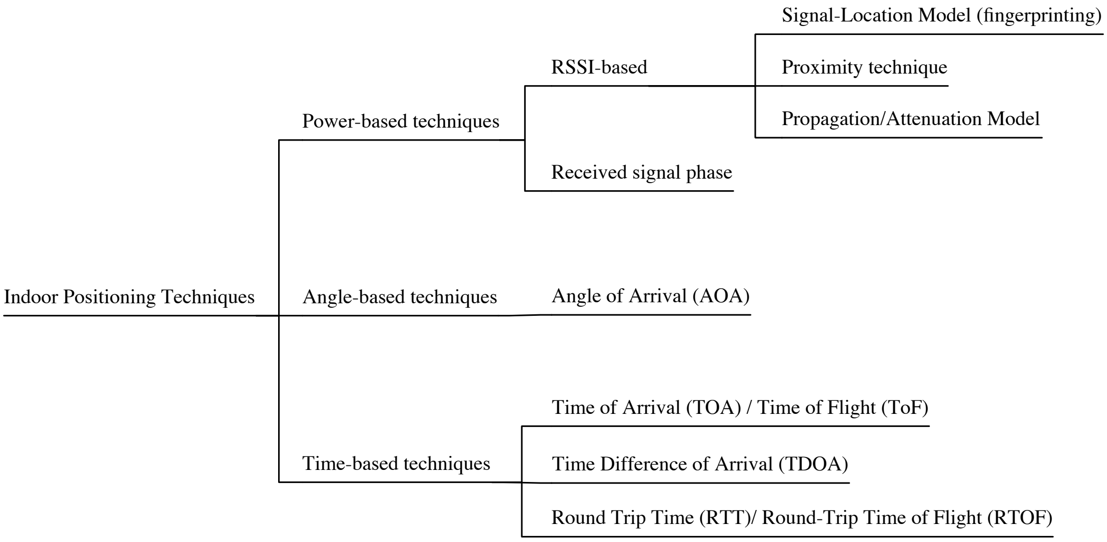

Indoor Positioning Techniques
Power-based techniques
RSSI-based
Signal-Location Model (fingerprinting)
Proximity technique
Propagation/Attenuation Model
Received signal phase
Angle-based techniques
Angle of Arrival (AOA)
Time-based techniques
Time of Arrival (TOA) / Time of Flight (ToF)
Time Difference of Arrival (TDOA)
Round Trip Time (RTT)/ Round-Trip Time of Flight (RTOF)

Figure 2.3: Indoor Positioning Techniques

system accuracy. On the other hand time-based techniques depend on highly accurate clock synchronisation. Both techniques, angle-based and time-based, require relatively large and complex hardware.

In summary the RSSI-based method was chosen to overcome real-world issues associated with angle and time-based methods.

When considering Figure 2.3 there are three main sub-areas to this area which are explained below.

2.3.1 Proximity Technique

This is a simple technique based on the proximity principle, where the closest beacon is considered the user's location. A drawback of this technique is that it is a requirement for the system to know the position of all the beacons. It is possible to estimate the location when the WAPs are correctly located at known places. But unfortunately this is not the case in “real-life” scenarios.

2.3.2 Attenuation Model

An attenuation model relies on a pre-defined characterisation of the signal strength attenuation. In other words, it is assumed that signal intensity detected by the receiver is the inverse of the distance squared; this applies within scenarios where there is no significant environmental effect to signal propagation. The work [CK02] investigates radio propagation modelling on a simulated experiment and compares results of different regressions models. It is concludes that WAPs should be "properly placed" in order to control the location error. In real life this scenario is unrealistic, where the WAP are previously placed for communication purposes. Also, an attenuation model can be complex, especially indoors, where the signal can be affected by other signals on the same frequency emitted by electronic devices, or obstructed by objects or people in the environment. Also, signal propagation errors, such as multi-path, interference, diffraction around corners, multiple reflections, and shadowing are commonly present. Other parameters affecting WiFi localisation are: sensor's quality, type of antenna, attenuation of the signal over time, ambient conditions and scene characteristics.

2.3.3 Fingerprinting

The "Fingerprinting" term in a positioning context was first coined by U.S. Wireless Corp. in 1998. It was named after its resemblance with the identification technique based on people's fingerprints so in a similar way, a geo-location can be identified based on a set of signal signatures ("fingerprints"). The Fingerprinting model can also calculate position taking into consideration the signal propagation issues as previously mentioned. Fingerprinting consists of collecting and storing a set of information that identifies uniquely a physical location. The actual concept of Fingerprints based on Radio Frequency signals, was first presented in 2000 in the work RADAR [BP00].

Within a WiFi context a Fingerprint is the relationship between the symbolic location and the set of BSSIDs and RSSIs received from each WAP within the user's device range. The symbolic location refers to a descriptive string that fits the application, and this could be a name, number or grid coordinates (x, y) within a large open space.

By comparing fingerprinting with other techniques, it is concluded that Fingerprinting is relatively easy to implement as it requires no extra physical infrastructure to be installed or specially aligned. It uses existing WiFi infrastructure and sensors that are built into most mobile phones. It can be said that the fingerprinting problem could be addressed and modelled as a Database problem.

2.3.4 Fingerprinting Challenges

Listed below are some of the challenges that fingerprinting currently faces.

- The effort required to build and maintain the fingerprint database.
- Efficiently managing a constantly growing database.
- Detecting changes in database due to changes in the WiFi infrastructure (e.g. removal, moving or addition of WAPs).
- Detecting changes in the physical environment (e.g. building renovations or refurbishing)
- System scalability for the database and database searches.
- Determining how often fingerprints/radio-map should be updated.
- Finding the optimal number of measurements required to have a trustworthy fingerprint.
- Defining the number of measurements required in order to provide an accurate fingerprint.

There has been a significant amount of research addressing these challenges [VA14] and state-of-the-art methodologies on Fingerprint-based WiFi Positioning Systems are described in the following section.

2.4 Fingerprint-based WiFi Positioning Systems

The Fingerprinting process of WiFi Positioning Systems takes place in two stages, the "off-line" or calibration stage, and the "on-line" or positioning stage.

The off-line stage consists of data collection such as a manual survey carried out by an expert. This survey involves making observations (BSSID and RSSI detected values from all WAPs in range) at every desired location. The observation then is tagged with the description of the location and the relationship between the WiFi data and the location creates a unique fingerprint. The complete set of fingerprints is called a "radio-map" and is stored into a database (DB) for use within the second stage.

The second stage is the "on-line" or user location stage. Assuming the system is a client-server application. The client (user) to be located makes an observation at a particular point. This observation consists of detecting the BSSID and RSSI from WAPs in the vicinity. The BSSID and RSSI are then compared with fingerprints in the radio-map/DB stored in the server. The fingerprint that best matches using some criteria

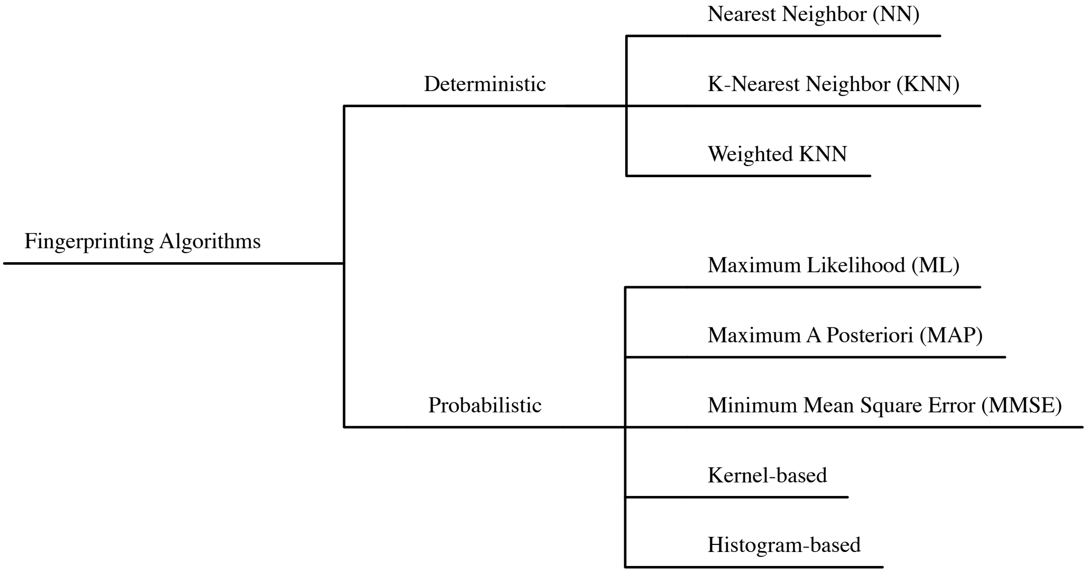


Figure 2.4: Fingerprinting Algorithms

the user observation is selected and presented as the possible current user's location. Several algorithms exist in order to find the best match from the DB in the fastest and lowest-computational cost possible [BB05] [DMM12] [HPALP09] [LWL$^+$05]. Figure 2.4 presents some of the main different algorithms used in fingerprinting.

Most of the state-of-the-art research focuses on improving the time and accuracy within the second stage of the process, and also in how to improve the data collection process. The data collection approaches available are: manual/wardriving, crowd-sourcing and beacons. These approaches will be described below.

2.4.1 Manual Fingerprinting

One of the basic approaches to fingerprint generation is to survey the location manually. This survey/data collection is also called "Wardriving". Wardriving consists of a person or vehicle with a WiFi enabled device scanning the wireless networks. Signal strength measurements are stored into a database to create fingerprints. Updates are then carried out by overwriting old fingerprints with new ones.

Some companies implementing this approach are: Ekahau® [urlb], Skyhook® [urlj], Ubisense® [urlk], and Google®. This manual process may involve a large amount of labour in order to re-survey the locations every time a new set of fingerprints is required. It represents a potentially large cost to ensure the DB is up-to-date and it is unclear when the locations should be surveyed again.

2.4.2 Crowdsourcing based Fingerprinting

The survey could be also done by crowdsourcing, where observations are collected by many final users. Crowdsourcing reduces the time for taking surveys and can be a faster approach for collecting data. Work presented by Gallagher et al [GLDR10] proposes updating the database through user feedback. As it maintains the fingerprint DB, the system monitors changes in the wireless environment. The method proposes a "points" based system where a WAP's score is incremented every time the WAP is seen in a particular location and decremented each time it is not seen in the location. This running count is the basis for adding or removing the WAP from the fingerprint/radio-map. By using a threshold value they determine whether the WAP should be used in the fingerprint/radio-map or not. A second threshold value can be applied to the WAP's signal strength. So for example when a WAP is detected with a signal strength lower than -80 dB, such a WAP is not included in the computations. Their objective is not just to update the fingerprint, but also to control the size of the fingerprint/radio-map DB. The system initially relies on a DB created by a skilled surveyor, then the system makes users to contribute for reducing survey effort. Users are asked their true position periodically. This information is the feedback and it is used for DB update.

Gallagher et al, do not consider the historical changes in the signal strength, instead they focus on the WAPs' frequency of appearance. Their updating methodology is based on the average between the previous value from the fingerprint and the latest value from user feedback.

Work presented by Jun-Sung et. al [LJYH13] introduces a more sophisticated scoring function. Their work uses an exponential approach for accelerating the inclusion of WAPs into fingerprints. This method modifies the simple counting based approach of the earlier work, by introducing a windowed range and an exponential term for the count reduction. The scoring function increments each time the WAP is seen and a threshold value is again used to decide if the WAP should be included in the fingerprint or not. In the instances that the WAP is not seen an exponential term is used to reduce the score which effectively reduces the score by smaller amounts initially. A second threshold value is used to decide when the WAP should be removed from the fingerprint. This approach results in fewer required observations to include a WAP into the fingerprint and requires more absences to remove the WAP. It is proposed that this results in a more stable fingerprint and takes into account more of the reality of WAPs absences. However similar to the previous works this method does not take into account the possibility of fluctuating signal strength values from the WAPs. Instead it relies on the average of the signal strength measurements.

Work in [LDGZ12] titled "WiFi Fingerprint Indoor Positioning System Using Probability Distribution Comparison" presents a fingerprint technique based on the probability distribution of the last n $(n = 100)$ points of the captured data. The off-line stage takes the historic measurements and creates a probability distribution for each WAP. This probability is then stored as the fingerprint value for each WAP. The quoted figure for the number of points used to create the distribution in this paper is 100. The effects of using a smaller or larger number of points is not discussed or evaluated. During the online stage, a smaller set of measurements is taken to create a new distribution. Therefore for each common access point to the user and the map, there exists a distribution. The comparison for likeness of these distributions is the basis for the classifier. Although [LDGZ12] shows that for the datasets used, their classifier does result in accurate results, there appears to be no consideration for the profile of the time history of signal strength measurements.

Barry et al. [BFC09] present a long-term analysis of a crowdsourced positioning system. In this study the dataset spans more than a year and focuses on analysing the system usage, such as: users locations, patterns on binding fingerprints with locations. They presented solid evidence regarding the practicality of using crowdsourcing as a method for data collection.

This work highlights three key areas for future work. The first is the need for a long term analysis on the ageing of fingerprints. Secondly they state the need for a weighted fingerprint update methodology, presumably to handle the effects of different ageing rates of fingerprints. Finally they note the need for a metric study on changes to the environment and how that affects the fingerprints in order to minimise localisation errors.

2.4.3 Beacons and Sensor Network

Finally, fixed beacons could be used instead of wardriving or crowdsourcing. The fixed beacons record observations periodically, and these beacons could be any client-like-device, or the actual WAPs as proposed in the work by Atia et al [ANK13]. These authors [ANK13] present an appropriate solution where the WAPs act as fixed beacons for positioning. This solution is a paradigm-change, and it requires re-programming the WAPs firmware, changing standards and WAPs manufacturing. If this scenario proposed is embraced by the community, it would represent a good scenario for the implementation of the work presented in this thesis.

The work considered in this thesis presents a novel approach that could be implemented on existing IPS. It involves the analysis of the temporal nature within historical data, and the presentation of statistical analysis for improving fingerprinting.

Most of the results from previous work on WiFi fingerprinting agreed that the fingerprinting calibration process should be periodically repeated to maintain an accurate radio-map/database. Although work has been done to overcome these issues, no previous work has investigated how often this calibration should take place. The previous researches also fail on:

- Understanding and detecting temporal signal patterns from WAPs
- Long-term analysis of fingerprints
- Specification of the radio-map updating time-interval
- Keeping radio-maps up-to-date
- Adaptive and more representative parameters for fingerprinting
- Identifying trustworthy WAPs
- Detection of unwanted behaviour on WAP's signal strength

This thesis aims to cover these gaps, proposing a technique that will allow a system to identify signal changes/environmental changes on the signal strength.

2.5 Relevant Statistical Methods

2.5.1 State Estimation

State estimation is the process of estimating the current internal state of a system given the data available. In robotics, for example, it could refer to the computation of a robot's pose. Knowing its current state is vital to determining the appropriate next action to take to accomplish its mission. Recursive state estimation is when the previous state is considered along with new data available. From the robot example, once the robot moves it is clear that its state changes. From this, a new state estimate should be carried out based on the previous state, recent action and new measurement. In other words the state is continuously updated based on input available. An iteration, considering time series, comprises the calculation of the state at time (t) from the state at time ($t-1$) along with the most recent measurement. This iteration is then repeated with every new measurement [TBF06].

Recursive state estimation is relevant for this thesis because a similar approach is employed. Using a recursive state estimator, the "state" of fingerprints is updated continuously based on new WiFi measurements.

Three relevant state estimators are covered in this section: the Weighted Arithmetic Mean a state estimator, the Exponentially Weighted Moving Average (EWMA) a recursive state estimator, and the Kalman Filter an optimal recursive state estimator.

Weighted Arithmetic Mean

The Weighted Arithmetic Mean, also called "Weighted Mean" is a measure of central tendency, it differentiates from the arithmetic mean (sum of all elements divided by element count) in the incorporation of the weighting of certain elements. From this, the arithmetic mean is an special case of the Weighted Arithmetic Mean for when the elements have the same weight, and these weights are normalised with a sum of all weight equal to one. The weighted arithmetic mean of a set of numbers $x_1, x_2, ..., x_n$ having weights $w_1, w_2, ..., w_n$ is presented in Equation 2.1 where the weights are non-negatives and normalised.

$$\bar{x} = \sum_{i=1}^{n} w_i x_i \tag{2.1}$$

Exponentially Weighted Moving Average

Exponentially Weighted Moving Average (EWMA) also known as exponential smoothing was presented by Robert Goodell Brown [BRO56] in 1956. It was designed for predicting demand within inventory-control and quality control for inspection processes. As presented in [BRO56], the EWMA statistics for a time series t=1,2,...,T, is:

$$\hat{x}(t) = \lambda x(t) + (1-\lambda)\hat{x}(t-1), \quad 0 < \lambda \leq 1 \tag{2.2}$$

where

- $\hat{x}(t)$ is the estimated average

- $x(t)$ is the observation/sample at time t
- $\hat{x}(t-1)$ is the previous estimated average
- λ is called the smoothing parameter or smoothing factor, it is a constant that determines the depth of memory of the EWMA

The larger the value of λ (e.g. close to 1) the higher the weight given to recent observations. The opposite occurs for small values of λ (e.g. close to 0), the smaller the value of λ the lower weight is given to recent observations, hence, more weight is attached to observations from distant past. Indifferently to the value assigned to λ, when fixed, the weight given to observations decreases exponentially as the observations get older.

The initial value given to $\hat{x}(t-1)$ from Equation 2.2 can be significant when using a small λ, for this it should be considered to initialise the algorithm by assigning to $\hat{x}(t-1)$ the average of a sample set. The number of samples used for computing this initial average can be selected according to the application.

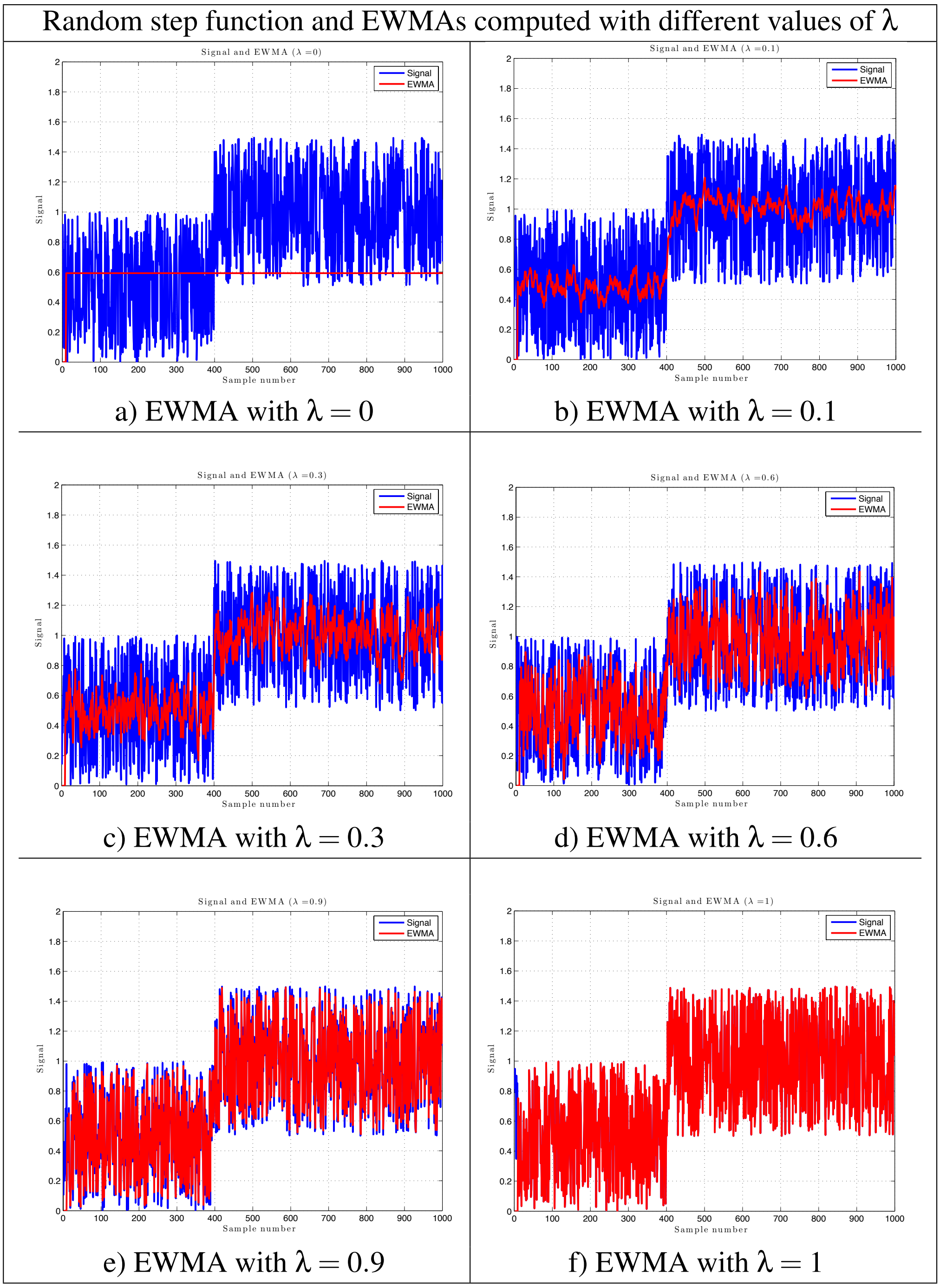
Figure 2.1 illustrates the effect of the parameter λ when EWMA is applied to a step function (step function generated with random numbers). Six EWMA were applied to the step function with the values $\lambda = 0$, $\lambda = 0.1$, $\lambda = 0.3$, $\lambda = 0.6$, $\lambda = 0.9$ and, $\lambda = 1$. In this example, for initialisation, a sample set of the first 10 data points is averaged and assigned to $\hat{x}(t-1)$, applying the EWMA from Equation 2.2 from sample 11 onwards.

It appears that Exponentially Weighted Moving Average (EWMA) has not been applied on fingerprinting techniques previously. Although a relevant work presented by Chen et al [CCLH08] proposes the use of Moving Average Convergence Divergence (MACD) for estimate user's motion. (The MACD scheme, developed by Gerald Appel [App85], is used for identifying market trends.) MACD involves two EWMA filters for computing an agile and a stable response.

In the work [CCLH08], the difference between the two EWMAs, a fast EWMA (large λ) and a slow EWMA, (small λ) is analysed on time series data with purpose of estimating when a user is walking towards or away from a particular WAP. Being [CCLH08] the only work know by the author where Exponentially Weighted Moving Average is applied on WiFi signal strengths.

Kalman Filter

Although the Kalman Filter is not used within the work proposed in this thesis it is briefly mention in this section since it is an optimal estimator. Proposed in 1960 in



Random step function and EWMAs computed with different values of λ

a) EWMA with $\lambda = 0$

b) EWMA with $\lambda = 0.1$

c) EWMA with $\lambda = 0.3$

d) EWMA with $\lambda = 0.6$

e) EWMA with $\lambda = 0.9$

f) EWMA with $\lambda = 1$

Table 2.1: Random step function and EWMAs computed with different values of λ The signal, a step function generated with random numbers (blue), is overlapped with EWMAs (red) computed with different weights (λ). Figure a) shows EWMA with $\lambda = 0$ the EWMA stays immovable form data point 11 onwards, since when using $\lambda = 0$ every new coming sample is ignored. Figures b), c), d) and e) illustrate how EWMA performs by changing the value of λ, the closer its value to 1, the closer the EWMA follows the samples. Figure f) shows the EWMA computed with $\lambda = 1$, for this value the resulting EWMA follows exactly the samples, ignoring completely any previous sample.

[Kal60], the Kalman filter is a recursive solution for estimate the state of a process. It could be seen as a cycle involving two main stages: “prediction” and “correction” [**?**], where with every time update a projection of the state ahead is predicted and with a measurement update, then a posteriori state estimate is updated.

Time update (“Predict”)

1. Project the state ahead

$$X_{t|t-1} = F_t X_{t-1|t-1} + B_t u_t + w_t \tag{2.3}$$

2. Project the error covariance ahead

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q \tag{2.4}$$

Measurement Update (“Correct”)

1. Compute the Kalman gain

$$y_t = z - H_t X_{t|t-1} \tag{2.5}$$

$$S_t = H_t P_{t|t-1} H_t^T + R_t \tag{2.6}$$

$$K_t = P_{t|t-1} H_t^T S_t^{-1} \tag{2.7}$$

2. Update estimate with measurement

$$X_{t|t} = X_{t|t-1} + K_t y_t \tag{2.8}$$

3. Update the error covariance

$$P_{t|t} = (I - K_t H_t) P_{t|t-1} \tag{2.9}$$

Where:

- X, is the system state vector.
- F, is the state transition matrix.
- B, is the control matrix, maps control input to the state.
- w, is the process noise vector.

- P, is the covariance matrix.
- Q, is the process noise covariance matrix.
- y, is the innovation.
- z, is a measurement from the system.
- H, is a matrix that relates the state to the measurement.
- S, is the innovation covariance.
- R, is the measurement noise covariance.
- K, is the Kalman Gain.

Two key concepts introduced in the Kalman filter are the *innovation* and the *Kalman gain*. The innovation is the difference between the measurement and the expected measurement. The gain is the weight given to the innovation, but differently to the weight in EWMA the gain is computed on every iteration. The evaluation into how EWMA performs versus Kalman filter is beyond the scope of this thesis and it could be proposed as further work.

Please refer to [WB95] for an introduction on Kalman Filter and Extended Kalman Filter (designed for nonlinear systems). For further details on Kalman filter please refer to [BH97] and [REI99]. Work applying Kalman filter on indoor positioning system is presented in [EM06].

Some other approaches that are used for state estimation are: Histogram Filter, Bayes Filter and Particle Filter. Details on these algorithms are covered in [Thr02, TBF06] and [GGB$^+$02] were the advantages of particle filters on positioning navigation and tracking are presented.

The work [MHYM13] proposes a localisation system that simultaneously generates 3D models, 2D floor plans and signatures from multiple sensors such as WiFi, cameras, laser scanners and Inertial Measurement Units (IMUs). The system consist on a man-portable backpack equipped with a laptop and sensors. Particle Filter algorithms are used to fuse data from the inertial sensors, WiFi readings and images to localise a mobile device with an average error of under 2m.

2.5.2 Statistical Quality Control

Statistical Quality Control (SQC) is the term referred to a collection of statistical tools that are used to identify quality problems in manufacture [RS05]. An important category within SQC is the Statistical Process Control (SPC). SPC refers to statistical methods used for monitoring the quality in processes and products. These methods are used to determine the amount of natural random variation in the process. Leading to the identification of problems in the process by detecting variations larger than the natural random variation.

Within SPC methods the Control Chart is the most common tool used for process monitoring [RS05]. The Control Chart is employed in this thesis for understanding and monitoring variation in the RSSI from WAPs. This method resulted in a useful tool for establishing a normal range of variation in the RSSI and to detect signal changes and outliers.

Control Charts

The standard Control Chart also called Shewhart Chart was proposed in 1924 by Walter Andrew Shewhart, who is also known for developing the basis of process quality control [She31] and [SD86]. A Control Chart is a quality control tool usually employed in manufacturing processes to determine when a process' variable is in statistical control. On monitoring the process' variable with a Control Chart it is possible to differentiate controlled variations within the process from uncontrolled variations non-intrinsic to the process. Controlled variations are the measurements that move away from the target value, but still are within control limits, on the other hand, uncontrolled variations are those that go outside the control limits. From this, the Control Chart is effective for detecting outliers, large shifts, change points and indicates when a process' variable is drifting unacceptably.

The elements within a Control Chart are the following:

- Initialisation data points. A set of data points that are used to establish baseline performance

- Target. Statistics such as mean, range, or proportion that are used as a baseline/target for the variable. this statistic is computed from the initialisation data points. It is considered the “Central Line” (CL) of the Control Chart. Ideally future observations should be around this value

- Standard deviation. The standard deviation of the initialisation data points is computed for the establishment of thresholds

- Control limits. Upper Control Limit (UCL) and Lower Control Limit (LCL) are equidistant from the central line

- Observations. Data points measured after the implementation of the Control Chart. An incoming observation/measurement is plotted on the Control Chart. If this value is located close to the Central Line and within the Control Limits, then it is considered that the process is under statistical control. On the other hand, if the plotted measurement lies outside the control limits, (over UCL or below LCL) then the variable is considered out of control

- Rules. Specification of what to undergo in the case that signal values drifts away from established mean.

The specific Control Chart employed in this thesis is the EWMA Control Chart. The EWMA Control Chart was initially proposed by [Rob59], described in [Cro87], [Shu08] and [CY08] It is differentiated form Shewhart in that the state of the system depends on the statistic EWMA (see Equation 2.2), rather that using only the latest measurement.

The weight given to λ can be tune according to the process to be monitoring, making the Control Chart able to detect gradual drifts away from the target, rather than react based solely on the latest observation. In the work [LSB+90] they suggest values of λ that can be useful for detecting small shifts.

The advantages of EWMA Control Chart are:

- It is effective for detecting small shifts in the mean or variance

- It combines historical data with the current data

- It can be used in forecasting

The EWMA initialisation involves the computation of the mean (μ) and standard deviation (σ) from preliminary data. These are used as baseline for the system where μ is the target. The variance is computed as shown in Equation 2.10

$$\hat{\sigma}^2 = \frac{\lambda}{(2-\lambda)}\sigma^2 \tag{2.10}$$

The upper and lower control limits are given by Equation 2.12

$$UCL = \mu + L\hat{\sigma} \tag{2.11}$$

$$LCL = \mu - L\hat{\sigma} \tag{2.12}$$

Where L is the number of standard deviations away from the target. In this thesis is presented for the first time the EWMA Control Chart applied to WiFi fingerprint IPS. The EWMA Control Chart is used for detecting changes and outliers in the WAP signal strength, this allows to identify and remove from the fingerprint those WAPs presenting outliers and other unwanted signal profiles. This will be clear in Section 6.1.6 in Chapter 6.

2.5.3 Positioning

The fingerprinting approach to positioning can be classified as a pattern recognition problem, where the fingerprints created from observed WiFi data are label with a location and stored in a database. Then a new observation (unlabelled) enters the system and it is compared against the labeled fingerprints. A metric is used to evaluate the similitude between the fingerprints and the new observation. It can be said that the use of the Nearest Neighbour and its variant k-Nearest Neighbour (KNN) [LDBL07], are a common approach in fingerprint-based indoor positioning systems [BP00], [BBP00] and [SCSB03] .

The Nearest Neighbour operates as follows; given a set of classified elements, find the most similar to a new and unlabelled element by determining their similitude based on a metric. Some metrics used for determining the similitude in NN are:

- Euclidian Distance
- Root Mean Square Error (RMSE)
- Manhattan Distance
- Mahalanobis

The Nearest Neighbour is relevant to this work since it was implemented on the positioning stage, using Root Mean Square Error (RMSE) as metric for fingerprint selection. The work [LYHS13] employs the NN technique for WiFi fingerprint-based indoor localisation.

Another relevant method that have being used in WiFi Fingerprint-based localisation is the Support Vector Machine (SVM). Introduced in [BGV92] developed from Statistical Learning Theory, SVM are a set of supervised learning algorithms, commonly used for classification and regression. A SVM consist basically in a hyperplane that divide two classes, this hyperplane is built to maximise its distance between both classes. This model then is used to determine the class for a new unclassified observation. A drawback of SVM is that it is applicable for two-class problems. WiFi positioning is a multi-class problem, SVM has to be used along with algorithms that reduced the data from multi-class to several two-class sub-problems. The work [BB05] implements Support Vector Machine (SVM) paradigm by mapping the input data into a higher-dimension employing Kernel functions. The research compares the SVM with three existing techniques used in WiFi fingerprinting such as: Weighted k nearest neighbours (WKNN), Bayesian modelling and Multi-layer perceptron. From comparison on the same data, their conclude that SVM outperforms the rest of the algorithms except for WKNN which matches closely their SVM results for spatial localisation.

Further references on the use of SVM can be found in [KPV07] and [VWG^{+}03].

Research [LSDR06] presents a comparison between deterministic and probabilistic methods. It concludes that the Bayesian method is more reliable for moving localisation where it outperforms NN. It also recommends Bayesian localisation since it facilitates the incorporation of probabilities from other sensors. Finally it concludes that the Bayesian based localisation outperforms only marginally the NN method when compare on static localisation. The static localisation is the one tested in this thesis.

Please refer to [DC11] and [HPALP09] for further comparative surveys including deterministic and probabilistic methods.

2.6 Chapter Summary

The concepts presented in this chapter show the importance of IPS and some of the current approaches were described. This chapter also presented some of the advantages that highlight WiFi technology as an existing infrastructure that provides the suitable conditions for an IPS to be built upon. The concept of fingerprinting was presented as well as some relevant fingerprinting methods and the following chapter will present the research Methodology that was undertaken.

Chapter 3

Methodology

"The only man I know who behaves sensibly is my tailor; he takes my measurements anew each time he sees me. The rest go on with their old measurements and expect me to fit them."

George Bernard Shaw

"Errors using inadequate data are much less than those using no data at all."

Charles Babbage

This chapter presents a review of the research method followed during this thesis, a discussion of the data collection method, and details on how the hypothesis was tested.

3.1 Data collection and measurements

The data collection method is designed as a passive scanning longitudinal study. In this case the quantity being measured is the observable occurrence of radio frequency signals from 802.11 networks.

The research study was designed to collect weekly measurements, carried out during an extended period of time lasting more than one year.

Custom software was created to capture the observations required. In order to maintain consistency across all the observations, a standard operational procedure involving the locations and hardware used was created. This procedure was adhered to for all the measurements taken.

3.1.1 Experimental Setting

The study was designed to capture data from real world scenarios, measuring WiFi signals from real and complex wireless environments with multiple 802.11 networks. In order to examine the variability of the WiFi environments presented in everyday settings some exploratory observations were carried out.

These exploratory observations were conducted in locations visited by the author during working days. Table 3.1 presents the schedule of times and locations visited during one of these days and indicates the period spent at each particular location. A set of measurements were then taken during these time periods.

Three significant locations were selected for in-depth analysis and evaluation.

These three locations are: Research Computing Services (Office 1.035) at The University of Manchester, Manchester Piccadilly Railway Station and a Private Residence.

These locations were selected based on the following criteria: convenience, consistency and repeatability. These locations also presented different types of WiFi environment characteristics.

This was of interest due to the potential differences in long term analysis of the different environment characteristics. The three locations can be broadly classified as follows: Manchester Piccadilly Railway Station as a highly crowded location, Research Computing Services (Office 1.035) as a medium crowded location and the Private Residence as a low crowded location.

During the extended period of data capture these three physical locations were measured with a higher degree of consistency.

Time	Selected Locations
8:00 - 8:50	Derby train station, Derby, UK
9:13 - 9:42	Train Derby-Sheffield, UK
9:45 - 10:05	Sheffield train station, Sheffield, UK
10:11 - 11:05	Train Sheffield-Manchester, UK
11:50 - 15:30	Research Computing Services, The University of Manchester, Manchester, UK
16:00 - 16:17	Piccadilly Railway Station, Manchester, UK
16:20 - 17:05	Train Manchester-Sheffield, UK
17:23 - 17:50	Train Sheffield-Derby, UK
18:30 - 24:00	Private residence, Derby, UK

Table 3.1: Schedule and locations: captured by the author of a typical working day.

3.1.2 Wi-Fi Data Sampling

All data capture was carried out using a MacBook (late 2008) running OS X 10.8 Mountain Lion. A unix script was designed and implemented to scan the 802.11 networks using the AirPort utility from Apple Private Frameworks.

The data capture script is started manually by the user. Once the script is running, the user is asked the name of the room, after 10 seconds if no name is provided a default *unknown* is set. A folder is created where new observations are going to be recorded. The folder is named after the current date, time and room name.

Then, a initial request is prompted by broadcasting a Probe Request Frame, that results in a response (Probe Response Frame) from every WAP within range. The script receives the Probe Response Frames from these WAPs and stores the returned information. This request process is repeated for a pre-defined number of times, with new requests prompted 5 seconds after the end of the previous one. The number of requests and the waiting time are configurable in the script.

The time it takes to detect and to record measurements varies based upon the number of WAPs detected. Hence, using a time interval once the previous measurement set has finished provides a higher certainty that all nearby WAPs are detected and their responses are recorded before requesting more measurements.

The script was tested in real world cases. From all the locations sampled, the location Piccadilly Railway Station presented the larger number of WAPs in a single observation (67 WAPs). A waiting time of 5 seconds was enough for receiving and storing the data from those 67 WAPs. For a location having significative larger number of WAPs, (e.g. more than 70) it is recommended to test the script, and potentially increase the waiting time, which is a configurable parameter in the script. This for allowing the data from all the surrounding WAPs to be received and stored before the following Probe Request Frame is broadcasted.

Once the specified number of request is complete, the responses are stored within the folder as an ASCII txt file. Finally the script indicates the end of the process and provides information about the route and name of the folder where the observations where stored. Please refer to appendix B.1 where this data collection script is provided.

Parameters are set, including the number of requests and waiting time in seconds, The user is asked the name of the room, after 10 seconds if no name is provided a default *unknown* is set. A folder is created where new observations are going to be recorded. The folder is named after the current date and time and the name of the room. A request is carried out, and the response is stored within the folder as an

ASCII txt file. After the response the script waits a pre-defined number of seconds before doing another request. This request process is carried out for a pre-defined number of times (numRequest). Then the script indicates the end of the process and provides information about the route and name of the folder where the observations where stored. Please refer to appendix B.1 where this data collection script is provided.

For each physical location the attributes of the measurement session undertaken is summarised in the list below:

- Measurements recorded at Piccadilly Railway Station were taken over a period between 15 and 20 mins, with an interval of 5 seconds between each measurement.

- Measurements recorded at Office 1.035 RCS, The University of Manchester, were carried out on periods of around 3 hours, also with intervals of 5 seconds between measurements.

- Measurements recorded at the Private Residence were extended to periods of around 5 hours with again 5 seconds delays between measurements.

3.1.3 Sampling System

All the data presented was collected using the same MacBook (late 2008) laptop and in order to provide consistent reproducibility the locations and position of the measuring device was carefully defined on every session. The observations were done statically, with the laptop placed in approximately the same position $\pm 1m$ for each of the subsequent surveys. The observations were made with minimal variation on device orientation and height above ground. The complete software and hardware system was made simple to use to minimise variation over the complete study period.

3.1.4 Variables

Each observation includes: timestamp, location and the set of WAPs detected. For each WAP the following parameters are recorded, *BSSID, RSSID, RSSI.*

3.1.5 Independent Variables

The Independent variables considered in this study are:

BSSID The BSSID is the identity of a particular WAP. The Basic Service Set Identification (BSSID), also known as the Media Access Control (MAC) address, is a key element for each WAP and is designed to be a unique identifier.

The Service Set Identification (SSID). Commonly known as the "name" of the network is a 32 bytes string that publicly identifies the network. This is designed to be human readable and understandable. Although it identifies the network, it is not necessarily unique and could be easily copied by other WAPs or changed.

3.1.6 Dependent Variables

Received Signal Strength (RSSI). The RSSI is a variable by which the WAP can be referenced. The RSSI is dependent on the WAPs location and the location where the measurement is taken and is one of the main parameters to be analysed. Generally speaking the closer to the WAP the higher the intensity with measured values for this parameter in the range -11dB to -96dB, with -11dB being the highest intensity (strongest signal) recorded in our datasets.

WAP score. This is the internal calculated metric to score the WAP trustworthiness. It is dependent on the RSSI, the EWMA (exponentially weighted moving average) values and the number of occurrences of the WAPs appearance.

The statistical procedures used to analyse the data are:

- *Descriptive Statistics*

- Measures of central tendency (mean, median and mode)

- Measures of variability about the average (range and standard deviation)

- *Exploratory Data Analysis*
- *Confirmatory Data Analysis*
- *Temporal order analysis*
- *Analyse relationships between variables*
- *Correlation between variables*
- *Identify changes over time*
- *Identify long term behaviour*

- *Analysing the mean and the distribution of the measurements*
- *Analysis in the spatial domain*
- *Analysis in the temporal domain*

3.2 Limitations

Some imitations in the research are:

- The extended period of time it requires for data collection
- Changes over small periods of time are not included (e.g. signals strength variations over one day)
- Only two dimensions coordinates for the space grid
- Only one orientation is considered (The orientation of the measuring device was fixed for each measurement session. Hence variations due to changes on the device orientation are not considered)

3.3 Summary

The purpose of this chapter was to describe the research methodology of this study, data collection details and define a plan for testing the hypothesis as well as present some of the limitations in the research.

Chapter 4

System Architecture

This chapter describes a Fingerprint-based WiFi Positioning System Architecture that has been implemented within this thesis. The components of the system, fundamental processes and interactions are all covered within this chapter.

4.1 Conceptual Overview

A Fingerprint-based WiFi Positioning System Architecture is formed by the following four components: WAPs, Client, Server and Database (DB). The system takes place in two stages, an on-line or calibration stage, and an off-line also called a positioning stage. Figure 4.1 is a conceptual overview diagram of the system. It presents the system divided by stages (rows) and components (columns), it also depicts the interactions between the components. The components highlighted on the diagram are covered in more detail in subsequent chapters.

4.2 Inter-Subsystem Data Flow

This section is a detailed description of the Research Architecture that was designed and implemented. The tasks carried out by each of the components are presented as subsystems. The Client is divided into three subsystems: Client Management, Data Acquisition and User Interface. The server is divided in five subsystems: Server Management, Data management, Fingerprinting, Positioning and Storage management. These subsystems are explained below. Figure 4.2 is the Inter-Subsystem Data Flow diagram. The subsystems are presented along with the interaction among them. The interactions or data flow is labelled with a number, each number represents the data

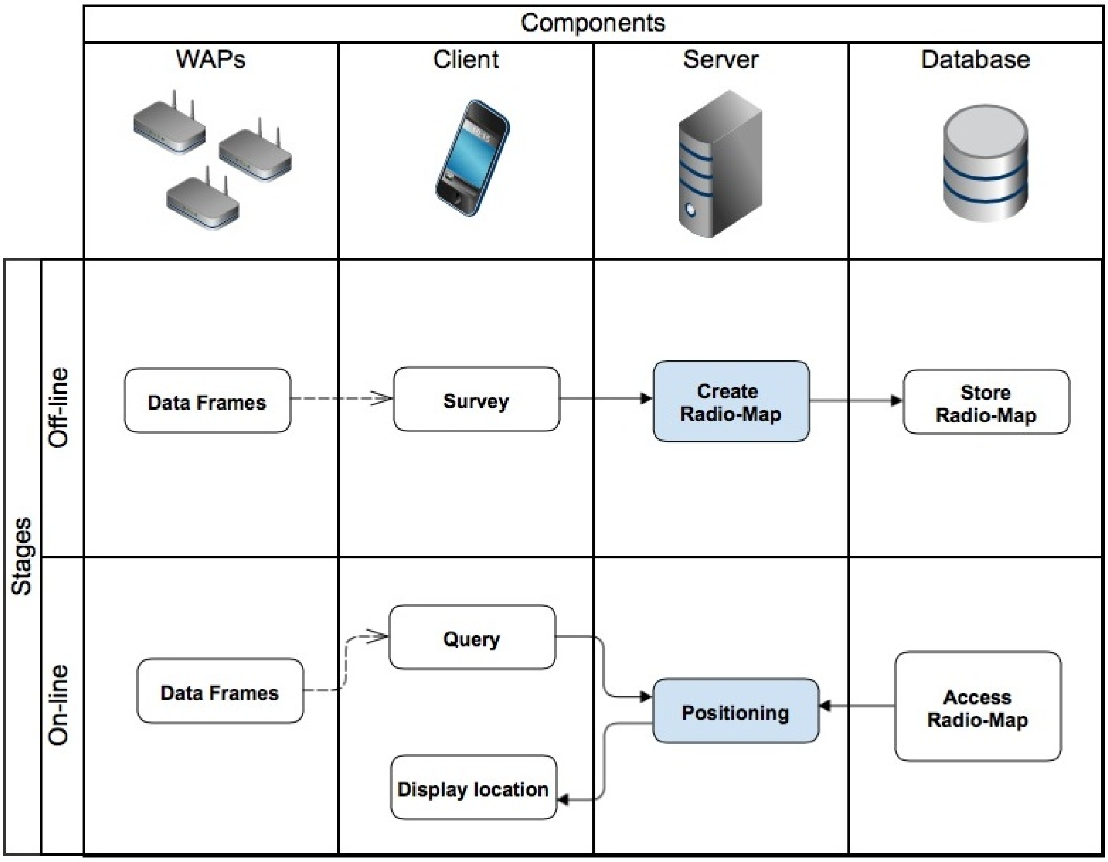


Figure 4.1: Conceptual Overview Diagram.

The key processes for each stage in a fingerprint-based positioning system are highlighted in blue. The process "Create Radio-Map" involves the creation of fingerprints from Survey observations. The algorithms for creating fingerprints are based on specific definitions (more detail on fingerprint definitions is discussed in chapter 6). The key process in the on-line stage is the "Positioning" component, this component involves algorithms for estimating a location by finding the best-match between the Query observation and the fingerprints stored in the radio-map. The algorithms used in finding the best match can follow a variety of probabilistic and/or deterministic approaches (see chapter 2).

elements flowing throughout the subsystems. Table 4.1 lists the names for each data element.

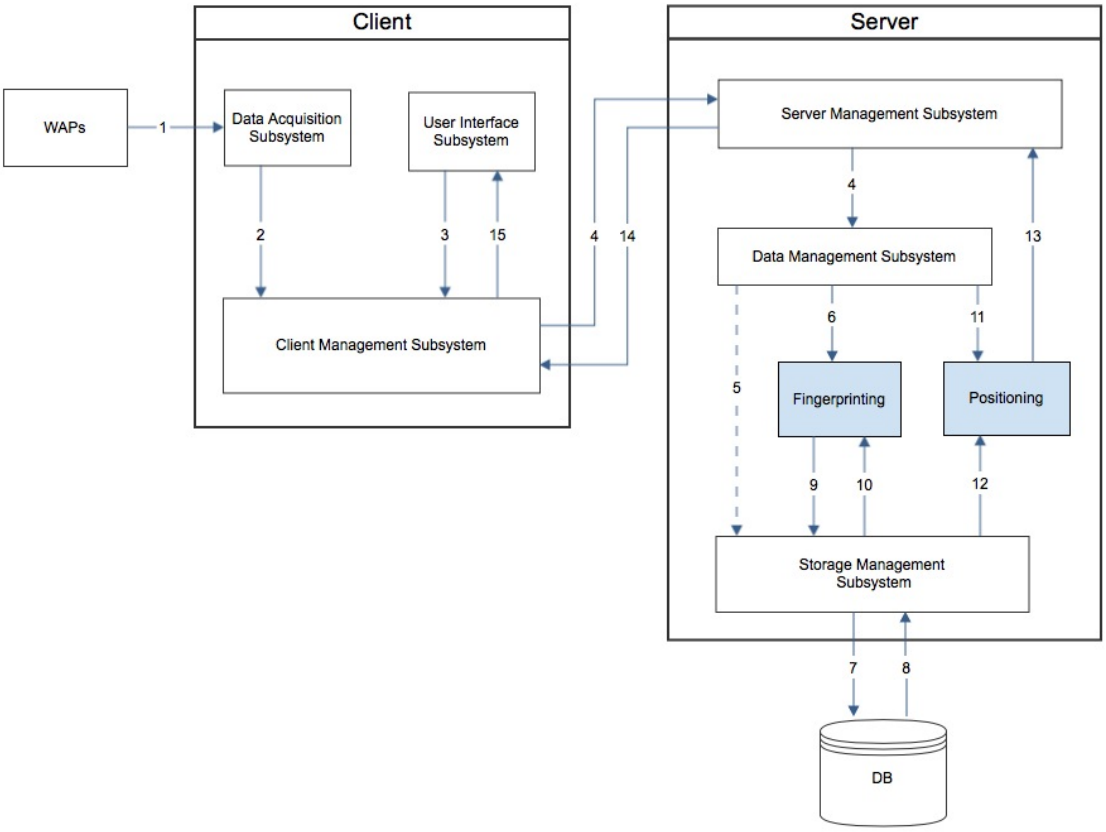

Client
Server
WAPs
1
Data Acquisition Subsystem
User Interface Subsystem
2
3
15
Client Management Subsystem
4
14
Server Management Subsystem
4
Data Management Subsystem
13
5
6
11
Fingerprinting
Positioning
9
10
12
Storage Management Subsystem
7
8
DB

Figure 4.2: Inter-Subsystem Data Flow.
This diagram represents the implementation of a fingerprinting-based positioning system. Key subsystems “Fingerprinting” and “Positioning” are highlighted in blue, these are equivalent to “Create Radio-Map” and “Positioning”, highlighted in diagram 4.1. The numbered lines in this diagram represent data elements transmitted between the subsystems. These data elements are named in Table 4.1 and described in section 4.2.1. Data element number 5 shown as a dashed line represents the storage of raw data for research analysis, and would not be required in an actual IPS implementation.

4.2.1 Inter-Subsystem Data Elements

This section covers in more detail the inter-subsystem data elements from Table 4.1.

Number: 1

Name: Probe Response Data Frame

Description: Information broadcasted by WAPs in response to the Probe Request Data

Data Elements
1. Probe Response Data Frame
2. Observations TXT Folder
3. Symbolic Location
4. tagged Observations TXT Folder
5. survey-Observations CSV Folder
6. survey-Observations CSV Folder
7. SQL statements
8. DB results
9. New/Updated Fingerprint
10. Fingerprint
11. query-Observation CSV Folder
12. Radio Map
13. Best Matched Fingerprint
14. Estimated Location
15. Estimated Location

Table 4.1: Inter-Subsystem Data Elements.

The abbreviation TXT used in data elements 2 and 4 refers to the filename extension, these files are stored as a plain text in ASCII file format. The abbreviation CSV used in data elements 5, 6 and 11 is the filename extension to files stored as character-separated values. The CSV file contains lines, these lines are considered records formed by fields. The fields within a record are separated by comma character. This CSV structured format facilitates the posterior database insertion.

Frame used by the client in an attempt to discover the available networks.

Number: 2
Name: ObservationsTXT Folder
Description: Folder containing one or several observations in txt format (Observation TXT). Each folder contains one or several observations taken at a specific time and location containing the following information per WAP detected: SSID, BSSI, RSSI, CHANNEL, CC, SECURITY. (refer to section acronyms if not recognised). Figure 4.3 is an example of an Observation TXT

Number: 3
Name: Symbolic Location
Description: A descriptive alphanumeric string, number or coordinates representing a physical location. This data element is a symbolic string, describing the location where measurements took place.

Number: 4
Name: tagged Observations TXT Folder
Description: Folder named after the location if it is known (survey), or after timestamp when location is unknown (query)

Number: 5,6
Name: survey-Observations CSV Folder
Description: Observations in Comma Separated Values (CSV), observations stored on DB for future analysis (number 5) and Observations for realtime fingerprinting (number 6). Figure 4.4 shows an example of an Observation CSV.

Number: 7
Name: SQL statements
Description: SQL queries for database interaction (e.g. insert, update and data retrieval)

Number: 8
Name: DB results
Description: Rows and tables from the database containing results requested thought

the SQL statements (data element number 7)

Number: 9
Name: New Fingerprint and Updated Fingerprint
Description: A data structure containing a new fingerprint or an updated fingerprint. These are the data structures resulting from the fingerprinting process. These fingerprints are formatted for database storage (to be included in the radio-map).

Number: 10
Name: Fingerprint
Description: Fingerprint stored in the database.

Number: 11
Name: query-Observation CSV Folder
Description: Folder containing observations from an unknown location

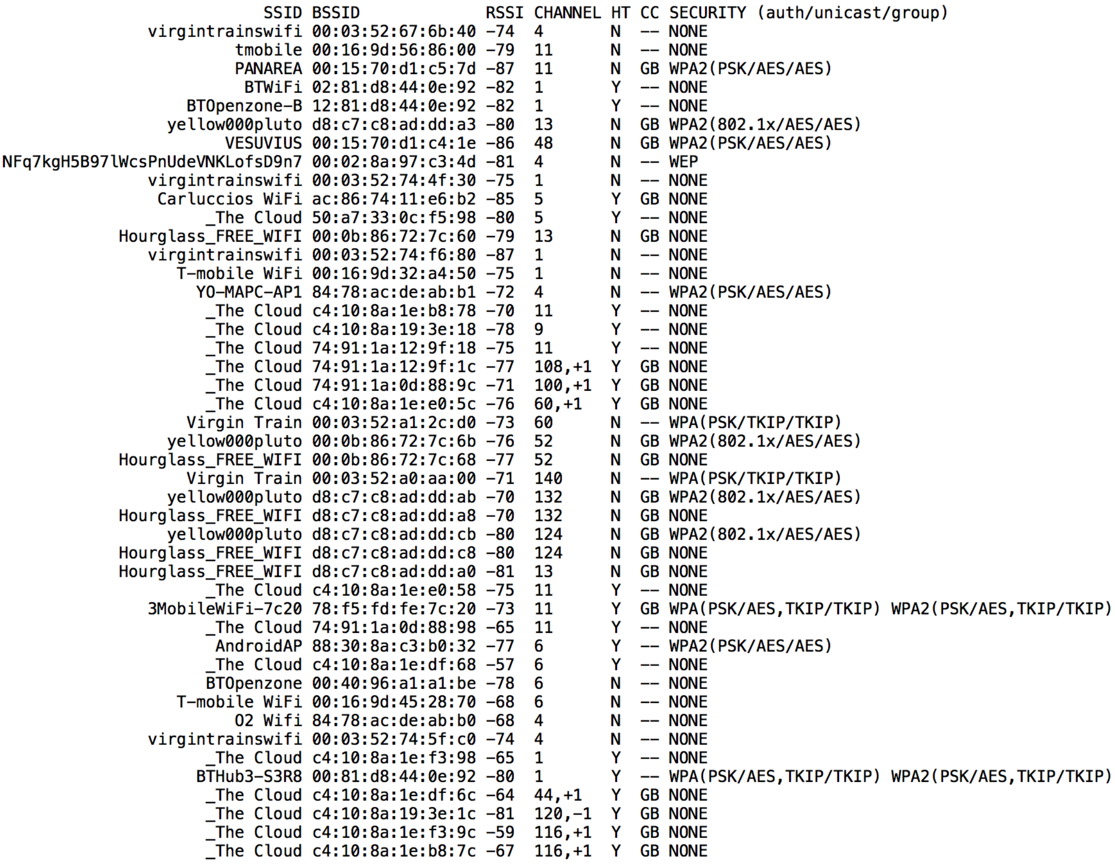
Number: 12
Name: Radio Map
Description: Selection of fingerprints to be matched with the query observation

Number: 13
Name: Best Matched Fingerprint
Description: Best matching fingerprint between query observations and the radio map fingerprints stored in the DB

Number: 14,15
Name: Estimated Location
Description: String representing the location of the best matched fingerprint.

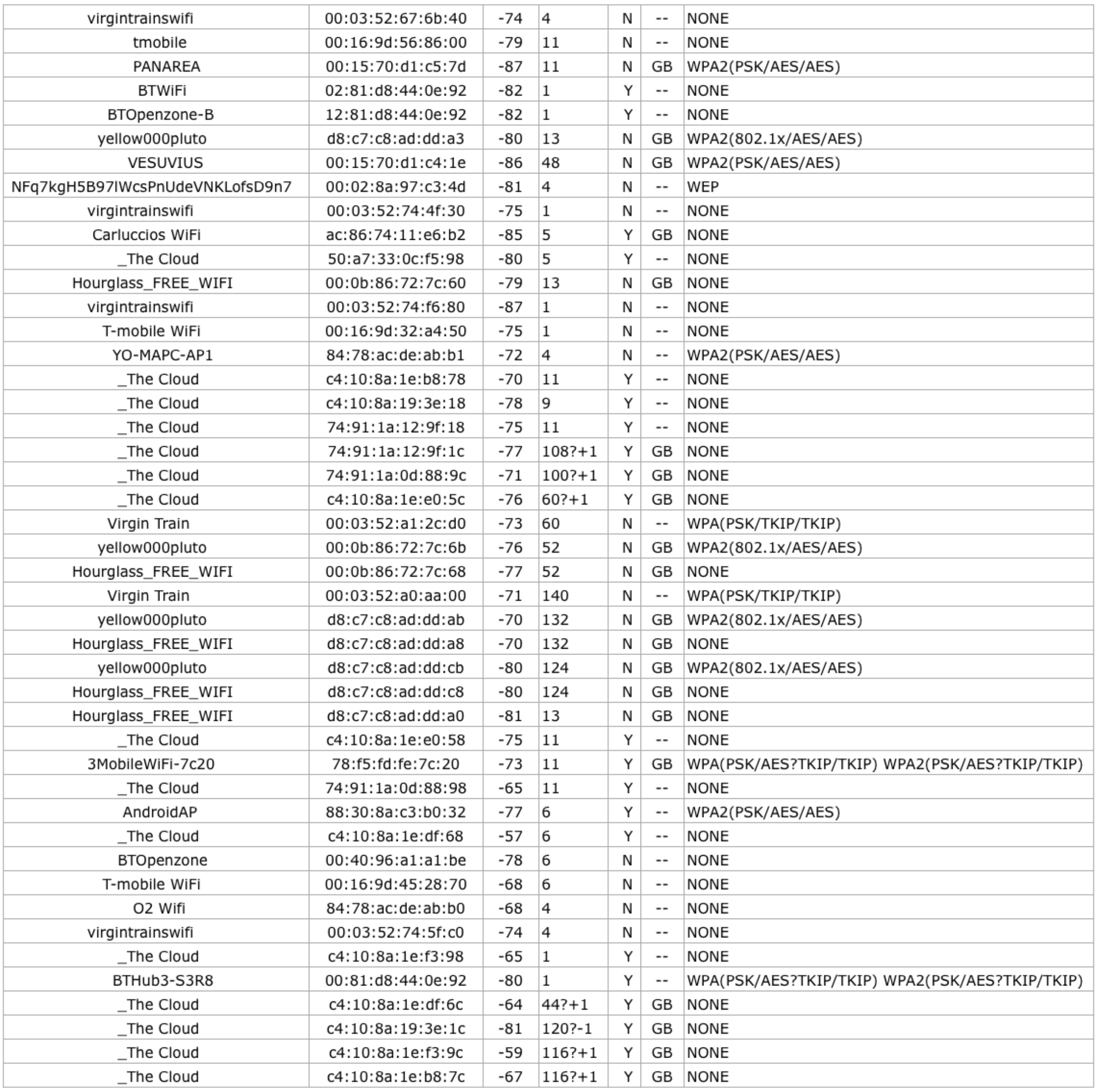
4.2.2 WAPs

The Wireless Access Points (WAPs), also called WiFi Access Points are a fundamental component on this architecture. The WAPs are electronic devices acting as transceivers. They form part of a pre-existing WiFi infrastructure whose function is primarily networking for communication.



```
                           SSID BSSID             RSSI CHANNEL HT CC SECURITY (auth/unicast/group)
               virgintrainswifi 00:03:52:67:6b:40 -74  4       N  -- NONE
                        tmobile 00:16:9d:56:86:00 -79  11      N  -- NONE
                        PANAREA 00:15:70:d1:c5:7d -87  11      N  GB WPA2(PSK/AES/AES)
                         BTWiFi 02:81:d8:44:0e:92 -82  1       Y  -- NONE
                   BTOpenzone-B 12:81:d8:44:0e:92 -82  1       Y  -- NONE
                 yellow000pluto d8:c7:c8:ad:dd:a3 -80  13      N  GB WPA2(802.1x/AES/AES)
                       VESUVIUS 00:15:70:d1:c4:1e -86  48      N  GB WPA2(PSK/AES/AES)
NFq7kgH5B97lWcsPnUdeVNKLofsD9n7 00:02:8a:97:c3:4d -81  4       N  -- WEP
               virgintrainswifi 00:03:52:74:4f:30 -75  1       N  -- NONE
                Carluccios WiFi ac:86:74:11:e6:b2 -85  5       Y  GB NONE
                     _The Cloud 50:a7:33:0c:f5:98 -80  5       Y  -- NONE
            Hourglass_FREE_WIFI 00:0b:86:72:7c:60 -79  13      N  GB NONE
               virgintrainswifi 00:03:52:74:f6:80 -87  1       N  -- NONE
                  T-mobile WiFi 00:16:9d:32:a4:50 -75  1       N  -- NONE
                    YO-MAPC-AP1 84:78:ac:de:ab:b1 -72  4       N  -- WPA2(PSK/AES/AES)
                     _The Cloud c4:10:8a:1e:b8:78 -70  11      Y  -- NONE
                     _The Cloud c4:10:8a:19:3e:18 -78  9       Y  -- NONE
                     _The Cloud 74:91:1a:12:9f:18 -75  11      Y  -- NONE
                     _The Cloud 74:91:1a:12:9f:1c -77  108,+1  Y  GB NONE
                     _The Cloud 74:91:1a:0d:88:9c -71  100,+1  Y  GB NONE
                     _The Cloud c4:10:8a:1e:e0:5c -76  60,+1   Y  GB NONE
                   Virgin Train 00:03:52:a1:2c:d0 -73  60      N  -- WPA(PSK/TKIP/TKIP)
                 yellow000pluto 00:0b:86:72:7c:6b -76  52      N  GB WPA2(802.1x/AES/AES)
            Hourglass_FREE_WIFI 00:0b:86:72:7c:68 -77  52      N  GB NONE
                   Virgin Train 00:03:52:a0:aa:00 -71  140     N  -- WPA(PSK/TKIP/TKIP)
                 yellow000pluto d8:c7:c8:ad:dd:ab -70  132     N  GB WPA2(802.1x/AES/AES)
            Hourglass_FREE_WIFI d8:c7:c8:ad:dd:a8 -70  132     N  GB NONE
                 yellow000pluto d8:c7:c8:ad:dd:cb -80  124     N  GB WPA2(802.1x/AES/AES)
            Hourglass_FREE_WIFI d8:c7:c8:ad:dd:c8 -80  124     N  GB NONE
            Hourglass_FREE_WIFI d8:c7:c8:ad:dd:a0 -81  13      N  GB NONE
                     _The Cloud c4:10:8a:1e:e0:58 -75  11      Y  -- NONE
               3MobileWiFi-7c20 78:f5:fd:fe:7c:20 -73  11      Y  GB WPA(PSK/AES,TKIP/TKIP) WPA2(PSK/AES,TKIP/TKIP)
                     _The Cloud 74:91:1a:0d:88:98 -65  11      Y  -- NONE
                      AndroidAP 88:30:8a:c3:b0:32 -77  6       Y  -- WPA2(PSK/AES/AES)
                     _The Cloud c4:10:8a:1e:df:68 -57  6       Y  -- NONE
                     BTOpenzone 00:40:96:a1:a1:be -78  6       N  -- NONE
                  T-mobile WiFi 00:16:9d:45:28:70 -68  6       N  -- NONE
                        O2 Wifi 84:78:ac:de:ab:b0 -68  4       N  -- NONE
               virgintrainswifi 00:03:52:74:5f:c0 -74  4       N  -- NONE
                     _The Cloud c4:10:8a:1e:f3:98 -65  1       Y  -- NONE
                    BTHub3-S3R8 00:81:d8:44:0e:92 -80  1       Y  -- WPA(PSK/AES,TKIP/TKIP) WPA2(PSK/AES,TKIP/TKIP)
                     _The Cloud c4:10:8a:1e:df:6c -64  44,+1   Y  GB NONE
                     _The Cloud c4:10:8a:19:3e:1c -81  120,-1  Y  GB NONE
                     _The Cloud c4:10:8a:1e:f3:9c -59  116,+1  Y  GB NONE
                     _The Cloud c4:10:8a:1e:b8:7c -67  116,+1  Y  GB NONE
```

Figure 4.3: Example of an ObservationTXT (file in .txt format). This example presents WiFi data recorded at Piccadilly train station.



virgintrainswifi	00:03:52:67:6b:40	-74	4	N	--	NONE
tmobile	00:16:9d:56:86:00	-79	11	N	--	NONE
PANAREA	00:15:70:d1:c5:7d	-87	11	N	GB	WPA2(PSK/AES/AES)
BTWiFi	02:81:d8:44:0e:92	-82	1	Y	--	NONE
BTOpenzone-B	12:81:d8:44:0e:92	-82	1	Y	--	NONE
yellow000pluto	d8:c7:c8:ad:dd:a3	-80	13	N	GB	WPA2(802.1x/AES/AES)
VESUVIUS	00:15:70:d1:c4:1e	-86	48	N	GB	WPA2(PSK/AES/AES)
NFq7kgH5B97lWcsPnUdeVNKLofsD9n7	00:02:8a:97:c3:4d	-81	4	N	--	WEP
virgintrainswifi	00:03:52:74:4f:30	-75	1	N	--	NONE
Carluccios WiFi	ac:86:74:11:e6:b2	-85	5	Y	GB	NONE
_The Cloud	50:a7:33:0c:f5:98	-80	5	Y	--	NONE
Hourglass_FREE_WIFI	00:0b:86:72:7c:60	-79	13	N	GB	NONE
virgintrainswifi	00:03:52:74:f6:80	-87	1	N	--	NONE
T-mobile WiFi	00:16:9d:32:a4:50	-75	1	N	--	NONE
YO-MAPC-AP1	84:78:ac:de:ab:b1	-72	4	N	--	WPA2(PSK/AES/AES)
_The Cloud	c4:10:8a:1e:b8:78	-70	11	Y	--	NONE
_The Cloud	c4:10:8a:19:3e:18	-78	9	Y	--	NONE
_The Cloud	74:91:1a:12:9f:18	-75	11	Y	--	NONE
_The Cloud	74:91:1a:12:9f:1c	-77	108?+1	Y	GB	NONE
_The Cloud	74:91:1a:0d:88:9c	-71	100?+1	Y	GB	NONE
_The Cloud	c4:10:8a:1e:e0:5c	-76	60?+1	Y	GB	NONE
Virgin Train	00:03:52:a1:2c:d0	-73	60	N	--	WPA(PSK/TKIP/TKIP)
yellow000pluto	00:0b:86:72:7c:6b	-76	52	N	GB	WPA2(802.1x/AES/AES)
Hourglass_FREE_WIFI	00:0b:86:72:7c:68	-77	52	N	GB	NONE
Virgin Train	00:03:52:a0:aa:00	-71	140	N	--	WPA(PSK/TKIP/TKIP)
yellow000pluto	d8:c7:c8:ad:dd:ab	-70	132	N	GB	WPA2(802.1x/AES/AES)
Hourglass_FREE_WIFI	d8:c7:c8:ad:dd:a8	-70	132	N	GB	NONE
yellow000pluto	d8:c7:c8:ad:dd:cb	-80	124	N	GB	WPA2(802.1x/AES/AES)
Hourglass_FREE_WIFI	d8:c7:c8:ad:dd:c8	-80	124	N	GB	NONE
Hourglass_FREE_WIFI	d8:c7:c8:ad:dd:a0	-81	13	N	GB	NONE
_The Cloud	c4:10:8a:1e:e0:58	-75	11	Y	--	NONE
3MobileWiFi-7c20	78:f5:fd:fe:7c:20	-73	11	Y	GB	WPA(PSK/AES?TKIP/TKIP) WPA2(PSK/AES?TKIP/TKIP)
_The Cloud	74:91:1a:0d:88:98	-65	11	Y	--	NONE
AndroidAP	88:30:8a:c3:b0:32	-77	6	Y	--	WPA2(PSK/AES/AES)
_The Cloud	c4:10:8a:1e:df:68	-57	6	Y	--	NONE
BTOpenzone	00:40:96:a1:a1:be	-78	6	N	--	NONE
T-mobile WiFi	00:16:9d:45:28:70	-68	6	N	--	NONE
O2 Wifi	84:78:ac:de:ab:b0	-68	4	N	--	NONE
virgintrainswifi	00:03:52:74:5f:c0	-74	4	N	--	NONE
_The Cloud	c4:10:8a:1e:f3:98	-65	1	Y	--	NONE
BTHub3-S3R8	00:81:d8:44:0e:92	-80	1	Y	--	WPA(PSK/AES?TKIP/TKIP) WPA2(PSK/AES?TKIP/TKIP)
_The Cloud	c4:10:8a:1e:df:6c	-64	44?+1	Y	GB	NONE
_The Cloud	c4:10:8a:19:3e:1c	-81	120?-1	Y	GB	NONE
_The Cloud	c4:10:8a:1e:f3:9c	-59	116?+1	Y	GB	NONE
_The Cloud	c4:10:8a:1e:b8:7c	-67	116?+1	Y	GB	NONE

Figure 4.4: Example of a ObservationCSV

The ObservationCSV (file in CSV format) is a collection of records (lines) formed by fields (columns). This particular example is the result after processing the ObservationTXT shown in figure 4.3, after processing, the fields are separated with commas and column titles are removed.

WAPs transmit radio frequency (RF) signals, where the information on the signal transmitted is structured based on a protocol. It can be received and recorded by any WiFi enabled device sharing the medium (air) and the protocol standard. Hence, any client device on the WAP coverage area can receive the signals.

The WAPs broadcast a specific data frame that contains the SSID, BSSI, RSSI, channel, HT, CC and security. This data and the strength (dB) at which the signal was received can be recorded by the client device.

Existing WAPs were used on the implementation stages and no WAPs were installed for this research. All the information here presented is based on "real-life" WAPs whose fundamental propose is communication.

4.2.3 Client

A client is a Wi-Fi enabled device, any electronic device with a built in 802.11 network interface. Example of diverse clients are: mobile phone/smart phone, tablet, laptop computer. The functions of the client device is to interact with the networks and to gather information about the WAPs in the vicinity. It is used on both stages of the system; used by experts for wardriving, or by users for requesting position.

The device employed as client device was a MacBook laptop with processor 2.4 GHz Intel Core 2 Duo, 8 GB in memory. The laptop has a built-in wireless network adapter with the following details:
Card Type: AirPort Extreme (0x14E4, 0x8D)
MAC Address: 00:23:12:53:ca:3a
Country Code: GB
Supported PHY Modes: 802.11 a/b/g/n
Supported Channels: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 36, 40, 44, 48, 52, 56, 60, 64, 100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140.

A script was implemented, using the software AirPort Utility version 6.2 (620.33) for scanning the 802.11 networks.

Devices having a wireless network adapter with distinct characteristics are expected to provide different measurements. However, if measurements are consistent using a single client device the results of the fingerprinting algorithms presented in this thesis are expected to be equivalent. Nevertheless, for measurements using a variety of devices (having distinct wireless network adapters), a calibration mechanism must be implemented for managing the potential offset in the data. The implementation of this calibration mechanism is out of the scope of this research.

4.2.4 Data Acquisition Subsystem

Data Acquisition Subsystem is the subsystem used to scan an area for available networks and recording the data. Table 4.2.4 provides the Data Acquisition input, functions and output for the subsystem.

Input	Request Response Data Frames
Functions	Accessing client's hardware Performing a wireless broadcast scan Recording Probe Response frames from near APs
Output	Observations TXT Folder

Table 4.2: Data Acquisition Subsystem

4.2.5 Client Management Subsystem

Input	Observations TXT Folder Symbolic Location (for tagging survey-observations) Estimated Location (result from positioning)
Functions	Interfacing between client systems Routing data and handling communication between client and server
Output	tagged Observations TXT Folder Estimated Location (positioning result)

Table 4.3: Client Management Subsystem

4.2.6 User Interface Subsystem

4.2.7 Server

The server is a dedicated computer running the fingerprinting and positioning related algorithms. It reduces the computational analysis on the client. Functions carried out on the server include: receiving data, processing observations, generating and updating fingerprints, interacting with the database and computing position. The server is a MacBook laptop with processor 2.4 GHz Intel Core 2 Duo, 8 GB in memory, running the following software: MATLAB ®, and Sequel Pro 1.0.2 .

Input	Estimated Location
Functions	Start/Stop survey Off-line observations capture (survey observations) On-line observations capture (query observations) Display Location (Positioning Results)
Output	Symbolic Location (string for tagging off-line observations)

Table 4.4: User Interface Subsystem

4.2.8 Server Management Subsystem

Input	tagged Observations TXT Folder Best Matched Fingerprint
Functions	Communication with client and routing within the server.
Output	tagged Observations TXT Folder Estimated Location

Table 4.5: Server Management Subsystem

4.2.9 Data Mangement Subsystem

The Data Management Subsystem processes the WiFi data by extracting the significant information. It generates the observations as a data element formatted for insertion on the database or for direct processing for other subsystems such fingerprint and positioning.

Input	query-Observation CSV Folder
Functions	Importing Observations Processing observations for appropriate format Routing observations to Fingerprinting or Positioning subsystem accordingly
Output	survey Observations CSV Folder query-Observation CSV Folder

Table 4.6: Data Mangement Subsystem

4.2.10 Fingerprinting Subsystem

A fingerprint is a unique identification of a physical location. It is formed by the relationship between a signature and its symbolic location. A large set of fingerprints is stored on the database. The fingerprint generation is an off-line process consisting on the analysis the survey-observations per location. The fingerprinting process is covered in depth in chapter 6.

Input	Survey-ObservationsCSV Folder Fingerprint
Functions	Fingerprint Generation Fingerprint Update
Output	New/Updated Fingerprint

Table 4.7: Fingerprinting Subsystem

4.2.11 Positioning Subsystem

Query-Observations are the input for Positioning. These observations do not include a location. The set of WAPs from a Query-Observation are compared against the fingerprints in order to identify the best matched fingerprint and link it with the estimated location for the query-observation. See chapter 6 for details on the positioning algorithms.

Input	Query-ObservationCSV Folder Radio Map
Functions	Running positioning algorithms
Output	Best Matched Fingerprint

Table 4.8: Positioning Subsystem

4.2.12 Storage Mangement Subsystem

The storage management subsystem's function is to act as a link between the database and the other subsystems on the server.

Input	SQL statements New/Updated Fingerprint DB results (Data rows, tables, and messages from the database)
Functions	Create SQL statements for interacting with the database such as: DB connection DB population (e.g. Insert observations, insert fingerprints, update fingerprints) DB retrievals (e.g. fetch observations, fetch fingerprints)
Output	SQL statements Fingerprint Radio Map

Table 4.9: Storage Mangement Subsystem

4.2.13 Database

The Database stores the system's data in a structured way, which allows the server to insert, store, and retrieve the data when required.

The database was carefully designed for research purposes. It provides the required structure for storing observations, fingerprints (from five different fingerprint definitions), and the logs of fingerprint's changes/updates (per fingerprint definition). Storage of observations is required for historical records and for a data analysis which is presented in chapter 5. Fingerprints from several definitions and their logs are kept on the database for analysis, comparison, and generation of results presented in chapters 6 to 8.

For clarity, the database's tables are presented in figures 4.5, 4.6 and 4.7 separated by functionality. The Enhanced Entity-Relationship Model of the Database in its entirety is shown in Appendix A.3.

Figure 4.5 includes tables used for storing the observations. These tables can stand alone as a separate database. The Entity-Relation Diagram and the metadata for database generation is publicly available (see `http://dx.doi.org/10.5281/zenodo.13793`). Also a database containing the observations collected during this research is publicly available (see `http://dx.doi.org/10.5281/zenodo.12913`).

Figure 4.6 presents the tables used for storing fingerprints. Although these tables are presented separately, they are not stand alone, they are directly related with tables presented in figure 4.5.

Figure 4.7 presents the tables used for storing the log of fingerprint's changes. These tables are not stand alone, they are related to tables presented in 4.6 and 4.6.

The database resides on the server, but it can be placed on a remote database dedicated server. It was implemented using the open-source relational database management system MySQL 5.5.28.

4.2.14 Database Subsystem

Input	SQL statements
Functions	Storage of observation and fingerprints/radio-map
Output	DB results

Table 4.10: Database Subsystem

4.3 Chapter Summary

This Chapter presented the system architecture of a Fingerprint-based WiFi Positioning System. The components, functions and relation were explained. The following chapter presents a detailed analysis of the data set collected using the components of this architecture.

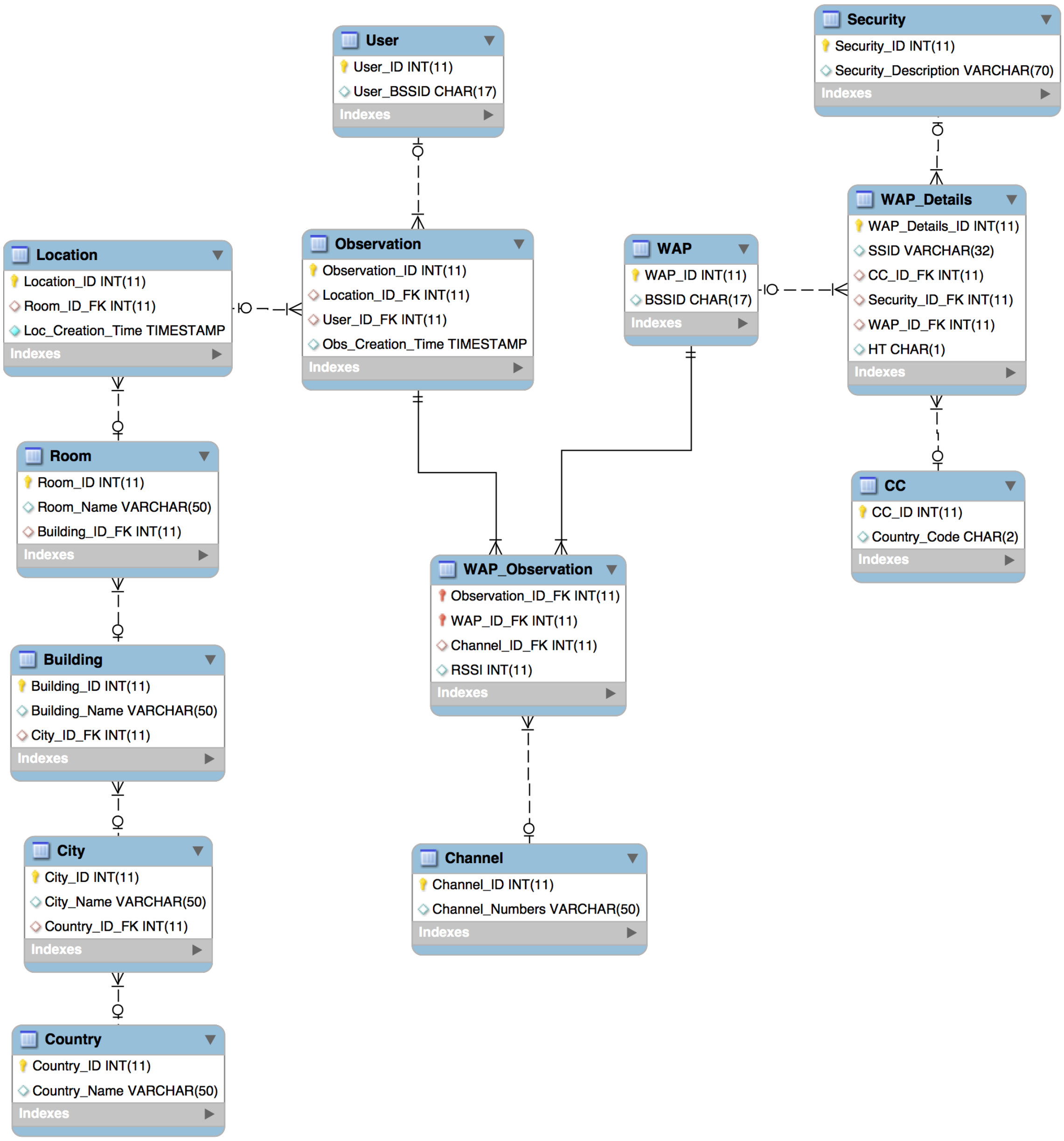


Figure 4.5: Database Entity-Relationship Diagram. The tables presented in this Figure are the ones used for the storage of the survey-observations and query-observations. Also the tables that have a direct relation are included (e.g. Location, Room, Building, City, etc.) Every observation remainds in the database for purposes of data analysis. In a final implementation, the storage of all the observations would not be requiered.

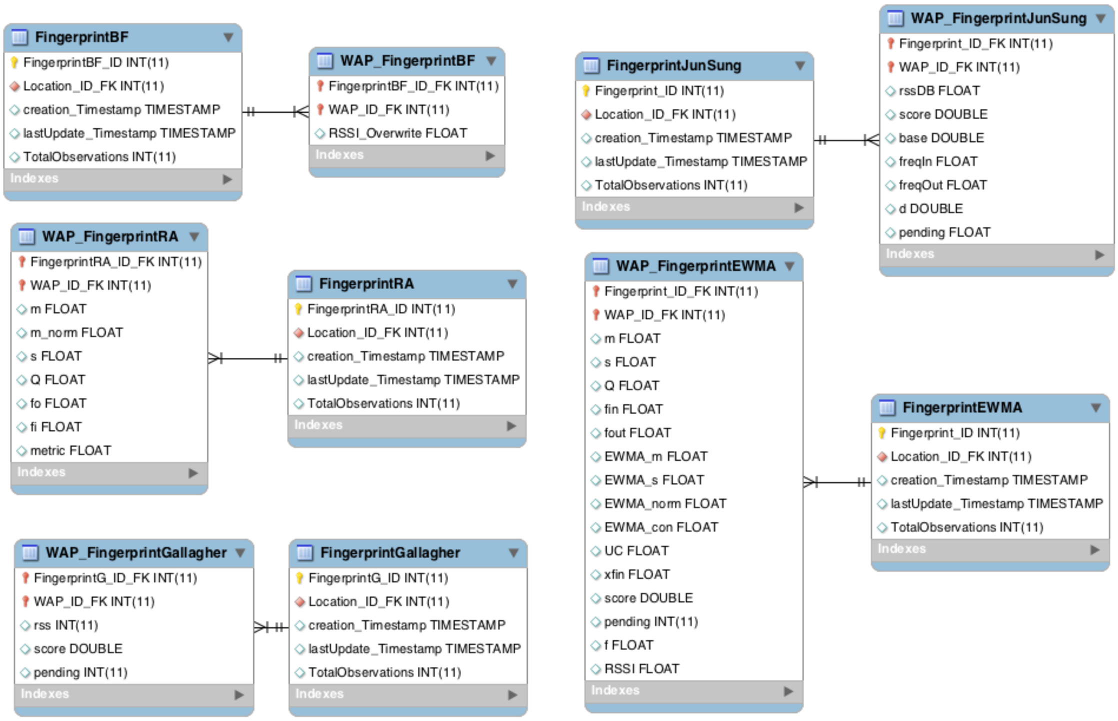


Figure 4.6: Database Entity-Relationship Diagram. The tables in this Figure are for the storage of the fingerprints generated with the five fingerprint definitions

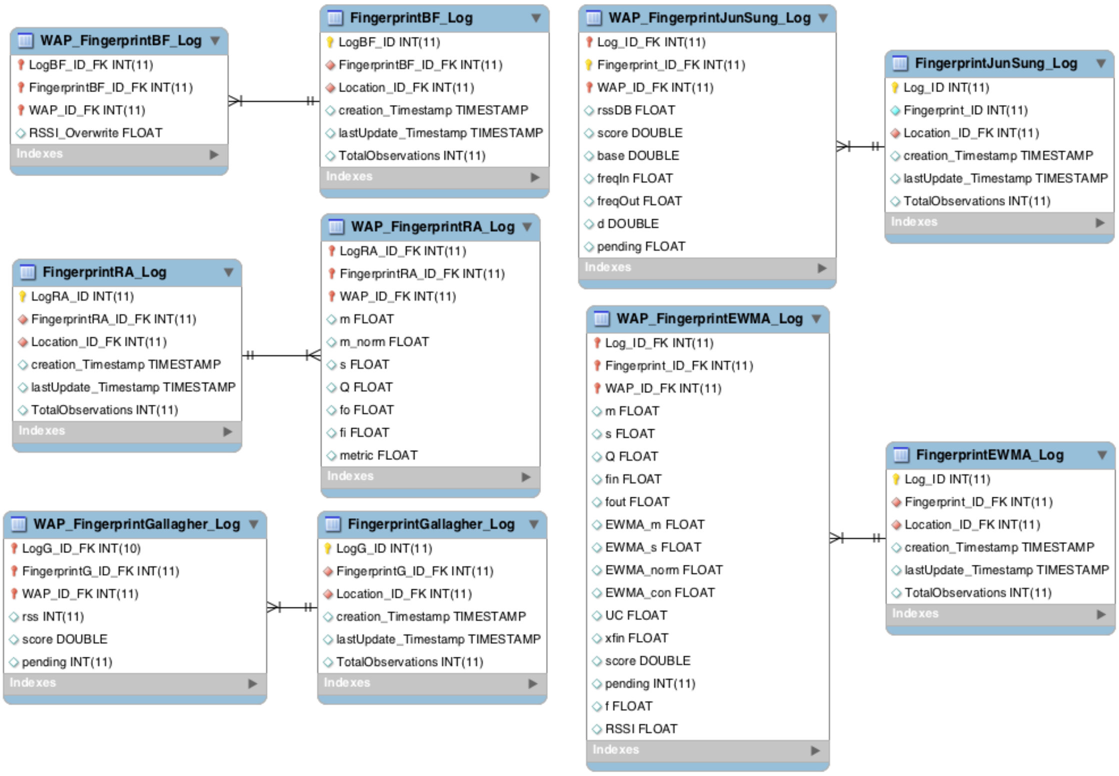


Figure 4.7: Database Entity-Relationship Diagram. The tables in this Figure are for the systematic recording of the fingerprints for the five fingerprint definitions implemented

Chapter 5

Data Analysis

This chapter presents an analysis of data from extensive temporal and spatial WiFi scanning that was designed specifically for this project and was undertaken for a period of over one year. The aim of this chapter is to present the data sets and provide a better understanding of the statistics and discuss certain features within them. Three different localities are the subject for this chapter. These localities represent three very different scenarios, each which can be used to understand the developments of a WiFi fingerprint system and assist in addressing the project's hypothesis. The following sections cover a description of the locations selected and the reason why they were chosen. Also descriptive statistics are presented for the top 10% most seen WAPs for each location and related graphs synthesising the data are shown.

5.1 Data Set

There are many WiFi data sets available for fingerprint analysis, with many being synthetic or set-up under controlled circumstances and often for a specific demonstration or paper. It was considered that a large temporally changing real environment dataset was required for this project. This data set would be useful for both answering the hypothesis questions raised as well as being a useful future resource for further studies and comparisons.

A large data set has been collected during a period of over one year (see chapter 3 for details on data collection). In brief this data set includes: 1,224 WAPs with unique MAC addresses, over 125,300 observations taken in 48 distinct locations(see Appendix A.1 for a display of all locations surveyed).

Within the entire data set, three locations were selected for extensive analysis, due to their differing characteristics, the selected locations are:

- Manchester Piccadilly Railway Station, Manchester, UK
- RCS offices within Devonshire House at The University of Manchester, Manchester, UK
- Private residence, Derby, UK.

The measurements gathered and statistics extracted from these three locations are presented and discussed in the subsequent sections.

5.1.1 Manchester Piccadilly Railway Station

Manchester Piccadilly railway station is the busiest station in Manchester. It comprises 14 rail platforms and has an estimated usage of 18.584 million rail passengers annually (information based on tickets sale in year 2011/12 [oRR]). The main station building was the subject of analysis, which is a large building formed by a lower and upper concourse areas with shops, food and drink outlets, ticket office and seats for people to rest.

Figure 5.1 is a modification of the Manchester Piccadilly Railway Station guide map from the Network Rail [web].

The data collection point was adjacent to the glass wall partitioning the main station concourse from the platforms at platform 6. Figure 5.2 presents a panoramic view of the Manchester Piccadilly rail station, which was taken from the same location that the WiFi data was collected.

In comparison with all the surveyed locations, Manchester Piccadilly Railway Station, presented the largest number of WAPs, a total of 555 distinct WAPs. Out of these only 20 WAPs were seen across 90% + of the observations. Describing the tail of the observations, 491 WAPs were seen for under 10% of the observations. Table 5.1 lists those high frequency WAPs that were detected within 90% to 100% of the observations. In the table the WAPs are ordered by the number of times each WAP was seen, called frequency of occurrence (f). As observed in the table, the WAP #1 has an SSID “Virgin Train” and was seen a total of 3,727 out of 3,728 observations, this is 99.97% of the total observations recorded in the railway station. On the other hand WAP #555 with SSID “virgintrainswifi” was seen in 1 out of 3,728 observations, which represents circa 0.03% of the total measurements.

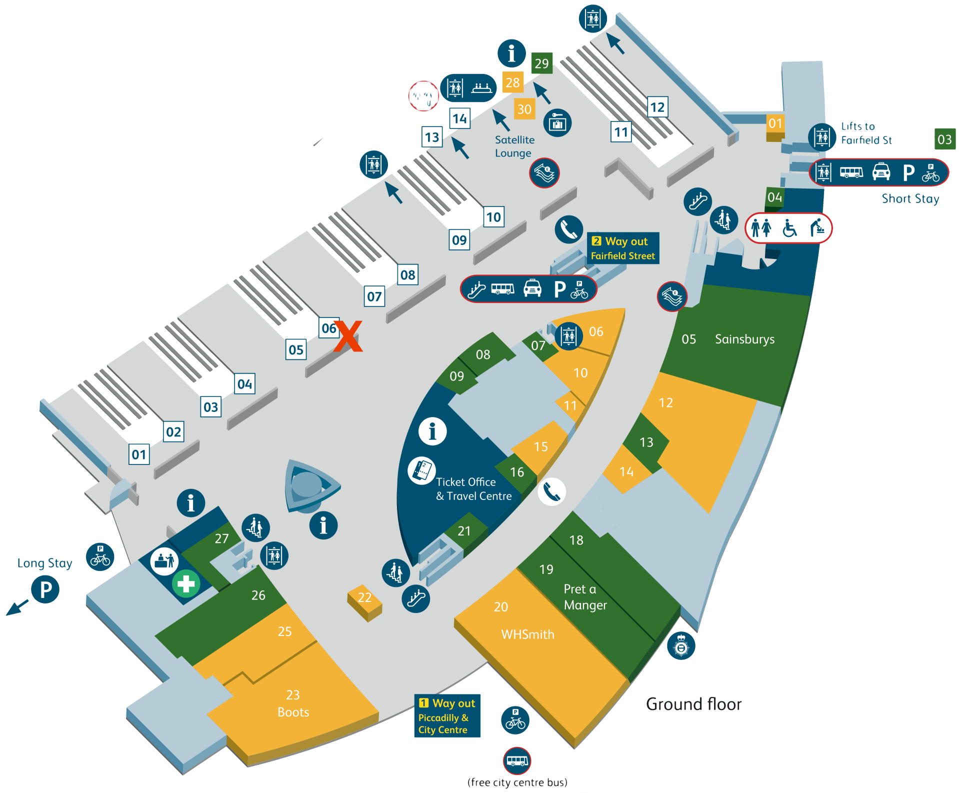


Figure 5.1: Data collection point at the Manchester Piccadilly Railway Station. This figure is a modification of a guide map taken from the Network Rail website

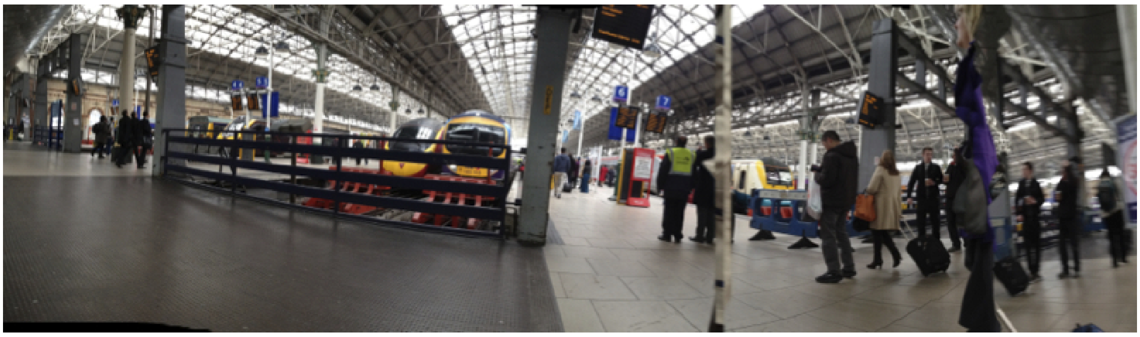


Figure 5.2: Manchester Piccadilly Railway Station, Manchester, UK

Table 5.1: WAPs observed at Manchester Piccadilly Rail Station.

#	WAP ID	BSSID	SSID	RSSI Mean (*dB*)	RSSI Std Dev.	Freq. of Occurrence	Relative Freq. (%)
1	137	00:03:52:a1:2c:d0	Virgin Train	-71.63	1.80	3727	100.00
2	127	00:03:52:a0:aa:00	Virgin Train	-70.34	1.91	3727	100.00
3	150	00:16:9d:45:28:70	T-mobile WiFi	-70.46	3.60	3725	99.95
4	161	00:16:9d:32:a4:50	T-mobile WiFi	-74.53	3.45	3710	99.54
5	194	c4:10:8a:1e:df:6c	The Cloud	-62.28	3.25	3672	98.52
6	199	c4:10:8a:1e:f3:9c	The Cloud	-63.67	2.85	3672	98.52
7	197	74:91:1a:0d:88:9c	The Cloud	-72.95	2.95	3672	98.52
8	195	74:91:1a:12:9f:1c	The Cloud	-77.52	3.15	3672	98.52
9	193	c4:10:8a:1e:e0:5c	The Cloud	-73.12	3.49	3672	98.52
10	146	00:16:a6:10:51:0a	BlisMobie Media	-54.24	3.24	3617	97.05
11	198	c4:10:8a:1e:b8:7c	The Cloud	-69.52	2.94	3591	96.35
12	192	c4:10:8a:1e:e0:58	The Cloud	-73.01	3.34	3541	95.01
13	246	d8:c7:c8:ad:dd:a8	Hourglass FREE WIFI	-70.01	2.31	3510	94.18
14	215	74:91:1a:0d:88:98	The Cloud	-70.43	3.95	3507	94.10
15	251	d8:c7:c8:ad:dd:a0	Hourglass FREE WIFI	-74.55	5.77	3490	93.64

Continued on next page

Table 5.1 – continued from previous page

#	WAP ID	BSSID	SSID	RSSI Mean (*dB*)	RSSI Std Dev.	Freq. of Occurrence	Relative Freq. (%)
16	444	d8:c7:c8:ad:dd:ab	yellow000pluto	-70.09	2.26	3438	92.25
17	442	00:0b:86:72:7c:68	Hourglass FREE WIFI	-76.77	1.91	3420	91.76
18	441	00:0b:86:72:7c:6b	yellow000pluto	-76.94	1.91	3420	91.76
19	163	00:16:9d:56:86:00	tmobile	-80.30	3.12	3380	90.69
20	245	d8:c7:c8:ad:dd:c8	Hourglass FREE WIFI	-79.68	2.04	3366	90.31

In analysis we can classify the WAPs from those detected into the following two types:

- Static WAPs. The static WAPs are the ones held approximately stationary, such as installed on the ceiling or walls. These WAPs are measured into the client receiving devices as a relatively stable signal strength (RSSI). A static WAP can have a low and fluctuating RSSI or can be a recently installed, or semi-temporary, appearing only in few recent or as a cluster of observations.

- Mobile WAP. Mobile WAP can have variable physical locations (e.g. mobile phones, tablets, portable computers and WAPs installed on board trains). The RSSI from mobile WAP almost always appears unsteady to the client device. They can be consistent in strength, for example train hosted WAPs will appear at very specific locations in the railway station and at specific times/dates.

Table 5.2 presents the descriptive statistics for the WAPs detected at Piccadilly Railway Station. Including the following information per WAP: ranking, WAP ID, mean, mode, median, minimum and maximum values of RSSI.

Table 5.2: Descriptive Statistics, Piccadilly Railway Station, Manchester, UK.

#	WAP ID	mean	mode	median	min	max
1	137	-71.63	-71	-72	-77	-66
2	127	-70.35	-71	-70	-77	-64
3	150	-70.46	-71	-70	-90	-42
4	161	-74.53	-74	-74	-90	-57
5	194	-62.28	-62	-62	-73	-52
6	199	-63.67	-63	-63	-72	-55
7	197	-72.95	-71	-73	-82	-60
8	195	-77.52	-76	-77	-87	-69
9	193	-73.12	-74	-73	-84	-62
10	146	-54.24	-55	-54	-87	-1
11	198	-69.52	-70	-70	-81	-59
12	192	-73.01	-72	-73	-89	-61

Continued on next page

Table 5.2 – continued from previous page

#	WAP ID	mean	mode	median	min	max
13	246	-70.01	-70	-70	-79	-62
14	215	-70.43	-69	-70	-90	-55
15	251	-74.55	-77	-75	-88	-55
16	444	-70.09	-70	-70	-78	-62
17	442	-76.77	-77	-77	-83	-70
18	441	-76.94	-77	-77	-82	-69
19	163	-80.3	-81	-81	-88	-63
20	245	-79.68	-80	-80	-86	-71

Figure 5.3 is a plot including all the WAPs detected at Piccadilly railway station sorted by frequency of occurrence. The long tail in the figure can be attributed to the WAPs with a low frequency of occurrence, such as mobile WAP.

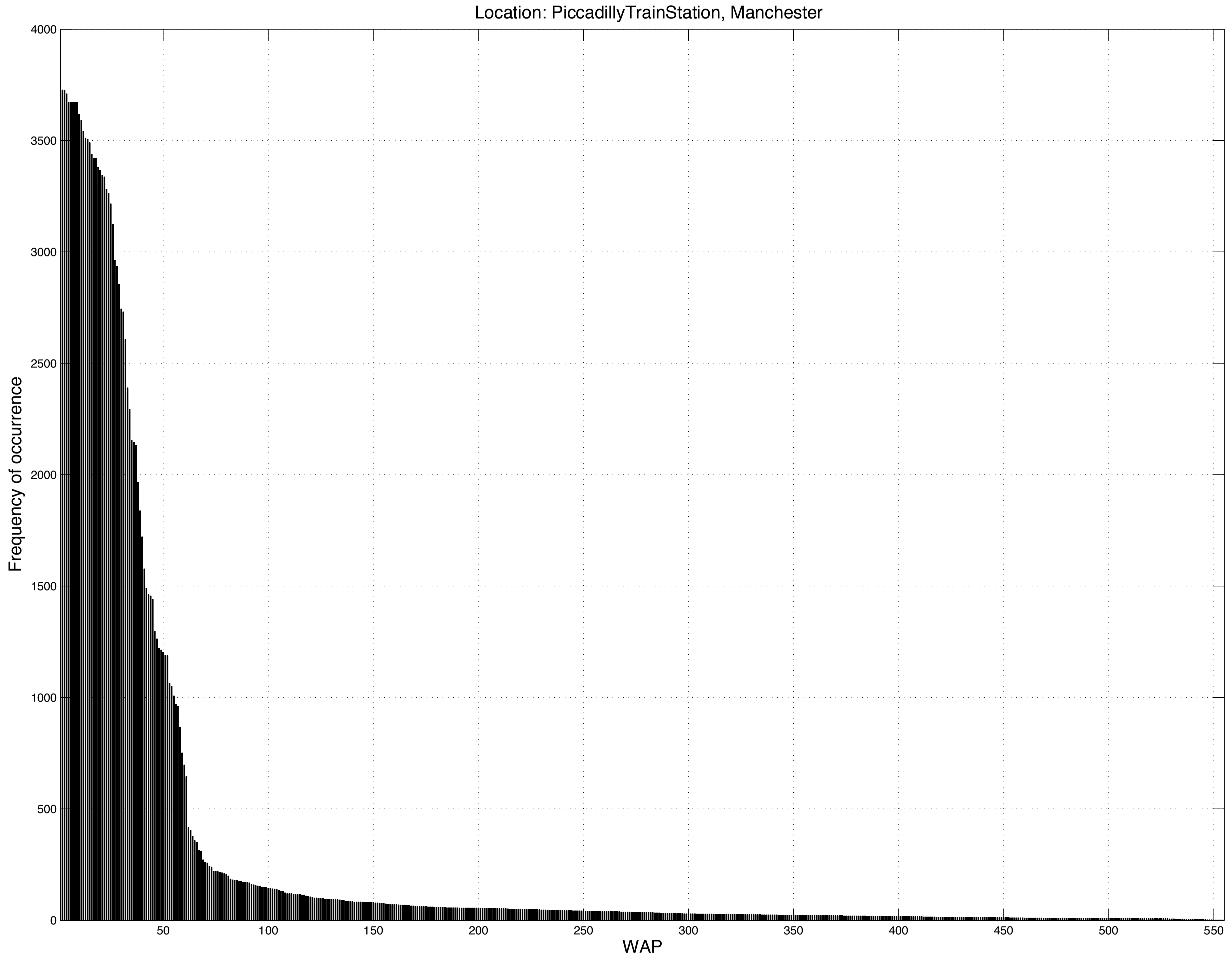


Figure 5.3: Frequency of occurrence for the WAPs detected at Piccadilly Railway Station. The *x* axis are the WAPs ordered from higher to lower frequency of occurrence.

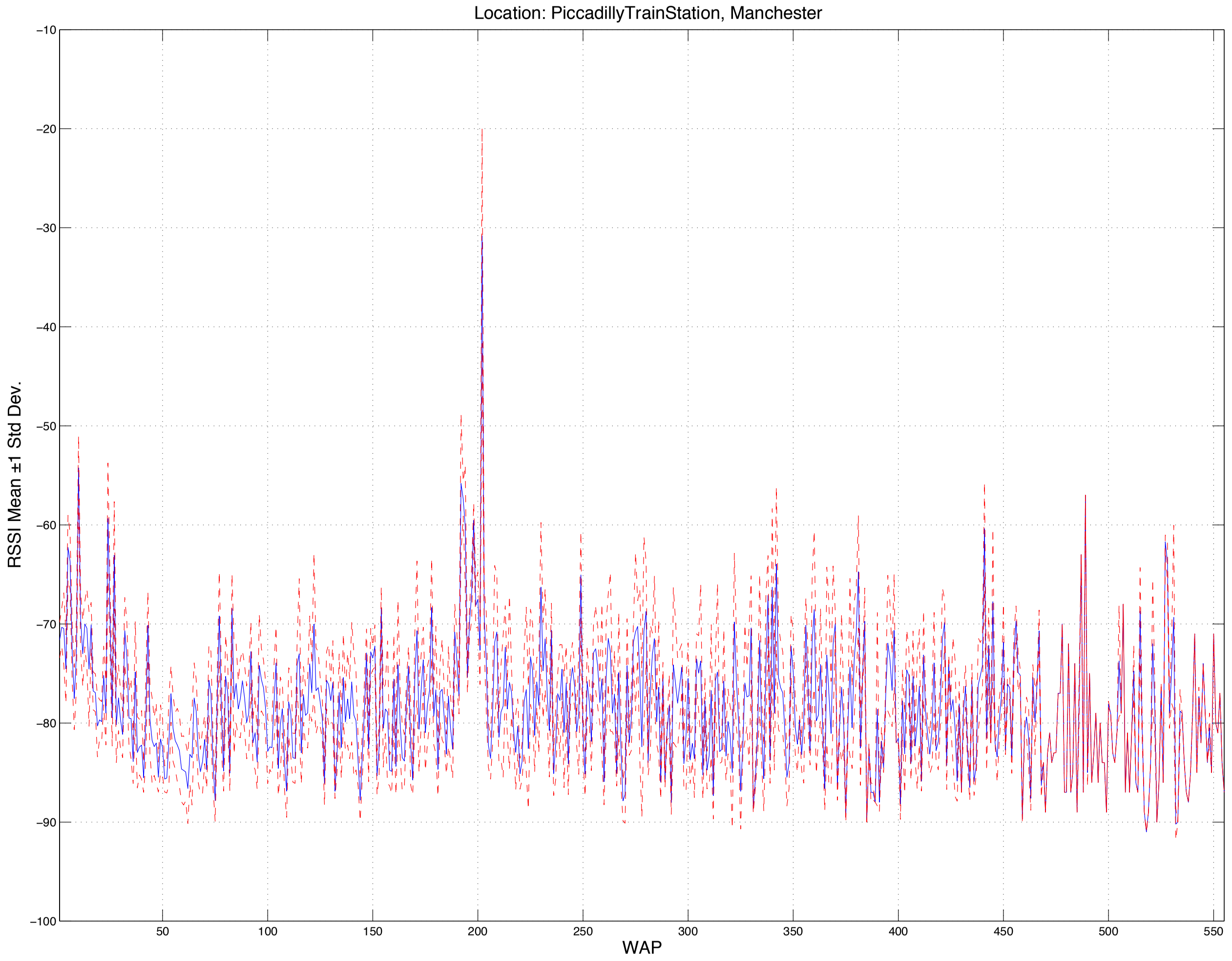


Figure 5.4: RSSI mean and ±1 standard deviation for WAPs detected at Manchester Piccadilly Railway Station. The x axis are the WAPs ordered from higher to lower frequency of occurrence.

5.1.2 Research Computing Services

The second location to be analysed is a university building. Located within Devonshire House, the RCS is part of The University of Manchester in Manchester, UK. RCS is a restricted access premises comprising offices mainly used by university staff.

Two sets of data were collected within RCS:

- Snapshot data collection. A "snapshot" data collection refers to a short-term collection of Wi-Fi data.
- Long-term data collection. The long-term data collection refers to a longstanding and repeated data collection at a particular office within RCS department.

Snapshot Data Collection

A "snapshot" data collection was carried in the majority of offices at RCS. The so called snapshot collection is the Wi-Fi scanning and recording of few observations (around 4 observations) per room/area. Two snapshot data collection sessions were executed in most of the RCS offices in two different dates; the first was at the beginning of data gathering period (October 2012) and the second one was done towards the end of data gathering period (August 2013). Figure 5.5 presents the blueprint of the floor where RCS is situated, highlighted in red are those rooms where snapshot data collection was carried out.

The objectives of collecting these snapshots are twofold; first to build up a WiFi radio-map (used for positioning) covering most of RCS offices, and second to analyse WiFi scenery changes occurring within a lapse of several months.

Long-term Data Collection

The long-term data collection consists of weekly data measurements. The measurements were carried out on a period covering over one year. The location subject to long-term data collection was an office (1.035) within the RCS department. Office 1.035 is allocated for the author's exclusive use. Therefore, it provides a controlled environment for repeatable data collection. Figure 5.6 shows the blueprint of the RCS department and office 1.035 is highlighted with a red mark.

Observations in office 1.035 were performed as similar to each other as possible. This was done by placing the receiving client device (MacBook laptop) on a static position on every data collection session.

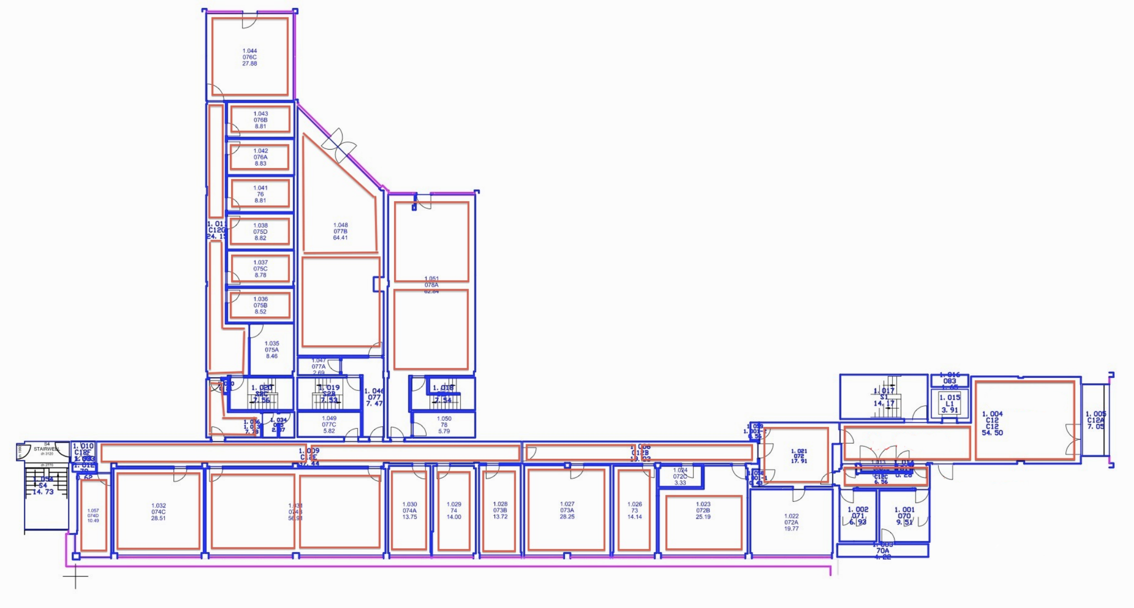


Figure 5.5: RCS department Blueprint, highlighted in red are those rooms where WiFi snapshot data was collected.

Figure 5.7 is a panorama image taken inside of office 1.035.

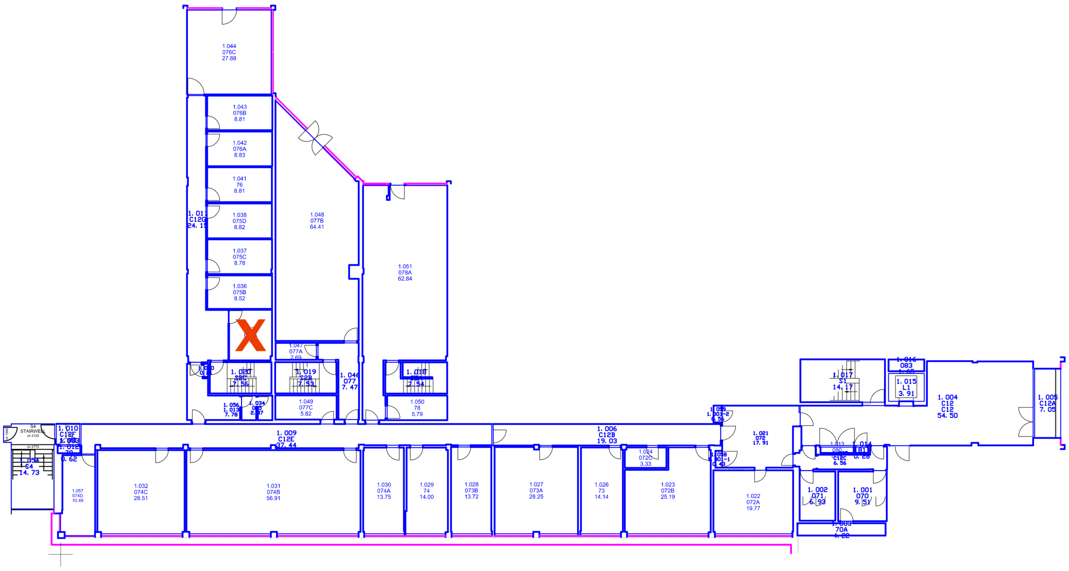


Figure 5.6: Office 1.035 at RCS, The University of Manchester, Manchester, UK

Analysis of data collected at office 1.035 revealed around 46,161 observations accommodating a total of 141 WAPs with distinct MAC addresses. Out of these 141

Figure 5.7: Panorama view from within Office 1.035 at RCS, The University of Manchester, Manchester, UK

WAPs, 13 were seen in 90%+ observations, and an average of 32 WAPs were detected per observation.

Table 5.3 presents details for the top 10% WAPs collected in office 1.035, which are ranked by frequency of occurrence. The columns on the table are: ranking number, WAP identifier, BSSID (MAC address), SSID (name of the network), RSSI average (μ), RSSI standard deviation (σ), frequency of occurrence (f) and percentage of relative frequency. The WAP that was seen in most of the observations is ranked #1, with SSID "UoM_WIFI" and BSSID "d8:c7:c8:ad:cd:b1", it was detected in 99.97% of the measurements. The WAP with lowest frequency of occurrence is ranked in position #141 with a SSID "Connectify-Jen" and BSSID "22:df:9a:38:b1:38", was detected in one single observation.

Table 5.3: WAPs ranked by relative frequency, detected at office 1.035, RCS department, The University of Manchester, Manchester, UK.

#	WAP ID	BSSID	SSID	RSSI Mean (*dB*)	RSSI Std Dev.	Freq. of Occurrence	Relative Freq. (%)
1	63	d8:c7:c8:ad:cd:b1	UoM WIFI	-62.51	5.30	46145	99.97
2	33	d8:c7:c8:ad:cc:99	UoM WIFI	-75.55	3.76	46036	99.73
3	42	d8:c7:c8:ad:cc:98	eduroam	-75.51	3.81	46032	99.72
4	49	d8:c7:c8:ad:cd:b8	eduroam	-63.57	4.84	46011	99.68
5	52	d8:c7:c8:ad:cd:b9	UoM WIFI	-63.62	4.86	46011	99.68
6	68	d8:c7:c8:ad:cc:d9	UoM WIFI	-70.69	5.22	45959	99.56
7	67	d8:c7:c8:ad:cc:d8	eduroam	-70.61	5.25	45936	99.51
8	83	d8:c7:c8:ad:cc:d1	UoM WIFI	-71.93	3.65	45901	99.44
9	65	d8:c7:c8:ad:c7:e1	UoM WIFI	-69.56	3.26	45872	99.37
10	55	d8:c7:c8:ad:c7:e9	UoM WIFI	-73.61	3.77	45837	99.30
11	54	d8:c7:c8:ad:c7:e8	eduroam	-73.52	3.75	45781	99.18
12	51	d8:c7:c8:ad:cc:91	UoM WIFI	-75.75	3.23	44280	95.93

Table 5.4 presents the descriptive statistics for the top 10% WAPs detected in office 1.035. The table presents the mean, mode, median, maximum value and minimum value of the RSSI measurements. WAPs are sorted by frequency of occurrence.

Table 5.4: Descriptive Statistics, RCS, Manchester, UK.

#	WAP ID	mean	mode	median	min	max
1	63	-62.51	-62	-62	-88	-49
2	33	-75.55	-75	-75	-89	-66
3	42	-75.51	-75	-75	-89	-65
4	49	-63.57	-61	-63	-89	-53
5	52	-63.62	-61	-63	-89	-53
6	68	-70.69	-68	-70	-88	-57
7	67	-70.61	-71	-70	-87	-58
8	83	-71.93	-72	-72	-88	-58
9	65	-69.56	-69	-69	-88	-55
10	55	-73.61	-72	-73	-89	-64
11	54	-73.52	-72	-73	-89	-64
12	51	-75.75	-77	-76	-88	-63

Figure 5.8 is an histogram of all the WAPs detected at office 1.035 sorted by frequency of occurrence. The number of WAPs on the tail of the figure can be attributed to the WAPs with a low frequency of occurrence, such as mobile WAPs.

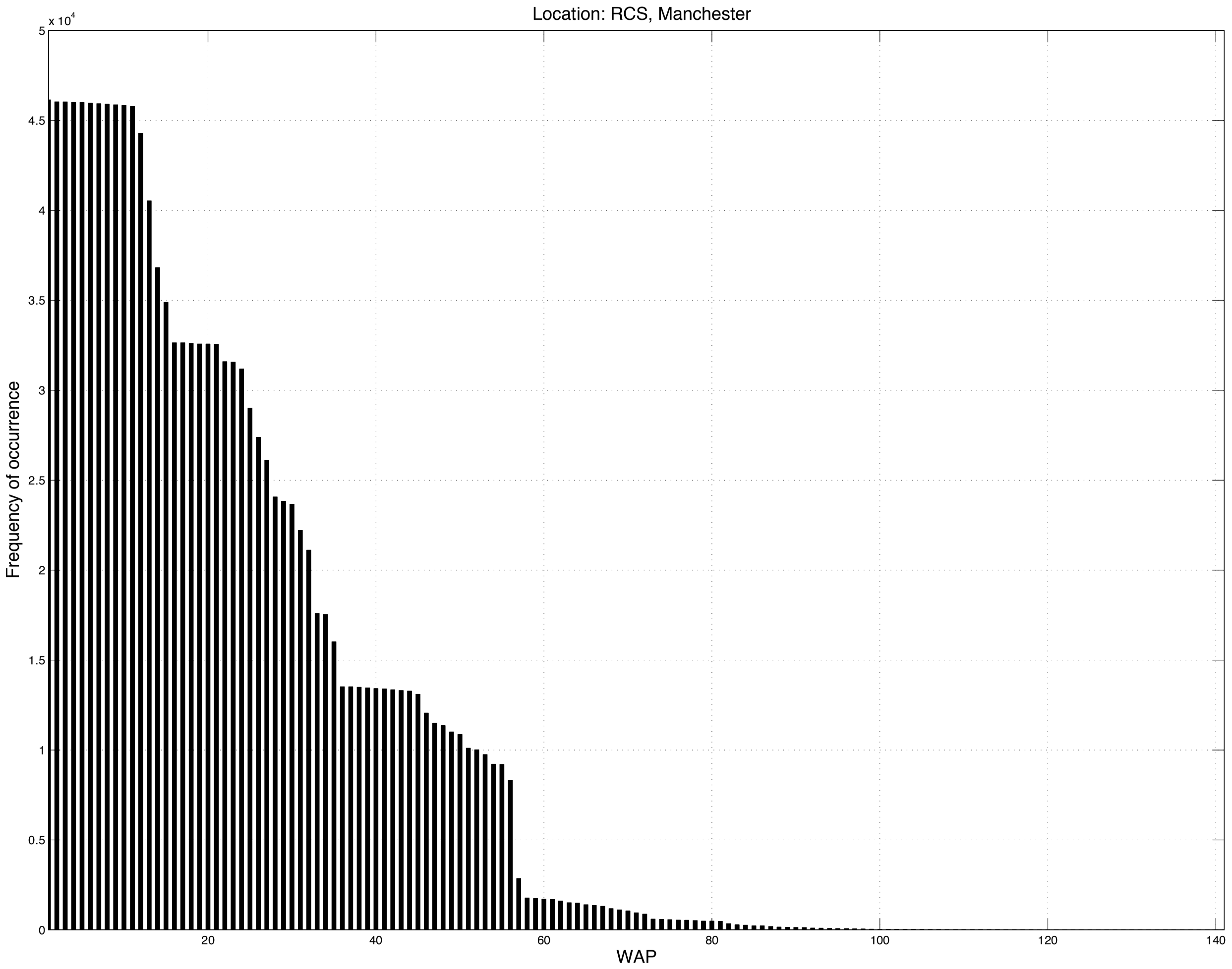


Figure 5.8: Frequency of Occurrence for the WAPs detected at RCS. The *x* axis are the WAPs ordered from higher to lower frequency of occurrence

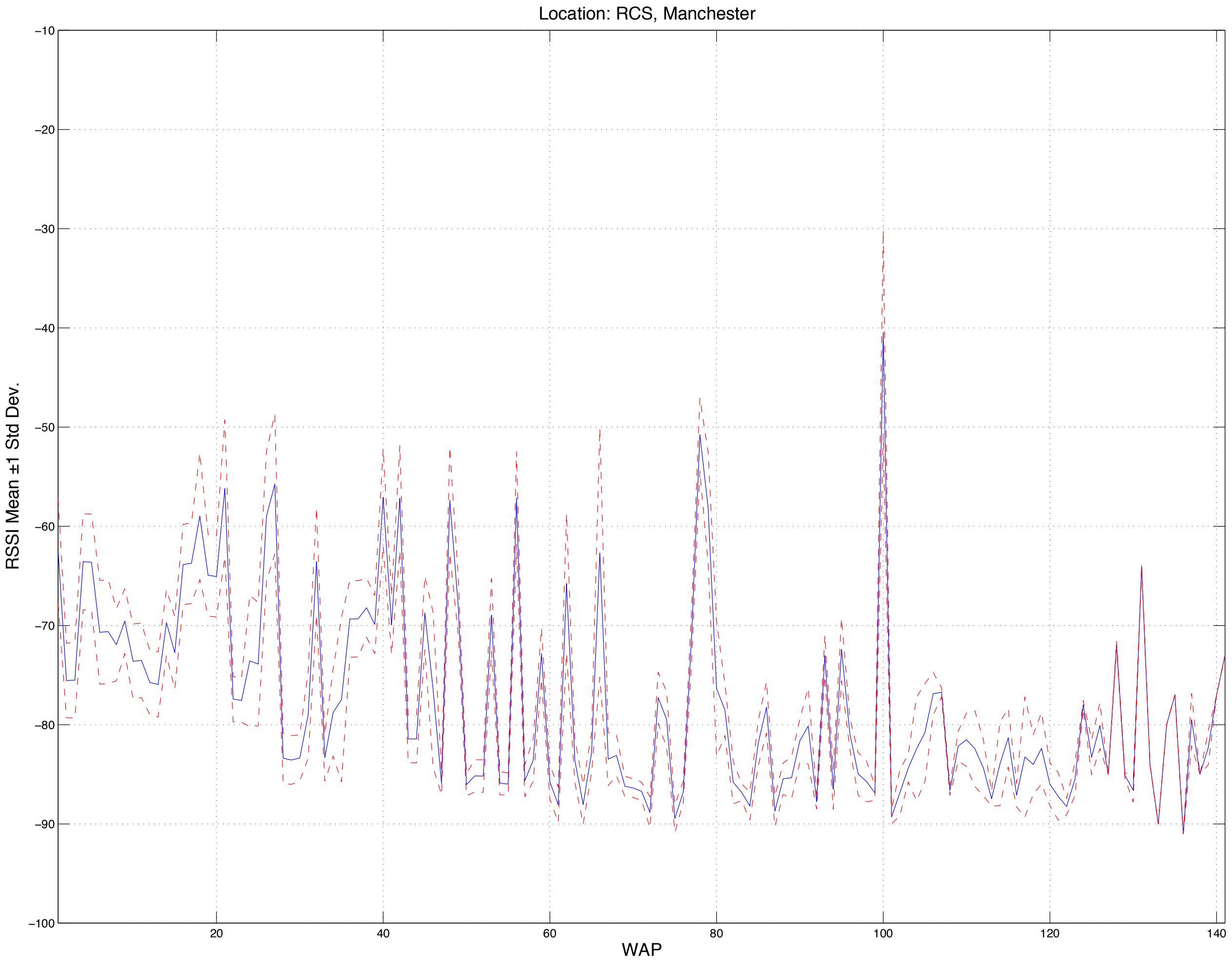


Figure 5.9: RSSI mean and ±1 standard deviation for WAPs detected at office 1.035. The *x* axis are the WAPs ordered from higher to lower frequency of occurrence.

5.1.3 Private Residence

The third location to be analysed is a private residence located close to the city centre in Derby, UK. The private residence is a stand alone detached maisonette.

In brief the residence comprises: Entrance Hallway, Lounge/Dining area, Kitchen, two double bedrooms and bathroom.

The dimensions of the property are:

- Lounge / Dining Kitchen 5.74m (max) x 4.33 (max)
- Master Bedroom 3.63m (extending) x 3.28 (max)
- Bedroom Two 2.79 (max) x 2.46m(extending)
- Bathroom 2.05m (max) x 1.66m (max)

Figure 5.10 presents the floor plan of the private residence, this figure shows the position where the WAP is installed within the master bedroom, position of the receiving client device (MacBook laptop) and a red camera icon indicating the place from where the panorama image (figure 5.11) was taken. Figure 5.11 is a panoramic image taken within the private residence.

Data collection at the property was done by a MacBook laptop. The positioning of the device was on the work-station placed at the corner of the lounge/dining area. At the residence there was only one WAP installed. Nevertheless, a total of 97 WAPs were detected over one year, which was the period of data collection. The number of observations recorded on the residence were 58,131, with an average of 23 WAPs detected per observation and 9 out of the 97 WAPs were seen in 90%+ of the observations.

Table 5.5 presents details for the top 10% WAPs collected at the private residence, these WAPs are are ranked by frequency of occurrence.

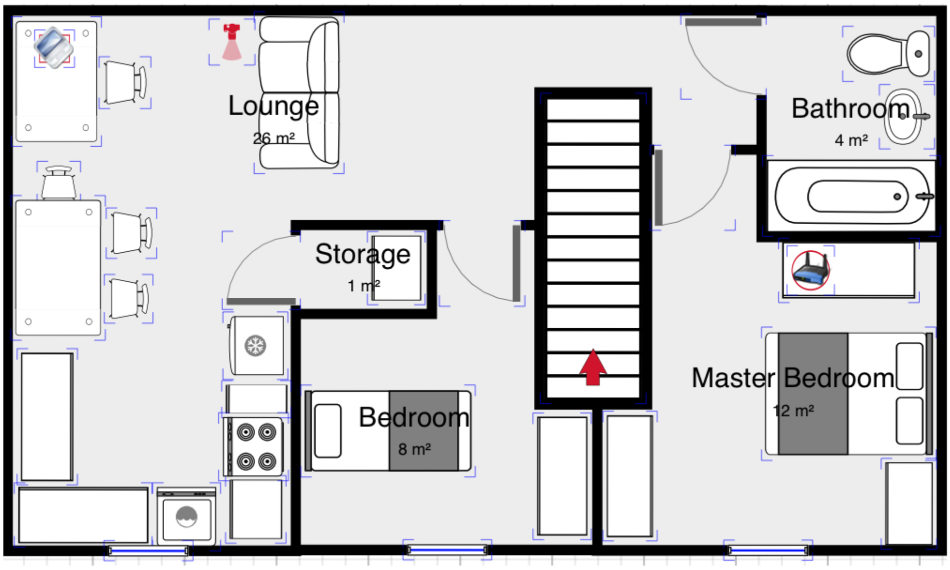


Figure 5.10: Private Residence floor plan, the position of the WAP is highlighted by a red circle, the client device recording observations is highlighted by a red square, and the position were the panoramic image was capture is indicated by a red camera.

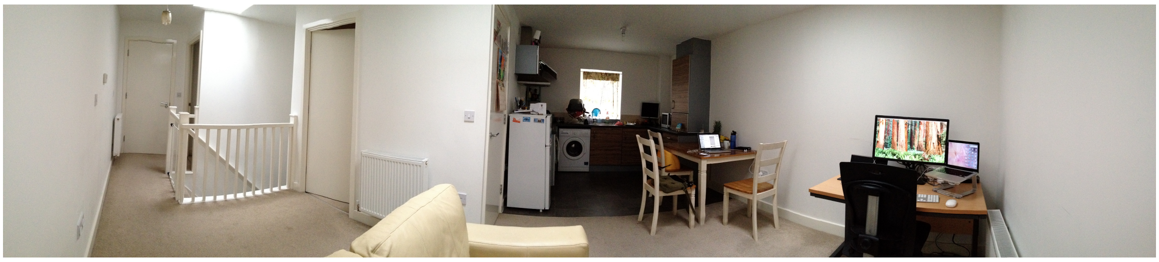


Figure 5.11: Panorama image within the Private Residence, Derby, UK

Table 5.5: WAPs detected at Private Residence, Derby, UK.

#	WAP ID	BSSID	SSID	RSSI Mean (*dB*)	RSSI Std Dev.	Freq. of Occurrence	Relative Freq. (%)
1	363	00:01:3b:ac:ca:e2	BTHub3-3458	-73.69	4.16	58062	99.88
2	360	02:01:3b:ac:ca:e2	BTWiFi	-73.67	4.10	58017	99.80
3	350	a0:21:b7:e8:f5:82	virginmediaGreen	-77.14	3.07	57956	99.70
4	361	12:01:3b:ac:ca:e2	BTWiFi-with-FON	-73.62	4.04	57881	99.57
5	353	a0:21:b7:ec:27:1e	virginmedia7604024	-74.98	3.69	57628	99.13
6	352	4c:17:eb:92:25:d1	SKY225D0	-79.59	3.37	54868	94.39
7	365	74:44:01:fc:dd:4c	virginmedia6578051	-78.33	4.21	54763	94.21
8	358	74:44:01:5f:8c:82	virginmedia2804029	-80.13	3.13	53160	91.45

Table 5.6 presents the descriptive statistics for the WAPs detected at the Private Residence.

Table 5.6: Descriptive Statistics, Private Residence, Derby, UK.

#	WAP ID	mean	mode	median	min	max
1	363	-73.69	-72	-73	-91	-35
2	360	-73.67	-73	-73	-91	-35
3	350	-77.14	-75	-77	-90	-43
4	361	-73.62	-73	-73	-92	-35
5	353	-74.98	-74	-74	-93	-37
6	352	-79.59	-78	-80	-96	-44
7	365	-78.33	-80	-79	-92	-42
8	358	-80.13	-80	-80	-94	-44

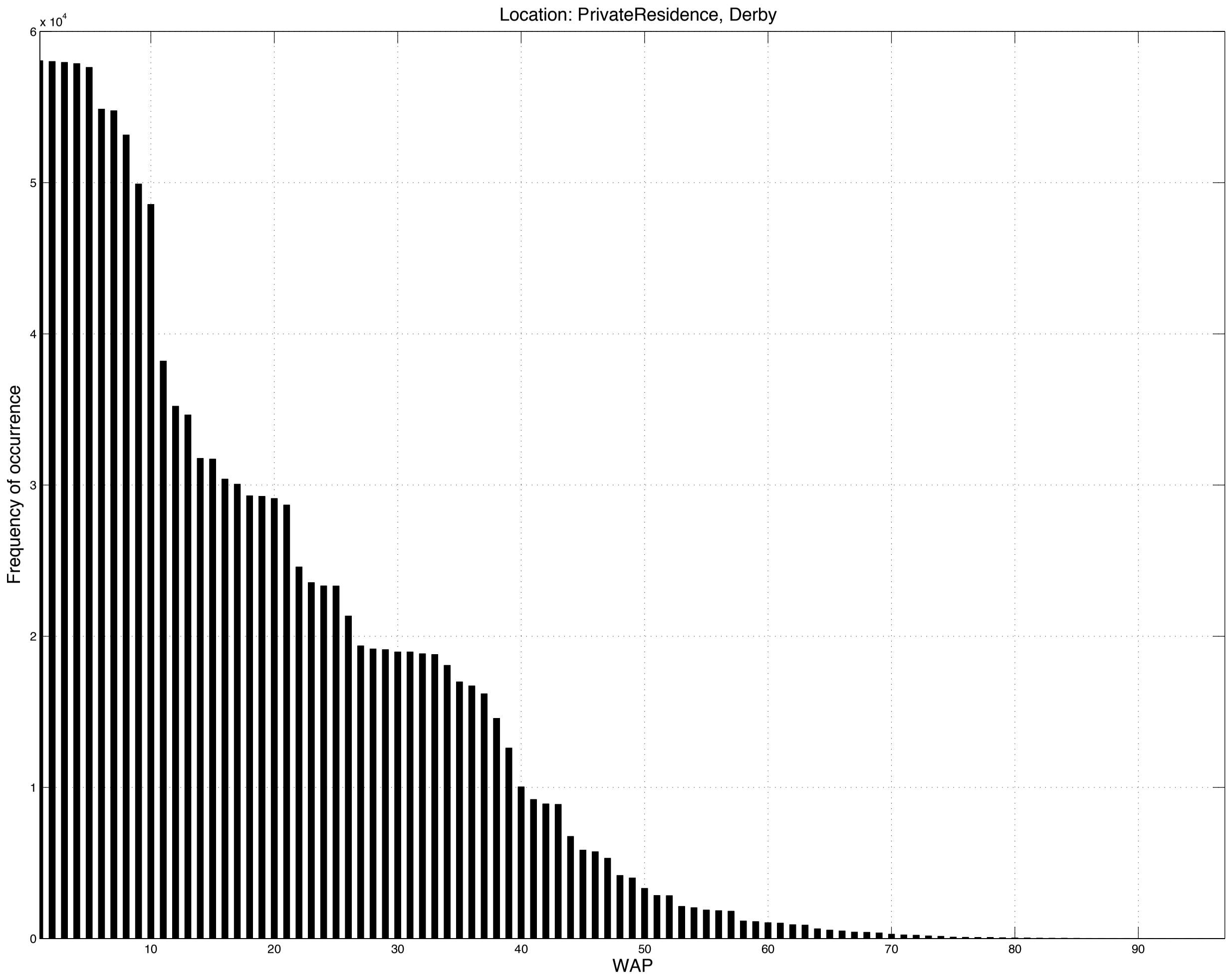


Figure 5.12: Frequency of Occurrence for the WAPs detected at the Private Residence. The *x* axis are the WAP ID ordered by higher to lower frequency of occurrence

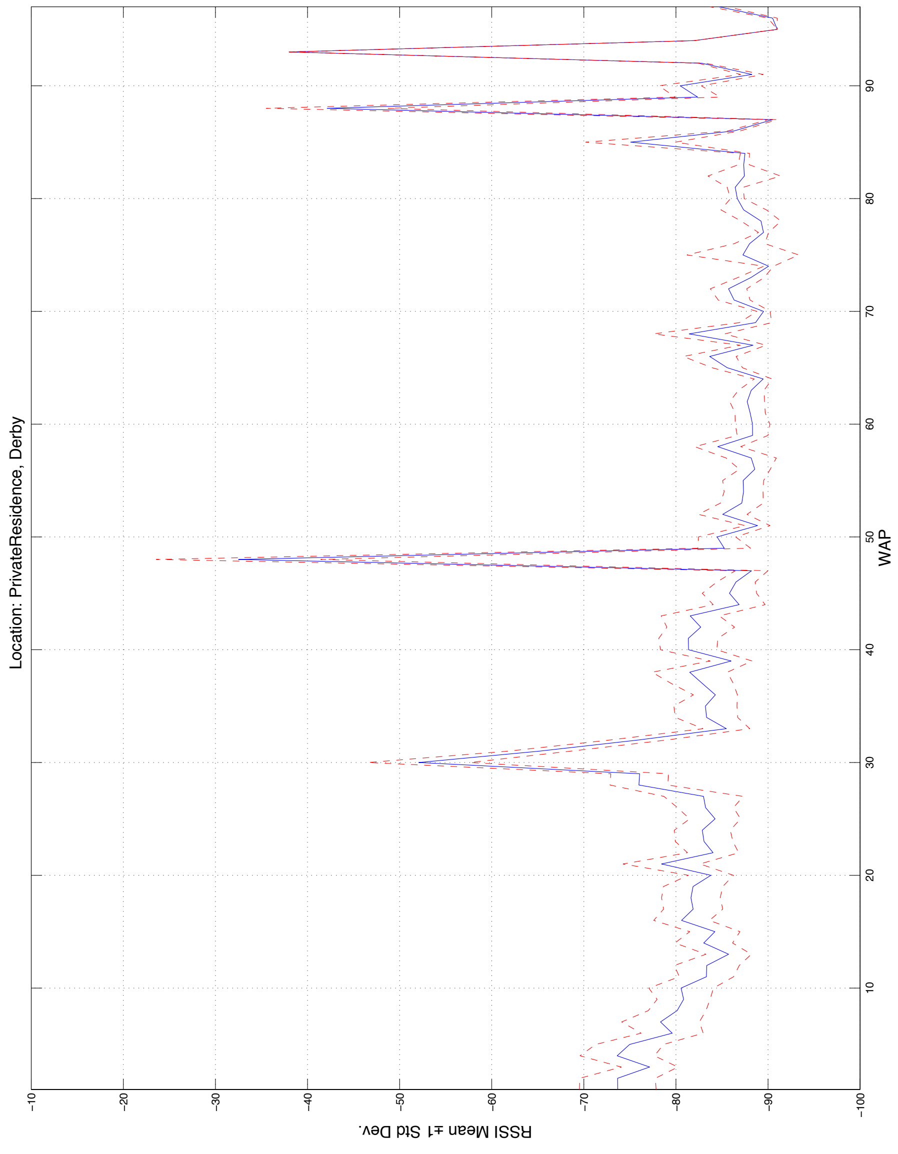


Figure 5.13: RSSI Mean and ±1 standard deviation for WAPs detected at the Private Residence. The x axis are the WAPs ordered from higher to lower frequency of occurrence.

When analysing data from the three locations, the location Platform 6 at Piccadilly Railway Station is different from the other two (Office 1.035 at RCS and the Private Residence) in the large number of WAPs and the probabilities presented in their frequency of occurrence.

This particular location (Platform 6 at Piccadilly Railway Station) is a highly crowded public space, not only regarding human presence, also highly crowded with WAPs. A total of 555 WAPs were detected, with an average count of 47 WAPs per observation. Nevertheless, as observed in Figure 5.3, most WAPs have a low frequency of occurrence. This is because of the detection of mobile WAPs (WAPs that are not fixed, such as mobile phones, or WAPs on board vehicles).

From analysing the data in Figure 5.3, 90% of the WAPs are unlikely to be observed, having a probability of occurrence lower than 0.25. From this it can be said that around 500 of the WAPs detected at this location provide little or no relevant information for positioning. With only 6% of the WAPs having a probability of occurrence higher than 0.5.

On the other hand, data from the locations Office 1.035 at RCS and Private Residence presented in Figure 5.8 and 5.12, showed a significative lower number of WAPs detected.

Office 1.035 had a total of 141 WAPs detected with an average WAP count of 32 per observation and location Private Residence had a total of 97 WAPs detected, with an average WAP count of 23 per observation. Both locations coincide in a 20% of WAPs with a probability higher than 0.5.

As result of comparing Figures 5.3, 5.8 and 5.12, it can be said that fingerprints for locations similar to the Piccadilly Railway Station would benefit from a ranking mechanism such as the proposed by the EWMA fingerprint presented in this thesis.

The EWMA fingerprint implements this ranking mechanism which takes into consideration the WAP frequency of occurrence (presence of the WAP), useful to generate light weight fingerprints by scoring WAPs on their “trustworthiness”. This is highly valuable for crowded locations such as Piccadilly Railway Station, where including all the WAPs when fingerprinting would result in larger fingerprints that require more computation.

Also, on analysing Figures 5.3, 5.8 and 5.12 an initial reasoning could be to expect a relation between the frequency of occurrence and signal strength. It could be expected that WAPs detected most of the time are those presenting a strong signal strength.

Figures 5.4, 5.9 and 5.13, corresponding to locations Platform 6 at Piccadilly Railway, Office 1.035 at RCS and Private residence respectively, present the mean and standard Dseviation of the RSSI per WAP. WAPs are arranged by frequency of occurrence (same order than presented in Figures 5.3, 5.8 and 5.12) where the WAP plotted at the far left is the one with the highest frequency of occurrence.

From Figures 5.4, 5.9 and 5.13 it is observed that there does not exist a substantial relationship between the mean of the signal strength and frequency of occurrence.

A misleading assumption of using only WAPs with the strongest signal for fingerprinting can result in the discrimination of valuable WAPs.

The proposed EWMA fingerprint takes into consideration not only the signal strength but also the frequency of occurrence. Furthermore, the EWMA fingerprint also considers historical values of the frequency of occurrence per WAP, which can result in robust better fitted fingerprints.

5.1.4 Timestamp Analysis

The figures 5.14, 5.15 and 5.16 present the RSSI (dB), mean, standard deviation, minimum and maximum values, over time for the WAP ranked #1 (most seen WAP) per location.

Table 5.7 presents a summary of the data collected in the three locations, including: the total observations recorded, the total number of WAPs detected, the top WAPs (number of WAPs detected in over 90%+ of the observations), and the average of WAPs detected per observation.

Location	**Total Observations**	**Total WAPs**	**Top WAPs (Present 90%+)**	**Average WAP count per Observation**
Platform 6, Piccadilly Railway Station	3,727	555	21	46
Office 1.035 RCS, UoM	46,161	141	13	32
Living Room Private Residence	58,132	97	9	23

Table 5.7: Summary of the dataset for the three locations subject to analysis

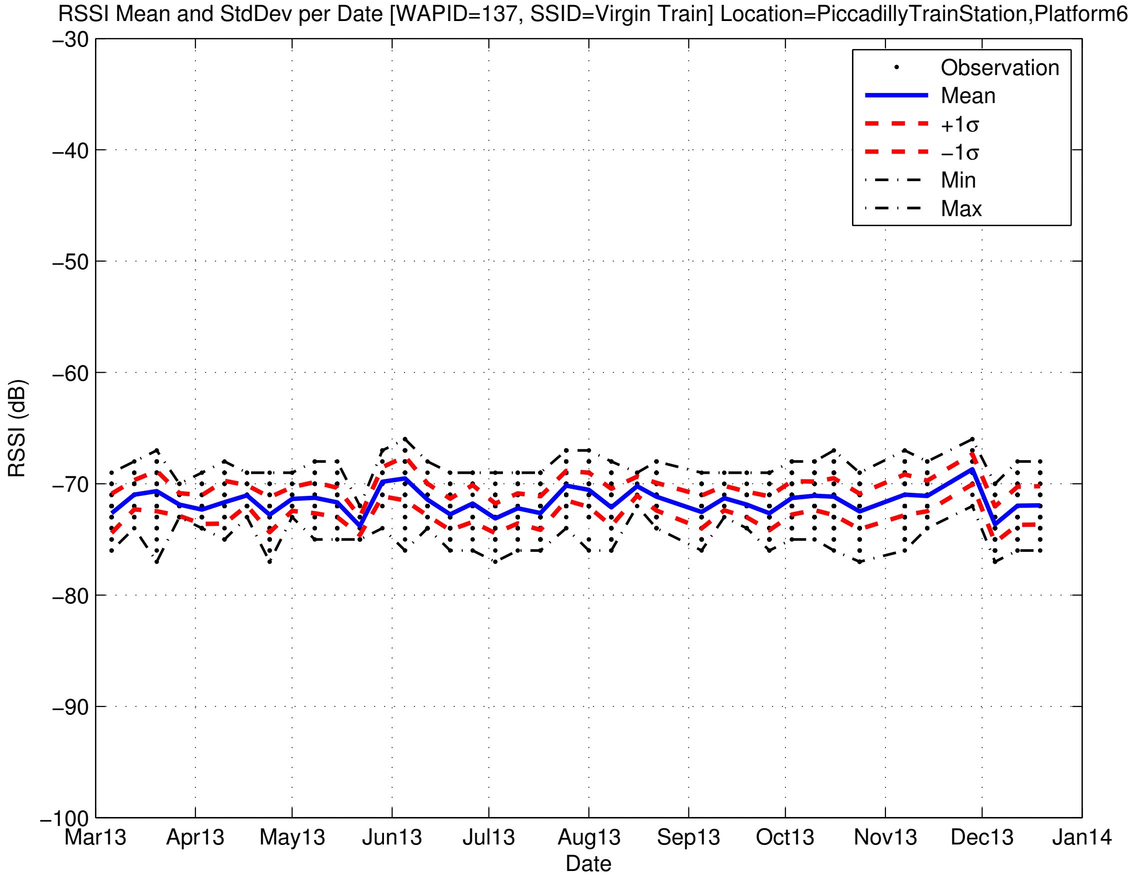


Figure 5.14: WAP 137. Piccadilly Railway Station, Platform 6

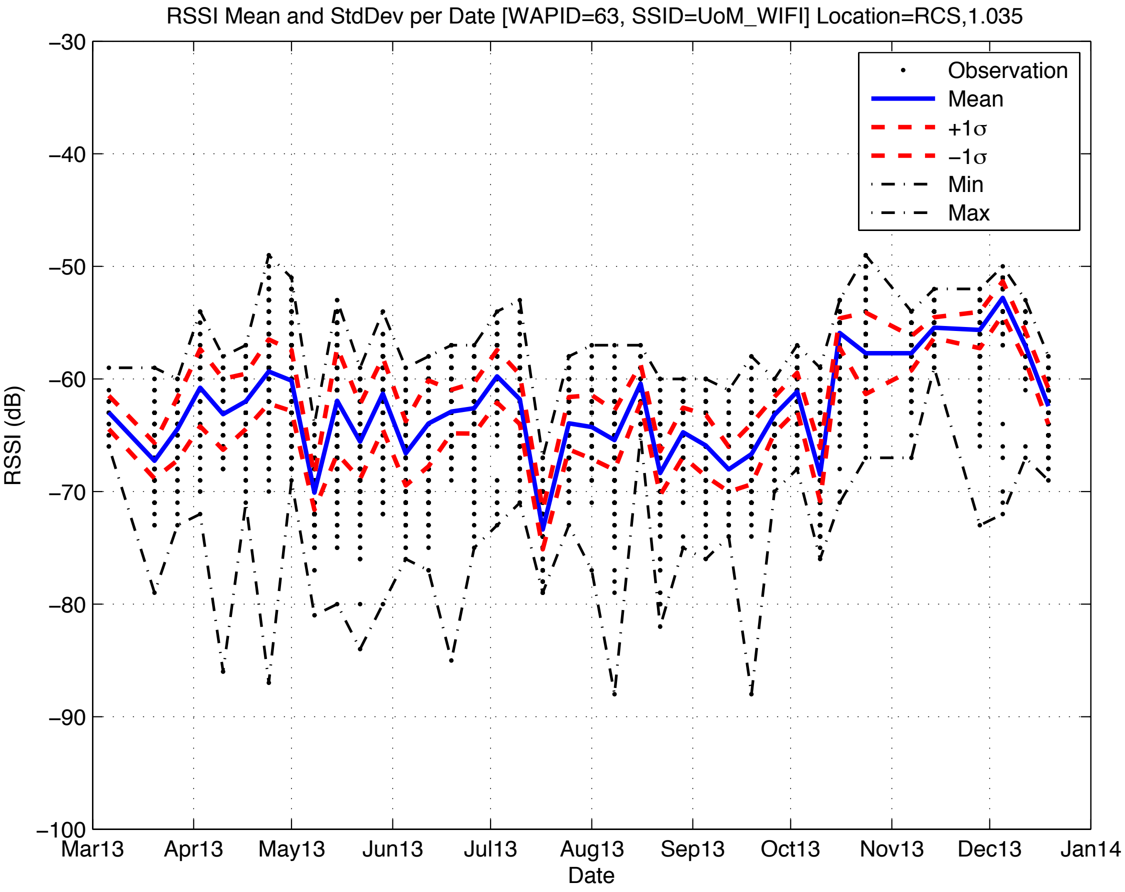


Figure 5.15: WAP 63. RCS, office 1.035.

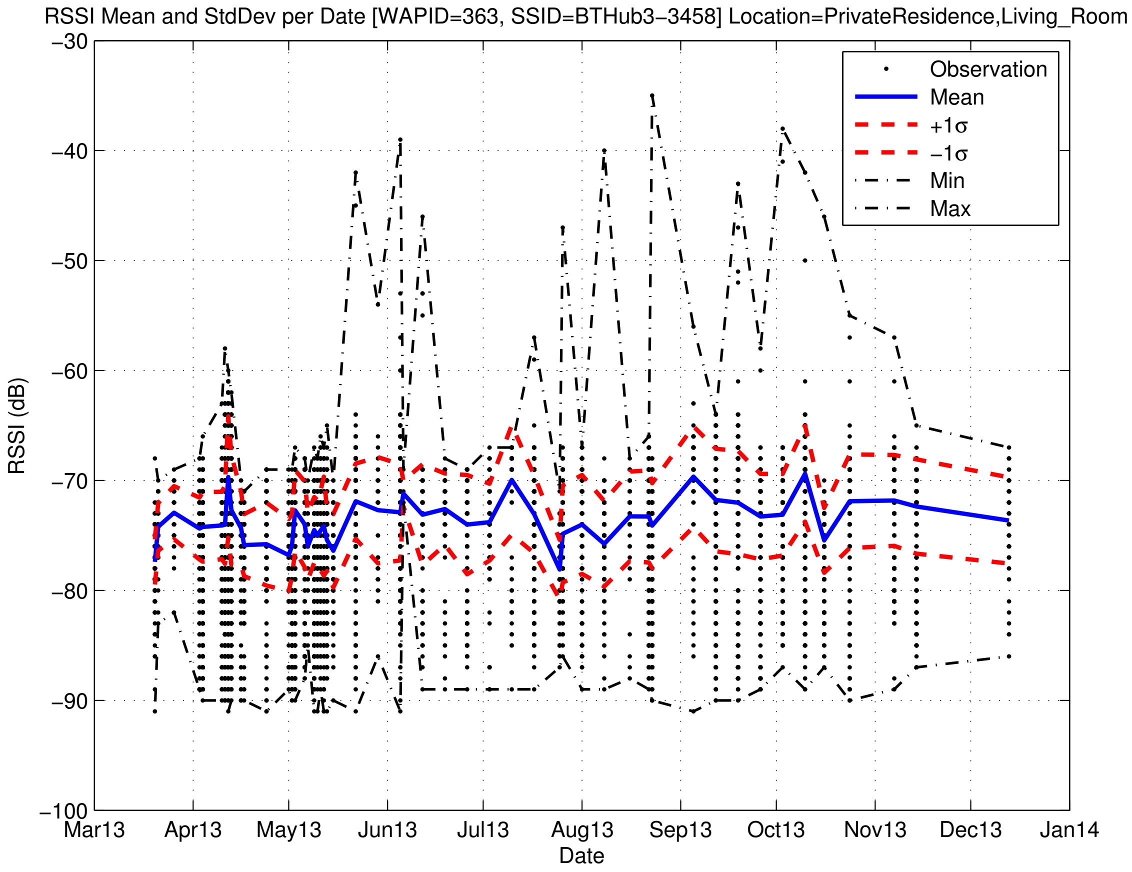


Figure 5.16: WAP 363. Private Residence

5.2 Chapter Summary

The chapter “Data Analysis” described in detail the locations selected for data collection. A data analysis was also presented, providing a better understanding of the real-life data collected. In the following chapter, the empirical technique called fingerprinting is explained. A variation of fingerprinting techniques is presented. This real-life data is used for demonstration and evaluation purposes.

Chapter 6

Fingerprinting and Positioning

As mentioned in Chapters 2 & 3, a fingerprint-based Indoor Positioning System consists of two stages; fingerprinting and positioning. This chapter develops the design and implementation for these two stages. The first section titled “Fingerprinting” explains the process for recording measurements, fingerprint generation and fingerprint updating. Also in this section, the EWMA Fingerprint is presented, an enhanced fingerprint definition that is adaptable to temporal changes in the WiFi environment. The second part of the chapter is the section titled “Positioning”, which presents three different but incremental positioning methods; these are based on WAP presence, WAP signal strength and a novel metric for ranking the WAPs.

Both processes, fingerprinting and positioning, require the input of observations. Therefore, observations are described before any further discussion.

Observations

The input for the WiFi based Indoor Positioning System is the data broadcast by WAPs within range of the measuring device. Every WAP broadcasts a Probe Response data frame containing information including: SSID, BSSI, RSSI, channel, HT, CC and security. The RSSI, in units of decibels (dB), indicates the strength at which the signal is received at the measuring device.

An observation includes: a location, a timestamp, and a set of WAPs. A formal definition of an observation is presented in Equation 6.1.

$$Observation = [l, t, \{WAP_{1,\ldots,N}\}] \tag{6.1}$$

$$WAP_n = [SSID, BSSID, RSSI, Channel, HT, CC, Security] \tag{6.2}$$

Descriptions for the parameters in an observation are as follows:

- Location (l). The location is the information about the place where the observation was recorded. This parameter is optional since it can be null in the situation where the physical place is unknown.

- Timestamp (t). The timestamp refers to the date and time with precision in seconds, defining when the observation was taken.

- WAPs ($\{WAP_{1,\ldots,N}\}$). Set of WAPs detected within range of the device recording the observation, where N is the number of WAPs within the set. As an example, in this study, the largest number of WAPs detected in a given observation is 67. Each WAP in the set includes the following parameters: SSID, BSSID (MAC address), RSSI, Channel, HT, CC and Security. A formal definition is presented in Equation 6.2, where n in WAP_n indicates the index of the WAP.

Recorded observations fit into one of two categories: "Survey-Observations" (so), observations tagged with a known location and "Query-Observations" (qo), observations from unknown location. Survey-observations are used to create and update fingerprints, whereas query-observations are processed for positioning. Fingerprinting and positioning are described in the following sections.

6.1 Fingerprinting

Fingerprinting, in the context of positioning, refers to the processing of data from the "WiFi scenery" recorded at a particular location. Fingerprinting is an iterative process and consists of the steps presented in Table 6.1. Steps 1 to 3 are related to initialisation also called training, these steps are executed for every new location to be added to the radio-map. Steps 4 to 6 are related to updating, these steps are repeated every time an update to the fingerprint is required. The fingerprint update is carried out when a survey-observation recorded at a location already fingerprinted is received into the system.

Fingerprinting
Initialisation
1. Collection of survey-observations for fingerprint initialisation
2. User validation
3. Fingerprint generation
Maintenance
4. Collection of survey-observations for fingerprint update
5. User validation
6. Fingerprint update

Table 6.1: Fingerprinting Process.

Fingerprint Initialisation

- **Collection of survey-observations for fingerprint initilisation.** The initial step for fingerprinting is to establish the locations to be scanned. This includes determining the size of the locations by considering: bulding layout and intended accuracy (for the final application). For example, within a building, a set of observations can be recorded in every room, then each room becomes a definable location. The RCS testbed presented in Chapter 5, is an example of a setting with room level accuracy. When scanning a big open space inside a building it can be partitioned into a grid where each square is considered a location. The sizes of the locations determines the resolution of the radio-map in the system. Once defined the locations, survey-observations are collected by an expert at the pre-defined locations. These survey-observations are labelled with an appropriate name (alphanumeric string) that describes each location. The label can be the name or number of a room (e.g. 1.005) and, if applicable, the grid coordinates. Survey-observations are required for fingerprint initialisation and updates once the user is validated.

- **User validation.** The user validation is an important aspect for fingerprinting. When survey-observations are received the user is validated against a set of authorised users. Survey-observations, coming from unauthorised users are not or may not be used for fingerprint initialisation nor for updates. The user is validated through the MAC address of the WiFi enabled device collecting the survey-observations. The input of deliberate or unintentional erroneous data is potentially reduced by carrying out user validation.

- **Fingerprint generation.** A fingerprint is initialised every time a survey-observation with a new location (not already existing in the database) is received. In order to carry on with the initialisation, firstly, new location details from the survey-observation are stored in the database. These details, embedded into the survey-observation by the user, can include: Country, City, Building, Room and Coordinates. Once the location is in the database, a new fingerprint is generated and linked with the location.

Fingerprint Maintenance

- **Collection of survery-observations for fingerprint update.** Fingerprints are maintained up-to-date by collecting further survey-observations. Fingerprints (fp) are updated with the incoming survey-observations (so) coinciding in location.

- **User validation.** The survey-observations used for updating fingerprint are also subject to user validation. Only those survey-observations coming from an authorised user are considered for updating existing fingerprints.

- **Fingerprint update.** An up-to-date fingerprint is computed based on the existing fingerprint and the survey-observation.

  This up-to-date fingerprint includes also a set of WAPs with their estimated expected signal strength value. This set is generated based on the WAPs in the fingerprint (fpWAPs) and the WAPs in the current survey-observation (soWAPs). The specific mechanism of how the fpWAPs and the soWAPs are "merged" for creating the up-to-date fingerprint, depends on the technique implemented.

  Including the proposed fingerprint (EWMA Fingerprint) there are five fingerprint techniques, (from here onwards called fingerprint definitions) explained and tested in this thesis.

  Once the WAPs for the up-to-date fingerprint are computed, the updating process is completed by replacing the existing/previous fingerprint with the up-to-date fingerprint.

The initialisation and maintenance processes explained are general purpose processes independent of the fingerprint definition. Differences and details of how the fingerprint is generated and updated per definition will be clear in the following section, where the different fingerprint definitions are explained.

6.1.1 Fingerprint

A fingerprint can be defined as a tuple formed by the location l and its set of WAPs. A fingerprint can be formally defined by,

$$FP = [l, \{WAP_{1,...,N}\}] \tag{6.3}$$

from where

- Location (l). An alphanumeric string that is considered the physical location "name". This location string can include the following: country, building (e.g. RCS), room (e.g. 1.035), and grid coordinates within the room (if appropriate)
- WAPs ($\{WAP_{1,...,N}\}$). The set of WAPs that were detected at the location. The number of WAPs within the set is indicated by N. Equation 6.4 presents the parameters stored per each WAP

$$WAP_n = [WAP\_ID_n, \hat{x}_n] \tag{6.4}$$

- WAP Identifier (WAP_ID). Parameter that identifies a WAP uniquely in the system. This identifier is directly related with the BSSID (MAC address) of the WAP
- Expected signal strength value ($\hat{x}$). Estimated expected value for the WAP signal strength. An estimated $\hat{x}$ is computed per each WAP within the set in the fingerprint
- Number of WAPs (N). Integer that indicates the number of WAPs included in the set $\{WAP_{1,...,N}\}$ within the fingerprint
- WAP index (n). Index for referencing individually the WAPs in the set $\{WAP_{1,...,N}\}$ from where $n = 1,...,N$

This fingerprint is the foundation on which other fingerprint definitions are based. Fingerprint definitions can be differentiated by the method used to calculate the expected signal strength value $\hat{x}$ per WAP. Another difference between fingerprints definitions is the number of parameters per WAP. Including less or more information that can be used for improving positioning. The estimate $\hat{x}$ is stored for each WAP in the fingerprint as an expected value of the WAP's RSSI. The computation of this value is critical since it is used for comparison in the positioning stage.

In order to understand the novel EWMA Fingerprint proposed in this thesis, the following existing methods of computing a fingerprint are described:

- Basic Fingerprint
- Default Fingerprint
- Gallagher Fingerprint
- Jung-Sung Fingerprint

Following is a description of every one of these listed techniques.

6.1.2 Basic Fingerprint

The simplest of the fingerprints techniques is referred by the author as a "Basic Fingerprint". This definition is a starting point for understanding fingerprint generation. A Basic Fingerprint is created based solely on the latest survey-observation. The set of WAPs in the Basic Fingerprint is exactly the same that the most recent survey-observation, furthermore the WAP's expected signal strength value $\hat{x}$ is simply the latest measure signal strength (RSSI). When using this technique, the historical information about the detected WAPs and their signal strength are lost. The fingerprint updating process consists in overwriting the fingerprint when a new observations is recorded. The Basic Fingerprint formal definition is identical to the one presented in Equations 6.3 and 6.4, the peculiarity about the Basic Fingerprint is the value assigned to $\hat{x}$, which is defined as follows:

$$\hat{x}_i = x_i = RSSI \tag{6.5}$$

Using the Basic fingerprinting method, all estimated values ($\hat{x}_i$) are equal to the current observed value (x_i). It is assumed that only the latest/current observation is important, and prior observations are not considered in the estimation of the expected value. This process can be seen as a weighted average, where the last observed value has all the weight.

The Basic Fingerprint was implemented. The data set described in Chapter 5 Data Analysis, was used as survey-observations for Basic Fingerprints initialisations and updates. Figure 6.1 presents the raw data for WAP63 from Location Office 1.035 at RCS (Location referred as Location 20 in the Figures).

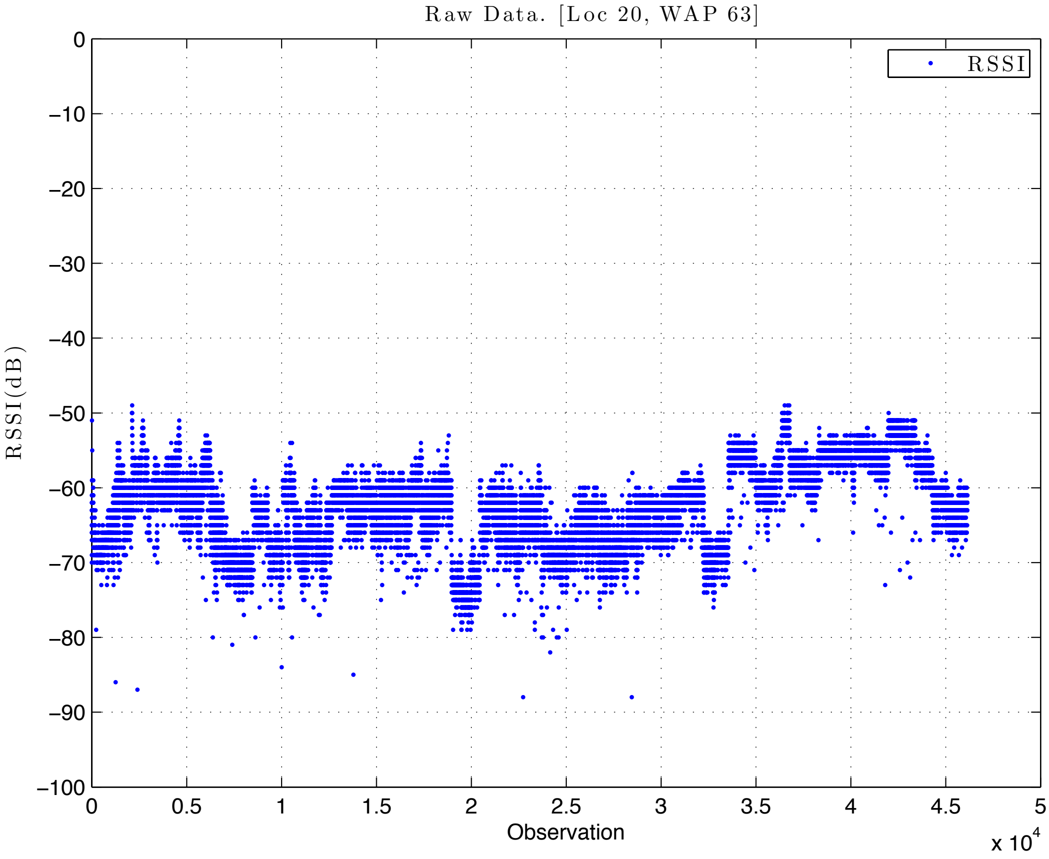


Figure 6.1: Raw data observed from WAP 63, Location Office 1.035, RCS (Location 20).

Figure 6.2 plots the values of $\hat{x}$ for WAP 63 when implementing the Basic Fingerprint approach on survey-observations collected at Location 20.

Figure 6.3 is the raw data plotted along with the Basic Fingerprint, as it is appreciated the $\hat{x}$ value (red line) overlaps the RSSI data (blue points).

The Basic Fingerprint approach is simple and the computation of the value $\hat{x}$ is remarkably straightforward, nevertheless this approach suffers from some limitations such as:

- Fingerprint is overwritten with every incoming observation
- Contains no knowledge of historical information
- It is not robust in the presence of signal changes and outliers

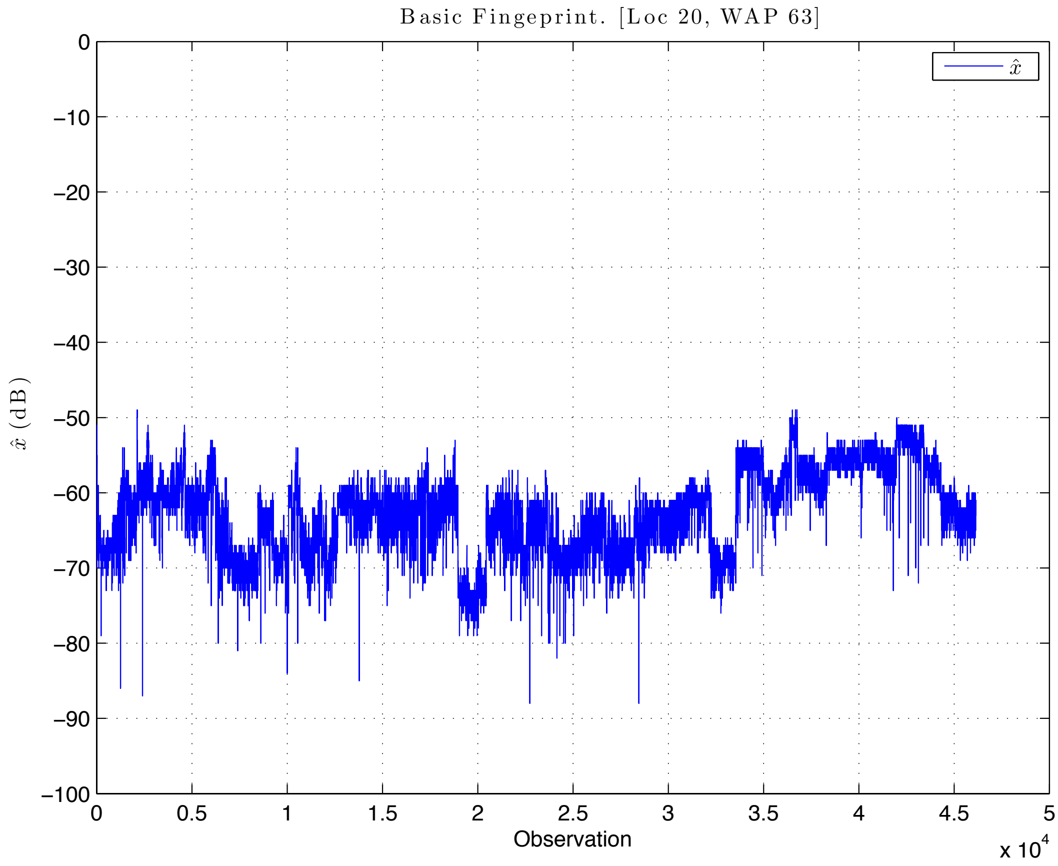


Figure 6.2: Expected RSSI computed with the Basic Fingerprint approach from WAP 63, Location Office 1.035, RCS (Location 20).

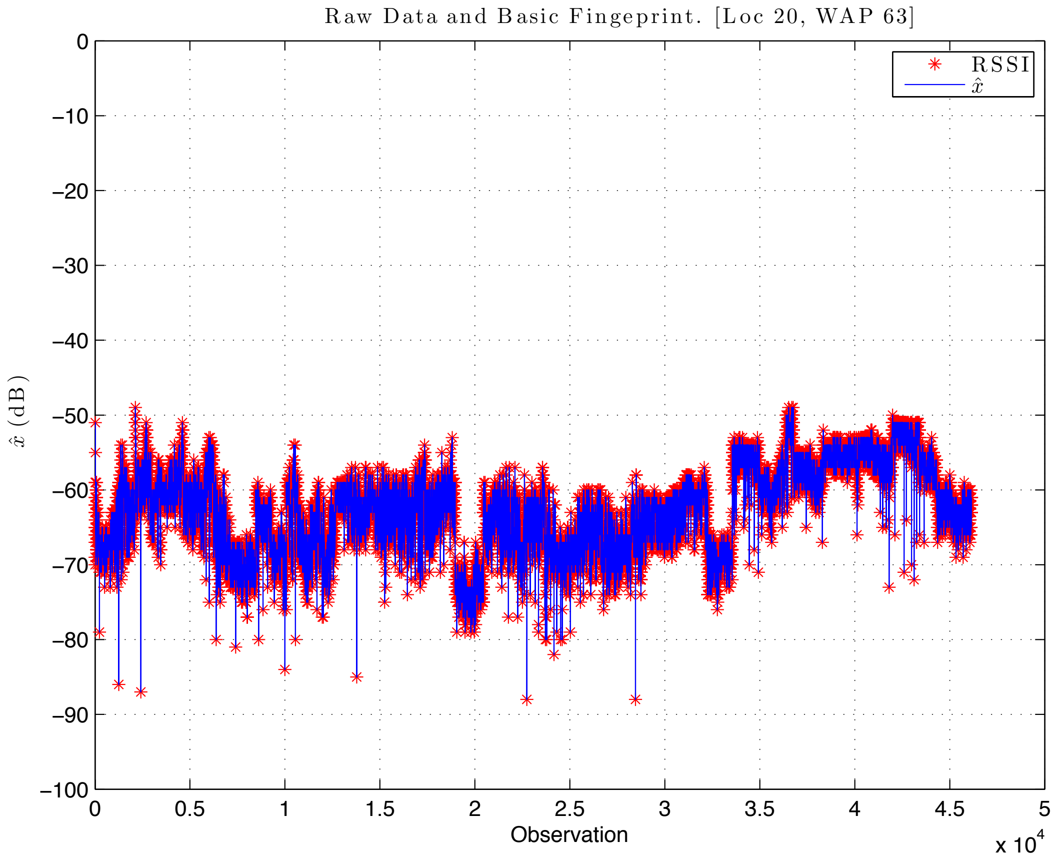


Figure 6.3: Raw data and Basic Fingerprint from WAP 63. It is observed from this figure that the Basic Fingerprint (blue) follows identically the RSSI (red).

6.1.3 Default Fingerprint

The Default Fingerprint, named as such in [FLMG13] and defined in [BP00], consists of a location and a list of WAPs with their corresponding expected signal strength value ($\hat{x}$). Similarly to the Basic Fingerprint, the Default Fingerprint is defined by Equations 6.3 and 6.4. The difference is in the computation of $\hat{x}$. The Default Fingerprint uses a statistical parameter (mean or median) [FLMG13] as $\hat{x}$. A common approach is to use a simple average. Equation 6.6 is the formal definition for the estimation of the expected signal strength value on the implementation of a Default Fingerprint.

$$\hat{x} = \mu = \frac{1}{I}\sum_{i=1}^{I} x_i \tag{6.6}$$

The estimated value $\hat{x}$ is the mean of all the RSSI values observed up-to-date per WAP.

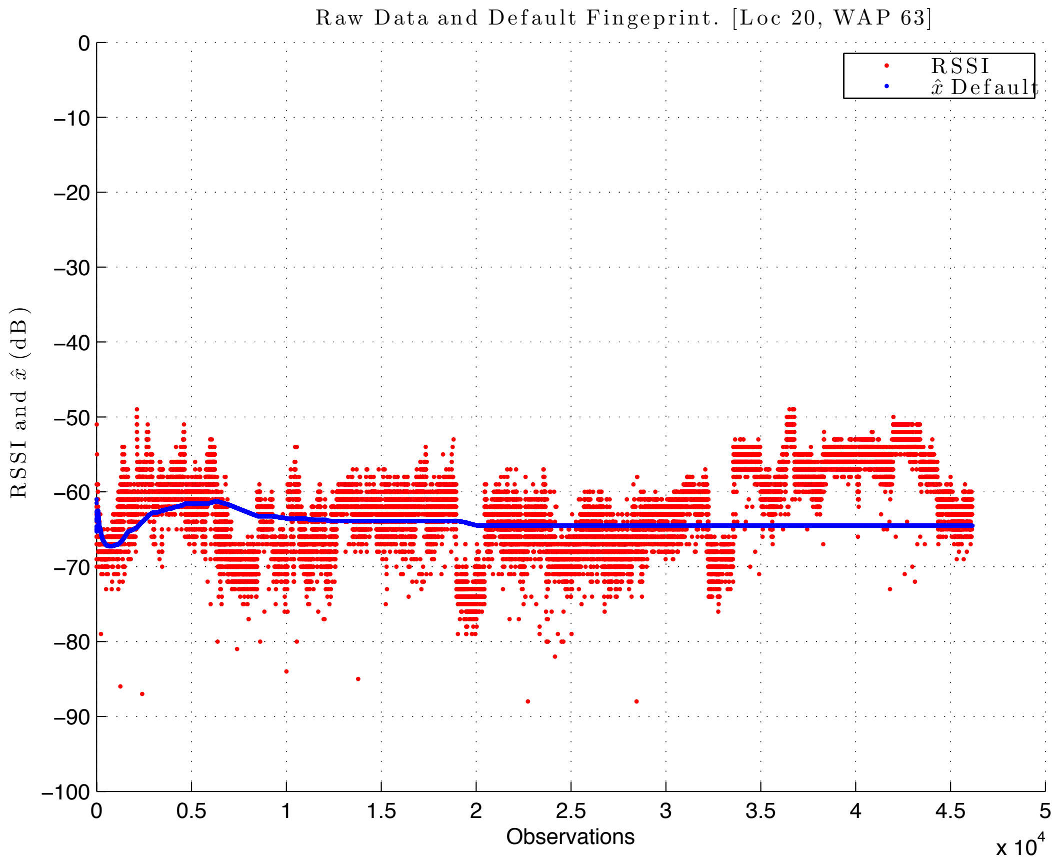


Figure 6.4: Raw data (red) and Default Fingerprint expected signal strength $\hat{x}$ (blue) from WAP 63, Location 20.

The computation presented in Equation 6.6 is not straightforward, a naïve approach is to store all the previous values of RSSI, and then compute the mean as required. The problem with storing every single value is that it requires a large storage that would increase with every new survey-observation received. Also, the amount of data received in a crowdsourced fingerprint based IPS can be large and it can increase over time. Therefore the use of a simple arithmetic mean and standard deviation would not be suitable. In order to remove the requirement of storing every one of the survey-observations a recursive method was employed on the implementation of the Default Fingerprint. The implemented method takes the new data point and the previous average to arrive at both the new average and standard deviation. In other words, rather than storing every one of the RSSI values per WAP per observation, it is only required to store the latest average and an intermediate variable used for computing the standard deviation. This method [Wel62], [Lin74] and [CGL83] consists of one-pass iterative calculation of the mean, where signal strength measurements are used once only and therefore measurements do not need to be stored. In this thesis the programming of the Default Fingerprinting algorithm implements this iterative approach to compute the mean and standard deviation.

The process is the following:

First, the mean μ and an intermediate value Q, are initialised to 0 (zero),

$$\mu_0 = 0 \tag{6.7}$$

$$Q_0 = 0 \tag{6.8}$$

Second, the mean is computed using Equation 6.9

$$\hat{x} = \mu_i = \mu_{i-1} + \frac{x_i - \mu_{i-1}}{i} \tag{6.9}$$

$$\tag{6.10}$$

For $i = 1...I$, where i is the number of observation, its value increases by one with every new survey-observation and I is the total number of observations made up-to-date. The value for x_i it the RSSI from incoming survey-observations,

$$x_i = RSSI \tag{6.11}$$

Then, the intermediate value is computed using Equation 6.12

$$Q_i = Q_{i-1} + (x_i - \mu_{i-1})(x_i - \mu_i) \tag{6.12}$$

Finally, the standard deviation s_i is computed with the value Q and the number of observations, as shown in Equation 6.13

$$s_i = \sqrt{\frac{Q_i}{i-1}} \tag{6.13}$$

From this, the value $\hat{x}$ per WAP within a Default Fingerprint is μ_i from equation 6.9, and it is updated every survey-observation.

Figure 6.4 presents the changes on the value $\hat{x}$. For this, survey-observations from WAP 63 (Location 20) were processed following the Default Fingerprint definition.

A Default Fingerprint includes every one of the WAPs detected at the location. This can be considered as a limitation since it has no mechanisms to filter the WAPs. Hence, Default Fingerprint does not handle the addition and removal of WAPs. Also, the average method used to compute the expected signal strength assumes that all observations are equally important. This can be thought as a weighted average where the first observation is weighted exactly the same as the latest observation, even if the first observation was made in the very distant past. (e.g. Considering the recorded dataset, an observation made in March 2013, is weighted the same as an observation made in January 2014).

6.1.4 Gallagher Fingerprint

The third fingerprinting technique to discuss is the one proposed by Gallagher et al [GLDR10]. Gallagher Fingerprint is formed by a tuple of location and a set of WAPs, like the previously explained definitions. Nevertheless, Gallagher Fingerprint includes (in addition to *WAP_ID* and $\hat{x}$) two extra parameters per WAP: score and pending. Thus, Gallagher Fingerprint can be formally defined as equation 6.14 and 6.15.

$$G_{FP} = [l, \{WAP_{1,\ldots,N}\}] \tag{6.14}$$

$$WAP_n = [WAP\_ID_n, \hat{x}_n, score_n, pending_n] \tag{6.15}$$

- Expected value ($\hat{x}$) In Gallagher Fingerprint the expected value per WAP is the latest RSSI value.

$$\hat{x}_i = x_i = RSSI \tag{6.16}$$

Similarly to Basic Fingerprint, Gallagher's $\hat{x}$ is identical to the RSSI, as it is shown in Figure 6.5 were it is presented the RSSI and the estimated $\hat{x}$ for WAP 63, Location 20.

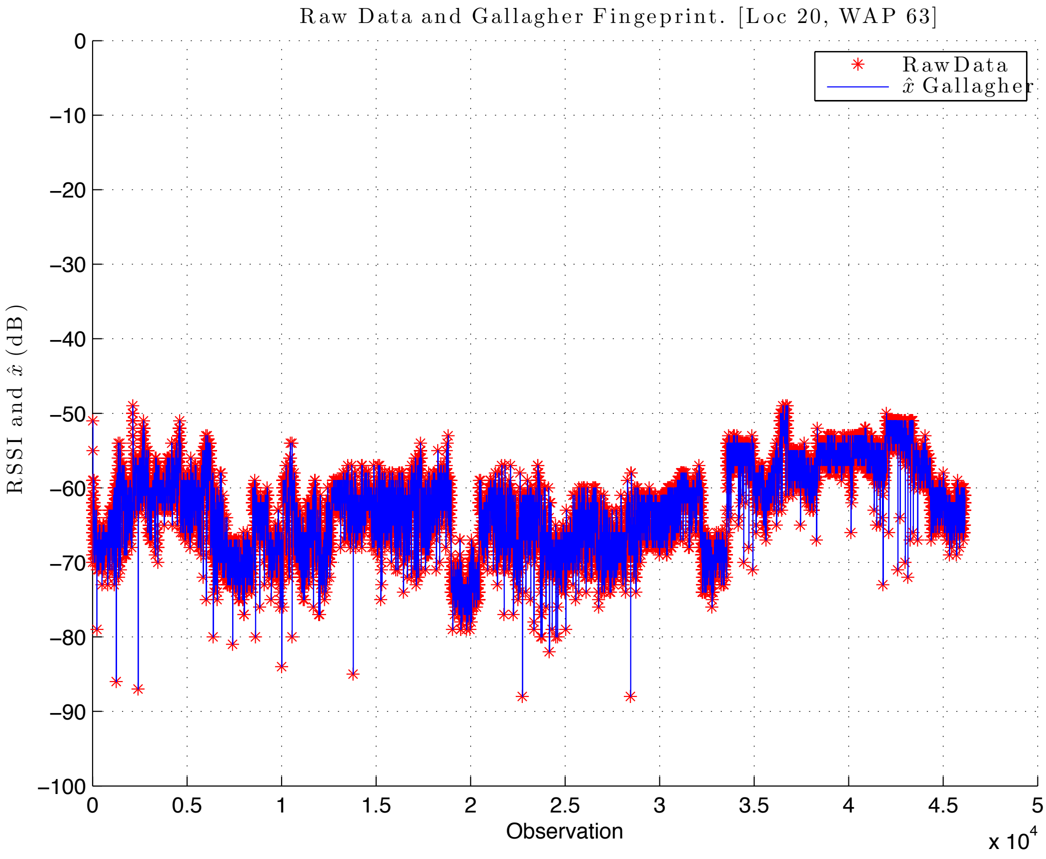


Figure 6.5: Raw data and Gallagher's $\hat{x}$ from WAP 63, Location 20.

- Score. Gallagher's fingerprint is based on 'points' per WAP, where the points are a *score* assigned to each WAP. The *score* increases by one every time the WAP is detected, and decreases by one when the WAP is not detected. It is established [GLDR10] that only WAPs with *score* over a threshold are included in the fingerprint. The threshold is an arbitrary value, [GLDR10] presents experiments using a threshold of 30. WAPs with a score over the threshold are identified using the boolean value *pending*.

- Pending. Parameter *pending* indicates when the WAP is included or not in the fingerprint, $pending = 0$ means the WAP is included in the fingerprint, $pending = 1$ means the WAP is not included in the fingerprint. In order to include a WAP in the fingerprint (setting $pending = 0$) the *score* should be over the specified threshold.

The implementation of Gallagher Fingerprinting technique is presented in the Algorithm 1.

Algorithm 1 Gallagher Fingerprinting

```
Require: $MAX\_SCORE$
$soWAPs = \{$WAP | WAP in survey-observation, where $WAP = (WAP\_ID, RSSI)\}$
$fpWAPs = \{$WAP | WAP in fingerprint, where $WAP = (WAP\_ID$, $\hat{x}$ , score, pending)$\}$
for all $soWAPs$ such that $WAP \in fpWAPs$ do                ▷ Stage 1
    if $WAP \in soWAP$ then
        if $pending = 1$ then
            $score++$
        else
            $score = MAX\_SCORE$
    else
        $score--$
        if $score < 0$ then
            $pending = 1$
            $fpWAPs \leftarrow fpWAPs - \{WAP\}$                ▷ Remove WAP
for all $fpWAP$ such that $WAP \in soWAPs$ do                 ▷ Stage 2
    if $WAP \in fpWAPs$ then
        if $score \leq MAX\_SCORE$ and $pending = 1$ then
            $pending = 0$
        else
            $pending = 1$
    else
        $pending = 1$
        $score = 1$
        $fpWAPs \leftarrow fpWAPs \cup \{WAP\}$                 ▷ Insert WAP
```

Figure 6.6 presents the *score* and *pending* values for WAP 63 at Location 20, computed using the Gallagher Fingerprinting algorithm. This graph shows the increase and decrease for the *score* when the WAP is detected or not detected. The threshold in the *score*, for this particular implementation was set to 15 (Similarly to the threshold used in [LJYH13]).

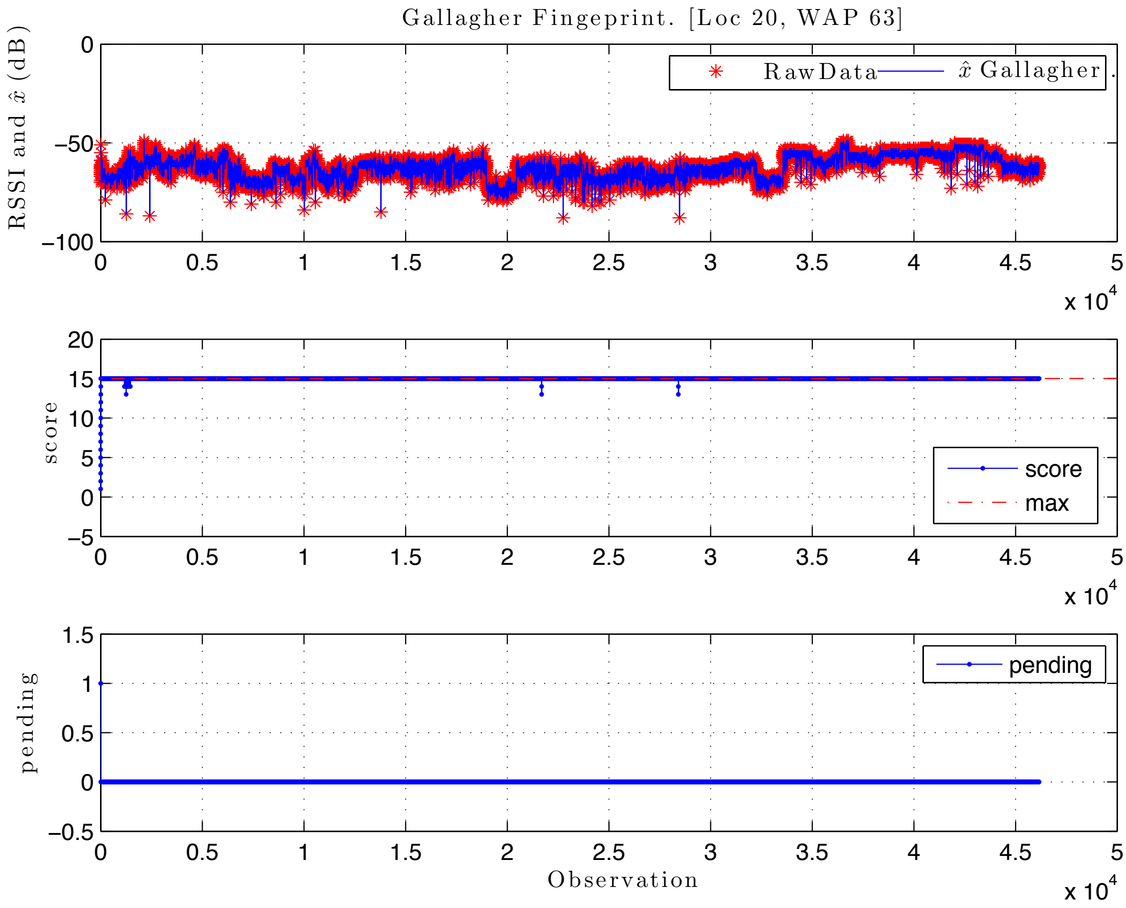


Figure 6.6: Gallagher's fingerprint technique applied to the RSS data from WAP 63. WAP 63 is detected in most of the observations at Location Office 1.035, RCS. The score increases and reaches the threshold rapidly, setting pending to 0 (zero). This includes the WAP in the fingerprint after 15 observations. Because WAP 63 is stable and observed most of the time, then the score stays on the maximum (max) value.

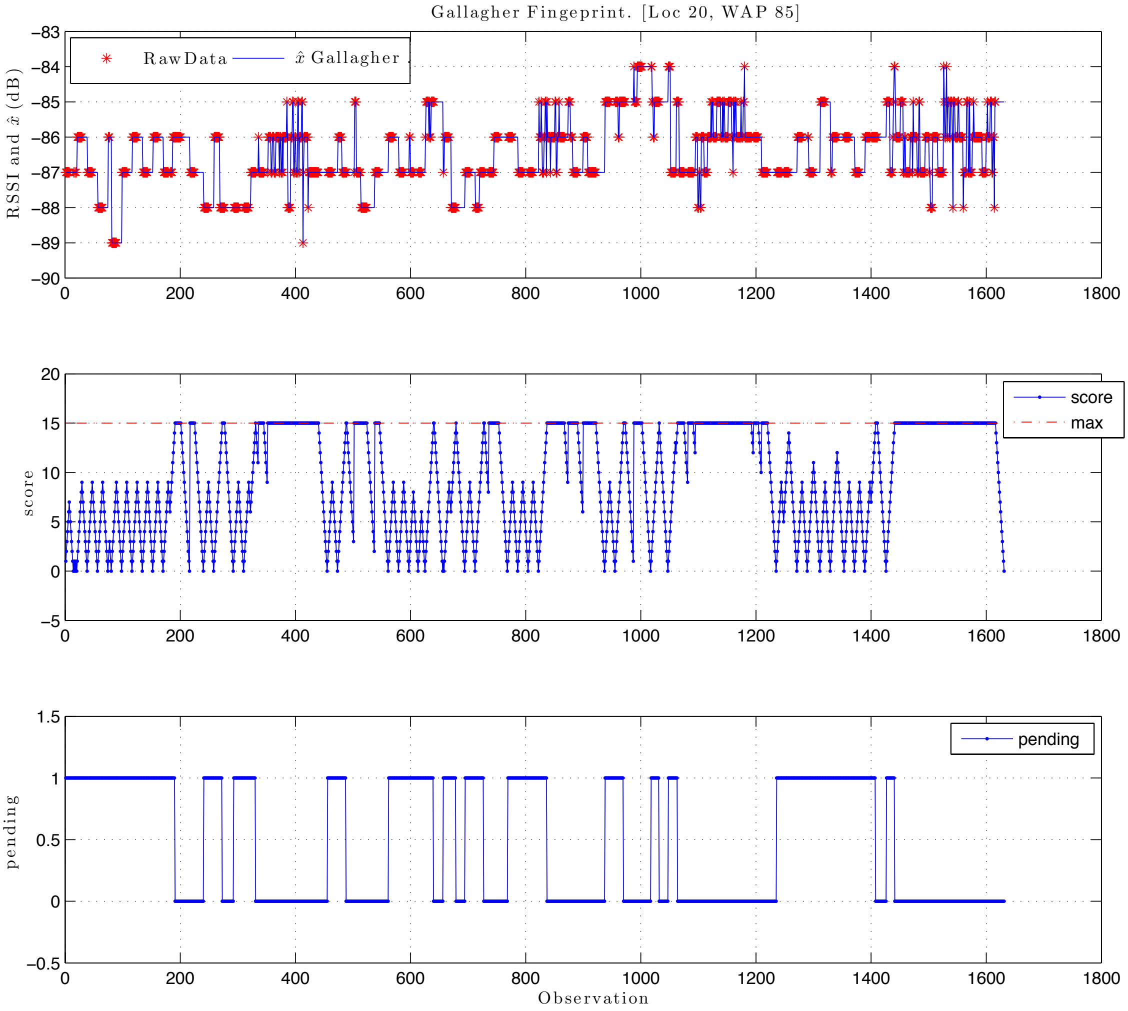


Figure 6.7: Gallagher's fingerprint from WAP 85, Location 20. When WAP 85 is absent the score decreases by one, when the score reaches 0 (zero), the value of pending is set to one indicating that this WAP is not in the fingerprint.

Although the technique presented by Gallagher in [GLDR10] does not process data from WAPs with RSSI values below -80 dB, meaning that WAPs in the survey-observation with a RSSI lower than -80 dB are ignored and are not considered for fingerprinting.

In this thesis, the -80 dB restriction was removed, thus in the implementation of the Gallagher Fingerprint every single one of the measurements is employed on the computation of the fingerprint no matter how low is the signal strength. This allows an equitable comparison of fingerprinting techniques.

Regarding the estimation of expected signal strength the Gallagher Fingerprint only considered the last observation. This is equivalent to give all the weight to the last observation, ignoring historical observations (similarly to Basic Fingerprint).

6.1.5 Jun-Sung Fingerprint

The fingerprinting technique presented by Jun-Sung et al. in the work [LJYH13], proposes a improvement on computing the expected signal strength, and includes extra parameters per WAP within the fingerprint. Jun-Sung Fingerprint is also a tuple of a location and a set of WAPs. A formal Jun-Sung Fingerprint definition is the following:

$$JS_{FP} = [l, \{WAP_{1,\ldots,N}\}] \tag{6.17}$$

$$WAP_n = [BSSID_n, \hat{x}_n, score_n, base_n, freq_{in_n}, freq_{out\,n}, \delta_n, pending_n] \tag{6.18}$$

It differs from previous definitions on the parameters stored per WAP. In addition to BSSID and $\hat{x}$, the following parameters are included: score, base, $freq_{in}$,$freq_{out}$, δ, $pending$.

- **Estimating Expected Signal Strength ($\hat{x}$).** In order to estimate the value $\hat{x}$ the following procedure is carried out for every WAP within the incoming survey-observation. The estimated $\hat{x}$ is initialised with the value of the first measurement RSSI from the corresponding WAP. The formal representation of this initialisation is the following,

$$\hat{x}_i = x_i = RSSI, \quad \text{for } i = 1 \tag{6.19}$$

Index i indicates the current state/iteration of the measurement x_i and the current expected value $\hat{x}_i$. For following measurements, the value of the $\hat{x}_i$ is computed with Equation 6.20.

$$\hat{x}_i = \frac{\hat{x}_{i-1} + x_i}{2} \tag{6.20}$$

Where x_i is considered the current RSSI per WAP. The $\hat{x}_i$ is the recently estimated signal strength and $\hat{x}_{i-1}$ is the previous one. The previous estimated value is retrieved from the database, and it is overwritten with the new estimated $\hat{x}_i$. When there is no $\hat{x}_{i-1}$ stored in the database, the WAP is new in the fingerprint. Figure 6.8 presents the raw data and the $\hat{x}$ computed with Jung-Sung's algorithm.

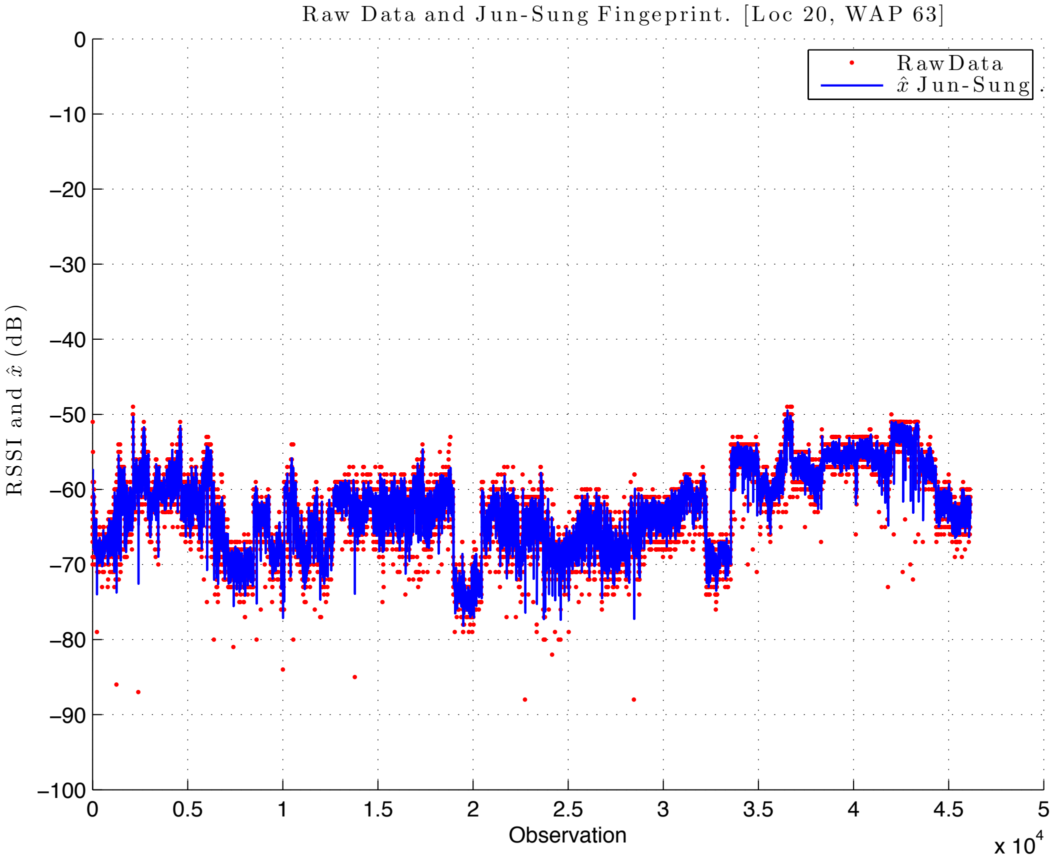


Figure 6.8: Jun-Sung's expected signal strength x_i and raw data from WAPs 63, Location 20.

- **Score.** The *score* is computed by an exponential function. Only a few observations are required to increase the *score* into a acceptance window. On the contrary, a large number of observations, where the WAP is absent, are required in order to remove the WAP from the fingerprint. The *score* increasing function is computed as Equation 6.21, and the *score* decreasing function is defined by

Equation 6.22.

$$score = base + (e^{freq_{in}++} - 1) \tag{6.21}$$

$$score = base - (e^{\frac{freq_{out}++}{\delta}} - 1) \tag{6.22}$$

Where the subtraction of 1 (one) in both Equations 6.21 and 6.22 is done to set the initial score value to 0 (zero).

- **Base (*base*).** Previous value of the *score*, the *base* is added to the exponential function for the situations when the WAP is present. When WAP is absent the corresponding exponential function is subtracted from *base*. The new value for *score* is the result of these computations, as shown in Equations 6.21 and 6.21

- **Frequency In ($freq_{in}$).** The parameter $freq_{in}$ is the number of consecutive observations that a WAP has been detected, and it is reset to 0 (zero) when the WAP is not detected

- **Frequency Out ($freq_{out}$).** Value $freq_{out}$ is the number of the consecutive observations of the WAP not being detected, and it is reset to zero once the WAP is detected again

- **Pending (*pending*).** The value *pending* is a flag that determines when the WAP is used in the fingerprint, pending is set to 0 (zero) when the WAP score is within a survival window, and 1 (one) when outside the same window

- **Acquisition Probability** (δ) **.** Value δ is the acquisition probability of a WAP to be detected, the document [LJYH13] (where the Jun-Sung Fingerprint is proposed) does not specify how this acquisition probability is computed. After an email request, the following information was provided directly by the authors of [LJYH13]:

  "The signal acquisition probability of an [W]AP varies depending on its RSS[I]. The lower the signal strength of an [W]AP is, the lower its acquisition probability is. This means that [W]APs with low signal strengths are not always captured when we collect fingerprints many times at the same location. We confirmed that from about 2,200 [W]AP signals (averaged from 20 fingerprints at each location) collected from a large-scale shopping mall (COEX, Seoul, Korea, $189,000m^2$*)."*

  They also provided a graph with their results of signal acquisition probability with different RSSI. This is shown in Figure 6.9

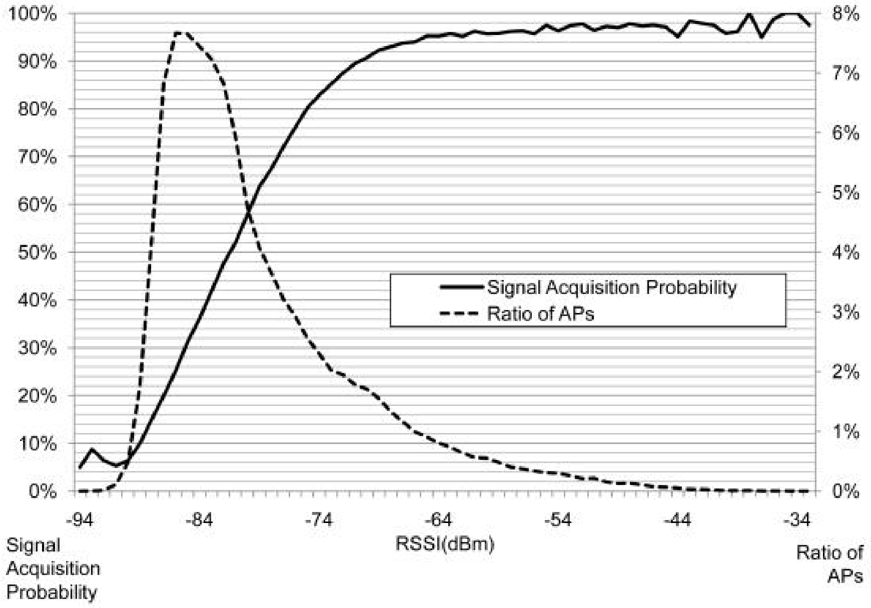


Figure 6.9: Acqusition Probability function and Ratio of WAPs. Where the Ratio of WAPs is the percentage of the presence count per RSSI. This image was provided by Jun-Sung Lim

The process of programming Jun-Sung Fingerprinting algorithm involved performing a mechanism to implement an Acquisition Probability based on information provided by Figure 6.9.

First, Jun-Sung's Acquisition Probability function from Figure 6.9 was sampled. The vector of sampled data points was then fitted using Matlab's function *plolyfit*. A 6th degree polynomial presented a fit approximation to sampled data points. The approximating 6th degree polynomial is presented with the respective coefficients in Equation 6.23.

$$\begin{aligned} f(x) = {} & 0.0292x^6 - 0.0611x^5 - 0.0658x^4 \\ & + 0.2461x^3 - 0.1976x^2 + 0.0661x + 0.9617 \end{aligned} \tag{6.23}$$

Jun-Sung's Signal Acquisition Probability, was fitted appropriately by the polynomial within the range (-93,-34). Figure 6.10 presents the sampled data from Jun-Sung's Acquisition Probability, the fitted polynomial, and the limits showing the range where the fitted polynomial was useful. The fitted polynomial was used for computing the probability for RSSI values between -93 to -34 (dBm).

Constant values were used for RSSI outside that range.

$$\delta = \begin{cases} 0.0379 & \text{for } x_i <= -93 \\ f(x_i) & \text{for } -93 < x_i < -34 \\ 0.9986 & \text{for } x_i >= -34 \end{cases} \tag{6.24}$$

The probability 0.0379 corresponding to -93 dB was used for RSSI equal to -93 dB and below. Probability 0.9986 corresponding to -34 dB was used for RSSI equal to -34 dB and over. From this, the Empirical Signal Acquisition Probability function is presented in Equation 6.24, and illustrated in Figure 6.10 by a blue solid line.

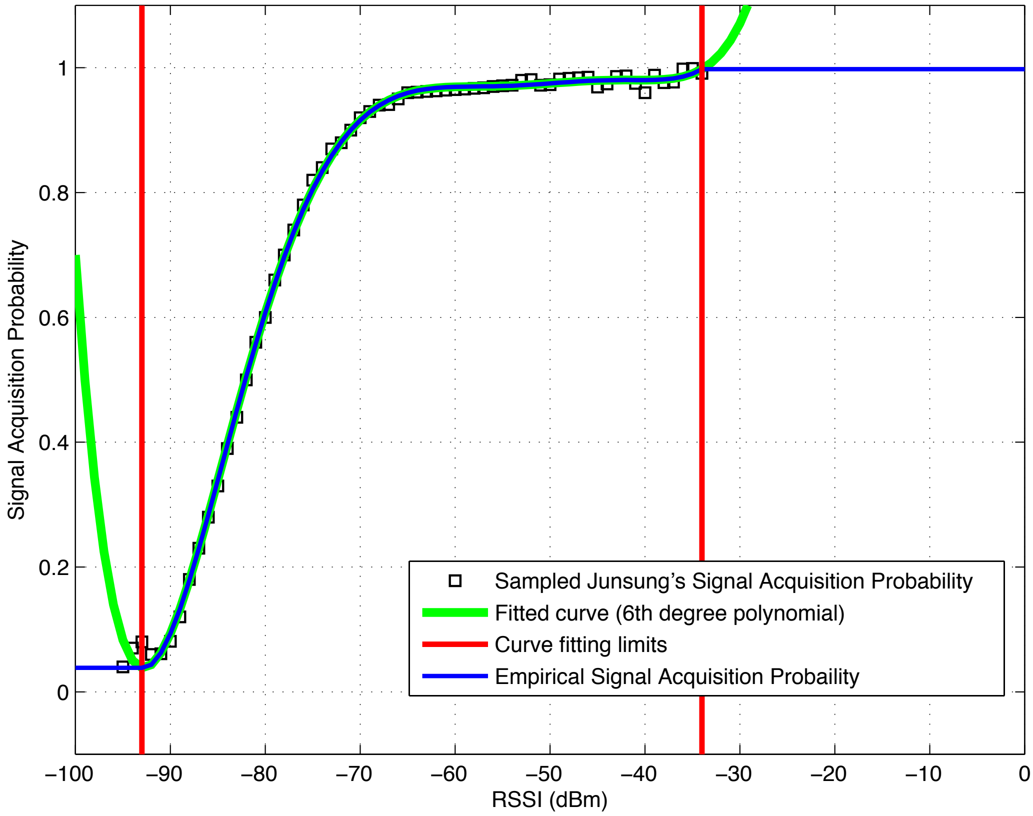


Figure 6.10: Empirical Signal Acqusition Probability.

Jun-Sung Fingerprinting technique was implemented for comparison. Algorithm 2 illustrates the pseudo code of this implementation. Figure 6.11 presents the Jun-Sung Fingerprint values for WAP 63 at Location 20 (Office 1.035, RCS). Figure 6.12 illustrates how the parameters for the Jun-Sung Fingerprint change for a less stable WAP.

Algorithm 2 Jun-Sung Fingerprinting

```
$soWAPs = \{$WAP | WAP in survey-observation, where $WAP = (WAP\_ID, RSSI)\}$
$fpWAPs = \{$WAP | WAP in fingerprint, where $WAP = (WAP\_ID$, $\hat{x}$ , score,base,
  $freq_{in}$,$freq_{out}$, $\delta$ , pending$)\}$
for all $WAP$ such that $WAP \in soWAPs \cup fpWAPs$ do
    if $WAP \in soWAP$ and $WAP \in fpWAP$ then                  ▷ WAP Update
        if $freq_{in} = 0$ then
            $freq_{in} \leftarrow 1$, $freq_{out} \leftarrow 0$, $base \leftarrow score$
        $\hat{x} \leftarrow (\hat{x} + RSSI)/2$
        $\delta \leftarrow 1/$acquisition probability$(\hat{x}) \times 100$
        $score \leftarrow base + (e^{freq_{in}++} - 1)$
        if $pending = true$ and $score \geq min$ then
            $pending \leftarrow false$
        else
            if $score > max$ then
                $score \leftarrow max$
    else
        if $WAP \in soWAPs$ and $WAP \notin fpWAPs$ then          ▷ WAP Addition
            $base \leftarrow 0$, $pending \leftarrow true$ , $freq_{in} \leftarrow 1$ ,$freq_{out} \leftarrow 0$
            $\hat{x} \leftarrow RSSI$
            $\delta \leftarrow 1/$acquisition probability$(\hat{x}) \times 100$
            $score \leftarrow base + (e^{freq_{in}++} - 1)$
            $fpWAPs \leftarrow fpWAPs \cup \{WAP\}$
        else
            if $WAP \notin soWAPs$ and $WAP \in fpWAPs$ then      ▷ WAP Removal
                if $freq_{out} = 0$ then
                    $freq_{out} \leftarrow 1$, $freq_{in} \leftarrow 0$, $base \leftarrow score$
                $score \leftarrow base - (e^{freq_{out}++/\delta} - 1)$
                if $pending = false$ and $score < min$ then
                    $pending \leftarrow true$
                else
                    if $score \leq 0$ then
                        $fpWAPs \leftarrow soWAP - \{WAP\}$
```

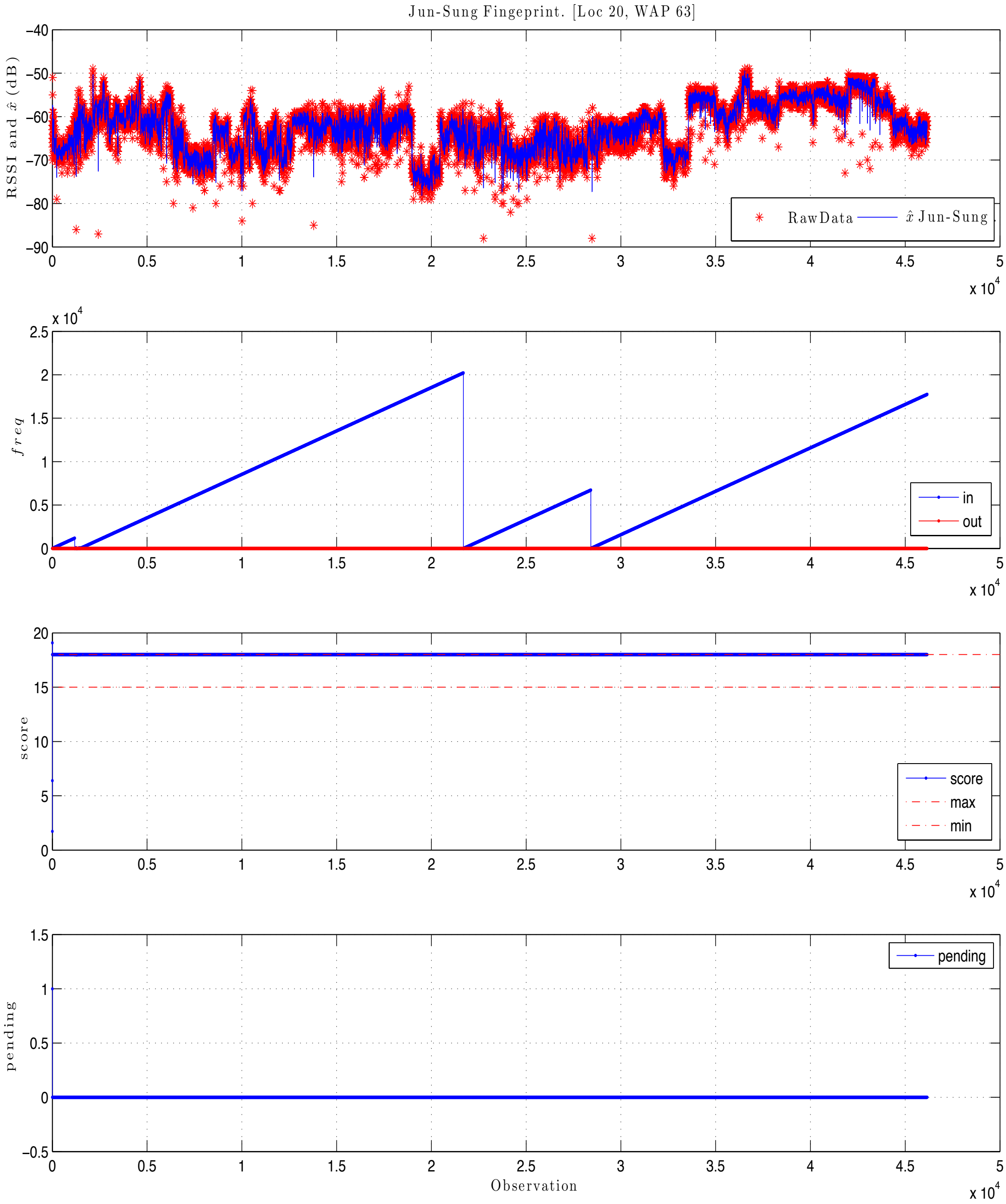


Figure 6.11: Jun-Sung Fingerprint technique applied to the RSSI data from WAPs 63, Location 20.

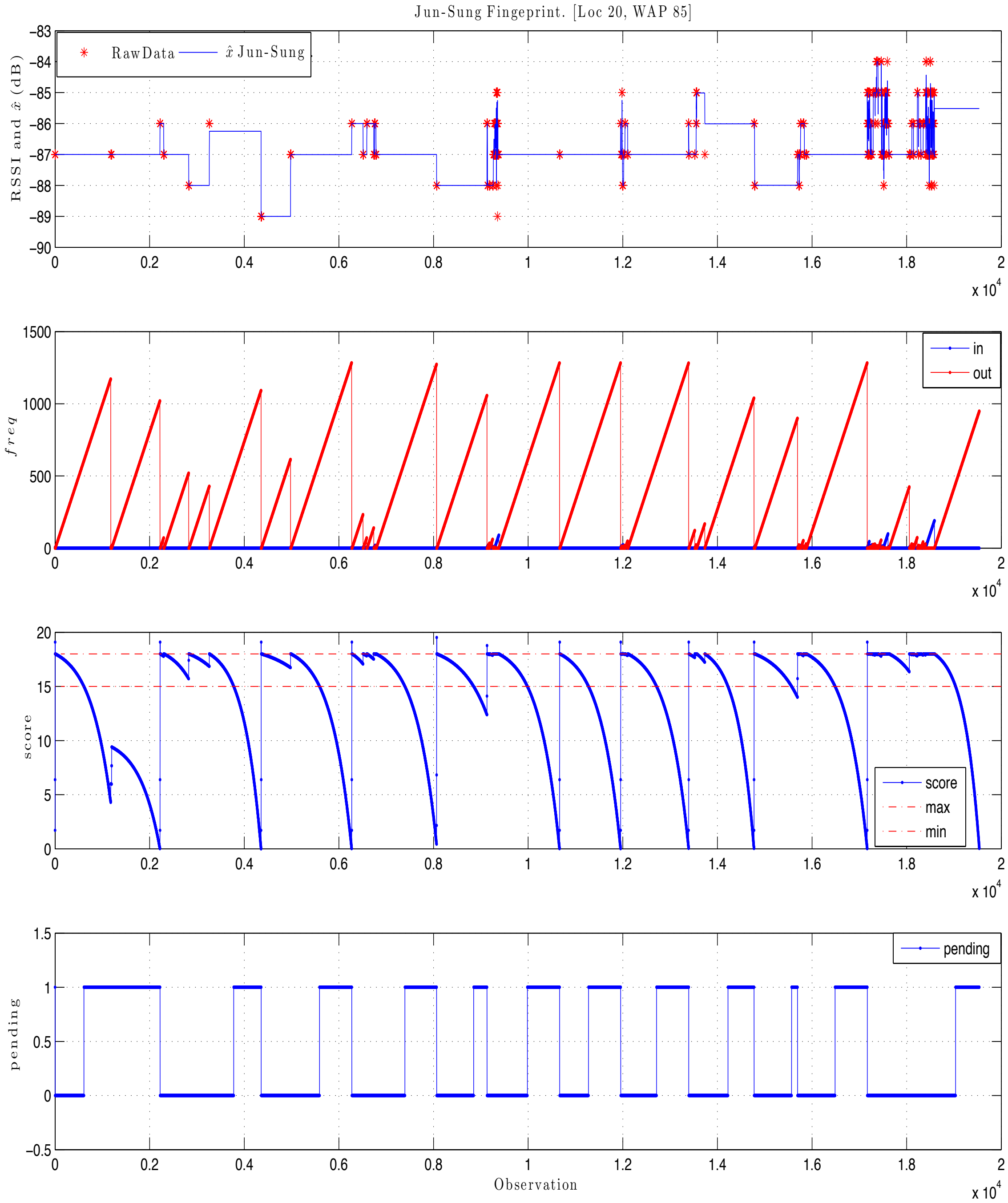


Figure 6.12: Jun-Sung Fingerprint technique applied to the RSSI data from WAPs 85, Location 20.

A disadvantage presented in Jun-Sung Fingerprinting algorithm is that WAPs stayed within the window for a large period of time. The estimation of the expected value in Jun-Sung Fingerprint is equivalent to a weighted average where all past observations are given half of the weight, and the latest observation is given the other half of the weight. In other words, the expected value computed with Jun-Sung algorithm is equivalent to a EWMA (see Equation 6.25) with a fixed value of 0.5 for the weight. A comparison between these techniques is presented in Chapter 7.

6.1.6 EWMA Fingerprint. An Improved Fingerprint Definition

This thesis introduces the EWMA Fingerprint definition designed by the author. EWMA stands for Exponentially Weighted Moving Average. The EWMA Fingerprinting was designed to address issues discussed previously by proposing the following improvements:

- An estimation of the expected signal strength value $\hat{x}$ based on EWMA, with weights computed per WAP that minimises error.
- WAP filtering based on a WAP ranking
- WAP filtering based on detection of changes and outliers in WAP's signal strength.

EWMA fingerprinting is a data driven approach proposing a WAP fitted computation of the estimated value, based on an analysis of WAP's signal profile.

The principles of EWMA (see Chapter 2.5) algorithm were used in EWMA Fingerprint for the following:

- Computing the estimated value $\hat{x}$

$$\hat{x}_i = (1-\lambda_x)\hat{x}_{i-1} + (\lambda_x)x_i \tag{6.25}$$

- Ranking WAPs

$$\hat{p}_j = (1-\lambda_p)\hat{p}_{j-1} + \lambda_p(p_j) \tag{6.26}$$

- Detecting changes and outliers in the WAPs' signal

$$\bar{\mu}_i = (1-\lambda_{c1})\mu_{\bar{i}-1} + (\lambda_{c1})\bar{\mu}_i \tag{6.27}$$

$$z_i = (1-\lambda_{c2})z_{i-1} + (\lambda_{c2})\bar{\mu}_i \tag{6.28}$$

A formal definition of the EWMA Fingerprint is the following:

$$EWMA_{FP} = [l, t_c, t_u, \{WAP_{1,\dots,N}\}] \tag{6.29}$$

where,

- l is the location
- t_c is the creation time, this parameter stores the timestamp when the fingerprint was first created and stored
- t_u is the last update time, every time a fingerprint is updated this parameter is also set with the timestamp from the latest observation used for the update
- $\{WAP_{1,\dots,N}\}$ is the set of Wireless Access Points

Within the set of WAPs, each WAP holds its own parameters. A formal definition of the parameters stored per WAP in a EWMA Fingerprint is given by

$$WAP_n = [WAP\_ID_n, \hat{x}_n, s_n, p_n, \hat{p}_n, a_n, uc_n, pending_n] \tag{6.30}$$

where,

- $\hat{x}$ is the expected signal strength value computed with Equation 6.31

$$\hat{x}_i = (1 - \lambda_x)\hat{x}_{i-1} + (\lambda_x)x_i \tag{6.31}$$

- s is the score, a parameter used for ranking WAPs, it is considered as the level of trustworthiness given to the WAP. The score range is (0-2) from which 2 is considered the highest. The WAP with highest score within a fingerprint is considered first in the positioning stage.
- uc is a binary parameter for identifying when $\hat{x}$ value of the WAP signal strength is within the upper and lower boundaries established by the control charts. The name uc stands for Under Control
- p is the WAP presence. It is the number of consecutive observations that a WAP has been detected, and it is reset to 0 (zero) when the WAP is not detected.
- a is the WAP absence. The value of a indicates the number of consecutive observations with the WAP absent (not being detected). Once the WAP is detected again the value of a is reset to 0 (zero).

- *pending* is the parameter that determines when the WAP is included in the fingerprint, *pending* is set to 0 (zero) when the WAP is in use within the fingerprint, and 1 (one) when not used.

The fingerprinting process using an EWMA Fingerprint definition can be divided into four components:

- **State estimate**. The estimated average for the expected signal strength value for a future observation. It is a prediction computed per WAP

- **Metric**. The metric is a value per WAP computed based on the estimated average and its variation. It is used as an initial ranking mechanism

- **Score**. The score is computed considering the metric and also considering WAP's occurrence. It can be said that the score is the metric affected by the presence/appearance or absence/disappearance of the WAP. The value of the score increases when the WAP is detected until it reaches its maximum. The maximum value per WAP corresponds to the value of its metric. The score decreases when the WAP is absent, until it reaches 0 (zero) then the WAP is removed from the EWMA Fingerprint

- **Control Chart**. The last component functionality is to detect changes and outliers in WAP's signal strength. This equips the system with the functionality of removing from the fingerprint those WAPs that are presenting unwanted or erratic behaviour. Leading to the mitigation of the positioning error by not including them until their profile is acceptable or once the new behaviour is considered as a permanent alteration

Figure 6.13 presents a block diagram of these components. The components within EWMA fingerprinting are depicted as blocks. The inputs and outputs are labeled with their respective parameters. A detailed explanation of these components in covered in the subsequent sections.

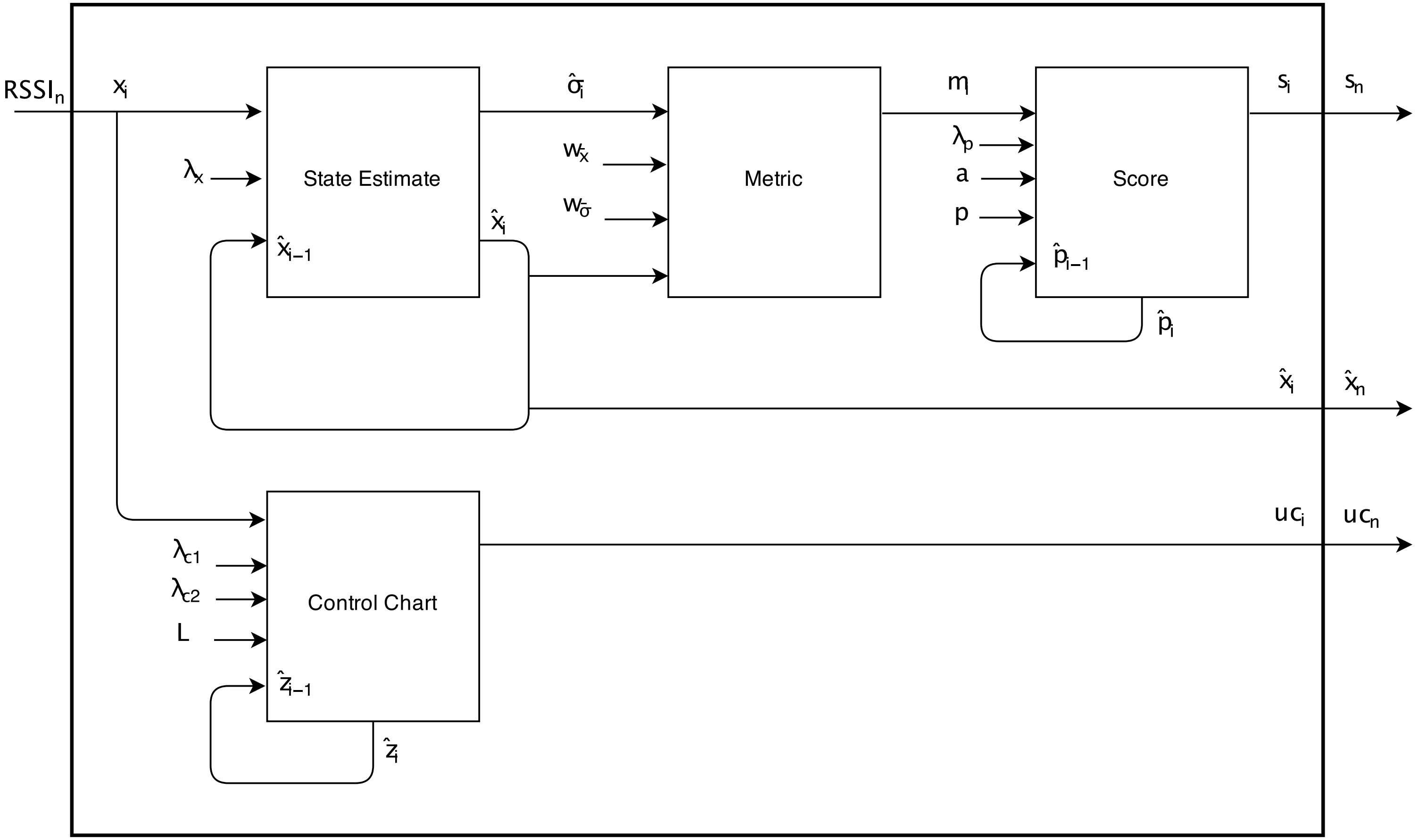


Figure 6.13: EWMA fingerprinting block diagram.
The processes and variables used within EWMA fingerprinting and their relation is depicted in this figure.This process is computed for every WAP in the incoming survey observation.

State Estimation

From previous techniques, the computation of the expected value $\hat{x}$ using only the last observation (Basic Fingerprint and Gallagher Fingerprint) fails in considering the valuable information provided by past observations. On the contrary, the expected value computed using a simple average (Default Fingerprint) gives the same weight to every single observation, this can present some disadvantages such as failing to provide an accurate expected average in situations when the signal is volatile or when it changes. Moreover, when considering the running average approach of computing the expected value $\hat{x}$, as performed by [LJYH13] (Jun-Sung Fingerprint), it can immediately be seen that this estimate value will be affected by outliers, step changes and ramp changes to the signal strength profile. Therefore we can recognise that using the running average is just one way of computing the expected signal strength in the presence of noisy measurements.

Although the aforementioned techniques could be suitable for particular signal's profiles, it is difficult to consider that a single estimation technique would be suitable for every WAP detected. Analysis from data collected in real-world scenarios showed that signals propagated by WAPs present distinct behaviours, hence they should be processed differently. The EWMA Fingerprint proposes a data driven computation of $\hat{x}$ per each WAP. Resulting in a specific lambda per WAP.

The formula for estimating the expected average employed by EWMA Fingerprint is the following:

$$\hat{x}_i = (\lambda_x)x_i + (1-\lambda_x)\hat{x}_{i-1}, \quad 0 < \lambda_x \leq 1 \tag{6.32}$$

where, the expected average $\hat{x}_i$ is computed as a recursive method that employs the latest observation x_i and the previous estimated average $\hat{x}_{i-1}$. The value of λ_x is the weight given to the current observation x_i, and λ_x's complement $(1-\lambda_x)$ is the weight given to the previous expected average.

The standard deviation is computed by

$$\hat{\sigma}_i = (\lambda_x)|\hat{x}_i - x_i| + (1-\lambda_x)\hat{\sigma}_{i-1} \tag{6.33}$$

Equations 6.32 and 6.33 are the basis for the State Estimate component. Following is explained the data driven approach used to obtain an appropriate λ_x per WAP.

Algorithm 3 presents the Matlab script implemented for computing the EWMA state estimate.

Algorithm 3 EWMA State Estimate

```
function [x_hatNew,s_hatNew] = IPSv2_fnEWMAse(RSSI,x_hatOld,
    s_hatOld,LambdaX)
     DefaultStdDev = 0.001;
     x(2) = RSSI;
     x_hat = x_hatOld;
     s_hat = s_hatOld;

     if (isempty(s_hat))
             s_hat = DefaultStdDev;
     elseif (s_hat == 0)
             s_hat = DefaultStdDev;
     end

     i=2;

     %Calculate the new expected value
     x_hat(i) = (x_hat(i-1)*(1-LambdaX)) + (x(i)*LambdaX);

     %Calculate the new standard deviation
     s_hat(i) = (s_hat(i-1)*(1-LambdaX)) + (abs((x_hat(i) - x(i)))*
         LambdaX);

     x_hatNew = x_hat(2);
     s_hatNew = s_hat(2);
end
```

Identifying a suitable λ_x per WAP The computation of an adaptive λ_x value per WAP is one of the key features that makes EWMA Fingerprint distinguishable from fingerprints previously mentioned.

The EWMA Fingerprint performance depends on the appropriate selection of the EWMA parameters. It is hypothesised that the optimal value for λ_x can be tuned according to the characteristics presented in the WAP's signal profile. For this, EWMA was applied to the WAP's RSSI, using linearly spaced values for λ_x in the range (0,1). The sum of the errors per λ_x was computed, and the λ_x generating the minimum error was identified.

$$\varepsilon_\lambda = \sum_{k_0 < i < n} |\hat{x}_{i-1} - x_i| \tag{6.34}$$

Equation 6.34 is the sum of errors per WAP per location. The error is the absolute difference between the estimated value from the previous iteration and the actual observed value in the current iteration. The value of λ_x which minimises the sum of errors, is selected as suitable for computing the expected value for the WAP.

The algorithm 4 illustrates how values for λ_x were computed per each WAP.

Algorithm 4 Computing Lambda

```
Identify WAPs per location observed longer than starting point $k_0$
Establish the number of $\lambda_x$ values to be tested. (e.g. 100 values of $\lambda_x$ linearly spaced within the restriction $0 < \lambda_x \leq 1$)
Evaluate every potential $\lambda_x$ using all observations
Compute the sum of differences between expected and actual signal strength per $\lambda_x$ per WAP is computed by Equation 6.34
Select the $\lambda_x$ generating the smallest error
```

A value of λ_x is stored per WAP, this value can become obsolete when the signal profile on the WAP goes through changes. A potential solution is to analyse the signal strength in order to identify changes. Adjustments of λ_x can be automatised according to the nature of the changes detected in the signal. This detection of changes and outliers is discussed in the section titled "Control Chart".

It is clear that signals from WAPs are different, then not all of the WAPs detected in a location are "worth" the same. In other words, not all the WAPs can be trusted equally when carry out positioning. Hence a mechanism to rank WAPs by their "trustworthiness" is proposed. This mechanism is the next feature to be discussed within EWMA Fingerprint, and it is composed by parameters: metric and the score.

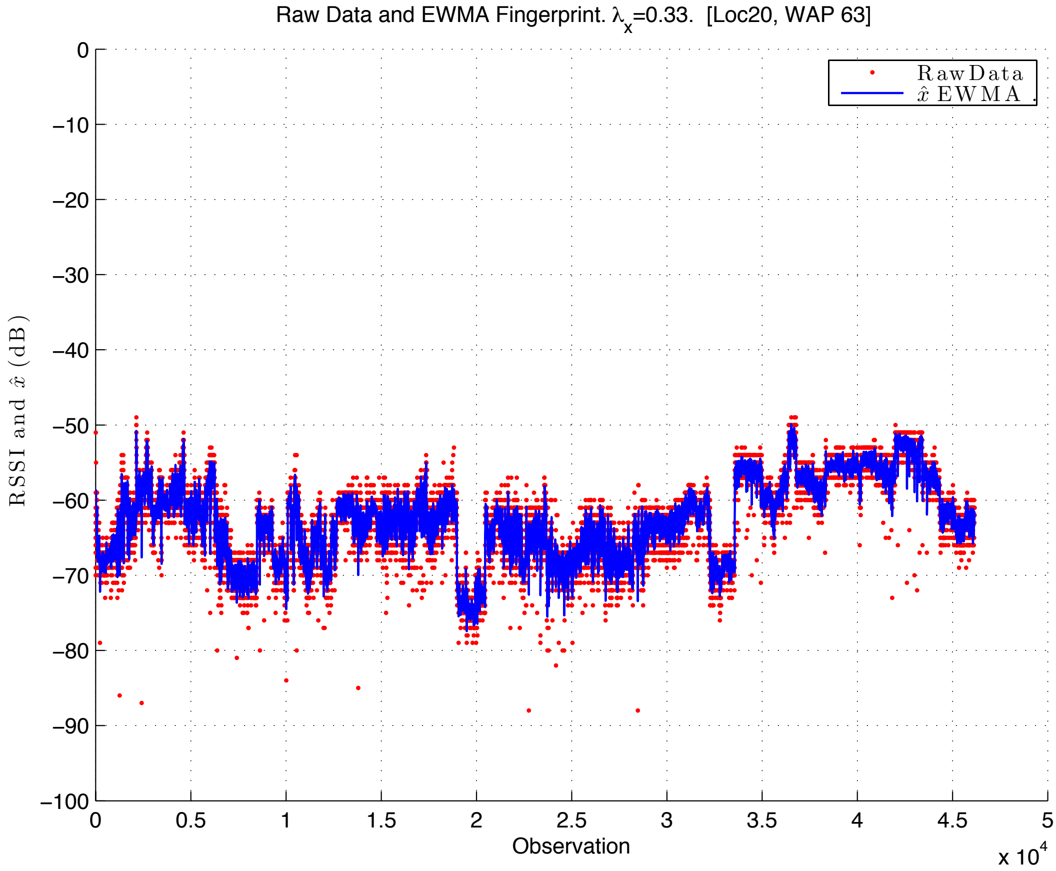


Figure 6.14: State Estimate. Raw data (red), and Expected Signal Strength $\hat{x}$ (blue) computed using EWMA fingerprinting and $\lambda_x = 0.327$. Observations from WAP 63, Location 20.

Metric

The *metric* is a measure based on a WAP's average and standard deviation. As shown in Chapter 5 Data Analysis, the number of WAPs detected for any particular location can be large, potentially increasing the database size and substantially increasing the computational time for positioning. In order to solve these problems, a WAP ranking process is proposed. This ranking process provides a score for WAP classification. Once classified, only trustworthy WAPs can be included in EWMA Fingerprints, instead of including every WAP. Ranking the WAPs is an important part of the algorithm in order to create a practical fingerprinting positioning system. The metric per WAP is given by

$$m_i = W_{\bar{x}}\bar{x}_i + W_{\bar{\sigma}}\bar{\sigma}_i \tag{6.35}$$

where

- $\bar{x}_i$ is the expected signal strength normalised. It is computed using Equation 6.36 from where $\hat{x}_i$ is the expected signal strength value (output from the state estimate process), $RSSI_{min}$ and $RSSI_{max}$ are the minimum and maximum empirical values registered for RSSI. Both values where retrieved from the dataset discussed in Chapter 5.

$$\bar{x}_i = \frac{\hat{x}_i - RSSI_{max}}{RSSI_{max} - RSSI_{min}} \tag{6.36}$$
$$RSSI_{min} = -100, \quad RSSI_{max} = 0$$

- $\bar{\sigma}_i$ is the standard deviation normalised. It is computed using Equation 6.37 from where $\hat{\sigma}_i$ is the standard deviation (output from the state estimate process), σ_{max} is the maximum standard deviation value identified from the dataset, its value corresponds to WAP 596 which presented the highest standard deviation in overall measurements.

$$\bar{\sigma}_i = \frac{\sigma_{max} - \hat{\sigma}_i}{\sigma_{max}}, \quad \sigma_{max} = 12 \tag{6.37}$$

- $W_{\bar{x}}$ is a weight given to $\bar{x}_i$, its value is a number in the range $(0 < W_{\bar{x}} < 1)$. The higher its value, the higher the weight (importance) is assigned to the average value

- $W_{\bar{\sigma}}$ is a weight given to $\bar{\sigma}_i$, its value is a number within the range $(0 < W_{\bar{\sigma}} < 1)$ with the restriction $W_{\bar{\sigma}} = 1 - W_{\bar{x}}$

The constant values $RSSI_{min}$, $RSSI_{max}$ and σ_{max} are fixed in the system settings, and are unique to the dataset analysed. The weights $W_{\bar{x}}$ and $W_{\bar{\sigma}}$ are constants defined in the system settings. Their values are established by the system administrator. They represent the level of importance given to the signal strength and to the standard deviation. For the generation of the results in this thesis, both weights are set to 0.5, this means that the same importance is given to the signal strength and to the standard deviation.

The metric is used along with the WAP persistence, for generating the WAP's score. The scoring technique is explained in detail in the following section. The code implemented to compute the metric is the following:

Algorithm 5 EWMA Metric

```
function [metric]= IPSv1_fnMetric(mu,s)
    [settings]=IPSv1_Settings();
    minSS = settings.minSS;
    maxSS = settings.maxSS;
    w1=settings.w1;
    w2=settings.w2;
    maxs=settings.maxs;
    if (s > maxs)
        s=maxs;
    end
    mu_norm = (mu-minSS) / (maxSS-minSS);
    s_norm = (maxs-s)/maxs;
    metric = w1*mu_norm + w2*s_norm ;
end
```

Score

The *score* *s*, computed for each WAP, quantifies the level of trustworthiness. WAPs within the fingerprint are ordered based on this parameter. The *score* provides an insight into how each WAP affects the positioning performance, and also determines when a WAP should be removed from the fingerprint. The score is computed as a function of the metric and the WAP's presence (detected or undetected WAP). The score has three states:

- **Increasing Score**. When a new WAP is detected, it is added into the fingerprint.

Its score is initialised according to Equation 6.38.

$$s_i = s_{i-1} + \frac{m_i - s_{i-1}}{k_0 - p} \tag{6.38}$$

- **Steady Score**. When the number of observations that a WAP has been detected is larger than a initialisation value (k_0) the WAP is considered stable. Hence the score is equal to the metric as shown in Equation 6.39

  $$s_i = m_i \tag{6.39}$$

  When the WAP has been detected for more than k_0 consecutive observations its score is at its maximum value, this maximum value is the metric.

- **Decreasing Score**. When the WAP is in the fingerprint and is absent (not detected) in an observation, the score decreases. Then it decreases with each consecutive observation that is missing the WAP. The number of observations that it takes to remove the WAP is dependent on the average number of observations the WAP was consecutively detected. It takes longer to remove (from the fingerprint) a WAP that was present for long time, than a WAP that was present on just few observations. When the WAP is absent, its score decreases according to Equation 6.40 until it reaches the value of 0 (zero), then the WAP is removed from the fingerprint.

  $$s_i = s_{i-1} - \frac{s_{i-1}}{\hat{p} - a} \tag{6.40}$$

  From Equation 6.40, $\hat{p}$ is the average of continuous observations where the WAP was detected. This average is given by Equation 6.41, which is an EWMA, with weight λ_p.

  $$\hat{p}_j = (1 - \lambda_p)\hat{p}_{j-1} + (\lambda_p)(p_j) \tag{6.41}$$

  The value for the weight is $\lambda_p = 0.5$, this value can be configurable in the system's settings. An evaluation for different values of λ_p is left for future work.

The three functions used in computing the score are summarised by,

$$s_i = \begin{cases} s_{i-1} + \frac{m_i - s_{i-1}}{k_0 - p} & \text{for } 0 < p \leq k_0 \\ m_i & \text{for } k_0 < p \\ s_{i-1} - \frac{s_{i-1}}{\hat{p} - a} & \text{for } 0 < a < \hat{p} \end{cases} \tag{6.42}$$

where k_0 is configured in the system settings, p is the number of times the WAP was detected consecutively before being undetected, $\hat{p}$ is an average (EWMA with $\lambda_p = 0.5$) of the previous consecutive times the WAP was seen. The WAPs are ranked into the fingerprint according to their *score*, the higher the *score* the more trustworthy the WAP.

The implemented function computing the score is the following:

Algorithm 6 EWMA Score

```
function [score]= IPSv1_fnScore(metric,old_score,fin,fout,xfin,sp)
    p = fin;
    a = fout;
    tsp = sp+1;
    hat_p = (ceil(xfin))+1;
    if (fin>0)
        if (fin<tsp)  %If fin is lower than the Starting Point
            %Increasing
            score = old_score + ( (metric - old_score)/(tsp-p) );
        else
            %Stedy
            score= metric;
        end
    else
        %fout(a) is larger than 0
        %Decreasing
        if (tfout>=txfin)
            score = old_score;
        else
            score = old_score - (old_score/(hat_p-a));
        end
    end
end
```

EWMA Control Chart

The third feature that makes EWMA Fingerprint unique is the implementation of an EWMA control chart. In this thesis the author proposes for the first time a fingerprinting mechanism implementing a EWMA control chart that allows detecting outliers and changes in signal from WAPs. The EWMA control chart is a variation of a standard control chart (see Section 2.5.2 for a description on EWMA control chart). Control charts detect large shifts in the mean value of the process. However, the signal strength

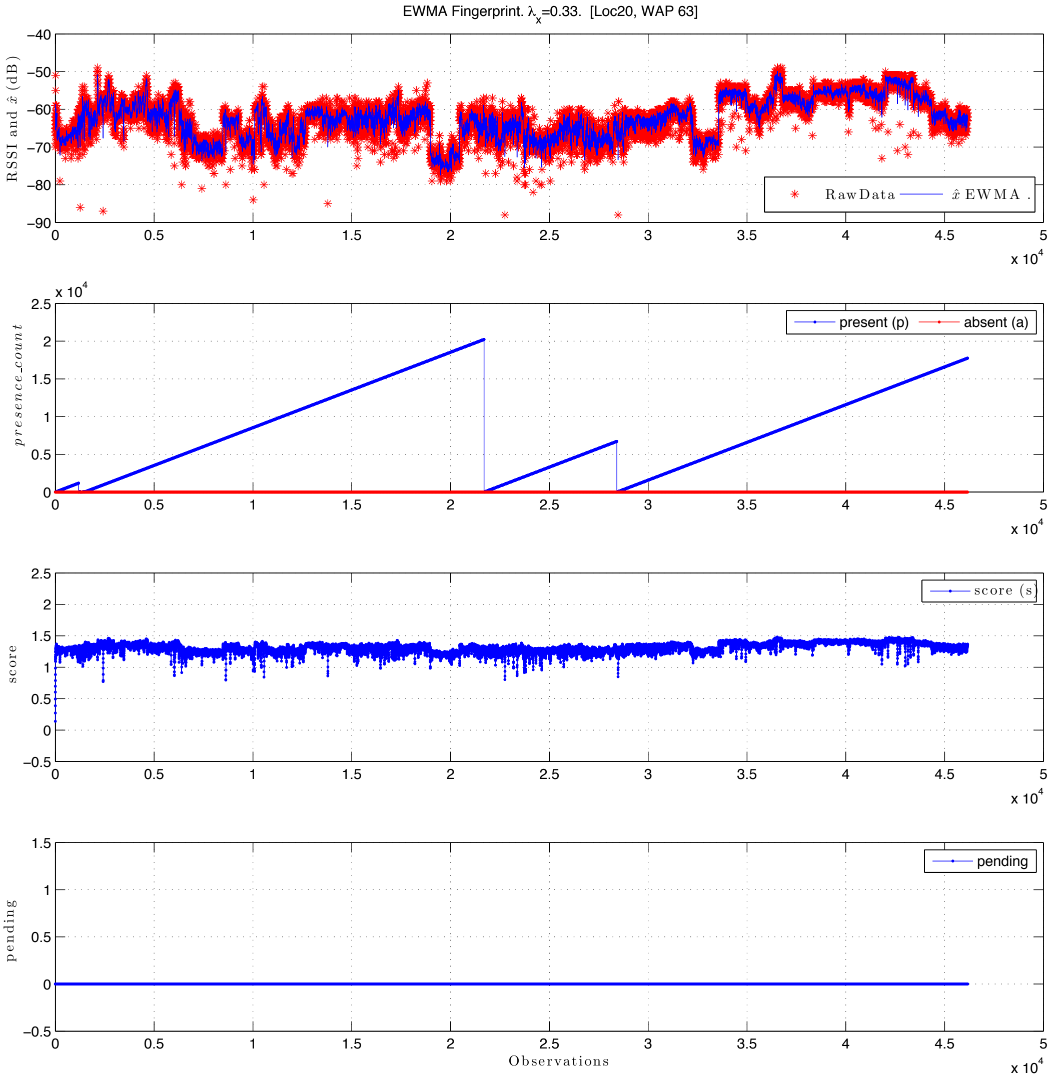


Figure 6.15: EWMA Fingerprint from WAP 63, Location 20.

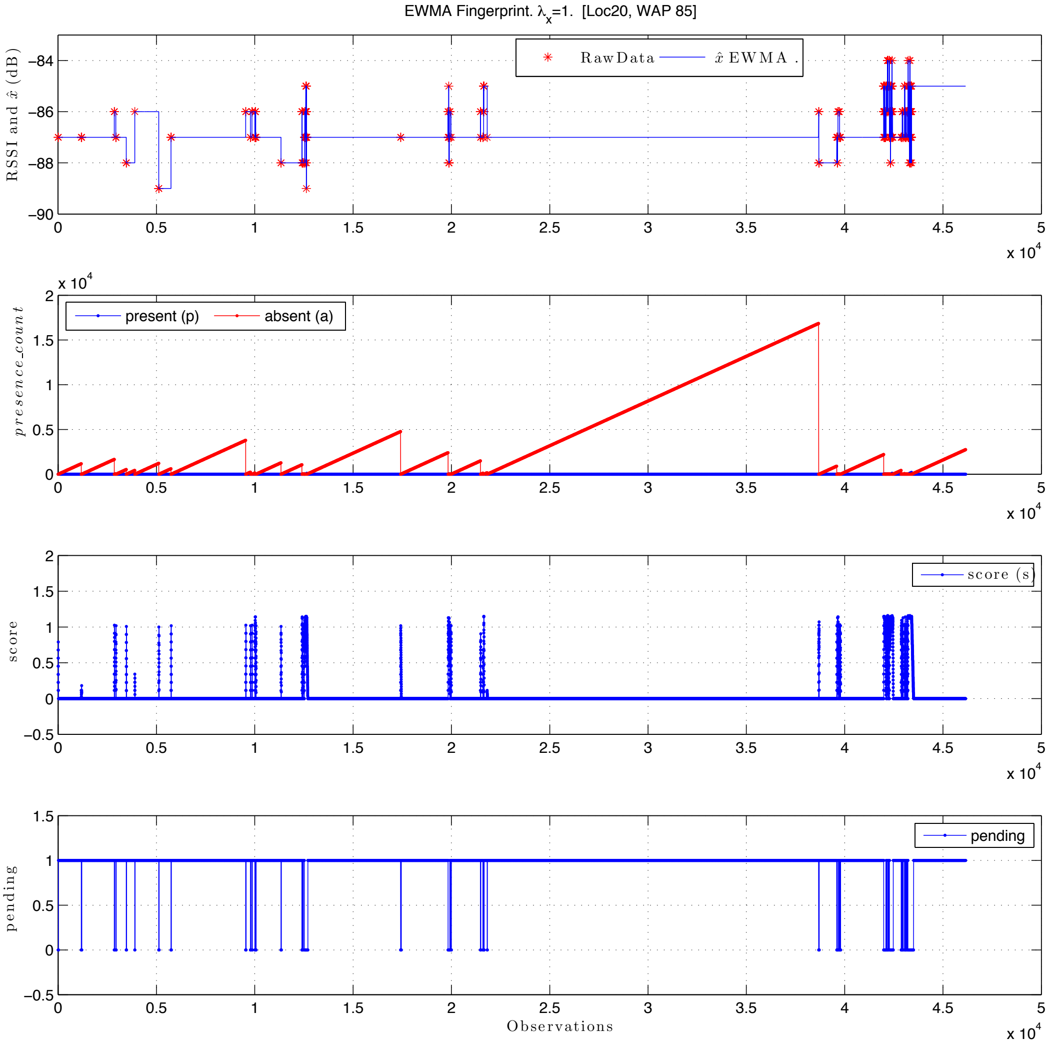


Figure 6.16: EWMA Fingerprint from WAP 85, Location 20

measurements recorded shows a variety of changes to the profile of the signal. To counter this, the Exponentially Weighted Moving Average (EWMA) control chart is used instead.

The detection of outliers, and drifts in the signals is a powerful mechanism that has not been explored for fingerprinting before. The use of control chart within the EWMA Fingerprint generates two significant advantages:

- WAP removal. Identify when a WAP's signal strength is out of control, this is being used to support the removal of WAP from the fingerprint for those WAPs presenting unwanted behaviour.
- WAP reincorporation. The control chart monitors the signal and it is used to reincorporate a WAP into the fingerprint in the situation for when the signal is stable.

Also, the control chart is proposed as a mechanism to determine when it is required to update the value of λ_x, which is used in the estimation of the expected signal strength. Input parameters for the EWMA control chart are the following:

- The current observation (x_i), is the RSSI
- A static λ_{c1}, set as a constant value pre-defined in the system's settings.
- A static λ_{c2}, weight establish on the system settings
- Recursive average computed with EWMA $\hat{z}_{i-1}$
- Multiple of standard deviation (L)

The EWMA control chart was implemented according to the following process:

1. Smoothing the raw data. The control chart is computed when the number of data points (i) are larger than k_0.

The EWMA algorithm is designed to recursively smooth data around the current estimate of the process mean; Equation 6.43 shows the EWMA calculation, where the EWMA response can be affected by changing λ_{c1}. Then, by using equations 6.43 and 6.44 to perform the normalisation, then it can be controlled to what degree changes in the data are retained post normalisation. Therefore in our change point detection we follow this process to normalise the raw data.

$$\mu_i = (\lambda_{c1})x_i + (1-\lambda_{c1})\mu_{i-1} \tag{6.43}$$

$$\sigma_i = (\lambda_{c1})|\mu_i - x_i| + (1-\lambda_{c1})\sigma_{i-1} \tag{6.44}$$

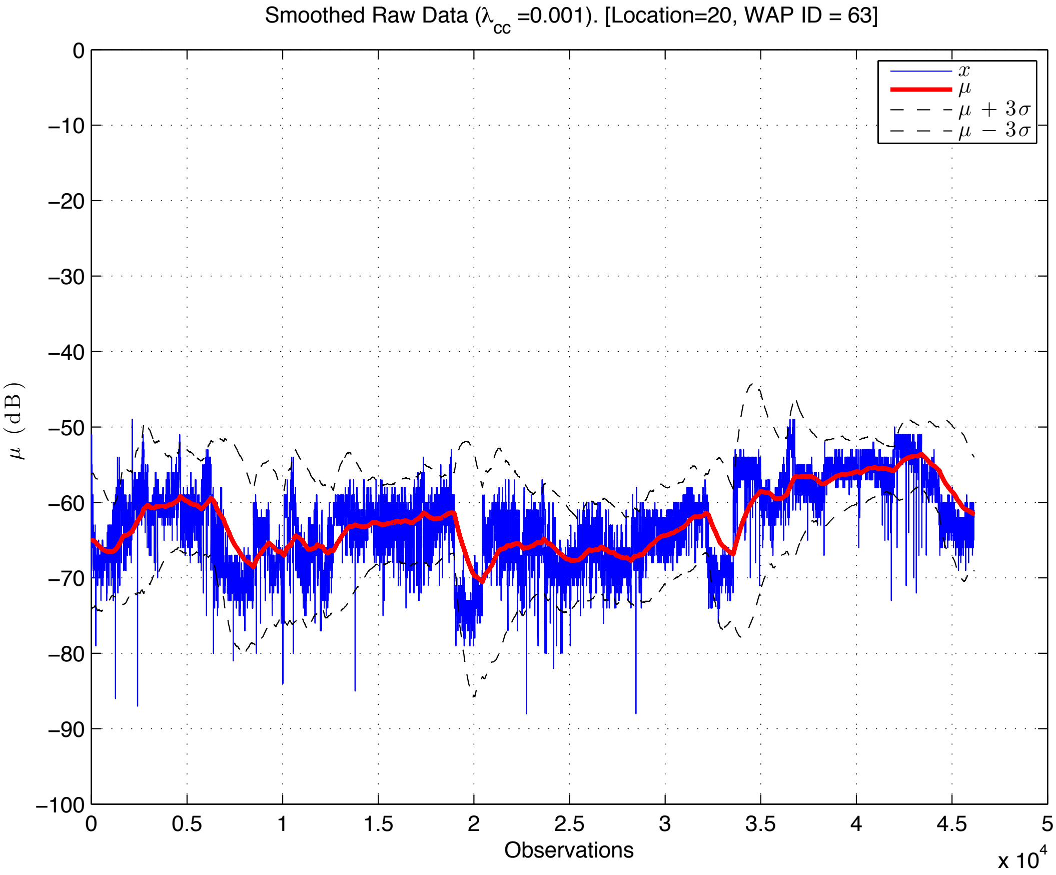


Figure 6.17: Smoothed raw data with $\lambda_{c1} = 0.001$ from WAP 63 at Location 20

2. Normalising smoothed data. From the body of work carried out on the EWMA control chart there are two main ways to apply it, either on the raw data or on a normalised version of the raw data. Essentially by applying normalisation it is possible to use the same settings for the EWMA control chart for data that originates from different distributions.

In general, the normalisation process is carried out by Equation 6.45; however there are a number of statistical methods to calculate the mean and standard deviation that could be used.

$$\bar{\mu}_i = \frac{x_i - \mu_i}{\sigma_i} \tag{6.45}$$

3. Smoothed normalised data using Weight (λ_{c2}) Once the data has been normalised a second EWMA with weight λ_{c2} is implemented. Equation 6.46 presents this EWMA.

$$z_i = (\lambda_{c2})\bar{\mu}_i + (1-\lambda_{c2})z_{i-1} \tag{6.46}$$

$$z_0 = \bar{\mu}_i \tag{6.47}$$

Where the initial value for z_{i-1} is z_0, the normalised mean $\bar{\mu}_i$ computed in the previous step by Equation6.45.

4. Calculate Control Limits The upper and lower limits are calculated using Equation 6.48, where L is the number of standard deviations.

$$CL = \pm L\sqrt{\frac{\lambda_{c2}}{2-\lambda_{c2}}} \tag{6.48}$$

5. Calculate the confidence The ratio between z_i and the control limit is found to allow us to set bounds between (-1,1) to detect when the change points occur.

$$C_i = \frac{z_i}{CL} \tag{6.49}$$

The output of the EWMA control chart is the parameter under-control (*uc*), this parameter indicates when the signal is within boundaries.

Selecting the parameters for the Control Chart

The values $\lambda_{c1} = 0.05$, $\lambda_{c1} = 0.10$ and $\lambda_{c1} = 0.20$ were suggested by [Mon08], with L=3 being the usual limit. It is claimed that particularly $\lambda_{c1} = 0.05$ and $L = 2.492$ performs very well against both normal and non-normal distributions. These values were used as a starting point for evaluating the EWMA control chart for RSSI-based fingerprinting. Results are presented in Chapter 7.

The implemented function for computing the control chart in Matlab is presented in Algorithm 7.

Figure 6.18, presents the parameters resulting from EWMA control chart with $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and L=6 for WAP 63, location Office 1.035, RCS, The University of Manchester. This parameters are the ones that showed suitable results on detecting outliers and changes for the three locations under analysis (their selection is based on results showed in Chapter 7).

Algorithm 7 EWMA Control Chart

```
function [EWMAnew,EWMANormNew,StdNew,ConfidenceNew] =
   IPSv2_fnEWMAcc(RSSI,EWMAold,STDold,EWMANormOld, FF,W,T)

    Lc1 = FF;
    Lc2 = W;
    L = T;

    rawData(2) =RSSI;
    %Initlize the algorithm
    EWMA = EWMAold;
    Std = STDold;

    if (isempty(Std))
        Std = 0.001;
    elseif (Std == 0)
         Std = 0.001;
    end

    i=2;

    %Step 1. Smooth the raw data and calcuate the standard
       deviation
    EWMA(i) = (EWMA(i-1)*(1-Lc1)) + (rawData(i)* Lc1);
    Std(i) = (Std(i-1)*(1-Lc1)) + (abs(rawData(i) - EWMA(i))*Lc1);

    %Step 2. Normalize the data
    if (Std(i) == 0)
            Std(i) = 0.001;
    end
    dataNorm(i) = (rawData(i) - EWMA(i)) / Std(i);


    %Step 3. Smooth the normalized data using the weight (Lc2)
    if isnan(EWMANormOld)
            EWMANorm = dataNorm(i);
    else
            EWMANorm = EWMANormOld;
    end
    EWMANorm(i) = EWMANorm(i-1)*(1-Lc2) + (dataNorm(i)*Lc2);

    %Step 4. Calculate the Control Limits
    CL = L * (sqrt(Lc2 / (2-Lc2)));

    %Step 5. Calculate the confidence
    Confidence(i) = EWMANorm(i)/CL;

    EWMAnew = EWMA(2);
    StdNew = Std(2);
    ConfidenceNew = Confidence(2);
    EWMANormNew = EWMANorm(2);

end
```

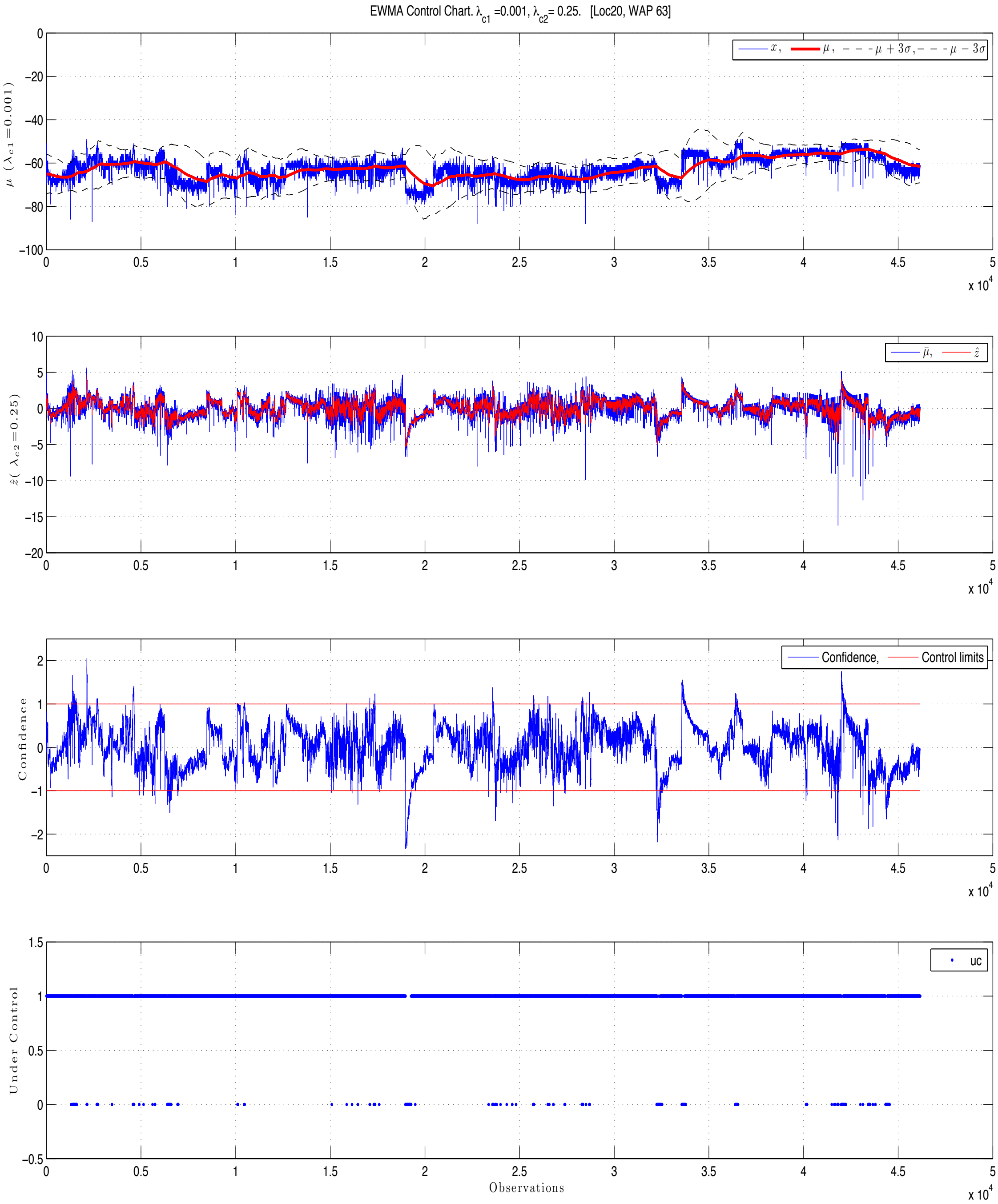


Figure 6.18: EWMA control chart with $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and L=6 for WAP 63, Location 20

Algorithm 8 EWMA fingerprinting

```
Require: $k_0$, $\lambda_x$, $\lambda_p$, $\lambda_{c1}$, $\lambda_{c2}$, $W_{\hat{x}}$, $W_{\hat{\sigma}}$, $dStdDev = 0.001$
$soWAPs = \{$WAP | WAP in survey-observation, where $WAP = (WAP\_ID, RSSI)\}$
$fpWAPs = \{$WAP | WAP in fingerprint, where $WAP = (WAP\_ID$, $\hat{x}$, s, $p$, $\hat{p}$, $a$, $uc$, $pending)\}$
for all $WAP$ such that $WAP \in soWAPs \cup fpWAPs$ do
    if $WAP \in soWAP$ and $WAP \in fpWAP$ then                                ▷ WAP Update
        $p++, a \leftarrow 0$
        if $p \leq k_o$ then
            $[\hat{x}, \hat{\sigma}, Q] \leftarrow function\_Default\_se(RSSI, \mu, Q, p)$
        else
            if $p > k_o$ then
                $[\hat{x}, \hat{\sigma}] \leftarrow function\_EWMAse(RSSI, \hat{x}, \hat{\sigma}, \lambda_x)$
                $[\hat{\mu}, \hat{\sigma}, z, uc] \leftarrow function\_EWMAcc(RSSI, \mu, \sigma, z, \lambda_{c1}, \lambda_{c2}, L)$
        $m \leftarrow function\_Metric(\hat{x}, \hat{\sigma})$
        $s \leftarrow function\_Score(m, s, p, a, \hat{p}, k_0)$
        $pending \leftarrow function\_Pending(uc, s)$
    else
        if $WAP \in soWAPs$ and $WAP \notin fpWAPs$ then
            $\hat{x} \leftarrow RSSI$, $Q \leftarrow 0$ , $\hat{\sigma} \leftarrow dStdDev$
            $p \leftarrow 1$, $a \leftarrow 0$, $\hat{p} \leftarrow 1$, $uc \leftarrow false$
            $m \leftarrow function\_Metric(\hat{x}, \hat{\sigma})$
            $s \leftarrow function\_Score(m, s, p, a, \hat{p}, k_0)$
            $pending \leftarrow function\_Pending(uc, s)$
            $fpWAPs \leftarrow fpWAPs \cup \{WAP\}$                             ▷ WAP Insertion
        else
            if $WAP \notin soWAPs$ and $WAP \in fpWAPs$ then                   ▷ WAP Absent
                if $a = 0$ then
                    $a \leftarrow 1$
                else
                    $a++$
                if $p > 0$ then
                    $\hat{p} \leftarrow function\_EWMA\_p(\hat{p}, p, \lambda_p)$
                    $p \leftarrow 0$
                $m \leftarrow function\_Metric(\hat{x}, \hat{\sigma})$
                $s \leftarrow function\_Score(m, s, p, a, \hat{p}, k_0)$
                $pending \leftarrow function\_Pending(uc, s)$
                if $pending = false$ and $s > 0$ and $uc = true$ then
                    $pending \leftarrow true$
                else
                    if $s \leq 0$ then
                        $fpWAPs \leftarrow soWAP - \{WAP\}$                    ▷ WAP Removal
```

6.2 Positioning. Location Estimation Algorithm

The data analysis and fingerprinting methods presented earlier have the ultimate purpose of improving positioning. Positioning refers to locating a person or asset by means of a WiFi enabled electronic and mobile device. This positioning can be carried out once an initial WiFi radio-map has been created. Input measurements from the user are compared against the radio-map, then after comparison, a particular or small set of fingerprints are selected as a possible best match.

This thesis aims to achieve a positioning resolution of room level accuracy (at say 2mx2m). The granularity is dependent on scale so in Piccadilly station granularity of a few metres would be for many application of little added value.

This section describes the methods for estimating the user's position based on the input measurements and EWMA Fingerprints. The Positioning process is listed in Table 6.2:

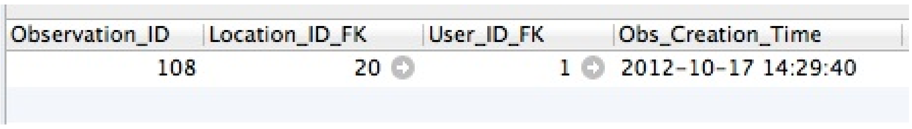
Positioning
1. Input Query-observation
2. Computing the best-Match fingerprint
3. Location Estimation Level 1
4. Location Estimation Level 2
5. Location Estimation Level 3
6. Display estimated location

Table 6.2: Positioning Process.

6.2.1 Query-Observation

An observation is called a "Query-Observation" when it is recorded by the user, from an "unfamiliar" position. The query-observations are recorded using the same script used for survey-observations (see script in Appendix B.1). The distinctive feature about query-observations is the fact that they are not tagged with a specific location but are tagged with the default location: "unknown".

An example of a query-observation (qo) from the database is shown in Figure 6.19 and its corresponding qoWAPs is depicted in Figure 6.20. The term fpWAPs is used to refer to the set of WAPs within a fingerprint (fp), and the term qoWAPs refers to the set of WAPs within a query-observation (qo). Query-Observations are compared against



Observation_ID	Location_ID_FK	User_ID_FK	Obs_Creation_Time
108	20	1	2012-10-17 14:29:40

Figure 6.19: The record of an observation in the DB

the fingerprints in the radio-map in order to find the best matching fingerprint. The best matched fingerprint can be considered coming from the same or a close location.

6.2.2 Location Estimation

The location estimation stage is the process of calculating the position from a query-observation. This is done by finding a fingerprint from the radio-map that has significantly close WiFi data. For this, the query-observation is compared against fingerprints in the DB. The “best fit” or “best matched” fingerprint according to specific criterion defined below is considered the location for the query-observation.

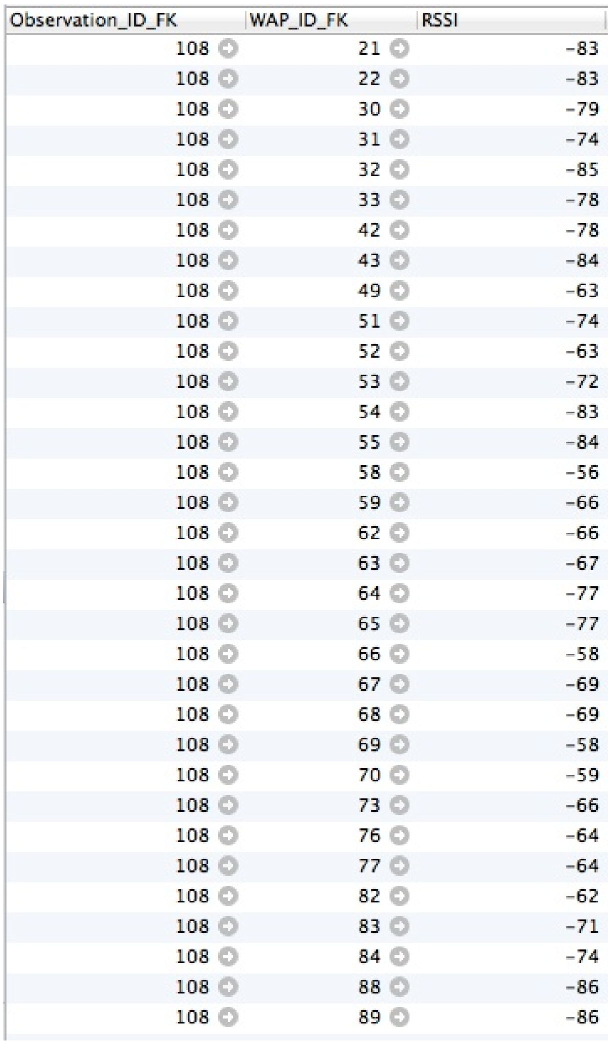
The following three levels of positioning were implemented for identifying the best matching fingerprint:

- Level 1: WAP-based Positioning
- Level 2: RSSI-based Positioning
- Level 3: Metric-based Positioning

6.2.3 Level 1: WAP-based Positioning

A basic approach to define a location when a query-observation is recorded is based on counting the number of coincident WAPs. This number gives us a first positioning level, and it is used for computing a rough estimation of where a query-observation was recorded. The methodology of this approach is as follows

- The classifier receives a query-observation.
- WAPs from the query-observation are compared with WAPs stored in the database.
- Fingerprints with matching WAPs are fetched from the database.
- The fingerprint with the highest number of matching WAPs is chosen as the best matching fingerprint.



Observation_ID_FK	WAP_ID_FK	RSSI
108	21	-83
108	22	-83
108	30	-79
108	31	-74
108	32	-85
108	33	-78
108	42	-78
108	43	-84
108	49	-63
108	51	-74
108	52	-63
108	53	-72
108	54	-83
108	55	-84
108	58	-56
108	59	-66
108	62	-66
108	63	-67
108	64	-77
108	65	-77
108	66	-58
108	67	-69
108	68	-69
108	69	-58
108	70	-59
108	73	-66
108	76	-64
108	77	-64
108	82	-62
108	83	-71
108	84	-74
108	88	-86
108	89	-86

Figure 6.20: List of WAPs and their respective RSSI from observation 108

This approach is easy to implement and can provide some level of accuracy but it suffers from one major drawback. Several fingerprints can result with the same number of matching WAPs, hence the final location can be ambiguous.

This first level of positioning can be suitable for particular applications where a rough positioning is good enough and it is not necessary to allocate more computational power to improve positioning. Nevertheless, there are situations when this approach is not good enough, and for these cases the second level of positioning has been developed.

6.2.4 Level 2: RSSI based Positioning

Positioning level 2 uses the result from positioning level 1, but rather than using only the WAPs presence it also uses the RSSI from the WAPs in the query-observation and the $\hat{x}$ from the WAPs in the coincident fingerprints.

The best fingerprint is selected by computing a distance metric, for example, the Root Mean Square Error (RMSE) between the WAPs in the query-observation (qo) and the WAPs from every fingerprint retrieved at positioning level 1. The fingerprint with the smallest error is considered as the most likely match. The RMSE (ε) is presented in Equation 6.50

$$\varepsilon = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \hat{x}_j)^2} \tag{6.50}$$

where $\hat{x}_j$ is the RSSI value for the jth WAP in the query-observation; $\hat{x_j}$ is the expected value for the jth WAP in the fingerprint; and n is the number of WAPs in the query-observation.

6.2.5 Level 3: Score based Positioning

Positioning level 3 is based on the score for selecting the WAPs to be included in the comparison. Only the top n ranked WAPs from the fingerprint are taken into consideration for comparison against the WAPs in the query-observation. Hence, the number of comparisons is reduced, so computational power and response time required is controlled. The RMSE and Euclidian distance is computed for the top n WAPs on every fingerprint, and the most likely matching fingerprint is the one with the smallest RMSE and minimum Euclidian distance.

$$d = \sqrt{\sum_{j=1}^{n} (x_j - \hat{x_j})^2} \tag{6.51}$$

If any further ambiguity exists, a comparison using the $n+1$ ranked WAPs is carried out and so on, until a final single matching fingerprint is estimated. Any WAP that is out of control (according to the EWMA upper and lower control limits) is not included within this positioning stage. Nevertheless, the WAP that is out of control is still updated with every new measurement, and once its $\hat{x}$ is under control it is again considered for positioning.

6.2.6 Displaying Estimated Location

Once the best-match is computed, the result is sent to the user's mobile device with the addition of graphical information such as a floor plan indicating the room and images from that particular room. The estimated location is displayed to the end user as the result from positioning algorithms.

Positioning was carried out, using the fingerprints generated from offices within RCS (as explained in 5.1.2 section Snapshot Data Collection).

A text example of an output after positioning is:

```
The best Matched location is :
Country: 'UK'
City: 'Manchester'
Building: 'RCS'
Room: '1.031b'
Number of matching WAPs: 24
RMSE: 31.7249
```

The display of the information can be customised according to the final application. Figures 6.21 and 6.22 illustrate results from positioning level 1 and 2 respectively.

6.2.7 Updating Map based on Query-Observations

User feedback and collaborative fingerprinting has been shown to be useful [LHC13]. It reduces the need for constant surveys by experts, and keeping the radio-map up-to-date.

The results from the location estimation can be positive (A location was matched,

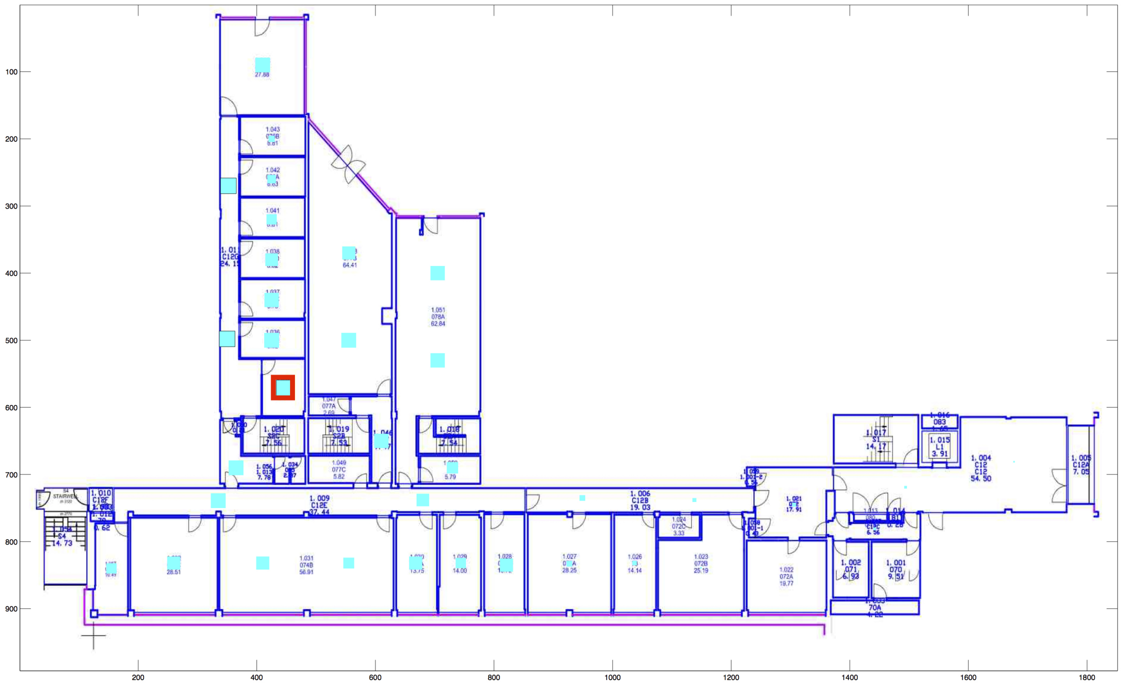


Figure 6.21: Position display based on WAPs presence. The pale blue squares indicate the rooms containing at least one matching WAP. The larger the size of the square, the larger the number of matching WAPs. The room with the largest number of matching WAPs is marked with a red edged square.

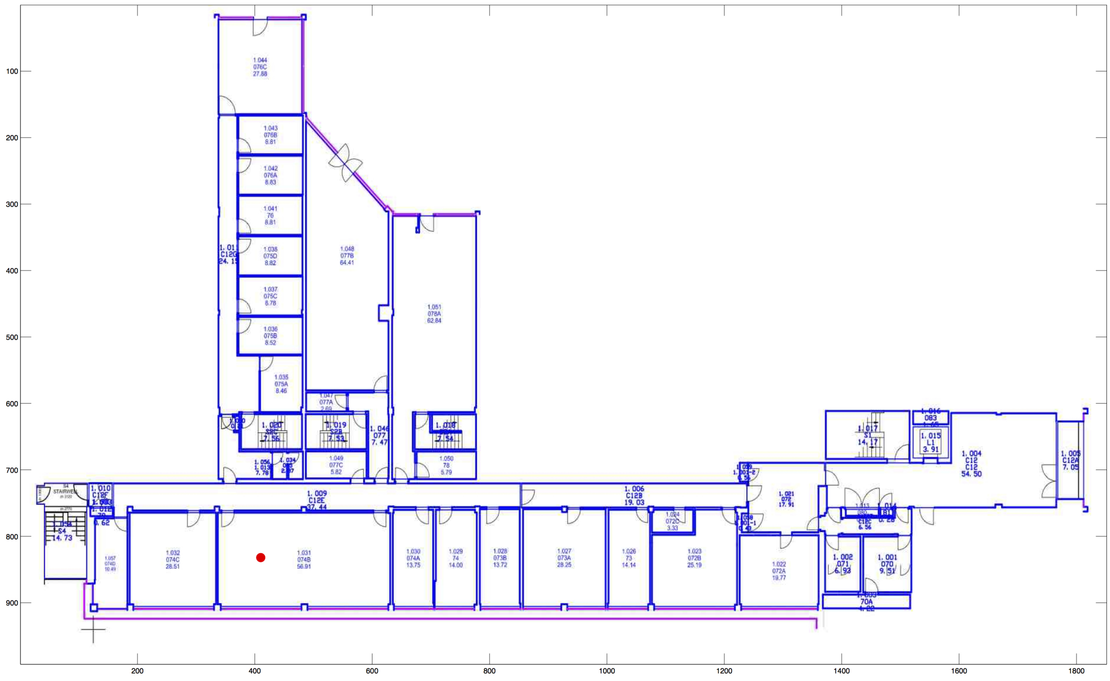


Figure 6.22: Position display based on WAPs RSSI. The room with the smallest RMSE is marked with a red dot on the floor plan.

with a certain degree of accuracy) or can be negative (No location was found at all). The potential for radio-map updates presented by query-observations is considered in the proposed system. After presenting the result to the user, the estimation parameters are analysed. In the case of a negative result, the query-observation is stored and labelled with an "unknown" location, and it is kept there for future analysis. On the other hand, when the location estimation result is positive, and the accuracy achieved has a high degree of certainty (an error smaller than a pre-defined threshold), the query-observation is then used for EWMA Fingerprint updating.

Chapter 7

Evaluation and Results

This chapter presents results on the selection of the parameters for computing the state estimate and control chart. It also presents results from implementing the five fingerprinting techniques, including the EWMA Fingerprint proposed by the author. The following fingerprinting algorithms are evaluated: Basic Fingerprint, Default Fingerprint, Gallagher Fingerprint, Jun-Sung Fingerprint and EWMA Fingerprint. These algorithms have been tested with the dataset described in Chapter 5.

The fingerprints are generated progressively, using observations from the three locations (Piccadilly Railway Station, RCS and Private Residence). Once the results are presented individually, the fingerprints are further analysed and compared with the EWMA Fingerprint method.

Also presented are the results of the scoring function that leads to the removal of certain WAPs from the EWMA Fingerprint, according to the signal strength, standard deviation and historical frequency of occurrence.

7.1 Results on Identifying a Suitable λ_x per WAP

It is hypothesised that the optimal value for λ_x can be tuned according to the characteristics presented in the WAP's signal profile. For this, EWMA was applied to the WAP's RSSI, using linearly spaced values for λ_x in the range (0,1). The sum of the errors per λ_x was computed, and the λ_x generating the minimum error was identified. The sum of the errors per λ_x is plotted in Figures 7.1, 7.2 and 7.3 for WAPs 137, 63 and 363 respectively. This process was performed for all WAPs detected in the three locations subject to analysis: Office 1.035, RCS (Location 20); Platform 6, Piccadilly (Location 36); Lounge, Private Residence (Location 39).

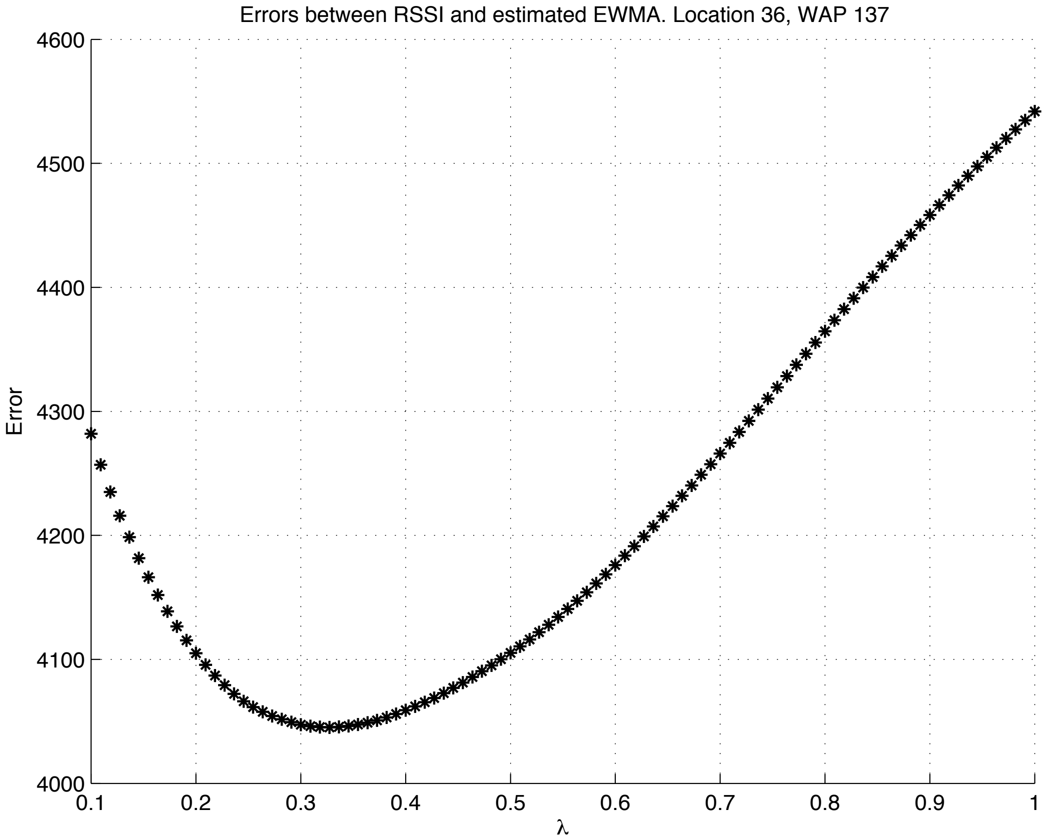


Figure 7.1: Sum of errors per λ_x for WAP 137 at Piccadilly

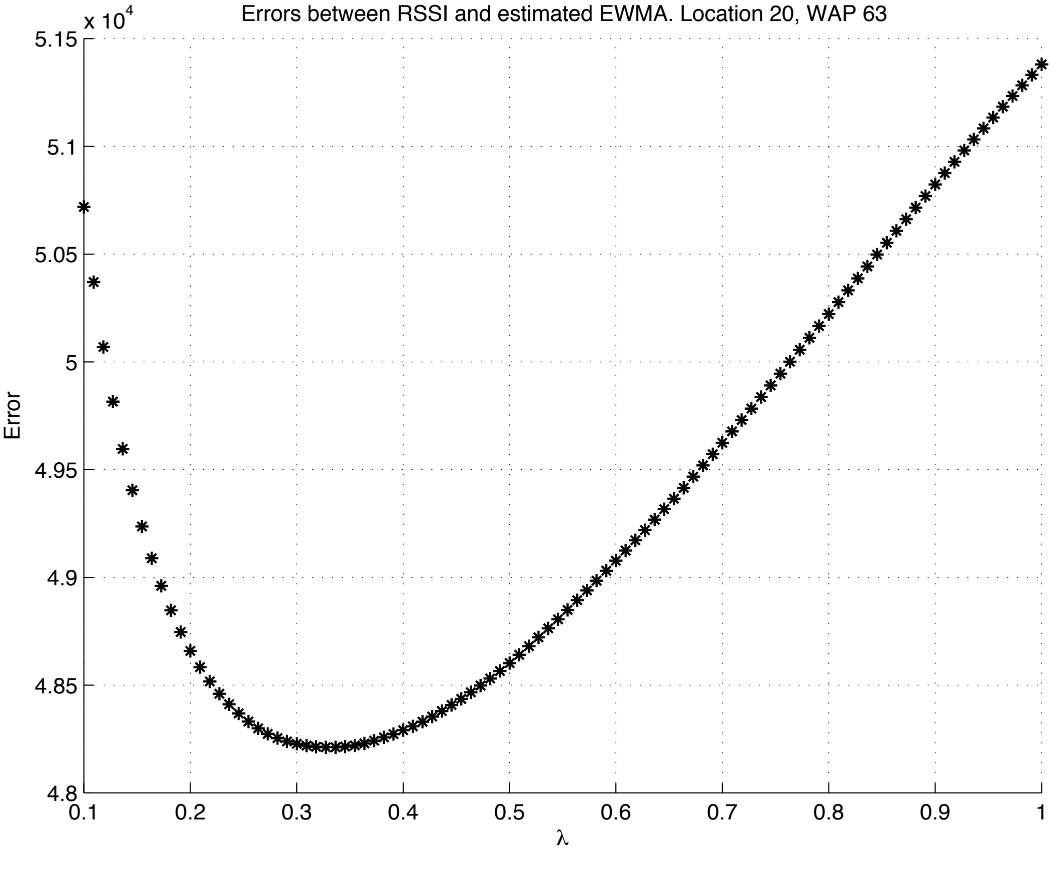


Figure 7.2: Sum of errors per λ_x for WAP 63 in Office 1.035, RCS

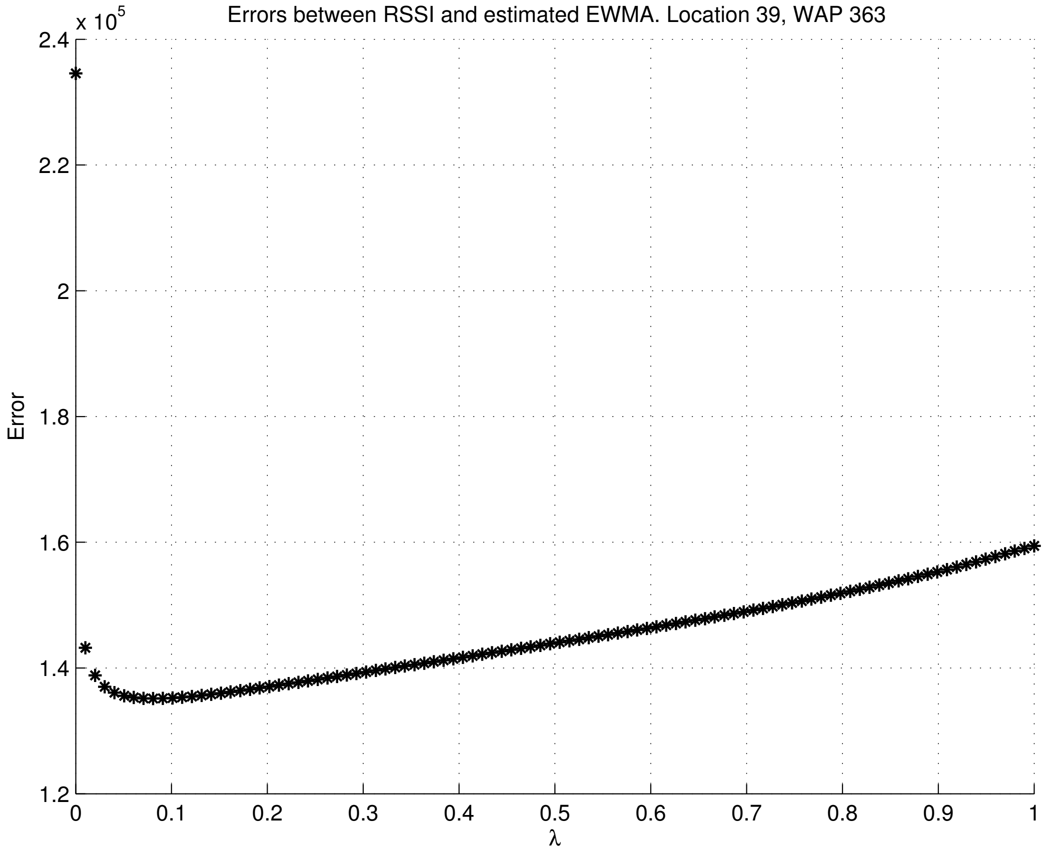


Figure 7.3: Sum of errors per λ_x for WAP 363 in Private Residence

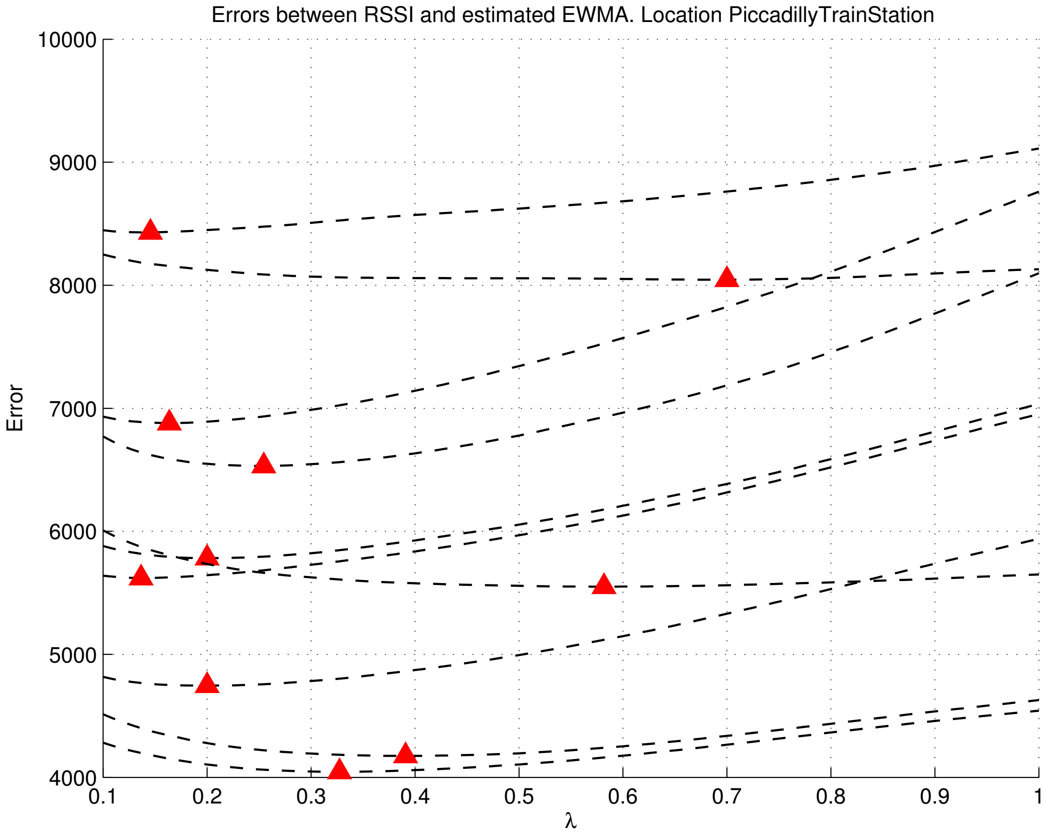


Figure 7.4: Sum of errors per λ_x for the top 10 WAPs, arranged by frequency of occurrence. Location Platform 6, Piccadilly Railway Station

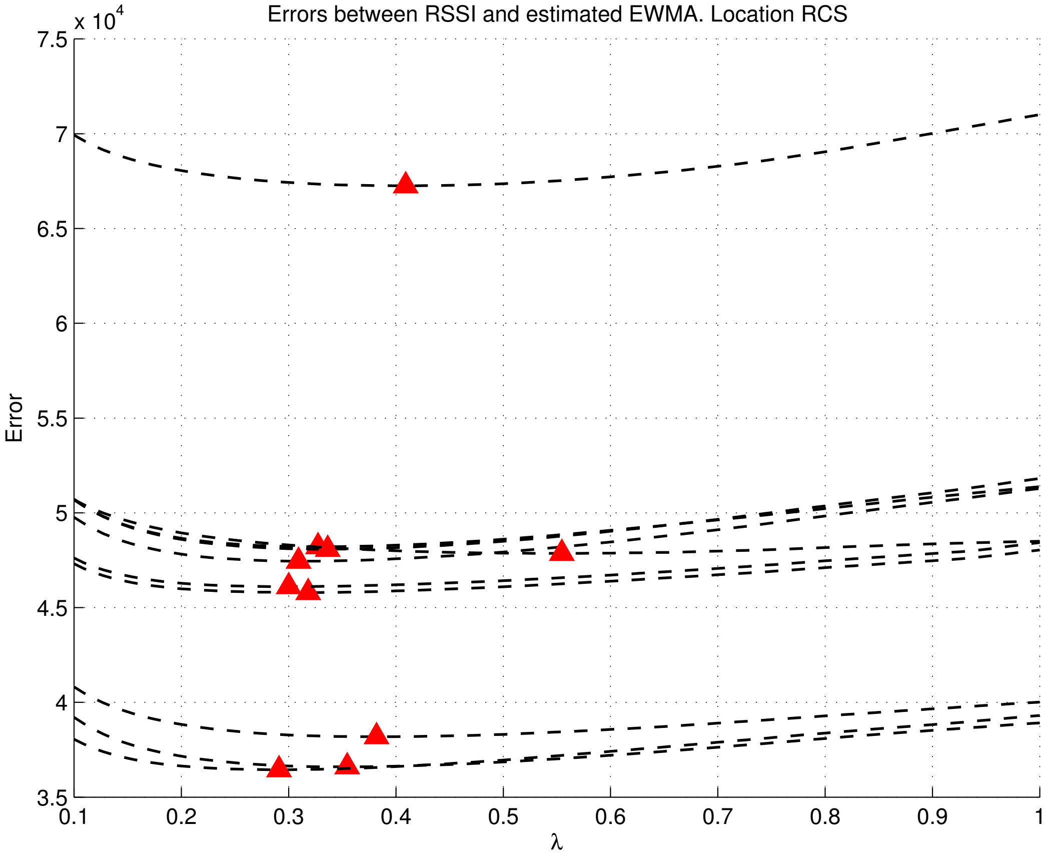


Figure 7.5: Sum of errors per λ_x for the top 10 WAPs, arranged by frequency of occurrence. Location Office 1.035, RCS

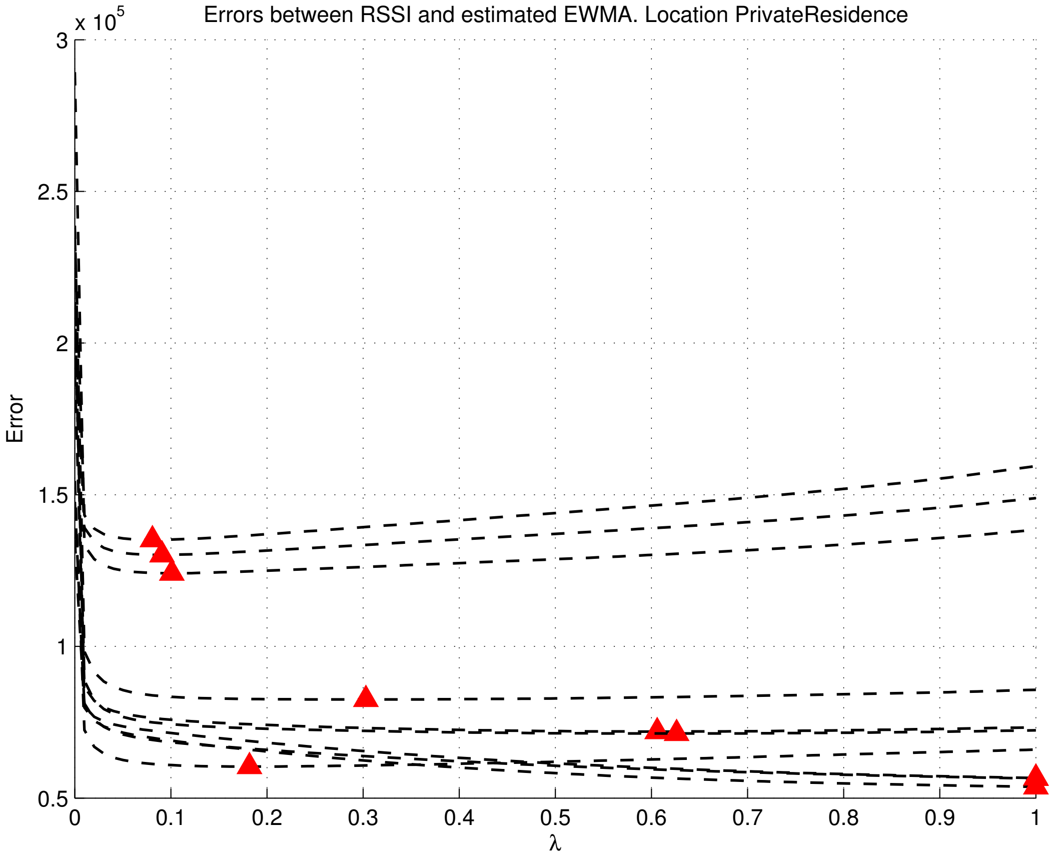


Figure 7.6: Sum of errors per λ_x for the top 10 WAPs, arranged by frequency of occurrence. Location Lounge, Private Residence

Figures 7.4, 7.5 and 7.6 show the errors for the 100 values using data from the top 10 WAPs at each location, the minimum error per WAP is highlighted by the red triangle.

Data resulting from processing the top 10 WAPs (WAPs with highest frequency of occurrence) are presented in Table 7.1 for WAPs at location Piccadilly (Platform 6), Table 7.2 for WAPs at location RCS (office 1.035) and Table 7.3 for WAPs at location Private Residence (Lounge). The values presented per column are the following: ranking number by frequency of occurrence (#), identifier of the WAPs (WAP ID), mean of RSSI (μ), standard deviation of RSSI (σ), the value registered as a minimum error (min Error), and finally the value of λ_x that generated the minimum Error.

Table 7.1: Generating a minimum error per WAP. Location Platform 6, Piccadilly Railway Station.

#	**WAP ID**	μ	σ	**min Error**	λ_x
1	137	-71.629	1.7995	4045.2	0.33
2	127	-70.345	1.9086	4174.8	0.39
3	150	-70.457	3.6036	8429.9	0.14
4	161	-74.532	3.4526	8045.3	0.7
5	194	-62.277	3.2507	6531.2	0.25
6	199	-63.667	2.8502	5781.9	0.2
7	197	-72.946	2.9551	4745.8	0.2
8	195	-77.525	3.1464	5620	0.14
9	193	-73.122	3.4918	6879.9	0.17
10	146	-54.239	3.2449	5550.1	0.59

Table 7.2: Generating a minimum error per WAP. Location Office 1.035, RCS.

#	**WAP ID**	μ	σ	**min Error**	λ_x
1	63	-62.506	5.303	48211	0.33
2	33	-75.555	3.7583	36599	0.35

Continued on next page

Table 7.2 – continued from previous page

#	WAP ID	μ	σ	min Error	λ_x
3	42	-75.507	3.8149	38188	0.38
4	49	-63.573	4.8422	45795	0.32
5	52	-63.619	4.8559	46108	0.3
6	68	-70.687	5.2232	47451	0.31
7	67	-70.611	5.2514	48080	0.34
8	83	-71.927	3.6536	67256	0.41
9	65	-69.565	3.2566	47864	0.55
10	55	-73.612	3.7673	36448	0.29

Table 7.3: Generating a minimum error per WAP. Location Lounge, Private Residence

#	WAP ID	μ	σ	min Error	λ_x
1	363	-73.693	4.1614	1.3514e+05	0.08
2	360	-73.669	4.1022	1.3017e+05	0.09
3	350	-77.141	3.0698	60374	0.18
4	361	-73.624	4.0404	1.2412e+05	0.1
5	353	-74.979	3.6876	82488	0.31
6	352	-79.594	3.3709	71932	0.61
7	365	-78.33	4.2151	71278	0.63
8	358	-80.134	3.1299	56462	1
9	362	-80.836	2.8749	53702	1
10	370	-80.566	3.512	56600	1

The used of a λ_x per WAP provides a more accurate state estimate of the RSSI. Therefore it might be possible to have a better match using existing fingerprint based positioning techniques. This is because the expected signal strength for all the WAPs in the fingerprint have been estimated with a λ_x that minimises the error.

The parameter λ_x is experimentally tuned for each WAP. The timing on computing

λ_x for a WAP per location is presented in Table D.1. The script used to compute the λ_x is presented in Appendix C. This script has been programmed aiming to proof the concept. Therefore, it can be optimised for reducing its processing time in the final implementation of the system.

Location	WAP Observations	WAP ID	Total time on computing λ_x (seconds)
Platform 6 Piccadilly Railway Station	3,727	137	1.18
Office 1.035 RCS, UoM	46,145	63	4.27
Private Residence Derby, UK	58,062	363	4.98

Table 7.4: Timing on computing λ_x for a single WAP per Location

The timing on computing λ_x of all the WAPs per locations is presented in Appendix D.

7.1.1 Updating Strategy for λ_x

The parameter λ_x is experimentally tuned for each WAP and this will have to be updated as the temporal signal of the WAP changes. Two strategies are discussed to accomplish this update:

- Time based. An updating strategy based on a fix time, or in a specific number of observations is an approach to solve the problem of the value of λ_x. Establish a mean number of observations between updates requires an in-depth analysis of the data set. Although it is a potential solution, a fix updating time does not provides the best updating strategy, since the WAPs signals strength profiles differs dramatically even in when recorded at the same location. For this the author proposes a second approach, a condition based strategy.

- Condition based. The Control Chart discussed previously detects when the data is out of control, and is used for managing the WAPs. This same data from the Control Chart can be used to determine when the λ_x should be updated. This provides a data driven mechanism that triggers alerts or automatically computes a new λ_x. The proposed strategy would consist of two steps. First, detecting when a WAP is out of control and removing it from the fingerprint. Second, monitoring the number of observations it takes the WAP to go back under control, if it takes larger than for example 100 observations, as shown in WAP 596 Figure 7.19, then the λ_x is updated using k_0 number of observations. The implementation of this updating strategy done is left as further work.

7.2 Comparing Estimated Value

A comparison of the estimated value $\hat{x}$ from the five fingerprinting techniques is presented in Figures 7.7, 7.8, 7.9 and 7.10. The estimated value corresponding to the EWMA Fingerprint was computed using λ_x that minimises the error per WAP. As it is observed the $\hat{x}$ values for EWMA method 7.10 are less volatile, nevertheless they respond to changes in the signal.

7.2.1 Comparing EWMA and Jun-Sung Fingerprints

The error between the raw data and the estimated value from the Jun-Sung algorithm was generated and compared with the error between raw data and the EWMA algorithm. Results showed that by selecting an optimal value of λ_x the error using EWMA can be lower than the error generated by Jun-Sung algorithm. Figures 7.11, 7.12 and 7.13 present the error differences between Jun-Sung and EWMA.

It is observed from the green sections on the curves that a smaller error can be achieved by implementing the EWMA algorithm with the optimal λ_x value, rather than the average as implemented by Jun-Sung. As explained in section 6.1.5 the Jun-Sung approach to compute the expected value performs as an EWMA with a value of $\lambda_x = 0.5$. This is illustrated in Figures 7.11,7.12 and 7.13 where the error difference is 0 (zero) for $\lambda_x = 0.5$.

7.3 EWMA on Detecting Changes and Outliers

This section presents the results of finding candidate parameters for detecting when the signal is out of control in the EWMA control chart. As explained in the section Control Chart in Chapter 6, for detection of changes two EWMA are employed. The parameter λ_{c1} has the function of smoothing the data, smaller values for λ_{c1} allow for

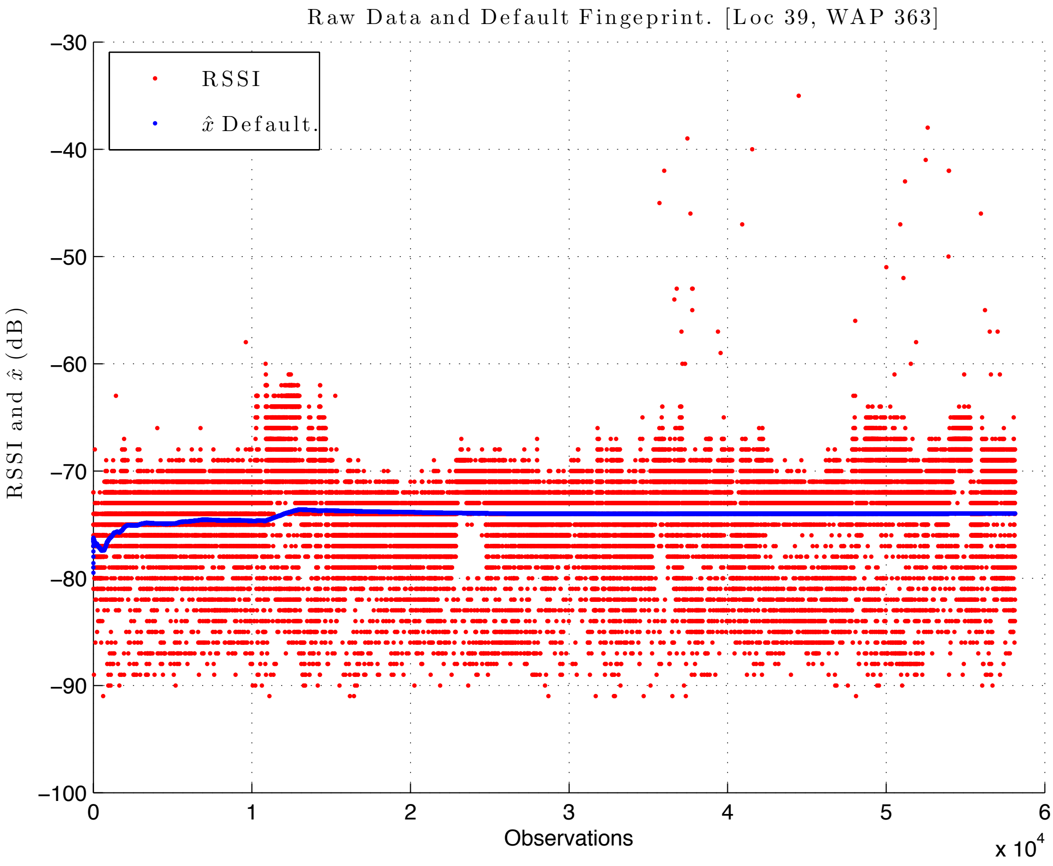


Figure 7.7: Comparison of the expected signal strength value $\hat{x}$ computed with the Basic Fingegerprint/raw data and the Default Fingerprint using observations taken from WAP 363.

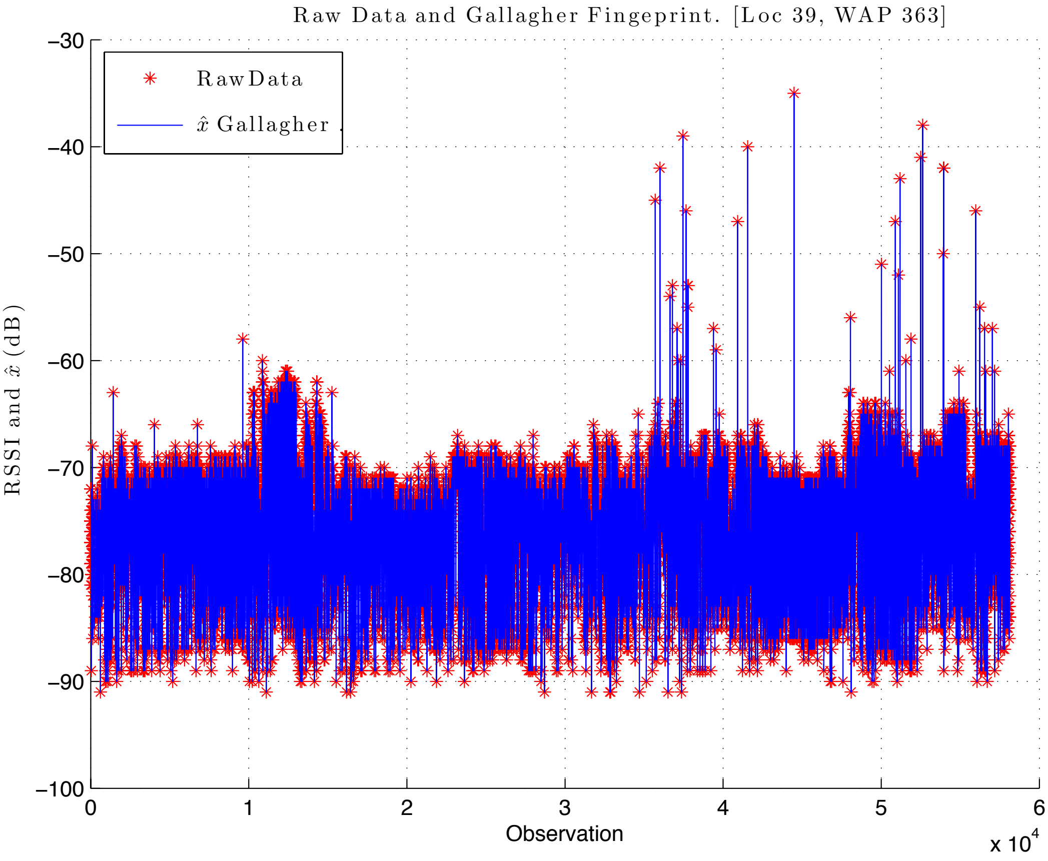


Figure 7.8: Comparison of the expected signal strength value $\hat{x}$ computed with the Basic Fingegerprint/raw data and the Gallagher Fingerprint using observations taken from WAP 363.

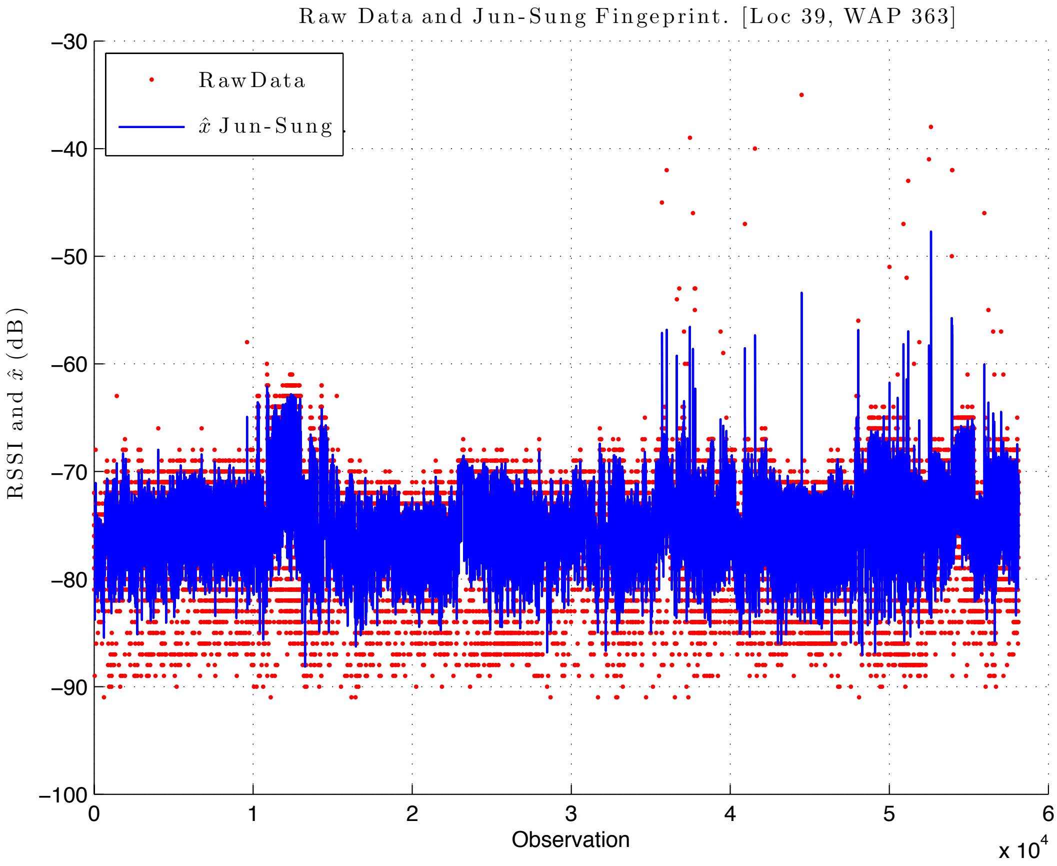


Figure 7.9: Comparison of the expected signal strength value $\hat{x}$ computed with the Basic Fingegerprint/raw data and the Jun-Sung Fingerprint using observations taken from WAP 363.

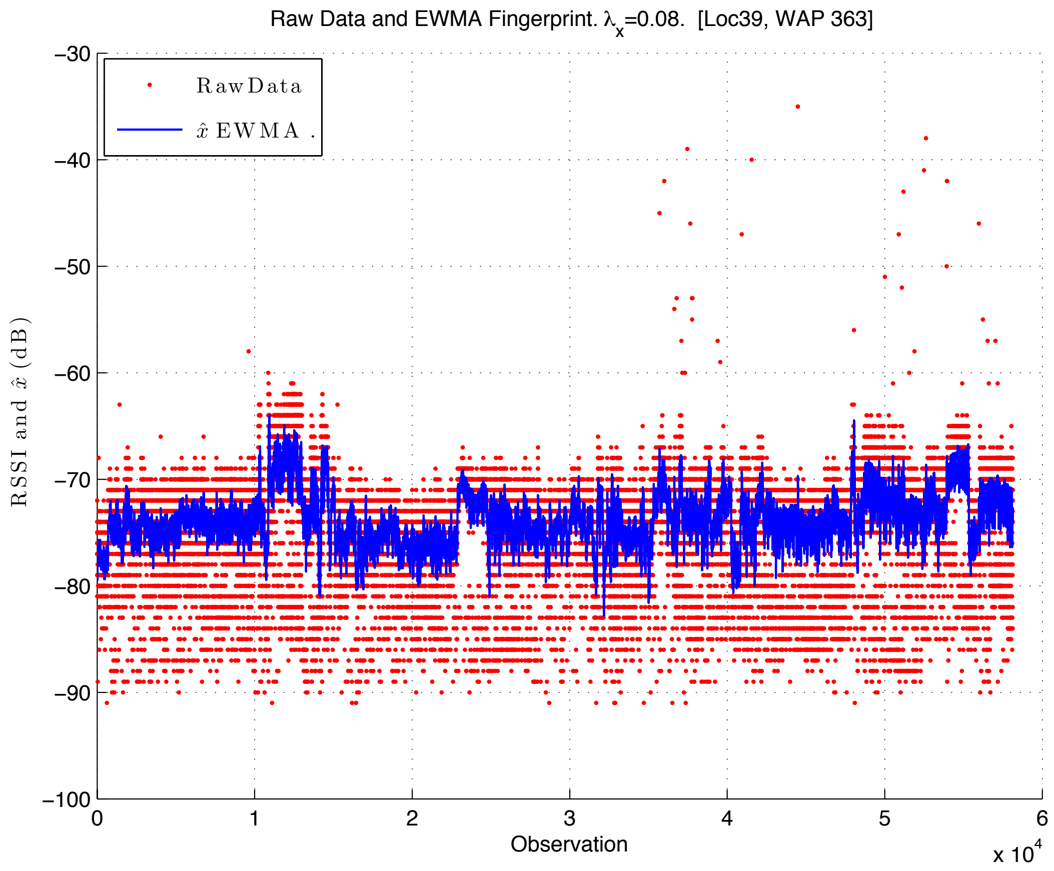


Figure 7.10: Comparison of the expected signal strength value $\hat{x}$ computed with the Basic Fingegerprint/raw data and the EWMA Fingerprint using observations taken from WAP 363.

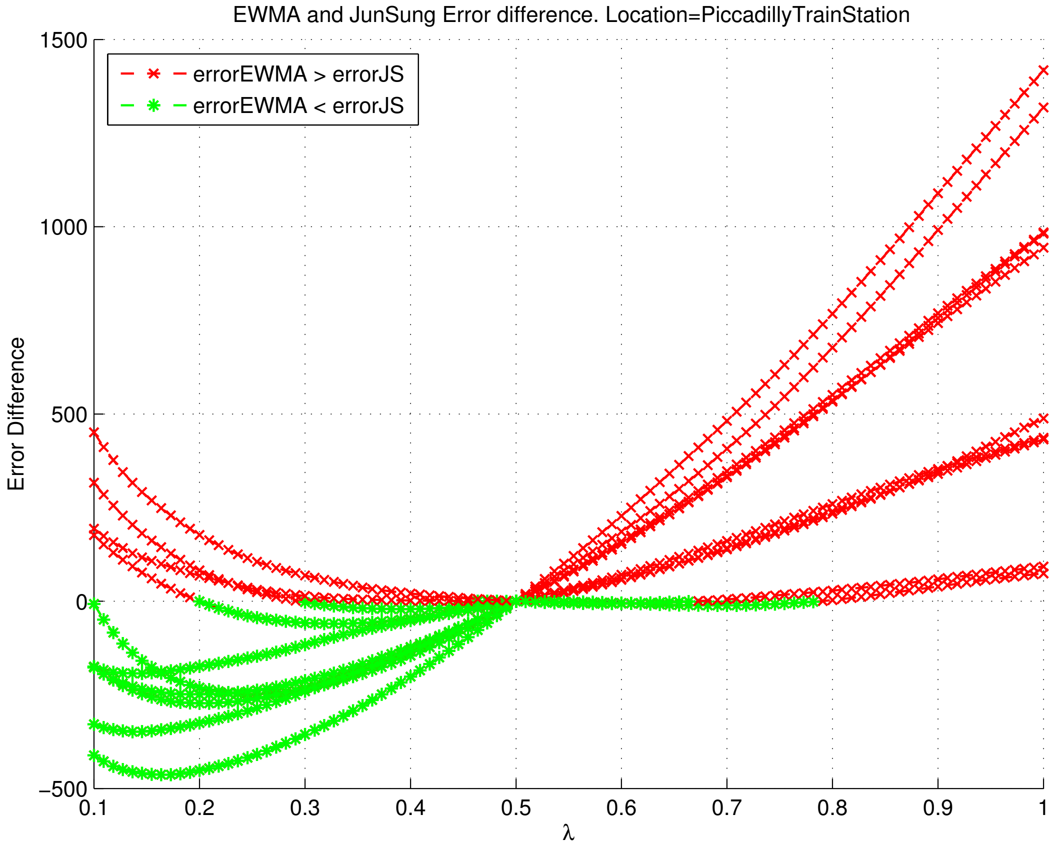


Figure 7.11: Location Piccadilly

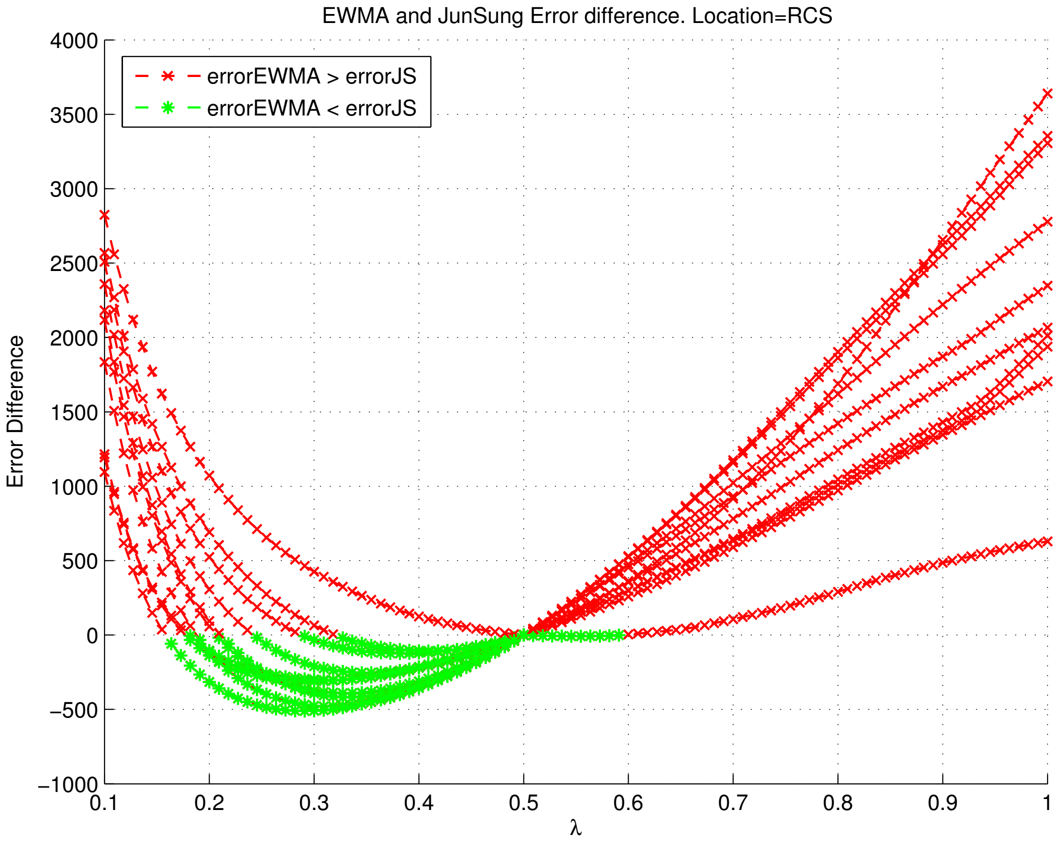


Figure 7.12: Location RCS

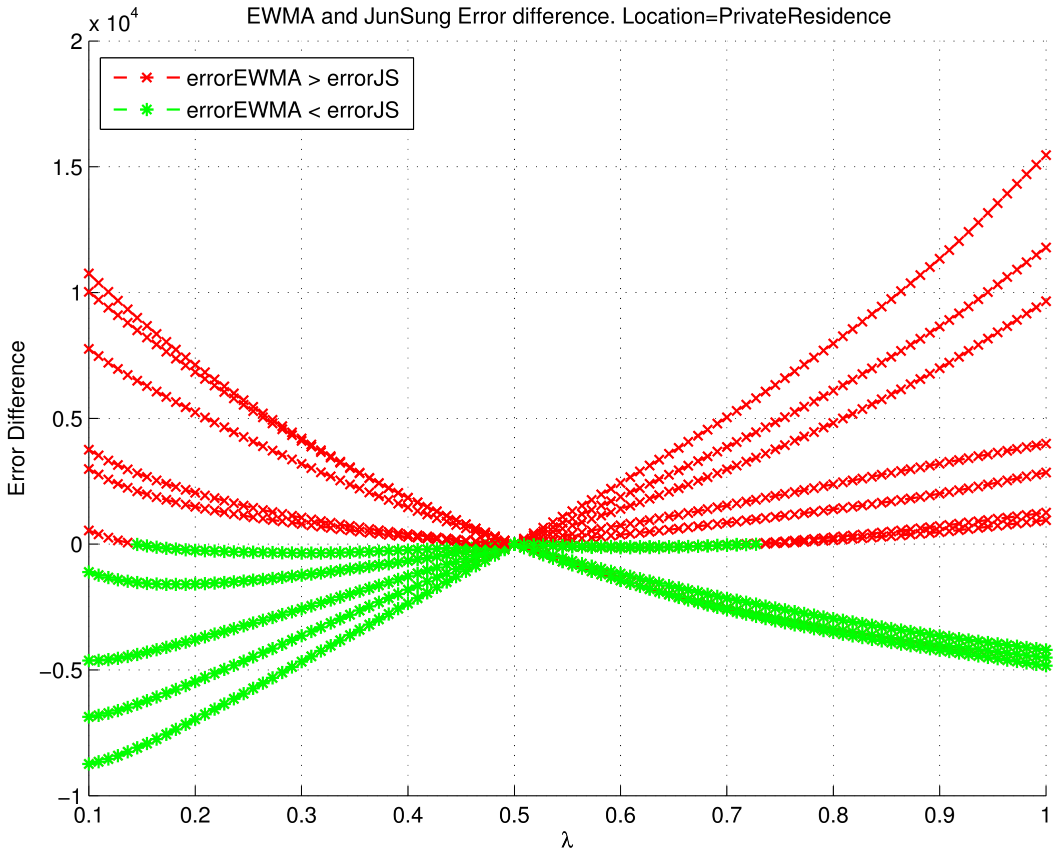


Figure 7.13: Location Private Residence

the shape of the data to be retained post normalisation, which is better for identifying changes.

7.3.1 Analysis on WAP 596

WAP 596 was selected due to its signal strength profile presenting a good example of a step change. Also, WAP 596 has the largest standard deviation within the WAPs detected at RCS. It was detected in a total of 1,363 observations with a standard deviation of $\sigma = 12.55$.

Figure 7.14 shows the signal strength measurements for WAP 596, a significant increase is shown at observation number 385.

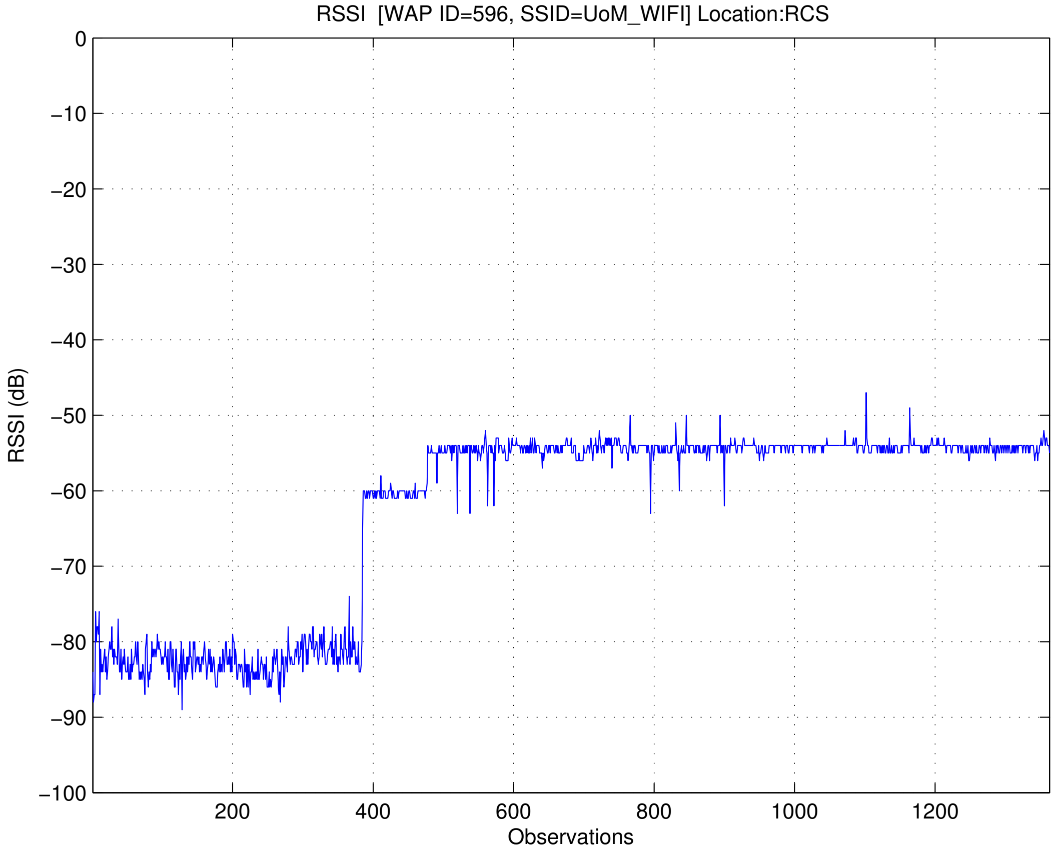


Figure 7.14: RSSI from WAP ID 596 at RCS

Varying parameter λ_{c1}

EWMA was applied to RSSI data from WAP 596, varying the values of λ_{c1} in the range (0,1) whilst fixing weight (λ_{c2}=0.5) and threshold (L=6). Acceptable results were found for small values of λ_{c1}. For example, Figure 7.15 was computed with a $\lambda_{c1} = 0.1$, $\lambda_{c2} = 0.5$ and $L = 6$, this parameter of λ_{c1} did not provide a desirable outcome since the confidence values are within boundaries when the step change is presented. It was noticed that for values of λ_{c1} larger than 0.1, the confidence was not large enough to detect the change in the signal. From this the range for λ_{c1} was narrowed to (0.001, 0.1).

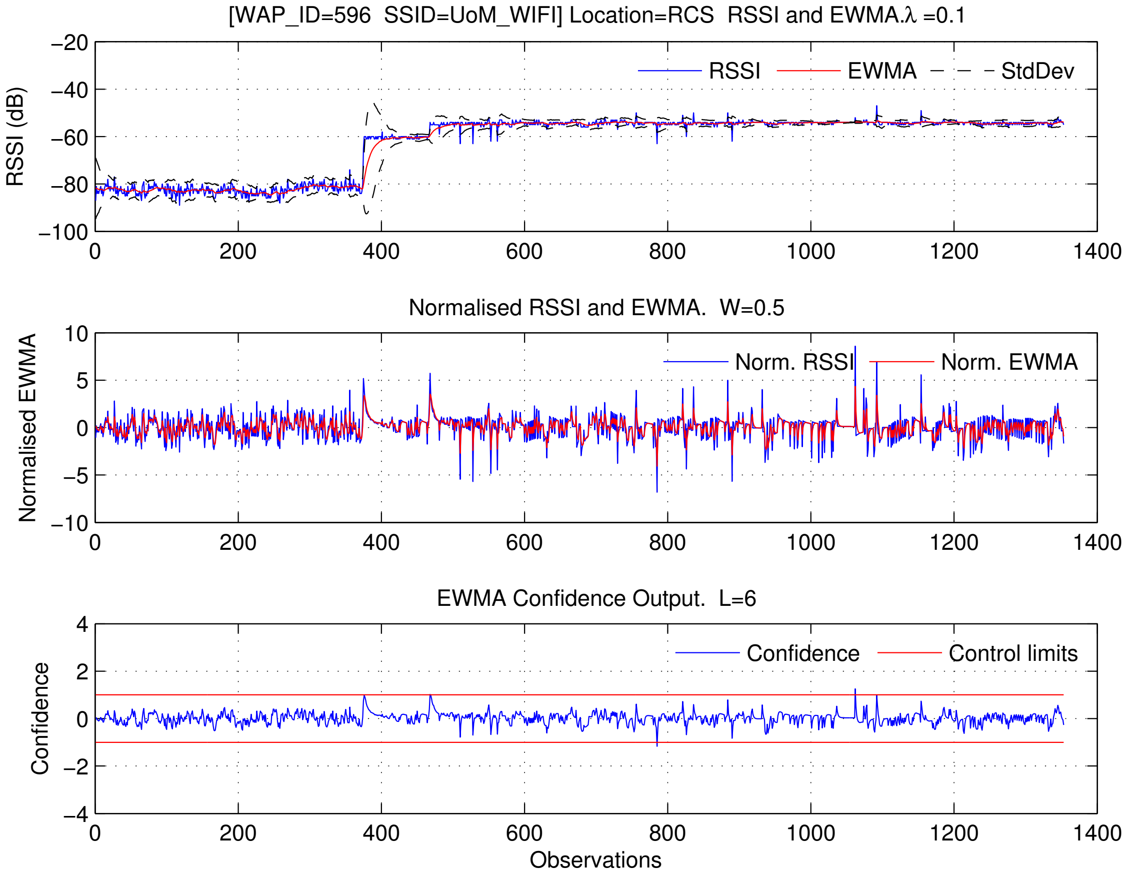


Figure 7.15: This Figure presents the EWMA control chart applied to data from WAP 596. Three plots are presented here: the RSSI, the Normalised EWMA and the confidence. The top plot presents the RSSI (blue line) and $EWMA_{c1}$ computed with a λ_{c1} = 0.1. The middle plot presents the Normalised EWMA (red line) which was computed with $EWMA_{c2}$ (the second EWMA in the control chart) having a weight (W) equal to 0.5 (W=λ_{c2}=0.5). Finally, the bottom plot presents the confidence and control limits with L=6

By reducing the value to $\lambda_{c1} = 0.01$ as shown in Figure 7.16 the change in RSSI at observation 385 is detected by the control limits, but the confidence values go out of control limits during just a few observations, since the signal in WAP 596 has the higher σ a more significant change in the *confidence* value is desirable.

Results found with a value of $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.5$ and $L = 6$ are presented in Figure 7.17, from the *confidence* plot at the bottom of the figure it can be appreciated that the step change is detected since the confidence value changes significantly enough. It can be said that the value $\lambda_{c1} = 0.001$ is a suitable candidate for step changes using WAP 596.

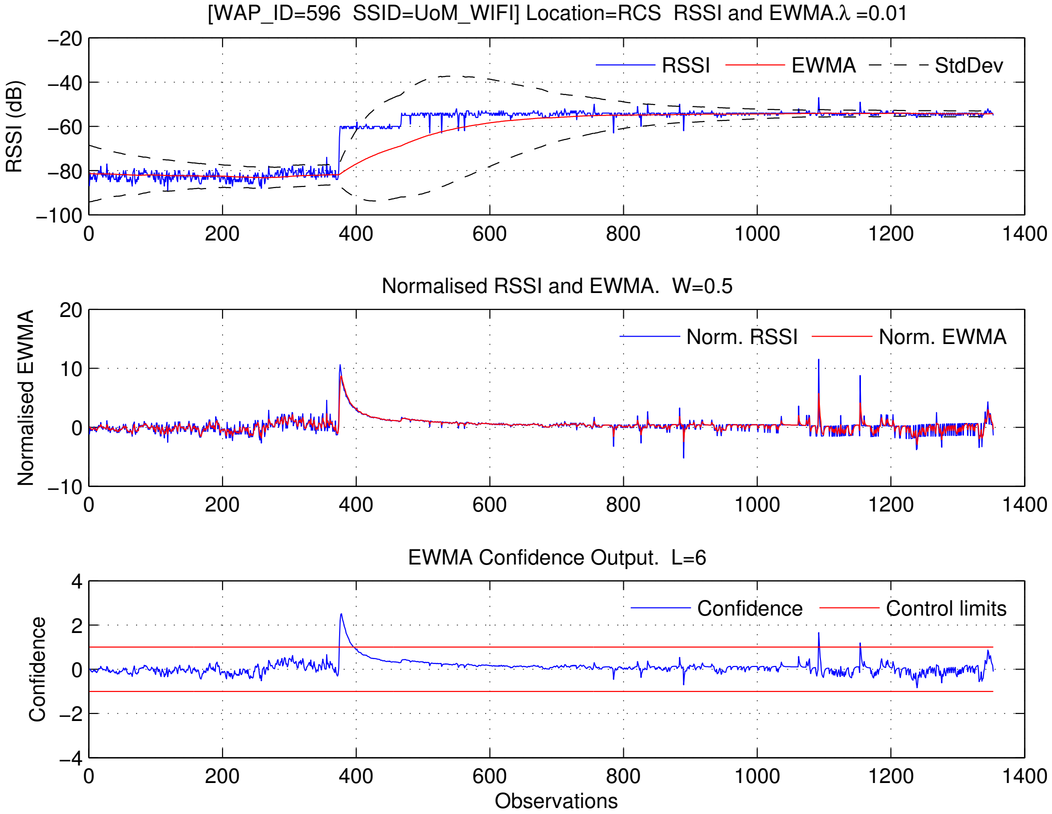


Figure 7.16: $\lambda_{c1} = 0.01$, λ_{c2}=0.5, L=6

Varying parameter λ_{c2}

Then changes to the parameter λ_{c2} were analysed, an initial approach was to implement EWMA with values suggested by [Mon08], which are $\lambda_{c2} = 0.05$ and L=2.492. Figures 7.18 presents results using the suggested parameters. From these results, it can be stated that the suggested values perform poorly for this particular RSSI data. It

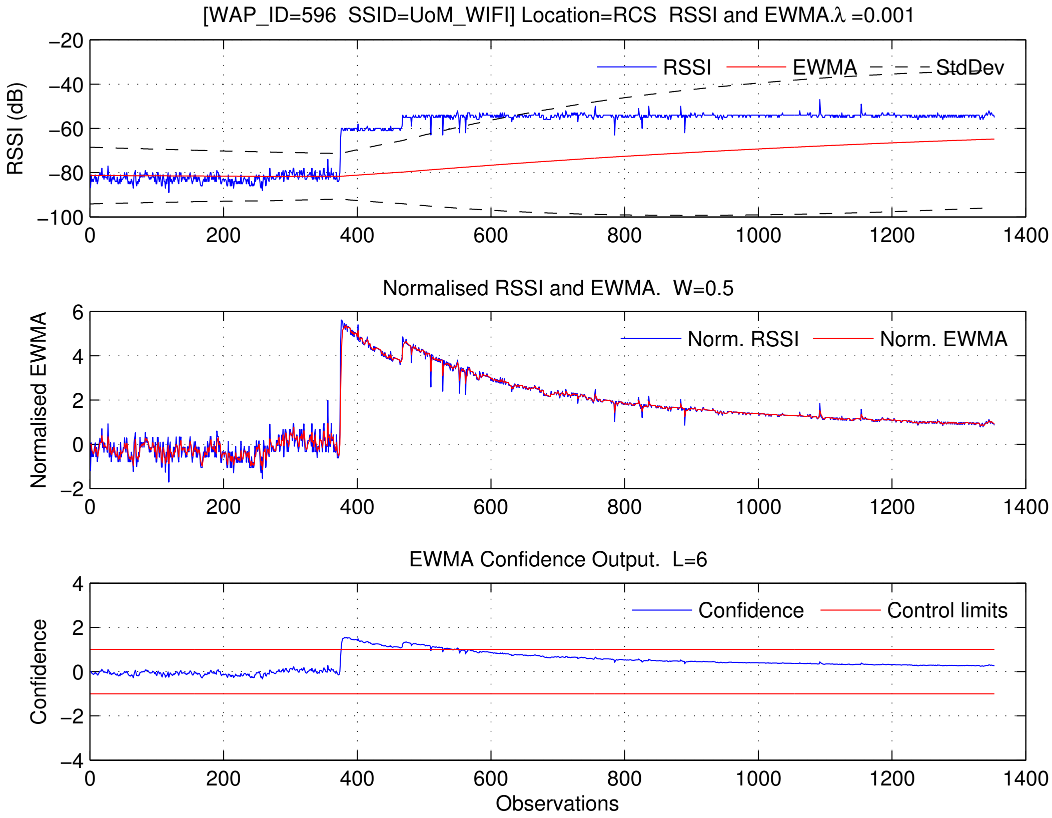


Figure 7.17: λ_{c1} = 0.001, λ_{c2}=0.5 ,L=6

is determined from this that other values should be identified. The second approach was to use $\lambda_{c2} = 0.34$ which is the optimal value found for estimating $\hat{x}$ for WAP 596 (see section 6.1.6 State Estimation). Results of implementing $\lambda_{c2} = 0.34$ are presented in Figure 7.19. A value of $\lambda_{c2} = 0.34$ presented a desirable outcome for detecting changes in the signal for WAP 596.

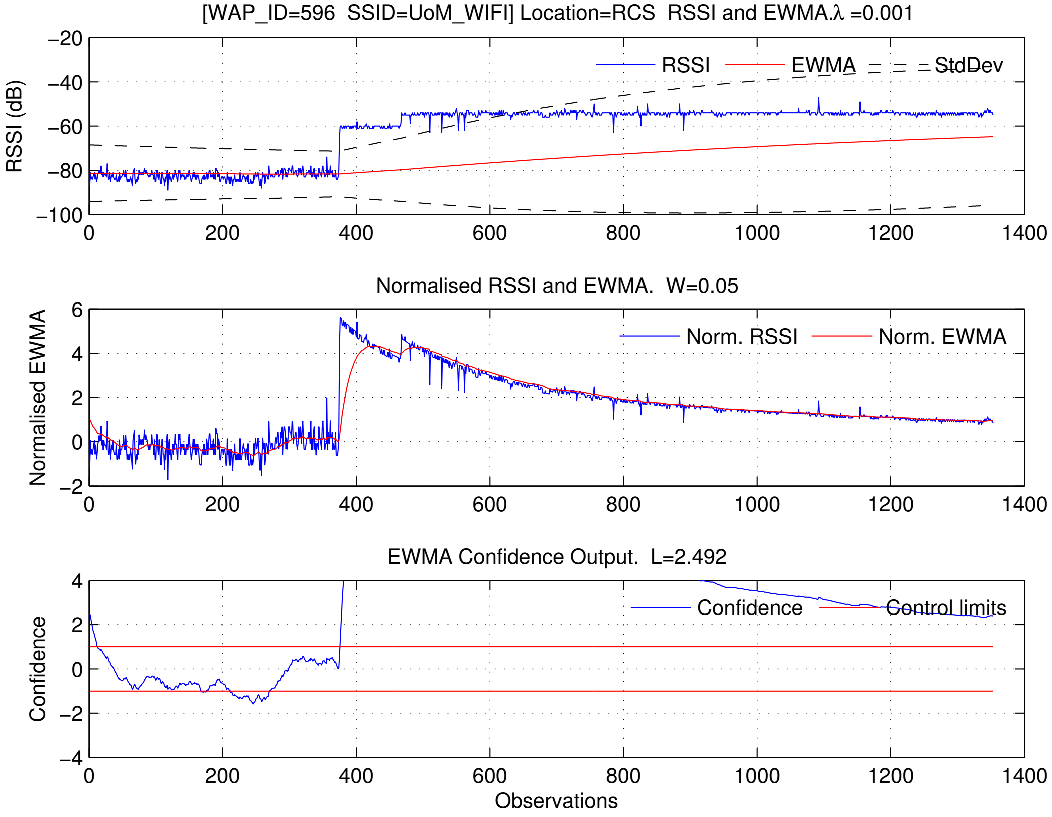


Figure 7.18: $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.05$ and $L = 2.492$. Values λ_{c2} and L sugested by [Mon08] applied for detecting changes and outliers in observation from WAP 596.

7.3.2 Analysis on WAP 63

WAP 63 was subject to analysis, results of fitting parameters ($\lambda_{c1} = 0.01$, $\lambda_{c2} = 0.5$ and $L = 6$) are observed in the confidence graph in Figure 7.20, these results are not the most desirable since the changes in the raw data are faintly reflected in the confidence. Also tested were the values within the range ($0.05 \le \lambda_{c2} \le 0.25$) suggested in [Mon08] suitable results were found using the $\lambda_{c2} = 0.25$. As presented in Figure 7.21, parameters ($\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and L=6) were found for WAP 63, resulting in a more accurate detection of changes by the confidence and control limits.

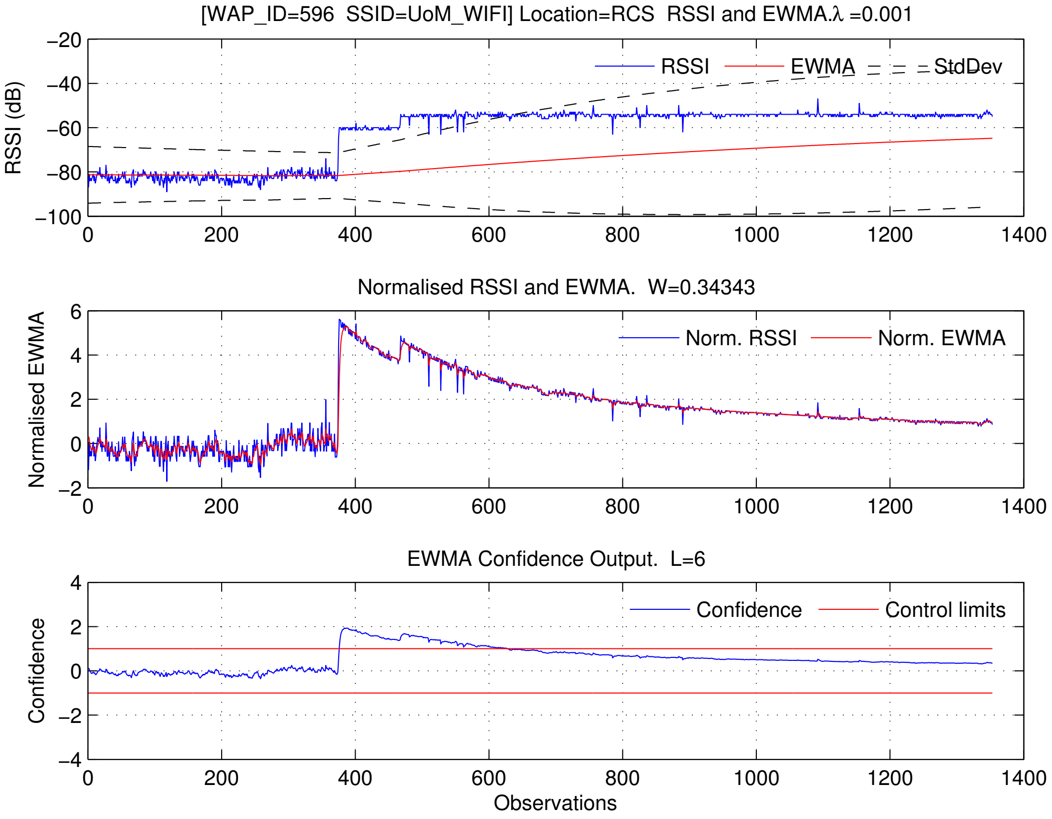


Figure 7.19: $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.34$ and $L = 6$. Varying λ_{c2} values for detecting changes and outliers for WAP 596.

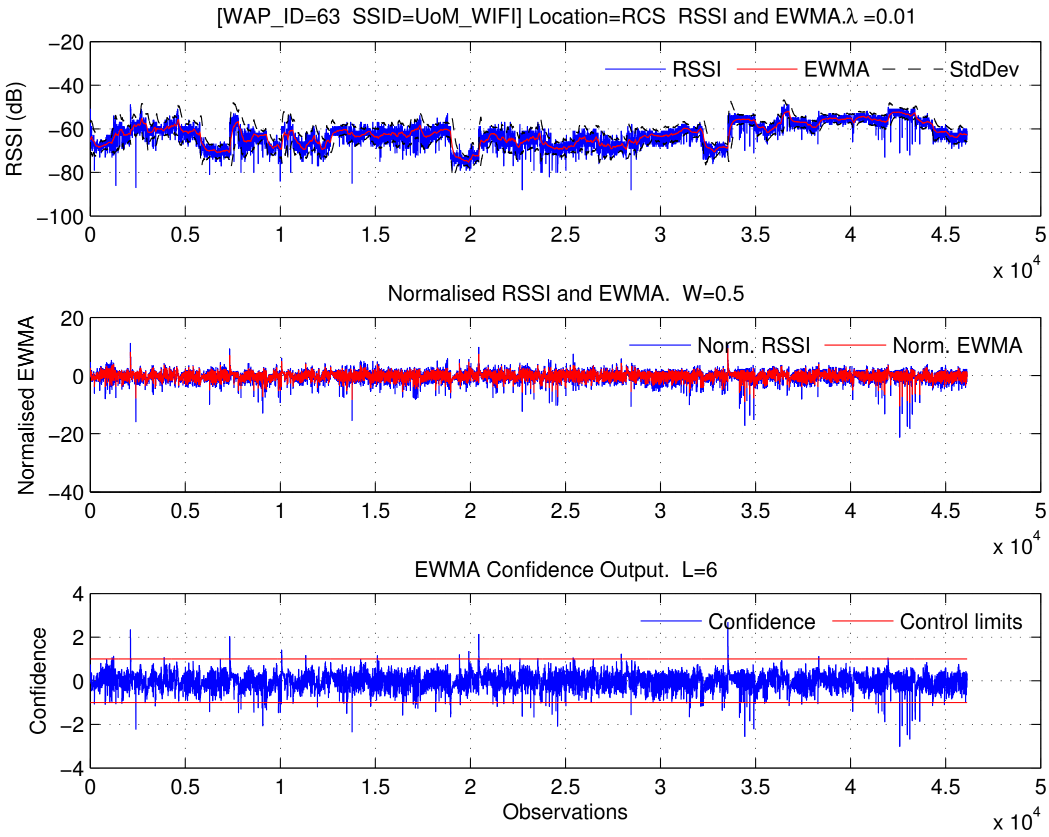


Figure 7.20: $\lambda_{c1} = 0.01$, $\lambda_{c2} = 0.5$ and $L = 6$. Varying parameters for detecting changes and outliers for WAP 63.

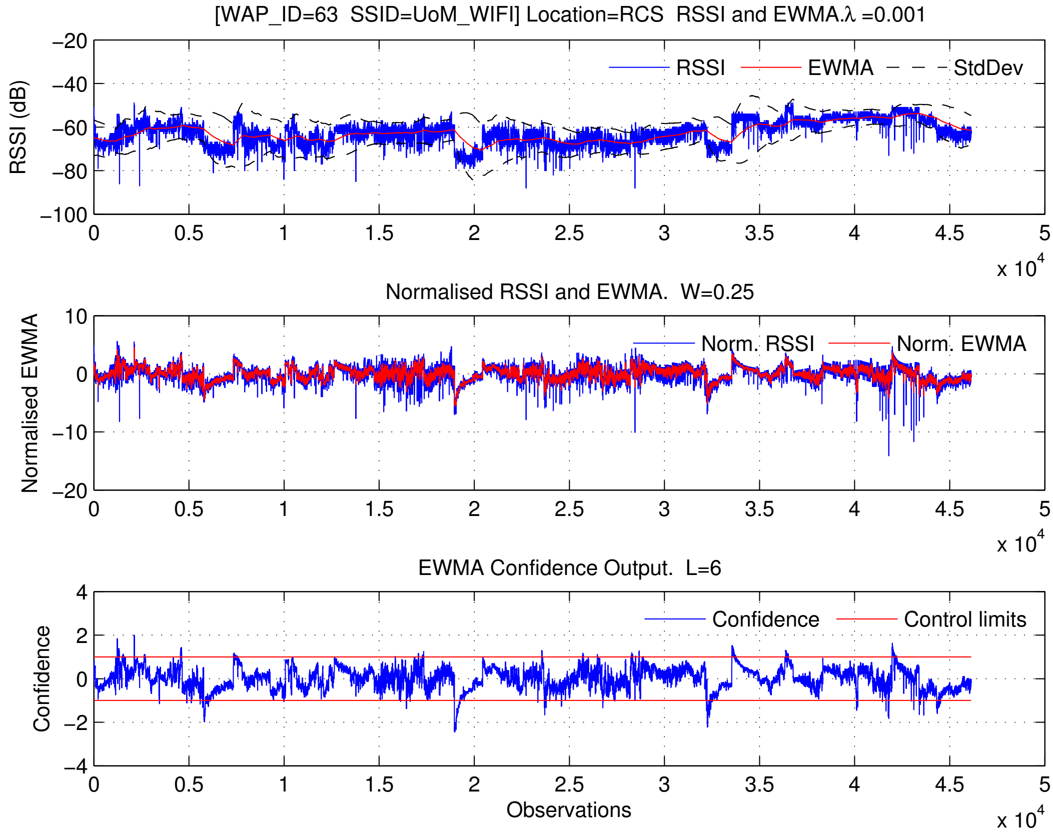


Figure 7.21: λ_{c1}= 0.001, λ_{c2}=0.25, L=6. Varying parameters for detecting changes and outliers for WAP 63.

7.3.3 Analysis on WAPs 137 and 363

These candidate values for λ_{c1}, λ_{c2} and L are then tested on other WAPs with varying signal strength profile in order to identify well suited parameters for all of the WAPs at the RCS location. EWMA was applied to WAP 137 (Piccadilly Railway Station, Platform 6), and to WAP 363 (Private Residence), using the following parameters $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and $L = 6$. Results are shown in Figures 7.23 and 7.24 respectively.

7.3.4 Analysis on WAPs 193, 197, 198 and 214

Results from WAPs 193, 197, 198 and 214 are shown in Figures 7.25, 7.26, 7.27 and 7.28. The parameters used are those that showed suitable results on previously analysed WAPs. These values are $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and $L = 6$. In the figures, the top subplot shows the raw measurements with the mean estimate overlaid. Control

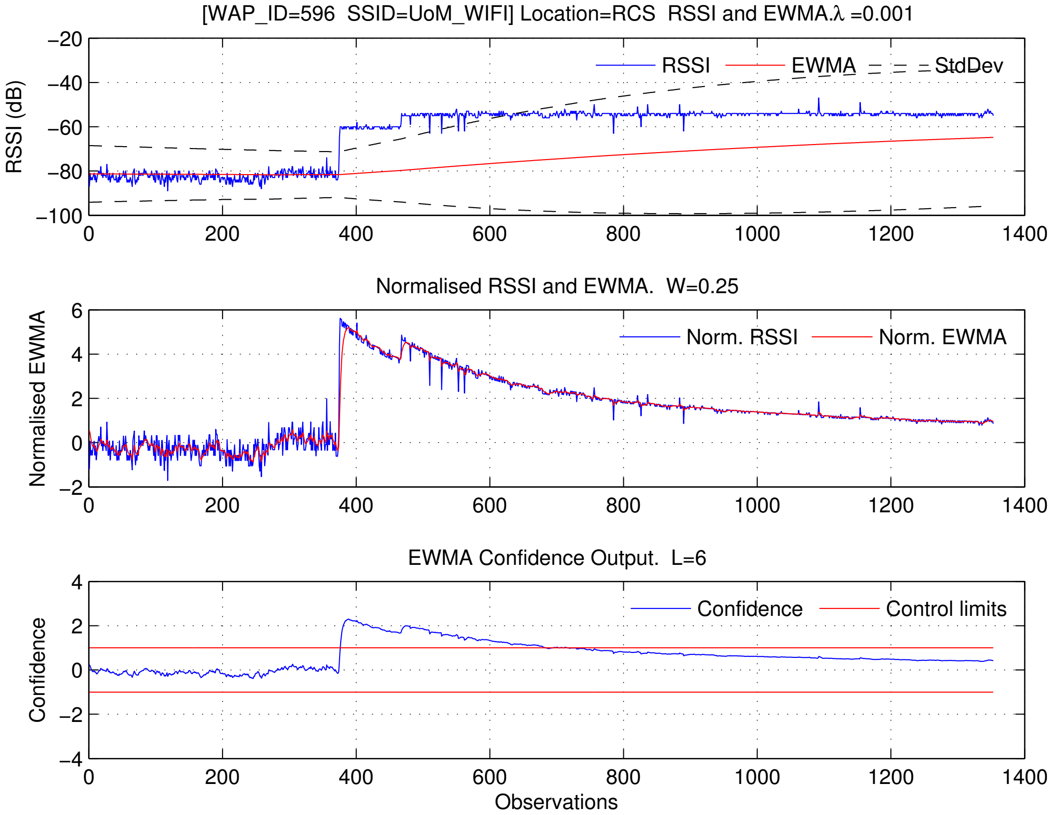


Figure 7.22: $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and $L = 6$ for EWMA control chart applied to WAP 596

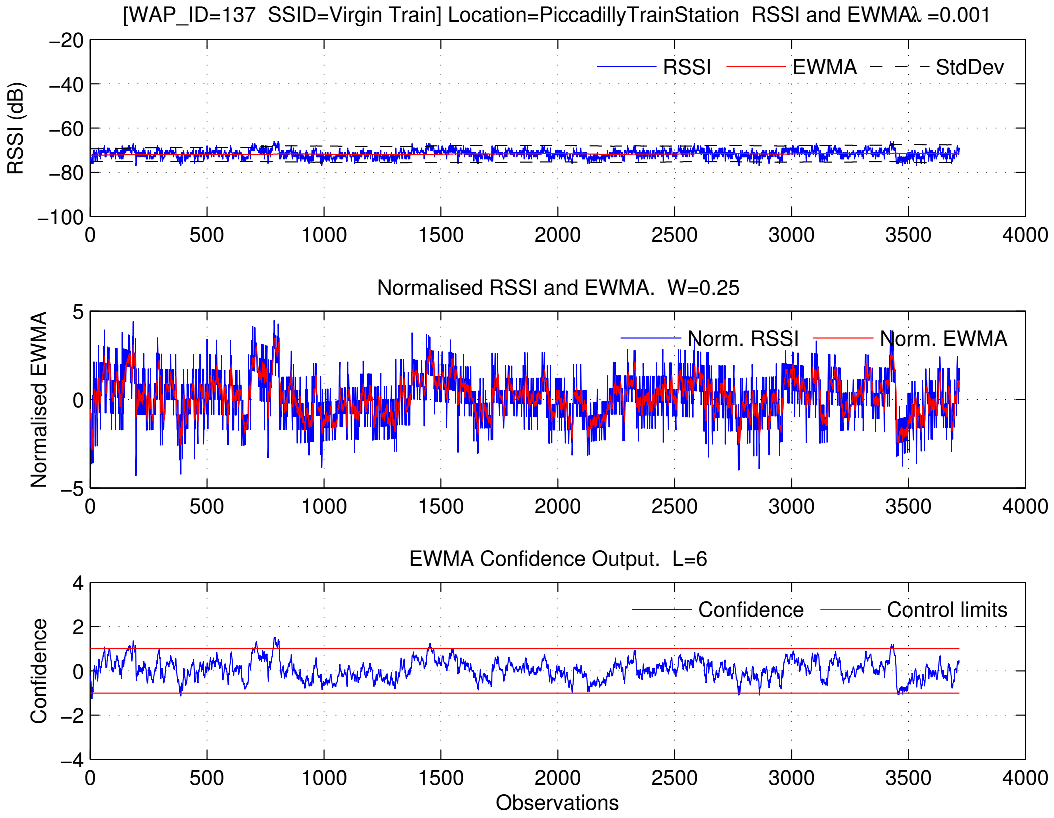


Figure 7.23: $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and $L = 6$ for EWMA control chart applied to WAP 137

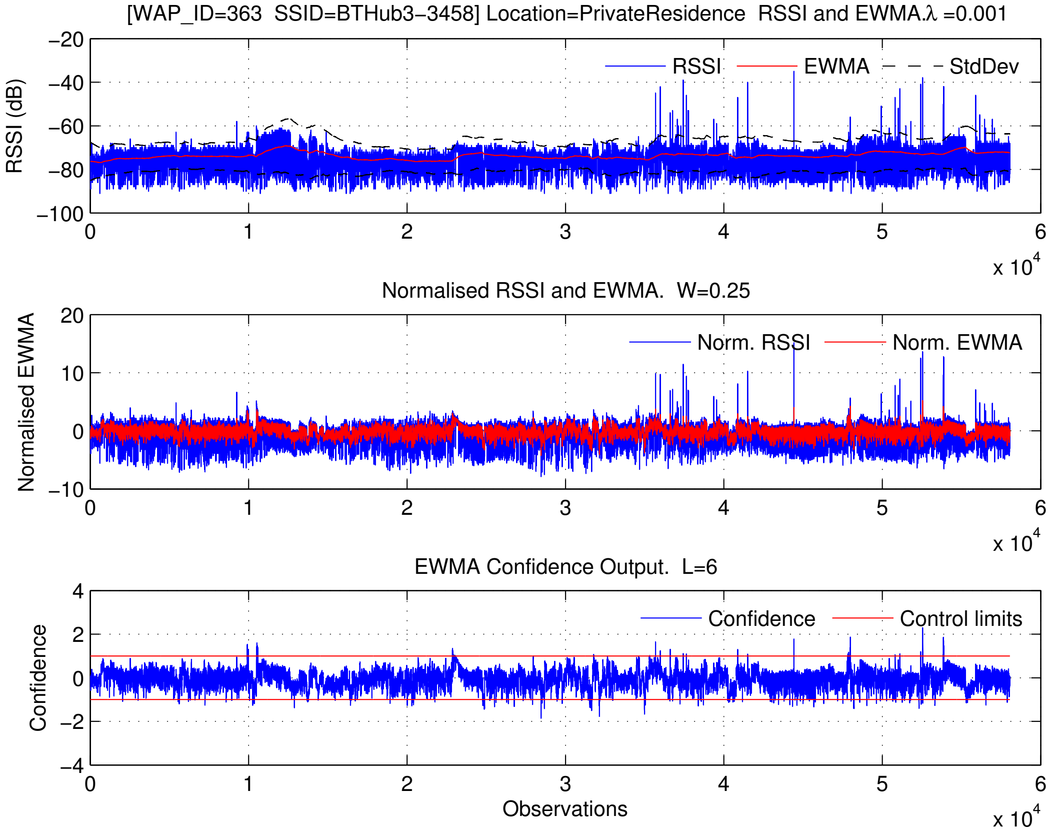


Figure 7.24: $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and $L = 6$ for EWMA control chart applied WAP 363

limits are placed at $\pm 6\sigma$ (L=6).

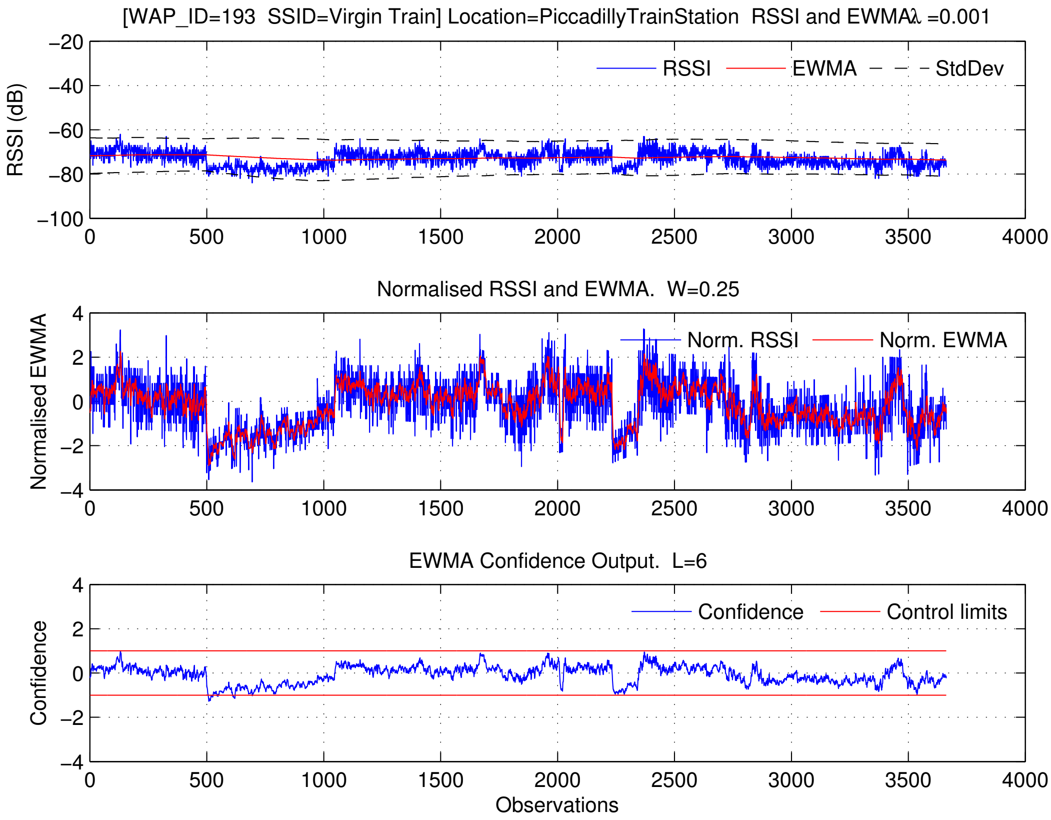


Figure 7.25: EWMA control chart with parameters $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.05$ and $L = 6$ applied to WAP 193

From the analysis and figures presented, it can be said that $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.25$ and $L = 6$ are a suitable a set of values for the EWMA control chart. Detecting changes and outliers succesfully for WAPs within this dataset.

It is clear from the visualisations in these figures, that whenever the confidence output values exceeds the boundaries (illustrated with the red lines in the third subplot) then the WAP is in transition or unstable and therefore should not be included when doing a fingerprint signature comparison. It is also clear that when stability is restored then the confidence output values revert to being within the boundary and thus indicate that this particular WAP could be and should be again used in a fingerprint signature comparison. These results illustrate that EWMA can be a useful tool for all the considered cases; of ramps (Figure 7.27) and steps (Figures 7.25 and 7.26). Outliers as shown in figure in Figure 7.28, can be accepted if isolated, even if there are many values, as shown, meaning that the confidence values remain within bounds.

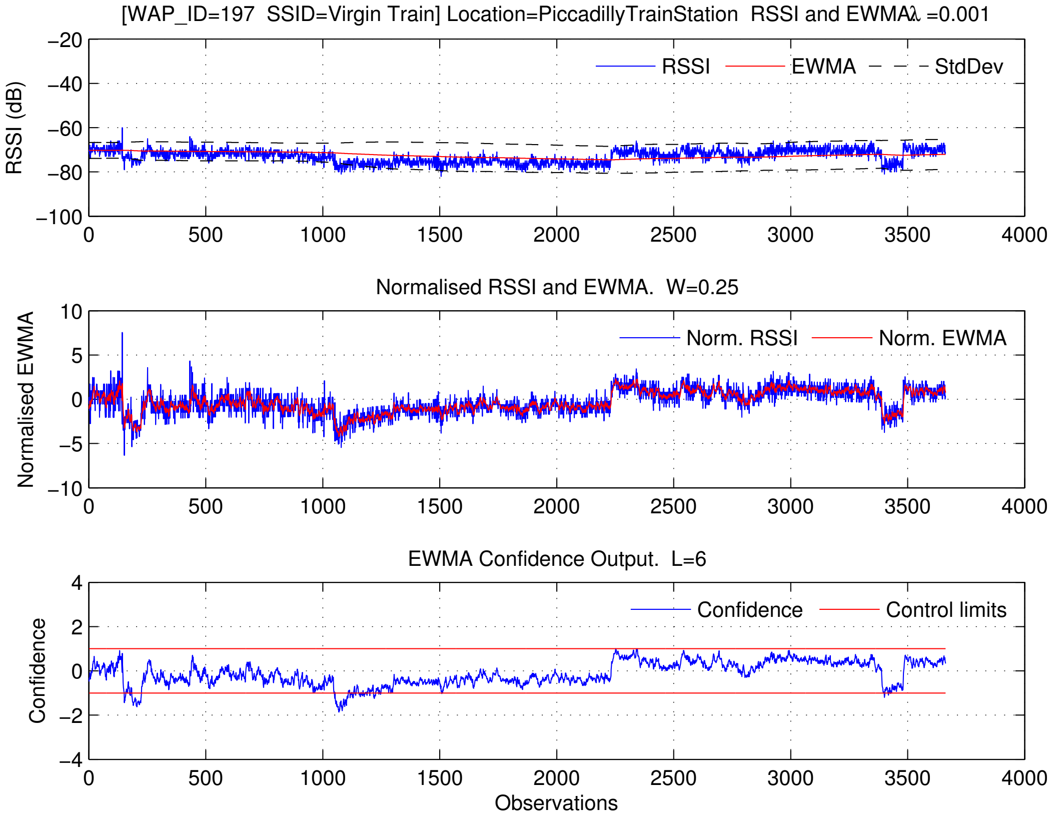


Figure 7.26: EWMA control chart with parameters $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.05$ and $L = 6$ applied to WAP 197

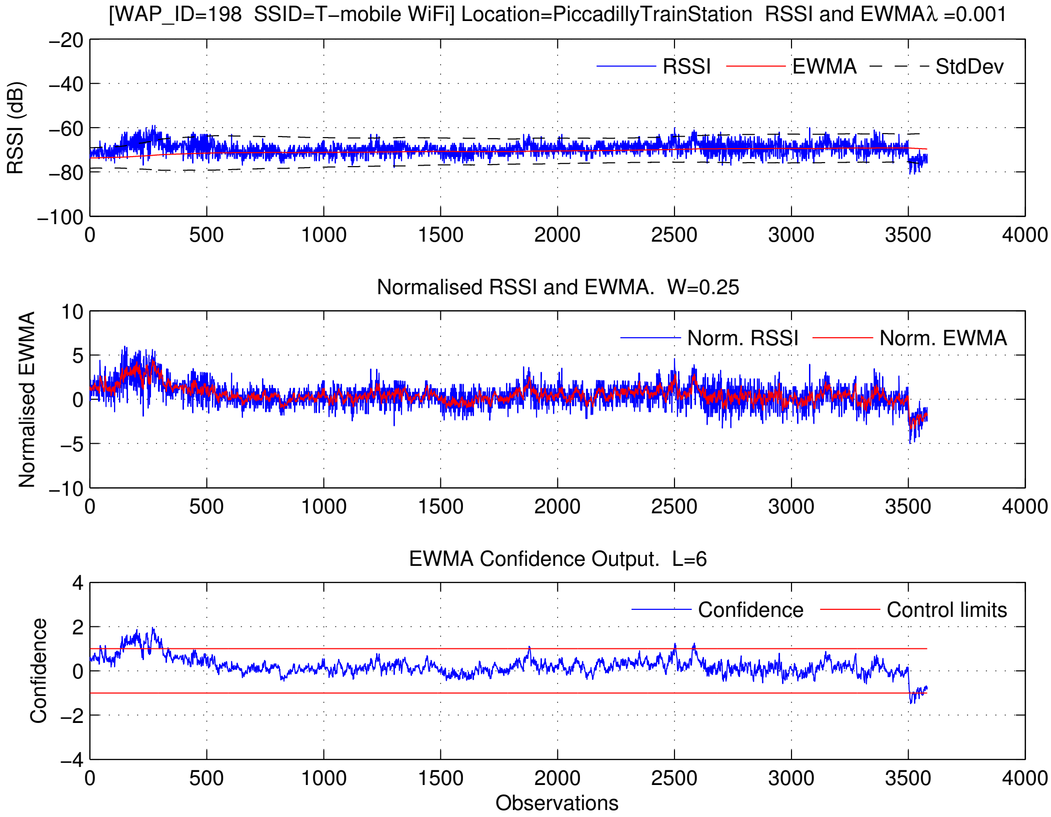


Figure 7.27: EWMA control chart with parameters $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.05$ and $L = 6$ applied to WAP 198

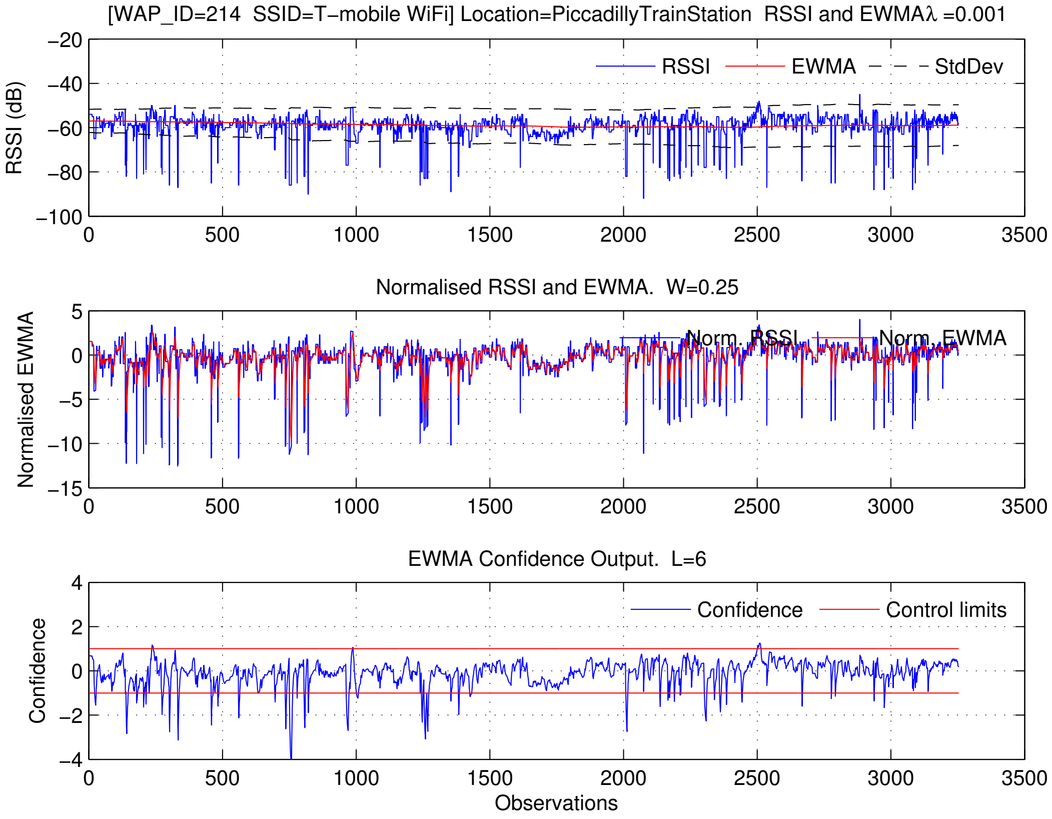


Figure 7.28: EWMA control chart with parameters $\lambda_{c1} = 0.001$, $\lambda_{c2} = 0.05$ and $L = 6$ applied to WAP 214

7.3.5 Filtering WAPs using EWMA Confidence

The addition or removal of a WAP can be determined by using the confidence boundaries. It can be detected when the WAP is out of control and should not be used for positioning until it is back under control. Even though it may be detected for the majority of the time. Effectively, the author is stating that appearance count alone does not provide robust decision making logic to manage the inclusion of a WAP into the fingerprint.

A solid example of the value of using control chart confidence levels is show in Figure 7.32.

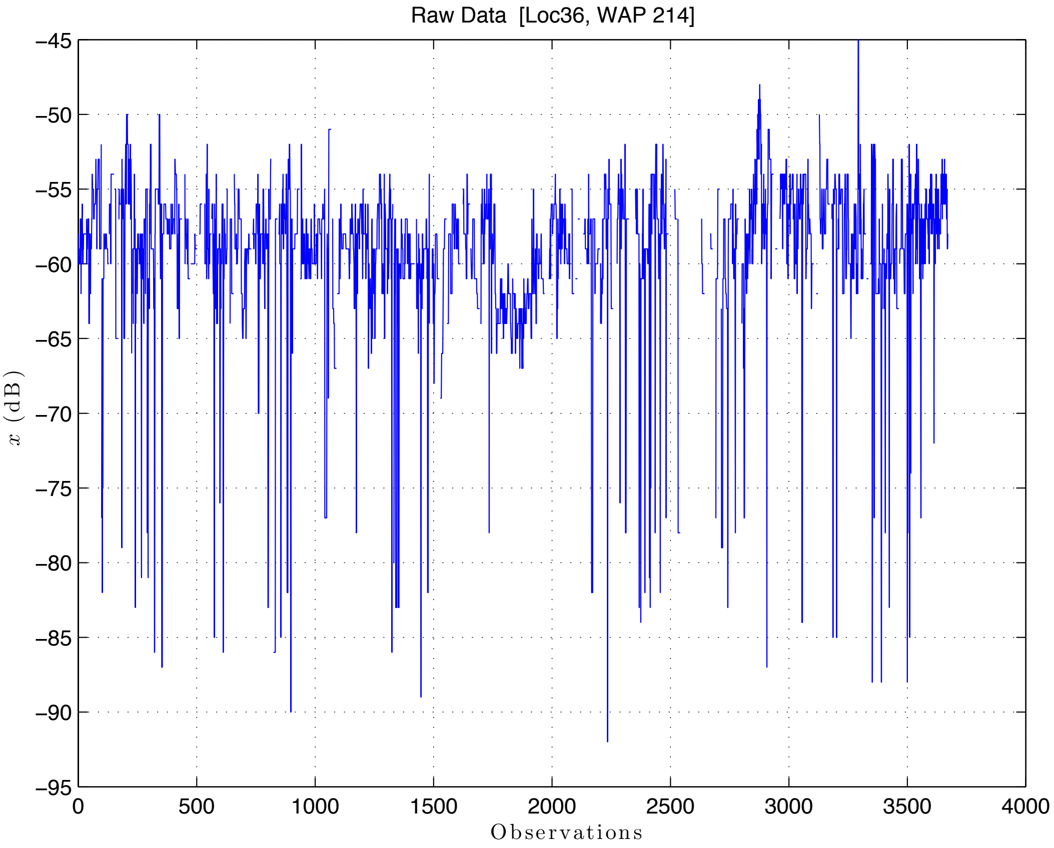


Figure 7.29: Raw data from WAP 214 at Location 36

Signal strength from WAP 214 (Figure 7.29) presents several data points that can be considered as outliers. In the presence of such characteristics the WAP should not be trusted. However, when analysing the scores computed by Gallagher (Figure 7.30) it is observed that the value *pending* only changes to 1 (one) briefly for when the WAP is not detected. Remembering that the value of *pending* determines when a WAP is included in the fingerprint, so when *pending* = 0 then the WAP is included in the fingerprint.

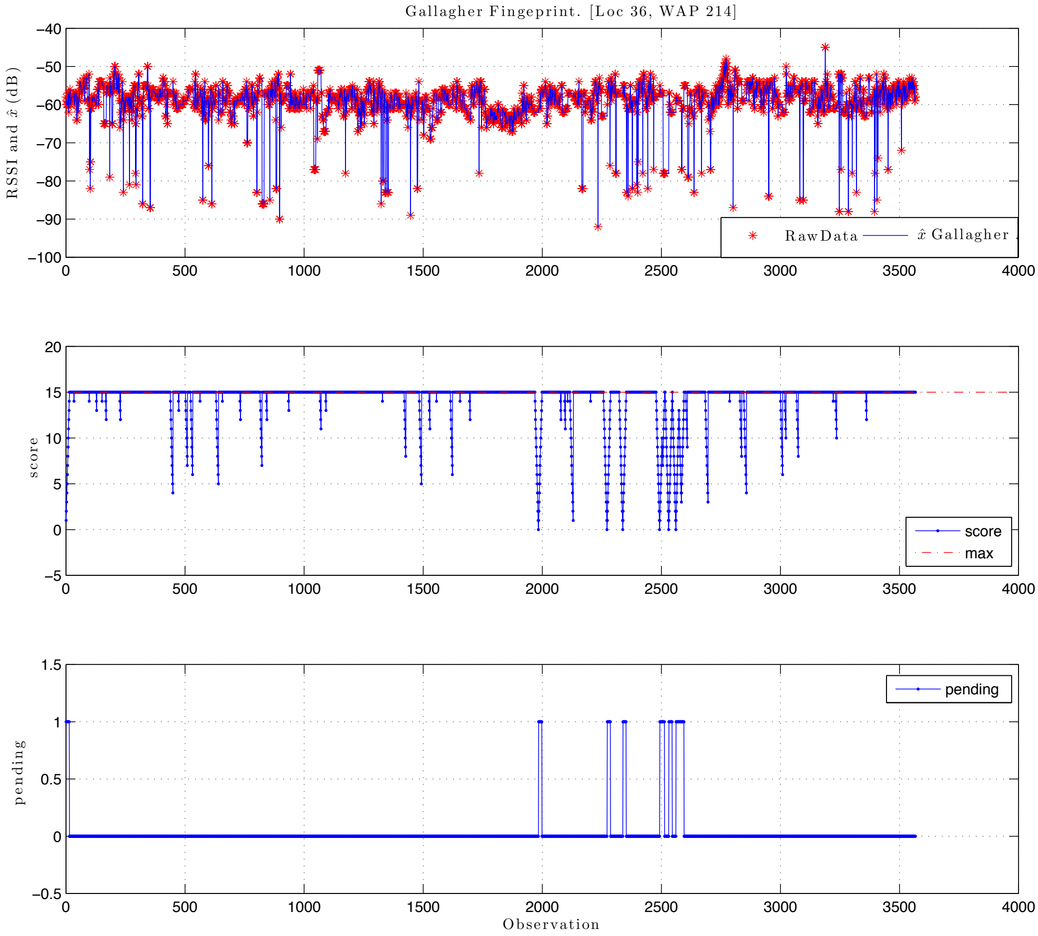


Figure 7.30: Gallagher Fingerprint for WAP 214 at Location 36

Considering Jun-Sung Fingerprint in Figure 7.31, this fingerprint technique has no mechanism for detecting a signal presenting outliers. Furthermore, even when the absence of the signal is reflected by the score the outliers in the signal strength are not reflected by the value of *pending*, as shown in the bottom plot from Figure 7.31. From these results it is clear that WAP 214 is indeed considered at all times for positioning when using the Jun-Sung algorithm.

Results from the EWMA control chart are presented in Figure 7.32. The *under control* parameter detects when data points shift away from the estimated value. The boundaries in this particular case are set at ±6σ. From the bottom plot in Figure 7.32 it is observed that parameter *under control* is actually detecting the outliers (shifts) in the signal from WAP 214. Here the control chart confidence states that the signal

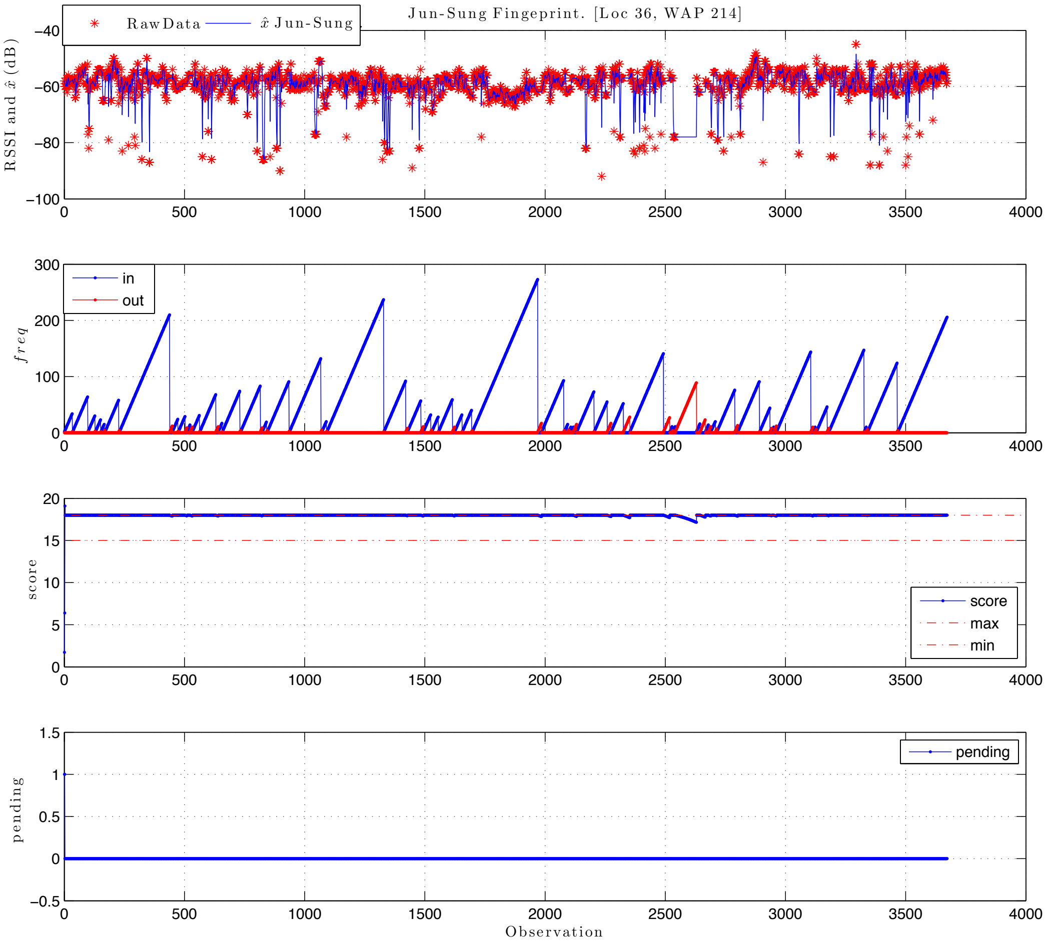


Figure 7.31: Jun-Sung Fingerprint for WAP 214 at Location 36

strength measurements have changed significantly and that the WAP should no longer be trusted. Furthermore, EWMA Fingerprints can detect when changes and outliers are occurring to a large number of WAPs in the fingerprint, which can be used as an indicator when a complete radio-map update is required.

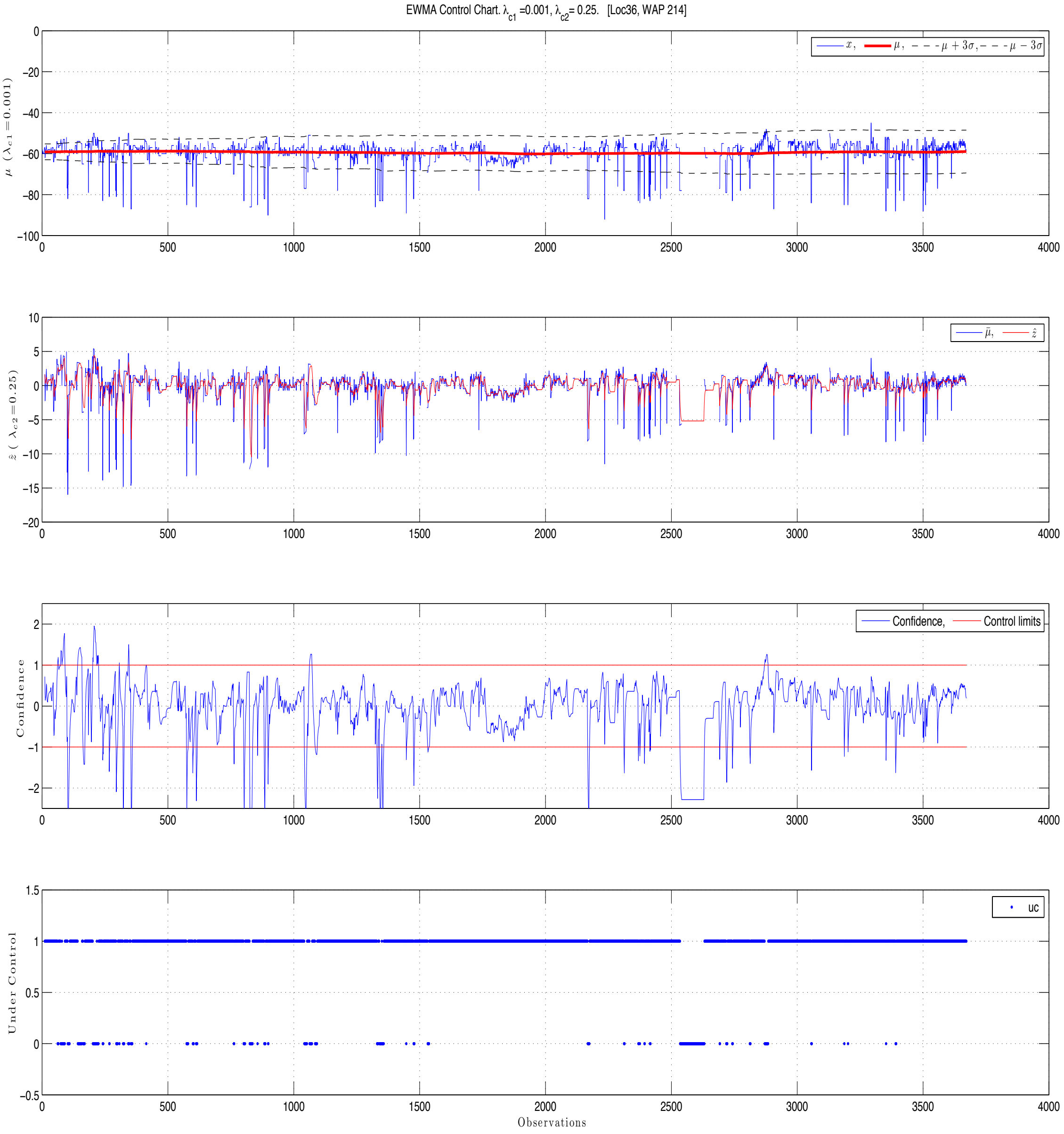


Figure 7.32: EWMA control chart computed with values $\lambda c1 = 0.001$,$\lambda c1 = 0.25$ and $L = 6$. The bottom plot is the *under control* parameter, which indicates when outliers are detected by EWMA control chart. Data from WAP 214, Location 36

Furthermore, Figure 7.33 presents the EWMA Fingerprint, where it is observed in the bottom plot that the score does respond to the outliers by decreasing it value, this will situate WAP 214 lower in the rank. The score along with the results from the EWMA control chart provide a more robust mechanism to filter out WAPs presenting outliers.

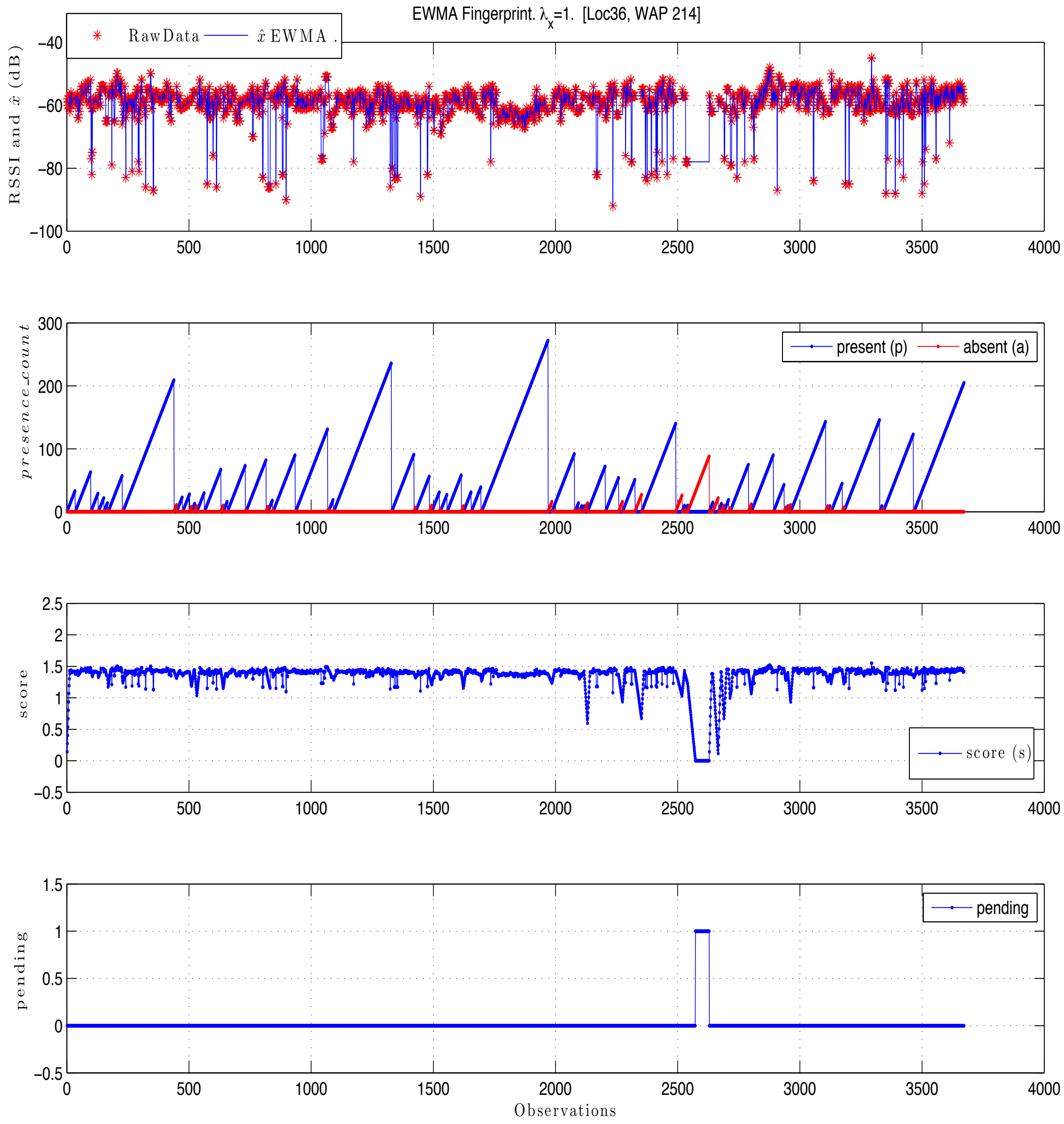


Figure 7.33: EWMA Fingerprint from WAP 214, Location 36

7.4 Comparing Score Functions

From analysing the algorithms presented by Gallagher et al. [GLDR10] and Jun-Sung et. al [LJYH13], it was noticed that the logic to add/remove the WAP to/from a fingerprint is based on a predefined count threshold. Although Jun-Sung goes further and introduces the exponential growth and decay, nevertheless it is still based on counting the number of appearances the WAP makes.

In order to address the addition and removal of the WAPs, two methods were proposed, the EWMA confidence value (discussed in previous section) and the score decay.

Accordingly, a simple version of the EWMA was employed for computing a gradient for score decay. In this case the weight (λ_p) is predefined as 0.5, with this same value being applied to all of the historical data and the current data. Nevertheless, this λ_p can be changed in the system settings. The analysis on modifying the weight λ_p is left for future work.

Figures 7.34, 7.35, 7.36 and 7.37 present a comparison of Gallagher, Jun-Sung and EWMA Fingerprints. Parameters *score* and *pending* from Gallagher and Jun-Sung Fingerprint and the parameter *uc* proposed in the EWMA control chart, are compared using the real-world data collected from WAP 527. Figure 7.34 shows the raw data for WAP 527, this WAP was detected in just a few, sparse observations. It illustrates clearly the advantage of the proposed scoring function.

EWMA Fingerprint for WAP 527 is shown in Figure 7.35, the *score* in EWMA Fingerprint is more responsive because the WAP signal strength, variation and averaged WAP presence $\hat{p}$ are considered. In comparison the score in Jun-Sung Fingerprint (Figure 7.37) reaches the acceptance window within a few observations, and it takes a larger amount of observations (minimum of 100) to remove it from the fingerprint. Therefore the proposed gradient function based on the metric m and averaged WAP presence $\hat{p}$ can be seen to provide a more data driven approach to reducing trust in the WAP.

It is observed in Figure 7.35 that EWMA score increases when the WAP is present, and decreases when the WAP is absent, but it goes beyond the inclusion o exclusion of the WAP. It is also a ranking mechanism within the fingerprint, and it affects the arrange of WAPs in the positioning stage. The score includes the WAP in the fingerprint for applications requiring as much WAPs as possible/available (e.g. Emergencies and rescue), but a low rank WAP can be leaved out for applications with not such requirements (e.g. finding a store within a mall).

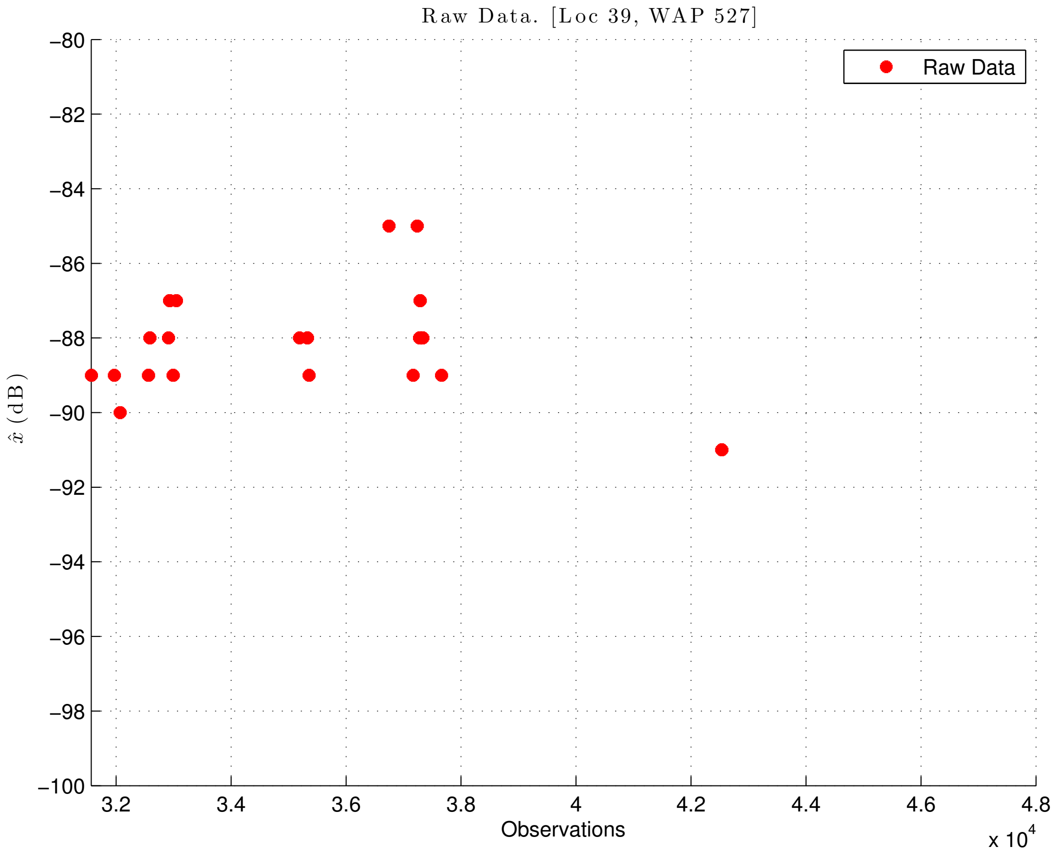


Figure 7.34: Raw data from WAP 527 at location 39

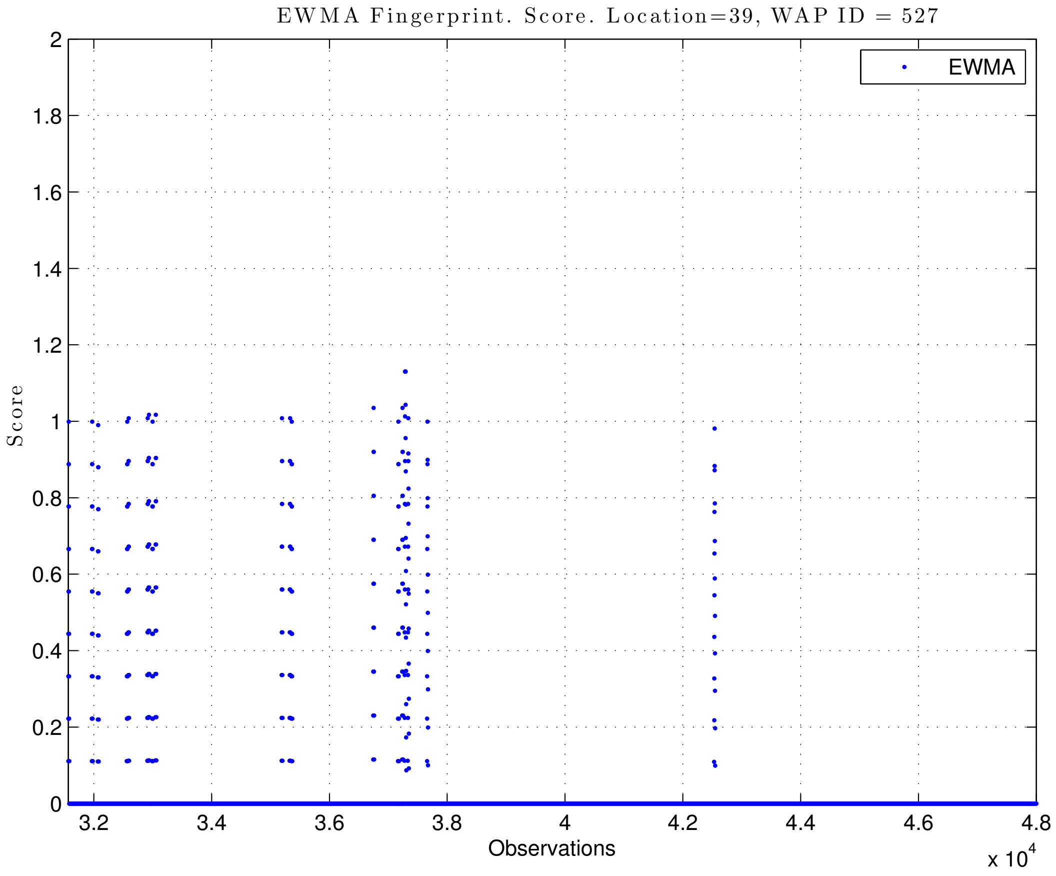


Figure 7.35: EWMA Fingerprint scoring function

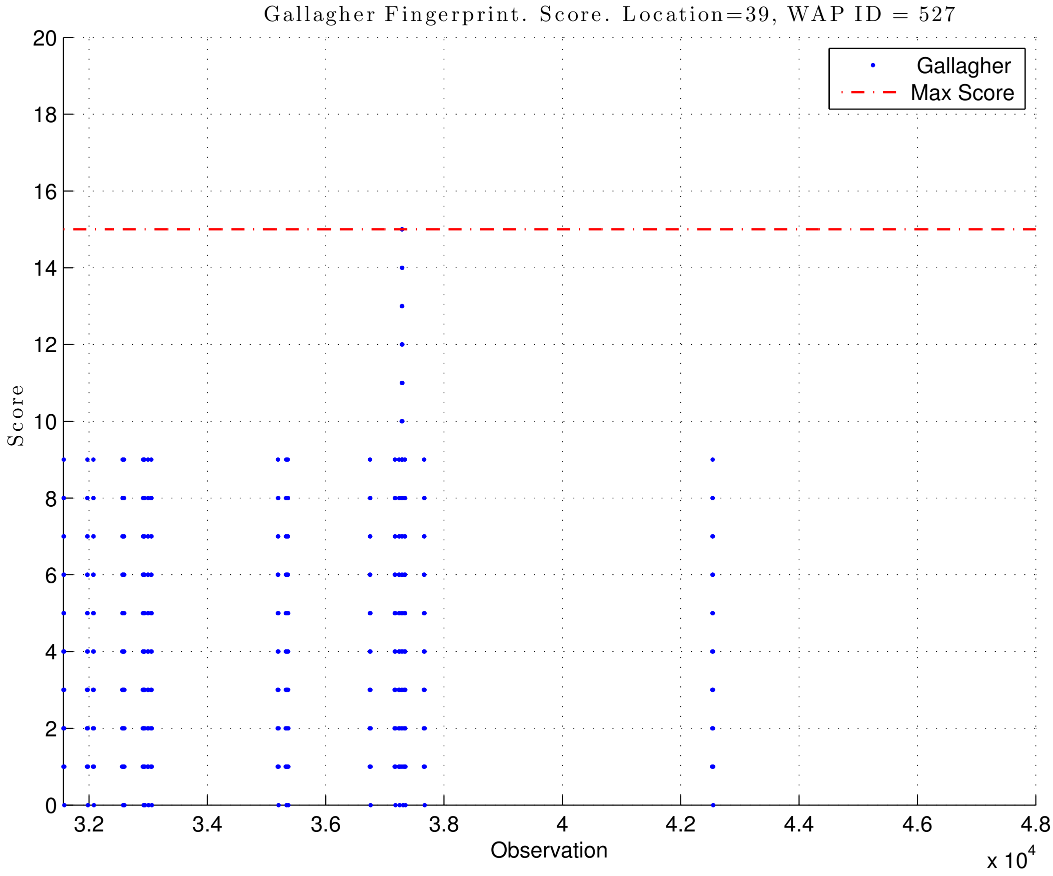


Figure 7.36: Gallagher score for WAP 527

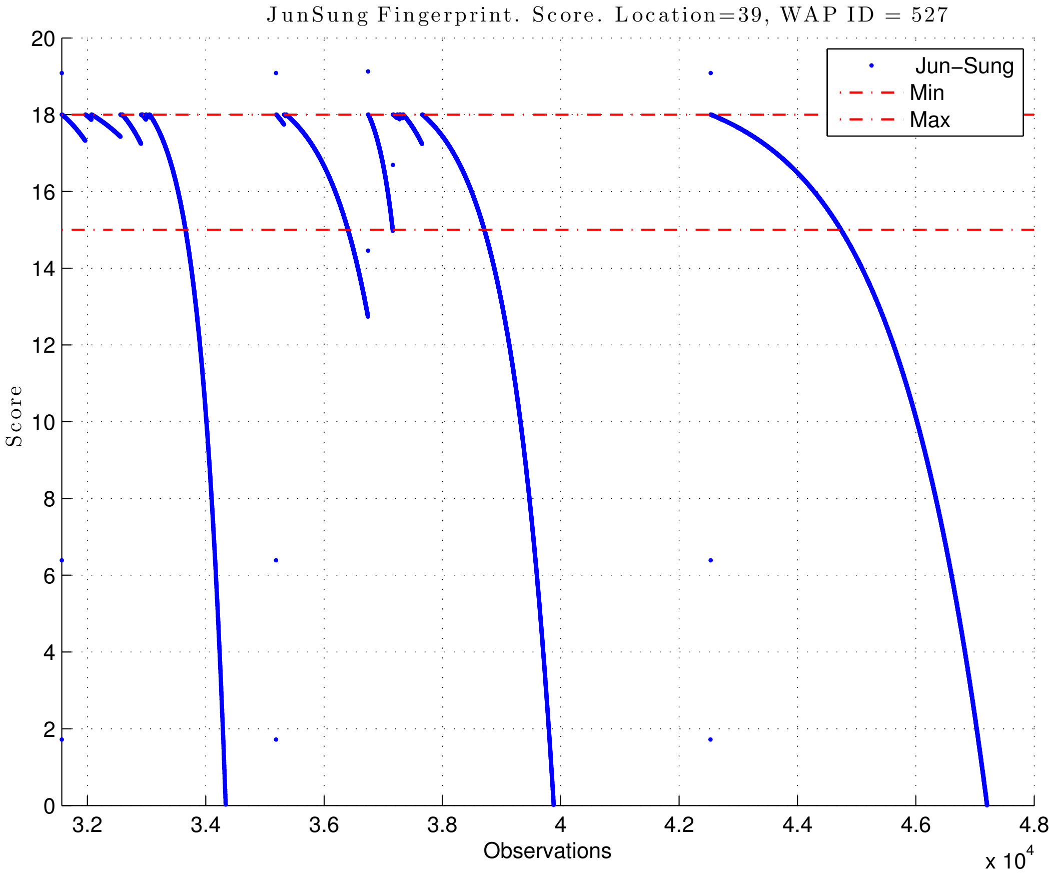


Figure 7.37: Jun-Sung scoring function
Jung-Sung's score decreasing slowly and staying in acceptance window for a rather long number of observations

On the other hand, the scores proposed by Jun-Sung and Gallagher have the single purpose of determine if the WAP should or not be included in the fingerprint by changing the value of parameter *pending*. As commented previously, a value of $pending = 0$ indicates that the WAP is used in the fingerprint. Figures 7.38, 7.39 and 7.40 reflect how the scores affect the WAP inclusion by modifying *pending*. Another disadvantage of Jun-Sung's technique is the large number of observations it takes to exclude the WAP from the fingerprint as seen in Figure 7.35. This potentially affects positioning results since the WAP is in the fingerprint although is has been absent for a large number of the observations.

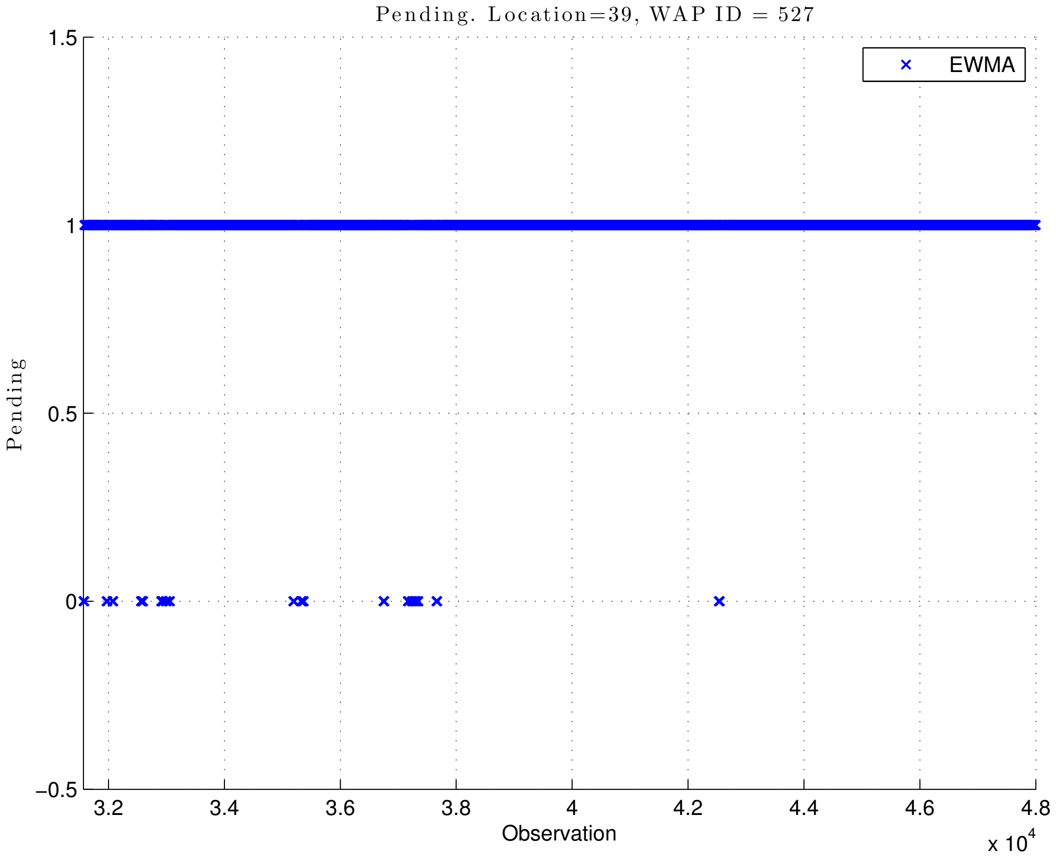


Figure 7.38: Historical values of parameter *pending* for WAP 527 computed with EWMA Fingerprint

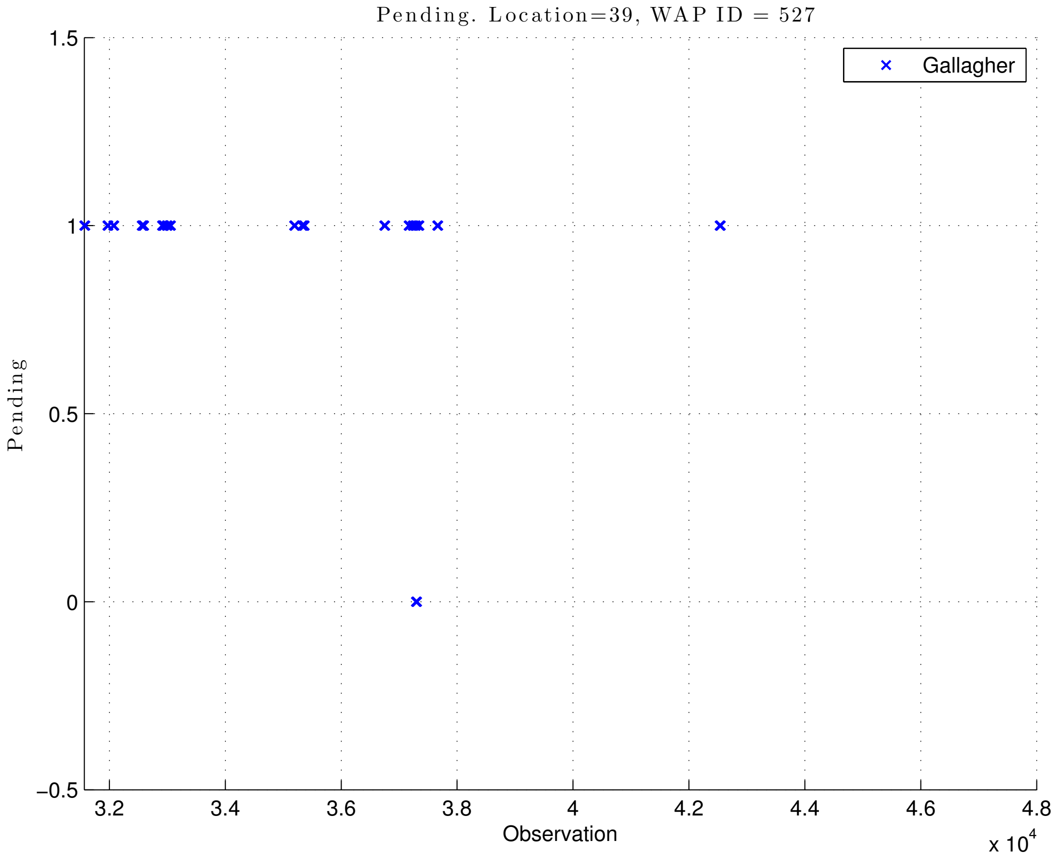


Figure 7.39: Historical values of parameter *pending* for WAP 527 computed with Gallagher Fingerprint

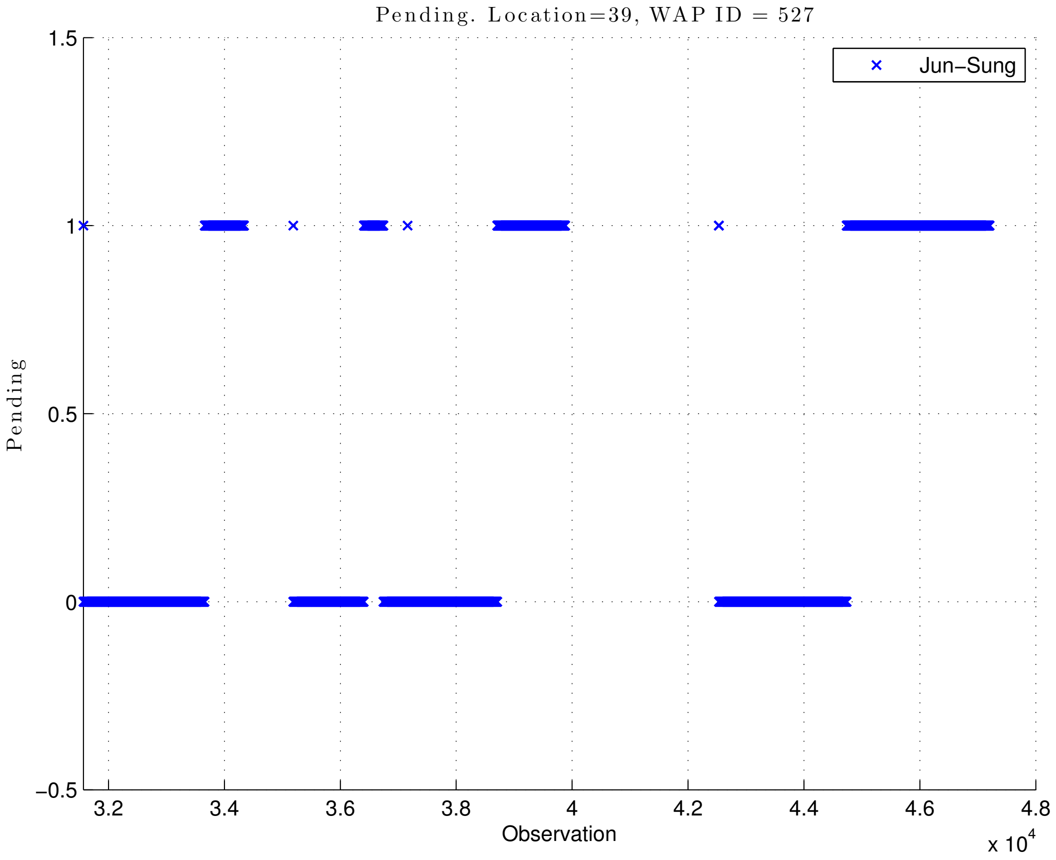


Figure 7.40: Historical values of parameter *pending* for WAP 527 computed with Jun-Sung Fingerprint

Chapter 8

Conclusions and Further Work

From analysing historical data streams from nearby WiFi Access Points (WAPs), captured over an extended period, a novel fingerprint definition based on control charts has been designed. Also, a metric for ranking WAPs was incorporated. This has led to an improved process of mapping and positioning over the tested WiFi indoor environments.

8.1 Contributions

This thesis makes the following contributions:

1. New method of evaluating the contribution of WAPs to fingerprinting. Consisting of the following:

    i. First time implementation of configurable EWMA for fingerprint-based Positioning Systems

    ii. Principle of WAP discrimination based on changes and outliers

    iii. Scoring function for ranking WAPs

2. Development of a unique and highly valuable dataset of observations

3. Database schema

4. Data capture scripts

8.1.1 Unique Data Set

From the collection of a large amount of data carried out over a long period of time we have shown graphically that there are time variant features in the signal strength profile of the WAPs. These features are likely to cause issues for IPS when considering the average as an estimated parameter. The findings presented increase the understanding of WiFi-fingerprint based IPS, having implications for their use to improve existing IPS.

8.1.2 EWMA and Control Charts

This work presents for the first time a system that incorporates a fingerprint algorithm based on a configurable EWMA control chart to improve fingerprint based IPS. The control chart is used primarily to include or remove WAPs from the fingerprints. Whereas an additional EWMA is used as a tuneable estimator to calculate the estimated value in the fingerprint. It was discovered that two implementations of EWMA were required in order to detect change points and estimate the fingerprint value, rather than a single implementation of EWMA attempting to accomplish both tasks.

8.1.3 Score Function

It was envisioned that the use of a metric for selecting the usable WAPs for fingerprinting is an efficient way to maintain a light weight and up-to-date radio-map. Developed initially in order to rank WAPs based on their signal strength, stability and variance there is a second advantage of using the score function as a discriminator. This states that even when the WAP is included into the fingerprint but its score is small it may not be considered; hence even if it has a strong signal if the score does not reach the potential good metric condition then the WAP is not included, due, for example, to low occurrence count.

8.1.4 General Conclusions

We conclude that timestamps and environment characteristics (scenarios) should be taken into account in order to develop an accurate, robust and scalable positioning system based on fingerprinting.

The control chart method that has been applied to the measurements has shown two key results. The first is that some process of smoothing the measurements is required in order to have an efficient and accurate estimation of the WAP signal strength. The second is that locations of 'change points' can be accurately detected by using the control chart. These results lay an important basis for the construction and comparison of fingerprints for an IPS. From this work it has been demonstrated that changes to the environment can be detected by identifying change points in signal strength behaviour. This in an important result that takes into consideration the history of the time series measurements.

It is concluded that the EWMA algorithm is an effective way to improve the detection of RSSI changes for fingerprinting based IPS.

This work investigates the possibility of localisation based on devices emitting electromagnetic signals, focusing mainly in the frequency used by WiFi communications. Although, it is argued that it is possible to achieve improvements in localisation by combining other data with the WiFi readings, for example using vision enabled devices and incorporating other bands of the electromagnetic spectrum to create a more accurate mapping.

Key results show that analysis of timestamps play a crucial role within WiFi-based positioning. These results are important for increasing the understanding of the performance of WiFi-based Indoor Positioning Systems that might be used over very long time periods and in large and complex environments.

8.2 Further Work

Following on from the results, an in-depth study of different smoothing constants and potentially different smoothing algorithms will be carried out. Also the methodology of using the change points will be explored. In order to maximise the potential of dynamically adjusting the fingerprint based on changes to signal strength trends.

Development of software for capturing observations from a mobile device.

Consideration of emerging IEEE standards, and its impact on existing fingerprint based positioning systems.

Exploring the inclusion of visual information into the fingerprint definition, to analyse the benefits and drawbacks of adding images into the fingerprinting and positioning processes.

As future work it is proposed to analyse the data collected on board trains whilst in motion, as this presents a challenging analysis of a dynamic environment.

Bibliography

[AKS04] Ankur Agiwal, Parakram Khandpur, and Huzur Saran. LOCATOR: Location Estimation System For Wireless LANs. In *Proceedings of the 2nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, pages 102–109, 2004.

[ANK13] M.M. Atia, A. Noureldin, and M.J. Korenberg. Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *Mobile Computing, IEEE Transactions on*, 12(9):1774–1787, 2013.

[App85] Gerald Appel. *The Moving Average Convergence-Divergence Trading Method.* Traders Press, June 1985.

[BB05] Mauro Brunato and Roberto Battiti. Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks*, 47(6):825–845, 2005.

[BBP00] P. Bahl, A. Balachandran, and V. Padmanabhan. Enhancements to the RADAR User Location and Tracking System. Technical report, 2000.

[BFC09] Andrew Barry, Benjamin Fisher, and Mark L. Chang. A long-duration study of user-trained 802.11 localization. In *Proceedings of the 2nd international conference on Mobile entity localization and tracking in GPS-less environments*, MELT'09, pages 197–212, Berlin, Heidelberg, 2009. Springer-Verlag.

[BGV92] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, pages 144–152, New York, NY, USA, 1992. ACM.

[BH97] Robert Grover Brown and Patrick Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering with MATLAB Exercises, 3rd Edition*. John Wiley & Sons, Inc., 1997.

[BP00] P. Bahl and V. N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2:775–784 vol.2, 2000.

[BRO56] RG BROWN. Exponential smoothing for predicting demand. *Cambridge, Massachusetts: Arthur D. Little Inc*, Philip Morris:15, Nov 1956.

[CCLH08] Yung-Mu Chen, Tein-Yaw Chung, Ming-Yen Lai, and Chih-Hung Hsu. Macd-based motion detection approach in heterogeneous networks. *EURASIP Journal on Wireless Communications and Networking*, 2008(1):540873, 2008.

[CGL83] Tony F. Chan, Gene H. Golub, and Randall J. LeVeque. Algorithms for computing the sample variance: Analysis and recommendations. *The American Statistician*, 37(3):242–247, 1983.

[CK02] Yongguang Chen and Hisashi Kobayashi. Signal strength based indoor geolocation. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 1, pages 436–439, 2002.

[Cro87] Stephen V. Crowder. A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics*, 29(4):pp. 401–407, 1987.

[CY08] Polona K. Carson and Arthur B. Yeh. Exponentially weighted moving average (ewma) control charts for monitoring an analytical process. *Industrial & Engineering Chemistry Research*, 47(2):405–411, 2008.

[DC11] Brett Dawes and Kwan-Wu Chin. A comparison of deterministic and probabilistic methods for indoor localization. *Journal of Systems and Software*, 84(3):442 – 451, 2011.

[DMM12] L.B. Del Mundo and R.S. Macatangga. Hybrid classifier for Wi-Fi fingerprinting system. In *ICT Convergence (ICTC), 2012 International Conference on*, pages 107 –112, October 2012.

[EM06] Frédéric Evennou and François Marx. Advanced integration of WIFI and inertial navigation systems for indoor mobile positioning. *EURASIP J. Appl. Signal Process.*, page 164, 2006.

[FCC10] Viacheslav Filonenko, Charlie Cullen, and James Carswell. Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pages 1–8, September 2010.

[FLMG13] A Farshad, Jiwei Li, M.K. Marina, and F.J. Garcia. A microscopic look at WiFi fingerprinting for indoor mobile phone localization in diverse environments. In *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, pages 1–10, October 2013.

[GGB$^+$02] F. Gustafsson, F. Gunnarsson, Niclas Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund. Particle filters for positioning, navigation, and tracking. *Signal Processing, IEEE Transactions on*, 50(2):425–437, Feb 2002.

[GJ04] Youngjune Gwon and Ravi Jain. Error Characteristics and Calibration-free Techniques for Wireless LAN-based Location Estimation. In *Proceedings of the Second International Workshop on Mobility Management &Amp; Wireless Access Protocols*, MobiWac '04, pages 2–9, New York, NY, USA, 2004. ACM.

[GKT10] Björn Gressmann, Helge Klimek, and Volker Turau. Towards ubiquitous indoor location based services and indoor navigation. In *Positioning Navigation and Communication (WPNC), 2010 7th Workshop on*, pages 107–112, March 2010.

[GLDR10] T. Gallagher, Binghao Li, A.G. Dempster, and C. Rizos. Database updating through user feedback in fingerprint-based Wi-Fi location systems. In *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010*, pages 1–8, 2010.

[Gos13] Subrata Goswami. *Indoor Location Technologies*. Springer New York, 1 edition, 2013.

[HHZ$^+$14] F. Hoflinger, J. Hoppe, R. Zhang, A. Ens, L. Reindl, J. Wendeberg, and C. Schindelhauer. Acoustic indoor-localization system for smart phones. In *Multi-Conference on Systems, Signals Devices (SSD), 2014 11th International*, pages 1–4, February 2014.

[HPALP09] Ville Honkavirta, Tommi Perala, Simo Ali-Loytty, and Robert Piche. A comparative survey of WLAN location fingerprinting methods. In *Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on*, pages 243–251, March 2009.

[HRF$^+$04] Andreas Haeberlen, Algis Rudys, Eliot Flannery, Dan S. Wallach, Andrew M. Ladd, and Lydia E. Kavraki. Practical robust localization over large-scale 802.11 wireless networks. In *in Proceedings of the 10th Annual International Conference on Mobile Computing and Networking - MOBICOM 2004*, pages 70–84, September 2004.

[JW08] Iris A. Junglas and Richard T. Watson. Location-based services. *Commun. ACM*, 51(3):65–69, March 2008.

[Kaa03] Eija Kaasinen. User needs for location-aware mobile services. *Personal Ubiquitous Comput.*, 7(1):70–79, May 2003.

[Kal60] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[KKJ$^+$04] P. Krishnan, A.S. Krishnakumar, Wen-Hua Ju, Colin. Mallows, and Sochin N. Gamt. A system for LEASE: location estimation assisted by stationary emitters for indoor RF wireless networks. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1001–1011, March 2004.

[KPV07] A. Kushki, K.N. Plataniotis, and A.N. Venetsanopoulos. Kernel-Based Positioning in Wireless Local Area Networks. *Mobile Computing, IEEE Transactions on*, 6(6):689–705, June 2007.

[LCC$^+$10] Timothy Liu, Matthew Carlberg, George Chen, Jacky Chen, John Kua, and Avideh Zakhor. Indoor localization and visualization using a human-operated backpack system. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pages 1–10, September 2010.

[LDBL07] Hui Liu, H. Darabi, P. Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, November 2007.

[LDGZ12] N. Le Dortz, F. Gain, and P. Zetterberg. WiFi fingerprint indoor positioning system using probability distribution comparison. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 2301–2304, 2012.

[LHC13] Yan Luo, Orland Hoeber, and Yuanzhu Chen. Enhancing Wi-Fi fingerprinting for indoor positioning using human-centric collaborative feedback. *Human-centric Computing and Information Sciences*, 3(1):2, 2013.

[Lin74] Robert F. Ling. Comparison of several algorithms for computing sample means and variances. *Journal of the American Statistical Association*, 69(348):pp. 859–866, 1974.

[LJYH13] Jun-Sung Lim, Woo-Hyuk Jang, Gi-Wan Yoon, and Dong-Soo Han. Radio map update automation for wifi positioning systems. *Communications Letters, IEEE*, 17(4):693–696, 2013.

[LKJ$^+$14] Sangwoo Lee, Bonhyun Koo, Myungjun Jin, Chansik Park, Min Joon Lee, and Sunwoo Kim. Range-free indoor positioning system using smartphone with bluetooth capability. In *Position, Location and Navigation Symposium - PLANS 2014, 2014 IEEE/ION*, pages 657–662, May 2014.

[LSB$^+$90] James M. Lucas, Michael S. Saccucci, Robert V. Baxley, Jr., William H. Woodall, Hazem D. Maragh, Fedrick W. Faltin, Gerald J. Hahn, William T. Tucker, J. Stuart Hunter, John F. MacGregor, and Thomas J. Harris. Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics*, 32(1):1–29, January 1990.

[LSDR06] B. Li, J. Salter, A. G. Dempster, and C. Rizos. Indoor positioning techniques based on wireless LAN. *First IEEE International*, 2006.

[LWL$^+$05] B. Li, Y. Wang, H.K. Lee, A. Dempster, and C. Rizos. Method for yielding a database of location fingerprints in wlan. *Communications, IEE Proceedings-*, 152(5):580–586, Oct 2005.

[LYHS13] Joo-Yub Lee, Cheal-Hwan Yoon, P Hyunjae, and Jungmin So. Analysis of location estimation algorithms for wifi fingerprint-based indoor localization. In *The 2nd international conference on software technology*, volume 19, pages 89–92, 2013.

[Mad14] Kasra Madadipouya. An Examination And Report On Potential Methods Of Strategic Location-Based Service Applications On Mobile Networks And Devices. *International Journal of Managing Public Sector Information and Communication Technologies (IJMPICT)*, 5(3), September 2014.

[MHYM13] P. Mirowski, Tin Kam Ho, Saehoon Yi, and M. Macdonald. SignalSLAM: Simultaneous localization and mapping with mixed WiFi, Bluetooth, LTE and magnetic signals. In *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, pages 1–10, October 2013.

[Mon08] D.C. Montgomery. *Introduction to Statistical Quality Control 6th Edition*. John Wiley & Sons Canada, Limited, 2008.

[MT11] Rainer Mautz and Sebastian Tilch. Survey of optical indoor positioning systems. In *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, pages 1–7, September 2011.

[Oks14] Irfan Oksar. A Bluetooth signal strength based indoor localization method. In *Systems, Signals and Image Processing (IWSSIP), 2014 International Conference on*, pages 251–254, May 2014.

[oRR] The Office of Rail Regulation. Estimates of station usage reports and data. [Online]. Available: http://www.railreg.gov.uk/server/show/nav.1529.

[PKL07] Shaun Phillips, Michael Katchabaw, and Hanan Lutfiyya. WLocator: An Indoor Positioning System. In *Wireless and Mobile Computing, Networking and Communications, 2007. WiMOB 2007. Third IEEE International Conference on*, pages 33–33, Oct 2007.

[PLM02] K. Pahlavan, Xinrong Li, and J.-P. Makela. Indoor geolocation science and technology. *Communications Magazine, IEEE*, 40(2):112–118, Feb 2002.

[Pri05] Nissanka B. Priyantha. *The cricket indoor location system*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2005.

[REI99] Konrad REIF. Stochastic stability of the discrete-time extended kalman filter. *IEEE Trans. Automatic Control*, 44:714–728, 1999.

[RFMT15a] Myrna M. Rodríguez-Frías, Tim Morris, and Martin J. Turner. WiFi Data Collection Script. [Online]. Available: http://dx.doi.org/10.5281/zenodo.18130, May 2015.

[RFMT15b] Myrna M. Rodríguez Frías, Tim Morris, and Martin J. Turner. WPSv1. [Online]. Available: http://dx.doi.org/10.5281/zenodo.19737, July 2015.

[RFTM14] Myrna M. Rodríguez-Frías, Martin J. Turner, and Tim Morris. Indoor Positioning Database. [Online]. Available: http://dx.doi.org/10.5281/zenodo.12913, Nov 2014. Wi-Fi measurements for Indoor Positioning Systems.

[RFTM15] Myrna M. Rodríguez-Frías, Martin J. Turner, and Tim Morris. WPS v1 Database schema and EER diagram. [Online]. Available: http://dx.doi.org/10.5281/zenodo.13793, January 2015.

[RK05] Ulf Rerrer and Odej Kao. Suitability of positioning techniques for location-based services in wireless lans. 2005.

[Rob59] S. W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1(3):pp. 239–250, 1959.

[RS05] Robert D. Reid and Nada R. Sanders. *Operations Management: An Integrated Approach*. John Wiley, 2005.

[SCSB03] Siddhartha Saha, Kamalika Chaudhuri, Dheeraj Sanghi, and Pravin Bhagwat. Location determination of a mobile device using IEEE 802.11b access point signals. In *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, volume 3, pages 1987–1992, March 2003.

[SD86] W. A. Shewhart and W. E. Deming. *Statistical method from the viewpoint of quality control*. Dover Publications, 1986.

[She31] W. A. Shewart. *Economic control of Quality of Manufactured Product*. Van Nostrand Reinhold Co., New York, 1931.

[Shu08] Lianjie Shu. An adaptive exponentially weighted moving average control chart for monitoring process variances. *Journal of Statistical Computation and Simulation*, 78(4):367–384, 2008.

[SL14] Aftab A. Memon Bhawani Shankar Chowdhry Ghazanfarullah Khan Shiraz Latif, Abdul Hadi. FaMWiFi: Fingerprinting/Sampling & Monitoring of WiFi signals for Indoor Positioning. First International Conference on Modern Communication & Computing Technologies (MCCT'14), 2014.

[TBF06] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT press, 2006.

[Thr02] Sebastian Thrun. Probabilistic robotics. *Commun. ACM*, 45(3):52–57, March 2002.

[urla] Beestar. [Online]. Available: http://www.beestar.eu. [Accessed: 15 June 2015].

[urlb] Ekahau Inc. Real-Time Location System (RTLS). [Online]. Available: http://www.ekahau.com. [Accessed: 15 June 2015].

[urlc] Fraunhofer IIS. [Online]. Available: http://www.iis.fraunhofer.de/en/ff/lok/tech/feldstaerke/rssi.html. [Accessed: 15 June 2015].

[urld] Here. [Online]. Available: http://here.com. [Accessed: 15 June 2015].

[urle] In-Location Alliance. [Online]. Available: http://www.in-location-alliance.com. [Accessed: 15 June 2015].

[urlf] MazeMap Inc. [Online]. Available: https://mazemap.com. [Accessed: 29 June 2015].

[urlg] Metaio Inc. http://www.junaio.com. [Accessed: 29 June 2015].

[urlh] Point Inside Inc. [Online]. Available: http://www.pointinside.com. [Accessed: 29 June 2015].

[urli] Qualcomm Atheros, Inc. [Online]. Available: http://www.qca.qualcomm.com. [Accessed: 29 June 2015].

[urlj] Skyhook. [Online]. Available: http://www.skyhookwireless.com. [Accessed: 29 Agust 2013].

[urlk] Ubisense. [Online]. Available: http://www.ubisense.net. [Accessed: 21 June 2015].

[VA14] B. Viel and M. Asplund. Why is fingerprint-based indoor localization still so hard? In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pages 443–448, March 2014.

[VDB76] A.J. Van Dierendonck and M. Birnbaum. Time Requirements in Tee NAVSTAR Positioning System (GPS) Global. In *30th Annual Symposium on Frequency Control. 1976*, pages 375–383, 1976.

[VSK$^+$10] Thorsten Vaupel, Jochen Seitz, Frédéric Kiefer, Stephan Haimerl, and Jörn Thielecke. Wi-Fi positioning: System considerations and device calibration. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pages 1–7, September 2010.

[VWG$^+$03] M. Vossiek, L. Wiebking, P. Gulden, J. Wieghardt, C. Hoffmann, and P. Heide. Wireless local positioning. *Microwave Magazine, IEEE*, 4(4):77–86, Dec 2003.

[WB95] Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. Technical report, Chapel Hill, NC, USA, 1995.

[web] Networkrail [online]. Available: http://www.networkrail.co.uk/manchester-piccadilly-station.

[Wel62] B. P. Welford. Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3):pp. 419–420, 1962.

[WHFaG92] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, January 1992.

[WJH97] Andy Ward, Alan Jones, and Andy Hopper. A new location technique for the active office. *Personal Communications, IEEE*, 4(5):42–47, 1997.

[XSC$^+$04] Z. Xiang, S. Song, J. Chen, H. Wang, J. Huang, and X. Gao. A wireless lan-based indoor positioning technology. *IBM J. Res. Dev.*, 48(5/6):617–626, September 2004.

[YAUS03] Moustafa A. Youssef, Ashok Agrawala, and A. Udaya Shankar. WLAN location determination via clustering and probability distributions. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 143–150, March 2003.

[ZFI13] Rosdiadee Nordin Zahid Farid and Mahamod Ismail. Recent advances in wireless indoor localization techniques and system. *Journal of Computer Networks and Communications*, vol. 2013(Article ID 185138, doi:10.1155/2013/185138):12 pages, 2013.

Appendix A

Database

A.1 Locations

Table A.1: Locations where data collection took place.

Location ID	Country	City	Building	Room
1	UK	Manchester	RCS	1.004a
2	UK	Manchester	RCS	1.004b
3	UK	Manchester	RCS	1.005_
4	UK	Manchester	RCS	1.006a
5	UK	Manchester	RCS	1.006b
6	UK	Manchester	RCS	1.009a
7	UK	Manchester	RCS	1.009b
8	UK	Manchester	RCS	1.011a
9	UK	Manchester	RCS	1.011b
10	UK	Manchester	RCS	1.021_
11	UK	Manchester	RCS	1.023_
12	UK	Manchester	RCS	1.026_
13	UK	Manchester	RCS	1.027_
14	UK	Manchester	RCS	1.028_
15	UK	Manchester	RCS	1.029_
16	UK	Manchester	RCS	1.030_
17	UK	Manchester	RCS	1.031a

Continued on next page

Table A.1 – continued from previous page

Location ID	Country	City	Building	Room
18	UK	Manchester	RCS	1.031b
19	UK	Manchester	RCS	1.032_
20	UK	Manchester	RCS	1.035_1.014
21	UK	Manchester	RCS	1.036_1.015
22	UK	Manchester	RCS	1.037_1.016
23	UK	Manchester	RCS	1.038_1.017
24	UK	Manchester	RCS	1.041_1.018
25	UK	Manchester	RCS	1.042_1.019
26	UK	Manchester	RCS	1.043_1.020
27	UK	Manchester	RCS	1.044_
28	UK	Manchester	RCS	1.046_
29	UK	Manchester	RCS	1.048a
30	UK	Manchester	RCS	1.048b
31	UK	Manchester	RCS	1.050_
32	UK	Manchester	RCS	1.051a
33	UK	Manchester	RCS	1.051b
34	UK	Manchester	RCS	1.056_
35	UK	Manchester	RCS	1.057_
36	UK	Manchester	PiccadillyTrainStation	Platform6
37	UK	Derby	DerbyTrainStation	Costa_Coffee_Premises
38	UK	Derby	Gym	Computer_Stations
39	UK	Derby	PrivateResidence	Living_Room
40	UK	Derby	Westfield	Food_Court
41	UK	Manchester	PiccadillyTrainStation	Platform5
42	UK	Manchester	KilburnBuilding	2.107
43	UK	Sheffield	SheffieldTrainStation	RoomPlatform8a
44	UK	Manchester	PiccadillyTrainStation	Platform4
45	UK	Manchester	PiccadillyTrainStation	Platform8
46	UK	Manchester	PiccadillyTrainStation	Platform9
47	UK	Manchester	University_Place	UoM_Food_Court
48	UK	Manchester	RCS	1.022_

A.2 Database Statistics

Table A.2: Database Statistics

Loc ID	Total Obs	Total WAPs	Average WAP count per Obs	Mean RSSI	StdDev RSSI	Mode RSSI	Median RSSI	Min RSSI	Max RSSI
1	4	26	24	-77.22	10.68	-85.00	-82.00	-88	-46
2	4	26	21	-79.82	7.20	-84.00	-82.50	-88	-58
3	8	42	24	-82.49	3.32	-86.00	-83.00	-88	-69
4	4	27	25	-72.96	8.56	-81.00	-74.00	-84	-49
5	4	32	30	-67.69	14.39	-67.00	-70.00	-87	-31
6	4	49	42	-74.35	9.37	-77.00	-77.00	-87	-46
7	4	54	47	-73.25	10.03	-86.00	-76.00	-88	-45
8	29	59	25	-73.23	13.22	-82.00	-78.00	-87	-25
9	31	58	27	-72.52	13.73	-84.00	-76.00	-90	-13
10	9	40	25	-77.39	8.50	-78.00	-80.00	-88	-48
11	16	37	24	-77.70	6.42	-86.00	-78.00	-89	-60
12	10	32	23	-71.30	10.29	-83.00	-73.00	-87	-52
13	11	40	26	-68.81	9.75	-67.00	-69.50	-87	-45
14	10	55	34	-74.68	6.33	-73.00	-74.00	-89	-58
15	15	53	32	-77.80	5.57	-77.00	-78.00	-87	-59
16	12	59	39	-76.90	7.81	-83.00	-79.00	-88	-57
17	10	59	37	-74.42	9.14	-83.00	-76.00	-88	-51
18	12	60	38	-74.30	9.89	-85.00	-77.00	-88	-44
19	22	56	31	-75.70	8.81	-85.00	-76.00	-88	-48
20	46161	141	32	-71.10	9.07	-69.00	-71.00	-92	-28
21	72	50	32	-70.78	10.68	-79.00	-73.00	-89	-25
22	80	49	29	-72.89	13.66	-77.00	-77.00	-90	-26
23	69	42	22	-72.19	14.74	-80.00	-78.00	-88	-25
24	67	42	20	-74.26	12.56	-85.00	-79.00	-89	-37
25	84	47	18	-75.36	10.38	-84.00	-78.00	-90	-44
26	86	45	18	-75.73	9.38	-85.00	-78.00	-91	-45
27	14	38	25	-75.89	8.10	-81.00	-77.50	-88	-49

Continued on next page

Table A.2 – continued from previous page

Loc ID	Total Obs	Total WAPs	Average WAP count per Obs	Mean RSSI	StdDev RSSI	Mode RSSI	Median RSSI	Min RSSI	Max RSSI
28	4	41	37	-74.08	9.05	-85.00	-76.00	-86	-53
29	4	36	36	-71.99	11.78	-68.00	-75.00	-86	-38
30	4	37	32	-75.21	10.64	-84.00	-79.00	-87	-48
31	5	43	41	-74.56	8.22	-78.00	-77.00	-89	-51
32	8	53	42	-75.93	9.03	-85.00	-79.00	-88	-50
33	5	52	48	-77.34	6.78	-80.00	-80.00	-87	-57
34	4	36	33	-72.08	8.42	-80.00	-71.00	-86	-52
35	18	42	26	-75.96	7.87	-86.00	-76.50	-88	-56
36	3727	555	47	-75.35	7.74	-78.00	-76.00	-92	-1
37	8954	113	11	-80.76	8.53	-87.00	-83.00	-98	-12
38	5331	14	5	-78.24	11.64	-89.00	-85.00	-92	-14
39	58132	97	23	-79.58	6.90	-81.00	-81.00	-96	-11
40	56	66	33	-66.88	11.18	-60.00	-68.00	-89	-39
41	210	111	50	-75.96	7.49	-80.00	-77.00	-92	-45
42	198	21	14	-70.04	15.17	-79.00	-77.00	-90	-33
43	1560	75	7	-78.12	10.07	-88.00	-80.00	-92	-18
44	15	58	47	-74.28	9.90	-83.00	-76.00	-88	-37
45	14	60	51	-74.21	9.11	-75.00	-75.00	-88	-42
46	15	58	47	-71.05	9.22	-75.00	-72.00	-87	-46
47	91	25	22	-70.49	14.08	-82.00	-74.50	-91	-47
48	6	24	16	-82.12	4.99	-83.00	-83.00	-89	-68

A.3 Enhanced Entity-Relationship Model

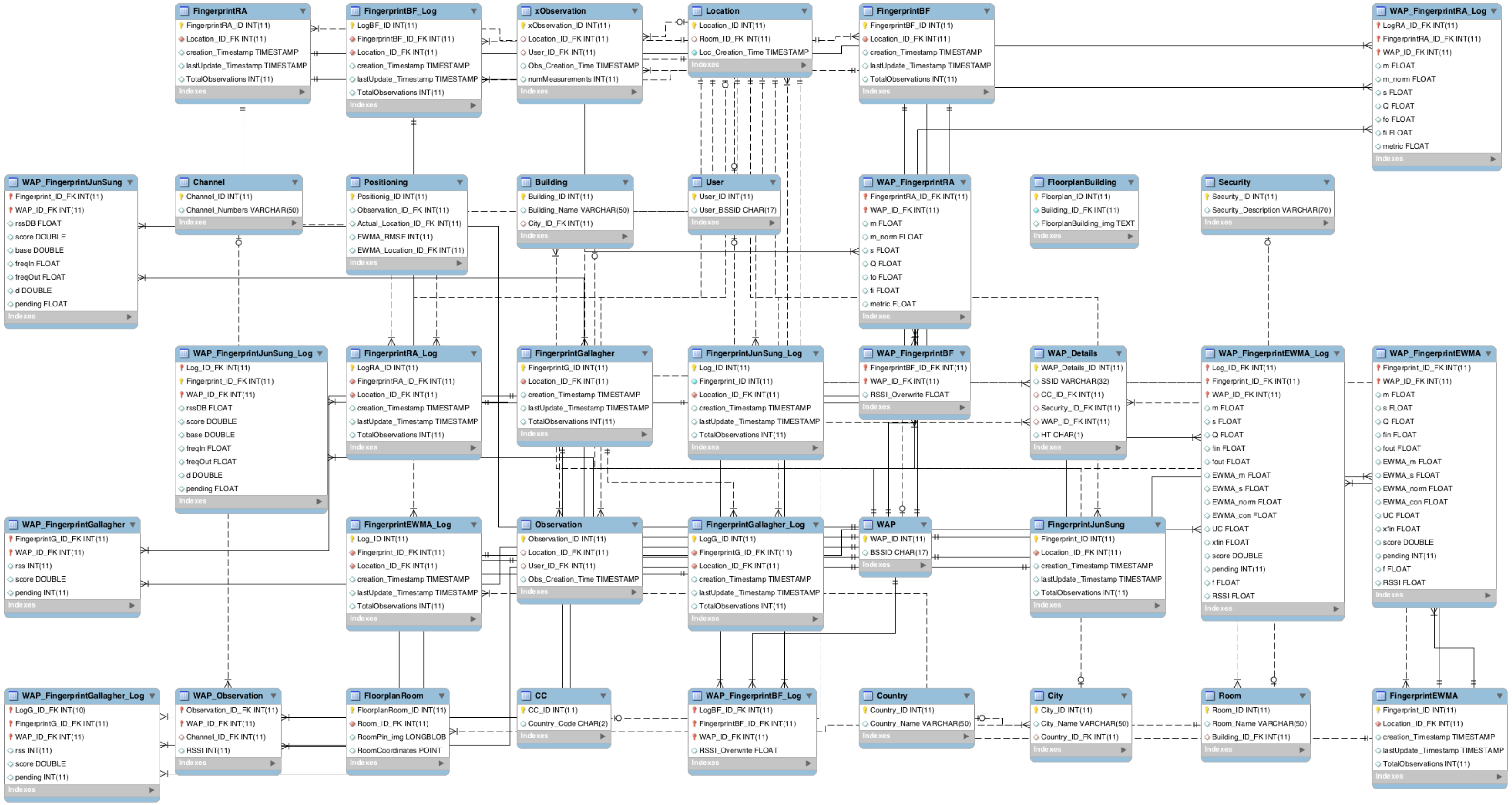


Figure A.1: Enhanced Entity-Relationship Model

Appendix B

Data Collection

B.1 Data Collection Script

```
#!/bin/bash
#WiFi Observations. Recording WiFi data from nearby wireless networks
# The Aiport Utility in OSX (10.10.1)
strAirport=”/System/Library/PrivateFrameworks/Apple80211.framework/Versions/Current/Resources/airport”
#Number of requests. It can be configurable according to the desired number of request
numRequests=4
# Waiting time (in seconds) between requests.
# It can be configurable according to the desired waiting time
waitInSeconds=5
# Obtain directory where script is installed, and where observations will be stored
SCRIPTPATH=$(dirname ”$BASH_SOURCE”)
# Change directory
cd $SCRIPTPATH
# Name of the folder for the session of observations
strFolderDate=$(date ”+%Y-%m-%d_%H-%M-%S”)
# Reads the name of the room (You have 10 seconds for entering the name of the room, otherwise the default ”Unknown” will be set)
echo -n ”Enter location name (with no spaces, e.g. Office_1) >”
if read -t 10 roomNameText; then #The -t option followed by a number of seconds provides an automatic timeout for the read command.
```

```
    echo ”You entered location name: $roomNameText”
else
    roomNameText=”Unknown”
    echo ”Time out”
fi
echo ”Room name: $roomNameText”
strName=$roomNameText@$strFolderDate
strFolder=”New_Observations/${strName}”
# Create the new folder
mkdir -p $strFolder
# Requests
for i in `seq 1 $numRequests`;
do
    echo ”Request $i ... wait”
    # Creating the string str with the date and the WiFi data
    strDate=$(date ”+%Y-%m-%d_%H-%M-%S”)
    # Probe Request
    strWiFi=$($strAirport -s)
    # Saving Response to a File
    echo ”$strWiFi” > $strFolder/”$strDate”.txt
    sleep $waitInSeconds
done
echo ”[- Done -] ”
echo ”A session of $numRequests Observations has been recorded.”
echo ”File $SCRIPTPATH/$strFolder”
```

B.2 Instructions for Data Collection Script

The script WiFiSamplingScript records WiFi data from nearby wireless networks.
This script employs the Aiport Utility in OSX (10.10.1)
Instructions

1. Crate in your computer a new directory (e.g. WiFiData), make sure the name of the file and its path contains no spaces. (e.g. /Users/Myrna/Desktop/WiFiData)
2. Copy WiFiSamplingScript.exe into the directory
3. Double click on the script WiFiSamplingScript.exe will execute the program

4. A new terminal will open
5. It will ask immediately to enter the location name. (This version of the script does not support spaces in the name. Enter a location with no spaces or use a underscore instead.) It will wait for 10 seconds, if no name is entered, it will be named unknown by default.
6. The script is set to do four requests. (This is configurable by modifying the variable numRequest within the script).
7. After every request, will be a waiting time of 5 seconds. (The waiting time is also configurable by modifying the variable waitInSeconds)
8. Once the requests have completed, the script will display [- Done -] and the number of observation that were recorded.
9. Also displayed is the path and directory created, which contains the WiFi observations recorded.
10. Repeat from step 3. for each location to be sampled.

Appendix C

Script for λ_x Identification

```
function IPSv2_generateLambdaEWMAse
% Computes the optimal LambdaX per WAP for all the Locations

    tic
    k0 =10; % Initialization (The Lambda is computed for WAPs that
       are being in observed 10+ times)

    allLocations = false; %[true-Compute LambdasX for all locations]
        [false-Compute LambdasX for one location]
    allWAPs = true; %[true-Compute LambdasX for all WAPs per
       Location] [false-Compute LambdasX for a specific WAP per
       location]

    % LambdasX
    numberPoints = 100; %Number of Lambdas to be evaluated
    Y = linspace(0.0,1.0,numberPoints);
    allLambdas = roundn(Y,-2);

    if (allLocations==true)
      for im =1:48
          strLocation_ID = num2str(im);
      end
    else
       % Selecting the Location to be analysed
       % strLocation_ID = '20'; %Office 1.035, RCS
       strLocation_ID = '36'; %Platform6, Piccadilly
       % strLocation_ID = '39'; % Living Room, Private Residence
```

```
end


if (allWAPs == true)
    %% All WAPs
    [allWAP_IDs_perLoc] = fnSQL_v2_LoadWAPsPerLocation(
       strLocation_ID);

    %% All the APs Sorted By Observation Count (DESC)
    WAPs = allWAP_IDs_perLoc;

else
    % Specific WAPs
     if strcmp(strLocation_ID, '36')
        % top10Piccadilly =
           [137;127;150;161;194;199;197;195;193;146];
        top1Pic=137;
        WAP_IDs_perLoc = top1Pic;
    elseif strcmp(strLocation_ID, '20')
        top1RCS = 63;
        %top10RCS =[63;33;42;49;52;68;67;83;65;55];
        WAP_IDs_perLoc = top1RCS;
    elseif strcmp( strLocation_ID, '39')
        %top10PR =[363;360;350;361;353;352;365;358;362;370];
        top1PriRes = 363;
        WAP_IDs_perLoc = top1PriRes;
     end


    %% WAPs
    WAPs = WAP_IDs_perLoc;
end


numWAPs = size(WAPs,1); %Number of WAPs

matWAP_minError_lambda = zeros(1, 5); %initilizing the matrix
   for results
l=0;

for i=1:numWAPs
    currentWAP = WAPs(i);
```

```
        strAP_ID = sprintf('%d',WAPs(i));

        %% Loading Observations (Per Location and WAP)
        cellAllMeasurementsPerAP=
           fnSQL_ObservationsPerWAPorderedByCreationTime(
           strLocation_ID,strAP_ID);
        iRSSI = cell2mat(cellAllMeasurementsPerAP(:,1));
        rowsRSSI = size(iRSSI,1); %Size of RSSI vector

        if (rowsRSSI > k0) % Checking that the size of the vector is
             bigger than k0
            matSumErrors =  zeros(1,numberPoints);

            for j=1:numberPoints
                        currentLambdaX = allLambdas(j);
                        [vecError]=IPSv2_fnEWMA_Lambda(iRSSI,
                           currentLambdaX, k0);
                        matSumErrors(j) = sum(vecError);
            end

            minError =min(matSumErrors); % Finding the lambda with
               the smaller error
            minErrorIndx = find(matSumErrors==minError);

            m = mean(iRSSI);
            s = std(iRSSI);

            l=l+1;
            matWAP_minError_lambda(l,:) = [currentWAP,m,s,minError
               (1),allLambdas(minErrorIndx(1))];

        end
    end
   %% Inserting the Lambda in the Database

    if (l>0)
      fnSQL_InsertNewLambdaEWMAseInDB(matWAP_minError_lambda,
         strLocation_ID);
    end

toc
end
```

Appendix D

Timing on computing λ_x for all WAPs per Location

Location	**Total Observations**	**Total WAPs**	**Number WAPs with λ_x**	**Total time on computing λ_x (seconds)**
Platform 6 Piccadilly Railway Station	3,727	555	459	276.99
Office 1.035 RCS, UoM	46,161	141	126	162.37
Private Residence Derby, UK	58,132	97	92	140.54

Table D.1: Timing on computing λ_x for all WAPs per Location