WI-FI BASED LOCALIZATION USING RSSI FINGERPRINTING

By

Ahsan Javid

A thesis submitted to the Faculty of Electrical Engineering Department, Military College of Signals, National University of Sciences & Technology, Pakistan in partial fulfillment of the requirements for the degree of MS in Electrical (Telecom)

July 2018

<u>THESIS ACCEPTANCE CERTIFICATE</u>

Certified that final copy of MS/MPhil thesis written by Mr/~~MS~~ **<u>Ahsan Javid</u>,** Registration No. **<u>NUST201463814MMCS25014F</u>** of **<u>Military College of Signals</u>** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: ___________________________________

Name of Supervisor: <u>Asst Prof Mir Yasir Umair, PhD</u>

Date: _______________________________________

Signature (HoD): _____________________________

Date: _______________________________________

Signature (Dean): ____________________________

Date: _______________________________________

ABSTRACT

Indoor localization has become inevitable in the technological sector especially when a majority of stand-alone as well as commercial applications require accurate location estimate for the users. Since GPS signals are insufficient for indoor localization of object Therefore, Wi-Fi based fingerprinting have received great attention for such localization, among the available Wi-Fi fingerprint-based localization techniques, K-Nearest Neighbor (KNN) is most popular. However, existing techniques failed to consider the fact that equal RSS spaces at different RSS level may not mean identical geometric distances in complex indoor environment in calculating the signal distance among the different RSS vectors. To address this issue, Feature-Scaling-based K-Nearest Neighbor (FS-KNN) algorithm is proposed for achieving improved localization accuracy.

In this research, we have proposed the Genetic algorithm (GA) based localization to minimize the error distance between the estimated position and actual position. Genetic algorithm solves problems through natural selection and evolutionary processes based on genes. It works with a population of solutions and attempts to guide the research toward improvement, using the survival of the fitness, which can improve the accuracy, stability and robustness of the localization gadget receiving RF signal from access point. A series of cumulative distribution plots rectify the performance measurement of this research and show that our proposed technique give better results in comparison to the previous techniques. The simulations were performed in MATLAB.

DEDICATION

In the name of ALLAH, the most Merciful and the most Beneficent.

Dedicated to my parents, who always have encouraged me and supported me and gave me strength in all phases of my life, and their sincere prayers led me to success.

ACKNOWLEDGEMENTS

First and foremost, I must pay gratitude to ALLAH Almighty for giving me access to new horizons in the literary research. I would like to thank Military College of Signals, NUST, for providing admirable research environment at the institute. This painstaking task was not possible without the support of my supervisor Dr. Mir Yasir Umair who ignited critical abilities in me. I am indebted to him for his timely guidance, profound encouragement and formative criticism. His affectionate and kind consideration towards my research helped me to carry on with my project in odd circumstances. I am also very grateful to my committee members, Dr. Abdul Wakeel and Dr. Attiq Ahmad for their involvement in this research work.

The never-ending support from my family and friends whose courage and helps enable me to complete my Master Degree in Electrical Engineering along with thesis. I am very grateful of the faculty members of Electrical Engineering Department and especially the Head of Electrical Engineering Department for their everlasting encouragement that has fueled my sense of continued determination over the years.

ACRONYMS

GPS	Global Positioning System
WLAN	Wireless Local Area Network
FP	Fingerprint
RP	Reference Point
AP	Access Point
RSS	Received Signal Strength
ToA	Time of Arrival
AOA	Angle of Arrival
TDOA	Time Difference of Arrival
WAF	Wall Attenuation Factor
LOS	Line of Sight
KNN	K Nearest Neighbor
FSKNN	Feature Scale K Nearest Neighbor
GA	Genetic Algorithm
SA	Simulated Annealing

Table of Contents

List of Figures

List of Tables

Chapter 1

1. INTRODUCTION

In the recent years, we have seen the increase in demand of location based services and Positioning systems for both people and objects. Due to this reason, the development in the field of positioning and localization has evolved immensely in the last decades. Localization is the process to find the position or location of a specific target based on some observable phenomena. Localization can be used for numerous applications like localizing soldiers in combat, collecting marketing data, tracking endangered species, navigating self-driving cars, robot movement, indoor localization for firefighters, hospitals and malls, to provide navigation aid, others want to use indoor positioning to better market to customers, automatic object location detection, location detection of products stored in a warehouse, location detection of medical personnel or equipment in a hospital, location detection of firemen in a building on fire, detecting the location of police dogs trained to find explosives in a building, and finding tagged maintenance tools and equipment scattered all over a plant, providing location-aware tour guide systems and many more.

Localization services have become popular with the development of modern communication technologies. The growing diversity of commercial applications has established demand for indoor location. Global Positioning System (GPS) provides adequate services for positioning and localization in outdoor environments, but the performance of GPS degrades in indoor environments and urban areas because of weak signal reception and the lack of line-of sight communication between the user and the satellites and thus researchers have proposed and developed different location systems for localization in indoor environments. Multiple solutions and systems have been developed for indoor localization; each has its own advantages and limitations. The most widely adopted is Fingerprinting technique as it provides better accuracy, although the practical implementation is quite laborious but the working is very simple or least complex as compared to other localization techniques. Also, it does not require any additional equipment and can be implemented using existing infrastructure.

1.1 Fingerprinting

In fingerprinting, the area of interest analyzed by the viewpoint of radio frequency. As the name is fingerprinting, same as each human has a unique fingerprint and can be identified by it similarly each area or region can be identified by a fingerprint of radio signals. Characteristics of radio signals mostly Received Signal Strength (RSS) are used for identification of each region. Usually the area of interest is divided into grids and at each grid point, the readings are noted and stored corresponding to its coordinates and later matched for localization purposes.

The basic idea of this research is that every location has a unique fingerprint based on the detected Base Stations (BS) or Access Points (AP) and the RSS or any other characteristics of BS/AP at each point. The fingerprinting algorithms use this idea to locate a mobile receiver by matching the BS/AP location fingerprint having the RSS for example to a database of location tagged RSS fingerprints collected in the data collection phase. The matching fingerprint is found by determining the Euclidean distance of each fingerprint. The location of the fingerprint that is the closest match to the detected fingerprint is estimated to be the current location of the mobile Receiver. We consider Wi-Fi access points as Base Stations in our work as Wi-Fi is the most widely used means of communication in indoor environments and the most reliable for localization purposes if compared to other sources like Cellular, Bluetooth, FM radio etc. Also, no extra equipment needs to be installed and the existing infrastructure could be utilized for indoor positioning or localization. Wi-Fi localization is feasible for indoor environments also because the deployment of Wi-Fi APs and Wi-Fi enabled mobile devices has become ubiquitous.

1.2 Literature Review

RADAR [12] is the first indoor positioning system that was based on Wi-Fi Fingerprinting developed by Microsoft's Research laboratory. Signal Strength information was gathered at reference points (RP) and later used for matching in the online phase. This work forms the basis of the later work in the same area. In [14] a hybrid technique is proposed which takes both fingerprinting and trilateration technique for the positioning purposes. The trilateration requires the position of minimum three APs for estimating the user location. In [14] the system proposed is the combination of Wi-Fi RSS based fingerprinting and trilateration, a combination of advantages of both systems is used for localization. But trilateration is a range based localization technique and the positioning is highly prone to error because of shadowing and indoor clutter environment. The research work in [2] proposes a scheme for localization using GSM signals in fingerprinting from the nearest Base Stations (BS). The signal strengths from different BS are taken at the reference point and stored in fingerprints. [17],[18]show the experimental work in indoor localization and also the impact of Aps deployment in the WLAN environment Another technique in [19] is proposed which uses both Wi-Fi and GSM signal strength for fingerprinting. The Nearest Neighbor algorithms were used for matching. Both [2] and [19] have certain limitations as the signal strengths of GSM Base stations are not stable as compared to WLAN signals in indoors and GSM networks differ for different users. The work in [13] uses Predicted K Nearest Neighbor technique for position estimation (PKNN). The PKNN algorithm uses the previous position and speed of the user along with the nearest neighbors. The drawback of this method is as it would increase the complexity and latency of the system. The work in [20] proposes a new method based on KNN, instead of using a fixed number of neighbors a technique is proposed which uses an adaptive method to take the number of users to be used in KNN. [4] Uses Particle filter along with KNN to improve accuracy and robustness of the positioning algorithm. The work in [21] says Euclidean distance is prone to error when there are unstable access points (APs) and the KNN algorithm is not capable of differentiating between weak and strong signals. A new scheme is proposed to find the nearest RPs. The researchers in [16] discuss the practical deployment challenges in fingerprinting and provide solutions for those challenges. In [22] an advanced fingerprinting method for wireless sensor networks (WSNs) proposed to reduce the interference of random movement of people and improve the accuracy of system, so Linearly Weighted Moving Average (LWMA) scheme is introduced which can recover back to a condition without any interference. In [23] a feature for floor detection using RSS measurements and Access points is introduced and a new metric for RSS variance problem using KL (Kullback–Leibler) Divergence is also proposed. [24] discusses several techniques to recover missing points in the fingerprint by interpolation and extrapolation techniques. The authors of [25] propose a method i.e. Access point selection and adaptive pattern matching for Wi-Fi indoor positioning, the algorithm they have developed removes the RSS outliers in samples that are received by APs to smooth the RSS and also selects the top APs to minimize the interference for which RSS variations are analyzed. The paper [26] uses Time of Arrival to estimate the stationary position. The ToA method utilizes the propagation time between sender and receiver for distance estimation. [10] uses ToA multilateration for position estimation and assumes that the clocks are synchronized. The drawback or limitations of ToA range based localizations is the effect of multipath and shadowing can cause a significant amount of positioning error. [27] [28] use Channel State Information (CSI) for localizations purposes. CSI can be used but the main problem faced is, not all of the network interface cards in mobile devices are capable of extracting the CSI information.[29] landmark algorithm is used with ToA and RSSI to determine the user location.[30] is addressed the importance of access point in indoor localization for the user location and also mention the selection and contribution of access point while locating of object. [31],[32] show the experimental work in indoor localization and also the impact of Aps deployment in the WLAN environment

1.3 Statement of Problems

- Nowadays, the positioning in outdoor environment, based on GNSS, is highly precise. However, positioning of mobile nodes, which may include people and robots, is still a challenge as signal propagation is not in the line of sight in indoor environment.
- Indoor Positioning System using Wi-Fi Techniques are very popular in recent years as the buildings are now equipped with the low cost Wi-Fi Aps that can be used for indoor positioning references. However, the accuracy and robustness of this technique is a challenge because the radio signals of Wi-Fi equipment are non-linear, non-line of sight and Multipath in nature.
- Many different techniques are proposed to improve the accuracy and robustness of the indoor positioning system in literature but they depend on the cost of equipment, computation loss of microprocessor, plan of the building and also the implementation.

1.4 Motivation

In today's world most of the applications require locating or tracking of equipment/physical belongings accurately in large buildings and other indoor environments; thus, the demand for a system to provide indoor localization has a key role. The GPS cannot provide adequate results in indoor scenarios so an alternative is required for indoors. Based on this, our research focuses on indoor localization techniques and aims to provide a solution, which deliver better results and accuracy. For the purpose of development of new technique the previous indoor localization and positioning systems, algorithms and techniques are studied and reviewed in detail. The results of the commonly used localization technique are also compared with the proposed technique.

Let us take a scenario, relevant to the current situation of the world and our country. If there is any unpleasant situation, or a terrorist activity somewhere, fire in a building or any other emergency situation, the security and law enforcement personnel need to know the number of people that are affected exactly and their exact locations in order to rescue them safely. Having simple knowledge of many people in the building or a large number of travelers at an airport is not enough to tackle such situation efficiently. The need is to know the exact locations of the individuals as this can be a matter of life and death. If a system with bidirectional communication is developed it can eliminate even more risk, people can be guided towards the safe exits or shelters. If a building needs to be vacated, it is required to know if anyone is in the building and their locations. In case of fire, gas leakage, chemical spill and many other unpleasant situations such systems can be used by the rescuers to safely evacuate the people trapped inside. So, radio systems capable of localization have emerging applications in homeland security, law enforcement, emergency response, defense command and control and battlefield command and control. Indoor location and positioning systems have become very popular in recent years because of the emerging technologies and inventions also. Internet of things, automation, directions and navigations, robotics, self-driven vehicles etc., all of these technologies need localization systems to be implemented within. Indoor localization systems also have a key role in automatic detection of objects or tracking products based on their locations. Location detection of a baggage in a large indoor environment, location detection of any lost product in a store, equipment scattered over a large plant or factory. Similarly, location tracking and updates of firefighters in a building caught by fire, trained dogs to find explosives, keeping track of a patient's movement in intensive care, tracking activities of a suspicious person and many more.

Several techniques of localization have been introduced for the purpose of positioning and navigation. During the literature review, it has been observed that each localization technique has its own advantages and satisfies the demands but on the other hand also has some drawbacks and limitations. Limitations can be in case of accuracy, range, system cost, complexity or other. So in the development of a new system or technique the pros and cons have to be considered, and characteristics or parameters should be very carefully chosen.

Various wireless technologies are used for wireless indoor localization. These may be classified based on the location positioning algorithm, i.e., the method of determining location, making use of various types of signal measurement techniques such as Time Of Arrival (TOA)[1][8][9], Time Difference of Arrival (TDOA) [2], Angle of arrival, and signal strength, CSI[3][7], the physical layer or location sensor infrastructure[1], i.e., the wireless technology used to communicate with the mobile devices or static devices and also Cellular networks[4] e-g GSM[5] [6] or LTE.

1.5 Goals and Objectives

Following are the objectives of this research:

1. To propose an indoor localization technique with better accuracy.

2. To achieve good results and maximum accuracy.

3. To propose a technique that is not complex to implement using existing infrastructure and is easy to understand.

1.6 Thesis outline

The Chapter 1, (current chapter) gives an introduction to the research conducted. It also justifies the importance of the research, motivation and its relevance to national needs. level of research already carried out in the field. The objectives of the research also addressed. The second chapter Basic Operation on Localization provides the complete details of localizations techniques and systems developed. Chapter 3, RSSI based Fingerprinting gives detailed introduction of the fingerprinting technique. All stages of the location Fingerprinting technique are presented in detail. Chapter 4 explains the Proposed technique, System Model of the simulations, mathematical models and equations used in the simulation. Chapter 5 Results and Discussions presents the results of the simulations, the basic RSS based technique and proposed techniques are tested in indoor scenarios and compared. Chapter 6 is Conclusion and Future work. Chapter 7 contains References and Bibliography.

Chapter 2

BASIC OPERATION OF LOCALIZATION

2.1 Introduction

In this chapter, we describe the basic process and classification of indoor localization systems based on the type of metric. The basic working principles and algorithms of indoor localization are introduced and the parameters used for localization will be discussed. The technique used in this research will be discussed and explained in detail in the upcoming chapters.

2.2 System for Localization

The basic operation of systems for localization is to collect the information about the location of a mobile device operating in a geographical area and process of that information to determined its location. A widespread technique is used to find out the relative direction is known as radiolocation. The position and movement of the object can be found using radio wave propagation characteristics (e.g. roundtrip delay, orientation of arrival, change of phase). The accuracy in the localization depends on multiple factors, such as noise, interference and the carrier frequency of the radio wave and topology of transmitters and receivers. Thus, one should apply the localization resources effectively, to have the least impact in the localization process.

2.3 Basic Definitions:

Before going into the details of localization techniques and algorithms, few basic concepts need to be cleared.

- **Signal Attenuation**: As the distance from the transmitter to the receiver increases the strength of the signal falls. This reduction in signal strength is logarithmic in nature[36].
- **Fading**: The attenuation of the signal that is caused by the changes in the transmission medium is called fading, or it can also be defined as the signal variations over time. It can occur because of multipath or shadowing etc.
- **Multipath**: When the radio signals are received from more than one path, the phenomenon is called as multipath. The signals can be in phase or out of phase at the receiving end, which can cause significant degradation in signal quality and strength.
- **Noise**: Any unwanted signal is called as noise. Whenever a signal is transmitted over a medium, it is modified by distortions or any unwanted signals because of the transmission medium. This affect is termed as noise or interference.
- **Channel Modeling**: A channel propagation model is required to be developed in order to know the behavior of the channel OR the transmission medium and it should describe the channel's overall behavior adequately. The channel modeling is also necessary in hardware implementation, because it is very necessary to evaluate the channel performance and signal attenuation before the practical deployment. Analyzing the channel through channel propagation model also helps in reducing the cost of the system, its complexity and enhances performance.

If we talk about indoor environment, the channels have multiple distortion factors. It can be walls, floors, partitions, movement of people etc. As these factors highly influence the signal, so a model according to the environment is highly needed. In our case if to judge the values of RSS in a specific environment, we would require a path loss model, which addresses the complete characteristics of the channel.

Basically a propagation model is a mathematical expression or an algorithm that represents the characteristics of the specific environment. In both indoor and outdoor environments the path loss between transmitter and receiver is inversely proportional to the received signal strength.

$$P(d) \propto \frac{1}{d} \qquad (1)$$

where 'd' is the distance among the transmitter and the receiver and P(d) is power received at distance 'd'. and the received signal strength can be used to estimate distance because the RSS is inversely proportional to the distance as shown in the above expression. Practically the RSS along with distance depends on other environmental factors as well. So, path loss exponent (or coefficient) is also added in the equation to present a complete propagation loss model. The propagation loss model coefficient represents the behavior of the environment and indicates the rate at which propagation loss increases with distance *d.*

The values of RSS are generated in simulation using the propagation loss model:

$$P(d) = Pd_0 + 10n\log(d/d_0) + X + WAF \qquad (2)$$

Here P(d) is the RSS value at a point at distance ' d ', $P(d_o)$ is the RSS at the reference distance 'd_o' which is taken 1 meter from the transmitter in the indoor environments. 'd' is the distance of the receiver from the transmitter, the distance between transmitter and receiver is calculated using the formula:

$$d = \sqrt{(xi - x)2 + (yi - y)2} \qquad (3)$$

X is the random noise element, which causes variance in the signals and has a different value at different instants. The X is mainly because of the human presence and movement, scattering effect and cluttering environment. This phenomenon is called as Log-Normal Shadowing. Variations in signal due to cluttering environment at different locations having same distance from the receiver, this effect is represented by X in the propagation model. WAF is the Wall Attenuation Factor which is the attenuation caused by any Wall to the signal.

2.4 Classification of localization system

The basic classification of the Localization system depends on the way in which a radio frequency can be used and includes systems that use methods as mentioned in the sub-sections. The radio frequency (RF) of the IEEE 802.11 WLAN, Infrared (IR) or ultrasound signals can be used in the localization system

2.4.1 Systems based on Received signal strength Indication (RSSI)

The Received Signal Strength based technique is used to estimate distance of the mobile nodes. The attenuation of radio signal is the basis of this technique, as explained in the earlier section the signal degrades with increasing distance and it can be exploited for distance estimations. Mathematical models for path loss and Propagation are used for finding distance, using the transmitted and received signal power. The main reason of its wide use is that most of the devices are furnished with the Wi-Fi module which is capable of extracting and displaying the RSS information. The RSSI is a calculation or estimation of the strength (quality is not necessary) of the received signal in wireless (RF) environment in the arbitrary units. Location detection system, which uses the signal strength of the receiving signal, uses a known mathematical model that determines the path loss attenuation to signal with a distance. The signal strength measurement gives an estimate of the distance between the mobile node and the access point. Thus, the mobile node must be located on a circle having the center, the access point and, as a radius, the distance between them. The distance can be estimated by measuring the path loss. The path loss can be found when the mobile node knows the power transmitted and the power received from the access point.

Generally, the quality of signal that a mobile node gets from an access point degrades as the node moves away from the access point. Incase three or more access points are used, the mobile node location can be estimated, utilizing the technique of triangulation. The prime source of the error for the location system based on signal power is multipath fading and shadowing. The shadow fading errors can be contested by using previously measured signal strength maps centered at the access point that defines the modifications of signal power for specified area. However, this requires a stable topological environment and sufficient preparation. In any instant, the results we take are extremely dependent on the environmental inconsistency.

2.4.2 Systems based on direction

The systems are based on the direction of a signal and calculate the position of the mobile node by first estimating the angle of arrival (AoA) of a signal from a mobile node at different Access points using the antenna arrays. The intersection of lines that describe the direction of the signal is to estimate the position of the mobile node as illustrated in (Figure 2.1). Dissipation and dispersion around the mobile node and the access point will modify the measured AoA.Without an observable line of sight (LOS) signal portion, the antenna array picks up a reflected signal that may not originate from the direction of the mobile node. Regardless of whether a line of sight LOS part is available, multipath wants to still affect the measurement angle. The accuracy of the angle of arrival(AOA) strategy decreases with increasing distance between the mobile node and the access point because of the central limitations of the devices used to measure the angles. In few estimations that have occurred, this method provides excellent results when used as part of the micro-cells because the signals arrive at the access point with a moderate narrow AoA spread. For micro-cells, this technique gives more worse results.

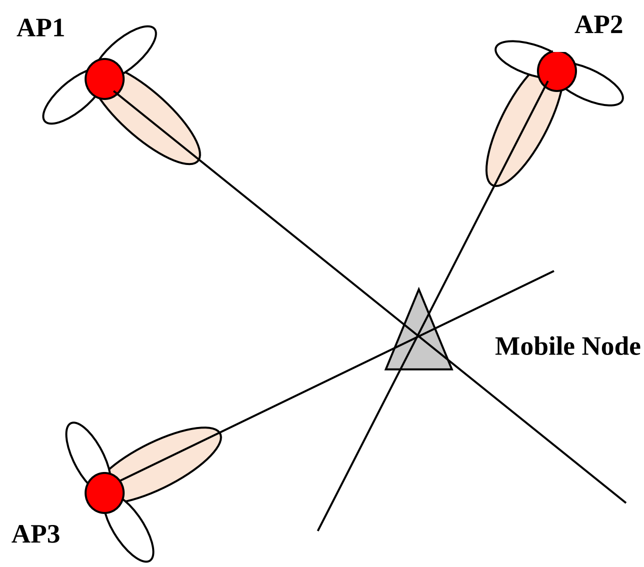

Figure 2.1 Location sensing using Angle of Arrival (AoA).

2.4.3 Systems based on distance

The distance estimation system measures the absolute or the distances difference between the mobile object from Access points (AP). The distance between a moving object and an AP is measured by finding the arrival time (ToA)[8][9] which is the one-way propagation time between them, assuming that the transmission time is known. Geometrically, this is represented by a circle around the AP on which the moving object must lie. By using no less than three APs to determine ambiguities, the position of the moveable object is at the intersection points of the circles (Figure 2.2)

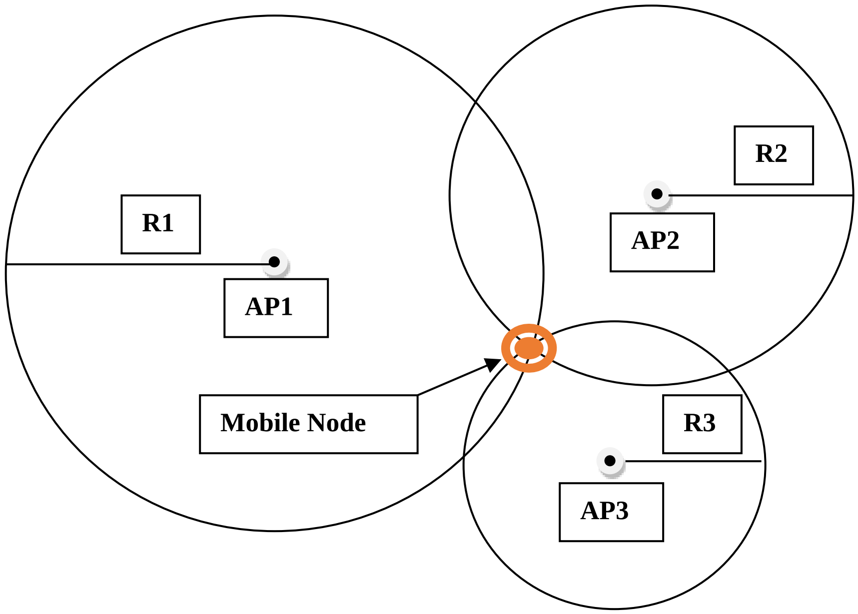

Figure 2.2: Location sensing Time of Arrival (ToA).

A basic requirement for the functioning of the system is the absolute synchronization of the movable object and the AP. When they are not synchronized, the time difference of arrival (TDoA)[37] is used instead of absolute time. Because it is a time differences of arrival of all aps. So it is much easier for the aps to be synchronized. Since the hyperbola is a constant curve for the arrival time difference of the two APs, the time difference defines the hyperbola that the mobile must exist.Hence The position of the moving object is at the intersection of hyperbolic curves, illustrated in Figure (2.3)."

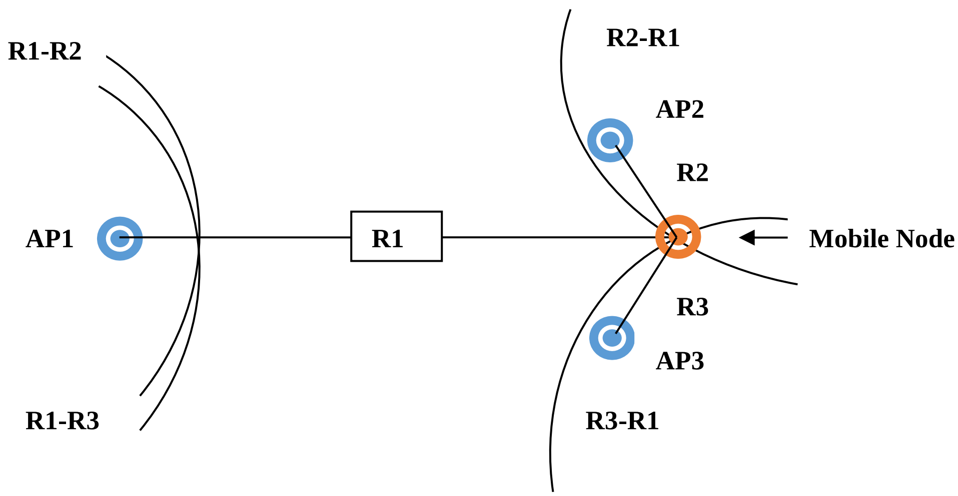

Figure 2.3: Location sensing Time Difference of Arrival (TDoA).

2.5 Localization using RF and ultrasound

There are systems for localization that use numerous modalities. For instance, they use both the ultrasound and RF tags. The better location accuracy can be provided by using more than one modalities therefore they need more hardware. In recent years, an increasing attention has been presented in the usage of radio frequency signals to achieve localization. The benefits of using radio-frequency signals for localization contain the capacity to localize, irrespective of the locating of walls or furniture, as well as the capability to use pretty low-cost hardware to resolve a problem that heretofore required highly specialized, difficult-to-obtain components. In addition, systems for localization can be characterized based on their scalability and their cost. Lastly, the privacy is a parameter which the systems should also take into consideration."

2.6 Accuracy and precision

Accuracy and precision are the two axes of a trade-off: less accuracy may be dealt for more precision. Therefore, use of only one of the two features of spatial position is not suitable for assessment of localization systems. Relatively, the location systems should be measured on the basis of the error distribution acquired when locating objects, taking into account any dependency, for example the required density of infrastructure (for example satellites, base stations, radio frequency readers and so on).The examples raise the question of accuracy naturally. More accurate system should be used if more resolution is required.

The basic definition of accuracy and precision are as follows.

2.6.1 Accuracy

The smaller distance's property that can be differentiated by the system is known as the accuracy of that system.

2.6.2 Precision

The percentage of the times that the prescribed accuracy is achieved is called precision of the system.

2.7 Location sensing techniques

In location sensing, the common techniques are triangulation, proximity and the scene analysis.

2.7.1 Triangulation

The geometric properties of triangles involve in the triangulation method to calculate positions of the object. This technique involves the lateration that uses several distance measurements amongst the known points, or by using the angulation method which computes angle relative to points with known separation.

2.7.2 Lateration

The procedure of lateration estimates the location of an object by calculating its distance from multiple reference positions. In two-dimensions, the position of an object is measured from 3 non-collinear points by using the distance measurements, as it is shown in Figure. 2.4 In three dimensions, the position of an object needs the distance measurements from 4 non-coplanar points.

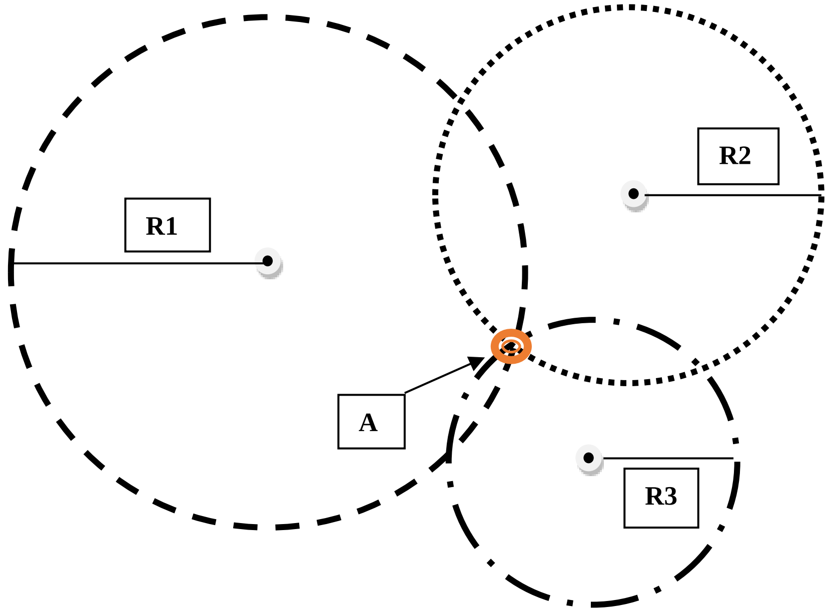

Figure 2.4: 'Determination of 2D position using lateration. It requires distance measurements between the object `A' and 3 non-collinear points.

The measurements of the distance can be calculated by using three techniques. Which are as follows.

2.7.2.1 Direct

A direct measurement of distance uses a physical action or movement.

2.7.2.2 Time-of-Flight

Estimating distance from an object to some point P using time-of-flight implies estimating the time it takes to move between the object and point P at a known velocity.

2.7.2.3 Attenuation

With the increase of distance from the emission source, the intensity or strength of the emitted signal decreases. The decrease relative to the original intensity is known as attenuation. For example, the radio signal emitted from an object is attenuate by a factor that is proportional to 1 / r when it reaches point P at distance r from the object .The time-of-flight is more accurate than attenuation.

2.7.3 Angulation

The angulation is similar to lateration but it uses the angles in determining of the location of the object instead. The two dimensional angulation needs one length and two angle measurements, i.e., the reference point distance as shown in Fig 2.5. In three dimensional angulation, one length, one azimuth and two angle measurements are involved in precise positioning.

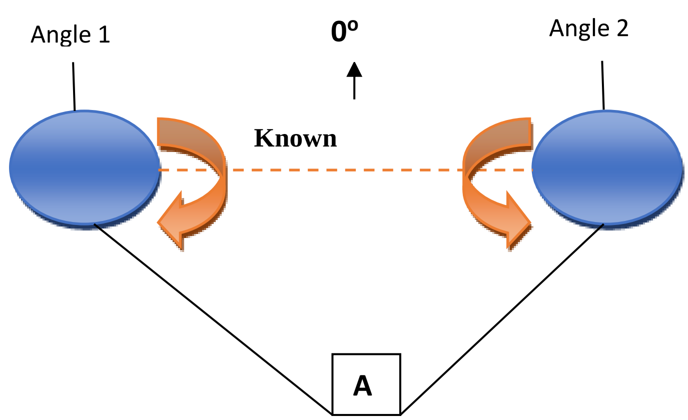

Figure 2.5: 2D position using angulation method requires one distance and two angle measurement.

2.7.4 Proximity

A technique measures the closeness of a known set of points knows as proximity. There are three methods generally used to measure the proximity, which are as follows.

2.7.4.1 Detection of the physical contact

The technologies such as capacitive field detectors, pressure, and touch sensors are used for the detection of the physical contact.

2.7.4.2 Measurement of cellular access point

The monitoring in the proximity location technique is used to monitor or locate the device when it moves from one access point, in a cellular network, to another.

2.7.4.3 Automatic Identification Systems

The third approach of the proximity technique is the use of Automatic Identification System (AIS) e.g., point of sale (POS) terminals, landline telephone records, computer logs, credit card and electronic card-lock logs.

2.8 Location sensing Systems

There are few location-sensing systems used in indoor localization that locate the position of the object, which are as follows.

2.8.1 Cellular

Cellular network can used for positioning and localization purposes in different methods,

Mainly Cell ID and OTDOA. [38]

2.8.2 Cell ID

In the Cell Identification positioning method the position of the user is localized to its

Serving Base Station, and if possible to the sector of the BS. This information is easily available as the Mobile device itself knows the BS it is connected to. The device can provide this information to the application or any service, which requires the location data of the user.

2.8.3 Enhanced Cell ID

The ECID system gives a more accurate measurement by using some extra parameters

Than just providing the serving BTS information. Additional parameters can be timing advance, which can provide the distance measurement between the mobile and the nearest BTS or BS. Furthermore, RSS and Angle of Arrival can also be utilized for positioning to give a more accurate result.

2.8.4 Bluetooth

Bluetooth can also be used for localization. It also operates in 2.4 GHz ISM band, but When compared to WLAN the bit rate and range is much lower. Bluetooth tags or Bluetooth low energy beacons can be used for indoor localization in three ways: proximity based localization, trilateration and fingerprinting based Most of the hand held devices today support Bluetooth. The drawback of Bluetooth based indoor localization system is the short range and extra installation of BLE beacons.

2.8.5 FM

FM radios are also considered for indoor localizations systems. They do not require any extra infrastructure, almost all mobile phones are equipped with FM receiver and multiple FM stations can be found easily anywhere in urban areas. Same techniques can be used for FM radio based localizations system as explained earlier. But the drawback for these systems is low accuracy.

2.8.6 WLAN

Wireless Local Area Network (WLAN) is a standard operation in ISM band of 2.4 GHz. It is has become very popular because of its numerous applications and high need. WLAN provides high data rates and reliable communication, which has made it an essential part of life nowadays. The high availability of WLAN access points in the indoor environments has made it the most appealing technique for indoor localization. Although it was not designed for localization purposes but it is widely adopted by researches for their techniques and systems implementation because of the above stated reasons. It can also be used for localization in three ways, i.e. proximity, triangulation/trilateration and fingerprinting.

Chapter 3

RSSI BASED LOCATION FINGERPRINTING

Wi-Fi base location fingerprinting is one of the range free technique for positioning and localization that is being used in all wireless networks. In literature, different fingerprinting based techniques have been proposed but all have the common fundamental. One of the fingerprinting technique is location matching that match the fingerprint of certain characteristics of a signal as a function of the position.

Fingerprint is a set of unique signal parameters that are location dependent, so each location has a unique fingerprint that is associated to that location. As analogous to the human fingerprints which are unique and on the basis of which a person can be identified, an unknown location can be found out by matching a unique set of RSS received, called the 'fingerprint' at each location, to a database of RSSI at known reference locations. The fingerprinting algorithm has two phases. One is called as training phase or offline phase, and second one is estimation/location determination or online phase. In the offline phase, an extensive site survey is carried and the area of interest is divided into grids or sections depending on the requirement of the system. At each point of the grid the signal characteristics mainly RSSI are noted and saved in the database, each point is called as Reference Point (RP). In the online phase or location estimation phase the location of the target is tracked by providing the set of values measured at the point and the server is asked to return the estimated position by matching the readings from the offline phase with those in database, and the best matching result is provided as the estimated location. Number of matching algorithms are used for accurate and efficient matching of the data with the query data.

Fingerprinting system is highly affected by factors like the constant changing of RSS due to multipath, shadowing, interference, signal variations etc. One of the main drawback of fingerprinting is the collection of RSS values at the reference points during the offline phase. It is a very laborious task and is vulnerable to environmental changes. For localization of the mobile object, following are the important concepts for fingerprint technique.

3.1 Grids and Reference points

The fingerprinting based localization system is highly dependent on the grids and RPs. The area of interested is divided into equally spaced grids, and RSS samples are gathered at the center point or corner point of the grid (shown in figure 3.1). The granularity of the RPs highly affects localization accuracy. The more the RPs higher the accuracy until it reaches a threshold. The grid points or RPs should not be that close to each other as no significant change in RSS level is observed. In literature, some systems have been developed with unequally spaced and identical grids but the area of interest is simply divided into regions, the objective in such cases is to only identify the region like room number, corridor, floor, office etc. For a better accuracy more RPs are required so more refined results can be obtained.

3.2 Radio Map

Radio Map is the backbone of the fingerprinting systems. Radio Map is basically achieved by making the database of fingerprints having the signatures of radio signals or fingerprints. Just as a simple map is used to identify a region on geographical basis, global coordinates or any other relative system similarly radio map can be used to identify a region based on RF signatures. Each RP has its own unique signature that is stored in the radio map and if similar results are found the target is most likely to be found in the same map is developed in the offline phase with the help of devices capable of showing the signal parameters for example smart phones, laptops and other handheld devices estimated location fully depends on the radio map. The radio map is built by dividing the area of interest into grids or cells with the help of floor plan. RSS values from all APs are collected inside the grid and stored in the database or we can also say stored in the radio map. The ith element from the radio map will be in the following form:

$$\{(x, y), (RSS1, RSS2... RSSn)\}$$

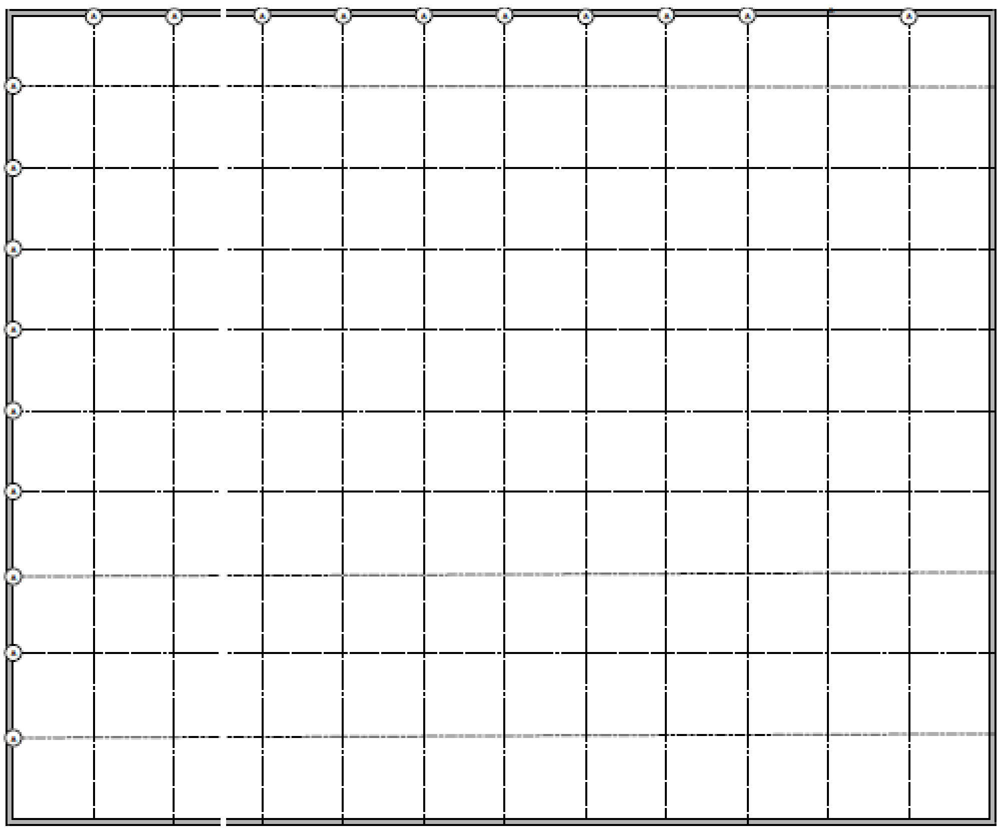

Figure 3.1: Area of Interest divided into equally spaced Grids

3.3 Variations in Signal Strength

The absolute value of signal strength is unstable and changes at every instant due to a large number of factors like shadowing, changes in environment, attenuation due to distance and many more as discussed in previous section. Signal strength noted at one instant is saved in the radio map but due to multipath, it is different at each instant of time for every location. In the offline phase, the matching can lead to high errors because most probably the signal strength received at the same point in the online phase will not be same. This variation in signal strength is one of the biggest challenges in fingerprint. Many ideas have been presented to tackle this problem. The signal strength variations can be divided mainly into two types.

- Temporal variations
- Spatial Variations

3.3.1 Temporal Variations

Temporal variations are those signal strength variations that are observed at a single point over the passing time. Temporal variations can also be defined as those signal strength variations which are observed over a large period. These variations are a major challenge in all type of localization systems and specially indoor localization systems because the temporal variations have a high effect in indoor environments. The main reasons of these type of signal strength variations are the surrounding environment, movement of people, temporary blockage of the path between transmitter and receiver etc. A system cannot rely on a single signal strength value observed at an instant and perform localization services corresponding to that observed value.

The signal strength values are very unstable so in the estimation phase the readings could not be the same, even could have a large difference. Therefore, to tackle this problem the signal strength of an access point at a specific reference point is recorded for a longer period of time, multiple RSS samples are collected and their average value is stored in the radio map. This average value gives an overall idea of the channel and RSS values for better accuracy RSS values can be collected at different times of the day. Therefore, it can be concluded that in order to increase accuracy, the system needs to use multiple samples from the same access point to account for the variation of signal strength.

3.3.2 Spatial Variations

Spatial variations are those signal strength variations that are caused because of change in the position of the receiver. The signal attenuates because of the increasing distance between transmitter and receiver as explained in previous chapter. These type of signal variations are desirable because it help to define a region by the values of signal strength from APs. The difference in two fingerprints for two different points is because of both spatial and temporal signal strength variations.

3.4 Matching Algorithms

After completion of offline phase, where a complete site survey is conducted and signal strength values from each access points at all reference points are saved in the radio map, we talk about techniques involved in the online phase. The position of the target is estimated from the data of radio map using different algorithms. The two most common algorithms or matching techniques are nearest neighbor (deterministic) and Bayesian (probabilistic). A simple difference between deterministic and probabilistic techniques is in the prior method the RSS values are represented as single values and the later uses probability distributions for representation. The probabilistic methods contain more information about the signal strength range but are much complex, whereas the deterministic methods are simple to implement and process.

3.4.1 Deterministic: K Nearest Neighbor (KNN)

KNN matching algorithm is one of the famous algorithm used for fingerprinting localization systems and generally for WLAN or Wi-Fi based systems in the online/location estimation phase.

The nearest neighbor algorithms choose those fingerprints, which has the minimum, distance from the test data. The test data is matched with the complete database or radio map and the target location is estimated. The distance used is Euclidean distance. If it is NN then only one value which has the minimum distance from the test data is returned as the estimated location, in KNN the algorithm returns 'K' nearest neighbors from the test data on the basis of minimum distance and an average value is calculated which is the estimated location. The fingerprints are incrementally sorted and the K reference points corresponding to those fingerprints are chosen.

Now the question arises what should be the optimal value of K in the KNN. The value of K is differently used in the literature [12] [13] [14] [16] [19]. But usually the values of K are taken 3 and 4, for large values of K the accuracy of the system degrades as outliers are also included as neighbors. Mathematical model and application of KNN algorithm will be explained in next chapter.

3.4.2 Probabilistic: Bayesian

The probabilistic method based on Bayes Theorem, the signal strength vector collected in the online phase say s = (s1, ..., sk), an d the set of locations is X = { x1 , x2 , …, xm}, we have to find the location in radio map that maximizes the conditional probability

$$P(xi/s\ ) \qquad (4)$$

Mathematically it can be presented as:

$$Argmaxx[P(x/s)] \qquad (5)$$

By Bayes theorem, we know that

$$P(x/s) = P(s/x)p(x)/P(s) \qquad (6)$$

The location with maximum probability of occurrence of the received signal strength vector will be the estimated location.

3.5 Famous Fingerprinting Based Indoor Localization Systems

The two famous techniques used for fingerprinting based localization systems are:

1. RADAR
2. HORUS

3.5.1 RADAR

RADAR [12] is the first indoor positioning system that was based on Wi-Fi Fingerprinting developed by Microsoft's Research laboratory. This system measures the RSS for the Wi-Fi signals and saves in the radio map. The estimation algorithm is deterministic i.e. K nearest algorithm. Mean value of RSS values of multiple samples is stored in the fingerprint database or radio map.

3.5.2 HORUS

The Horus system [39] is also a fingerprinting based WLAN system for location estimation. Unlike RADAR, it is a probabilistic system. The radio map saves the histogram of the received signal strength samples from each AP. Its location estimation phase is combination of two techniques i.e. discrete space estimator and continuous space estimator.

Chapter 4

SIMULATION MODEL AND PROPOSED TECHNIQUE

4.1 Introduction

This chapter explains the simulation model and proposed technique. The system model is simulated and propose technique is tested along with the existing and commonly used RSS based fingerprint for comparison in different scenarios. Now we explain the simulation model for test our proposed technique.

4.2 Simulation Model

Consider a single story indoor environment covered by certain number of WLAN APs, which are visible throughout the area of interest. In fingerprinting technique we need to build a Radio Map, so we divide our area of interest into grids, the two dimensional floor is divided into square grids of 2m by 2m. In the Offline phase, the readings from Wi-Fi Access Points (APs) are taken at the midpoints of those grids which are also called as calibration points or Reference Points (RPs) and stored in the fingerprinting database according to the coordinates of the grid points, so it can be used for matching purposes later in the online phase. The coordinates of any point can be represented as (x, y) where *x* and *y* represent the 2D coordinates on the floor.

Let,

χ = Two dimensional Space

(x, y) = coordinates in 2-D

$\hat{\boldsymbol{S}}$ = Vector having RSS reading

Ňap = number of APs

$\boldsymbol{\Psi}$ = samples of RSS from $\hat{S}$

Lx= dimensions of area along x-axis

Ly= dimensions of area along y-axis

4.3 Building of Offline Fingerprint Database

The pixel value (Pix) for division along x-axis and y-axis along the border to form uniform grid is taken 2. The readings are calculated at the mid points of each grid, so the x ,y coordinates of the Reference Points (RPs) are

Difx= [0.5: Pix: Lx]

Dify= [0.5: Pix: Ly]

It is known that the signal strength of received signal can be used to estimate distance because the power (or RSS) of electromagnetic waves is inversely proportional to the distance as shown in the following expression:

$$P \propto \frac{1}{d} \qquad (7)$$

Here P is the RSS value at a distance d from the transmitter. Practically the RSS value along with distance depends on other environmental factors also. So, path loss coefficient or exponent is added in the equation to present a complete path loss model. The path loss coefficient represents the behavior of the environments, specified in the path loss model along with RSS readings as given below:

$$P\,(d) = P\,(d_0) + 10.n.\log\left(\frac{d}{do}\right) + X + \hat{W} \qquad (8)$$

Here $P(d)$ is the RSS value at point at distance ' d ', P (d_0) is the RSS at the reference distance 'd_0' which is taken 1 meter in the indoor environments. 'd ' is the distance of the RP from the AP, distance between AP and RP was calculated using the formula:

$$d = \sqrt{(Difx(i) - AP(x))^2 + (Dify(j) - AP(y))^2} \qquad (9)$$

and in eq (8) 'n' is the path loss coefficient which is usually 3 in indoor environments. X is the random noise element which causes variance in the signals and has a different value at different instants. The X is mainly because of the human presence and movement, scattering effect and cluttering environment. $\hat{W}$ is the Wall Attenuation Factor which is the attenuation caused by any Wall to the signal. Multiple number of samples for RSS had to be taken, as values of RSS change at every instant. All the values of RSS for a certain reference point from a certain AP are averaged and stored in the fingerprint database. The more samples are averaged the better value of RSS is obtained.

$$RSS_{ij}(avg) = \frac{\sum_{n=1}^{N} RSS\ ij\ (n)}{N} \qquad (10)$$

In the equation (10) $RSS_{ij}(avg)$ is the average RSS value obtained, n is the number of sample, N is the total number of samples, so $RSS_{ij}(n)$ is the RSS of n^{th} sample of i^{th} RP from j^{th} AP.

The system automatically detects the walls in the channel between the AP and RP and includes the effect in the propagation model for RSS calculation. For creation of number of walls in simulation model between RP and AP, for each AP a matrix of zeroes is initialized.

$$
W_i = \begin{matrix} 0_{1,1} & \ldots\ldots & 0_{Lx,1} \\ \ldots & \ldots & \ldots\ldots \\ 0_{1,Ly} & \ldots & 0_{Lx,Ly} \end{matrix} \qquad (11)
$$

$$
W_i = zeros\ (Lx, Ly) \qquad (12)
$$

$$
W_i = [index = h] \qquad (13)
$$

Put value 'h' i.e. the number of walls between AP and RP on the indexes according to the map, and w_i is for the '*i*th' AP similarly for all Number of Aps ***Ňap***.

So, in w_i

$h = 0$, *no wall*

$h = \Omega$, where Ω= number of walls

Later in path loss model

$$
\hat{W} = W\,(x, y) * 2 \qquad (14)
$$

$$
\hat{W} = \Omega * 2 \qquad (15)
$$

The training data in the offline or calibration phase is stored in the database and now database of radio map will have the form.

$$
\lambda = \begin{matrix} dif\,x(1), dif\,y(1) & \hat{\mathbf{S}}_{1,1} = (\boldsymbol{\Psi}_1, \ldots\ \ldots., \boldsymbol{\Psi}_{\check{N}ap}) \\ \vdots & \ldots \qquad \vdots \\ dif\,x(\mathrm{Lx}), dif\,y(Ly) & \hat{\mathbf{S}}_{\mathrm{Lx,\,Ly}} = (\boldsymbol{\Psi}_1, \ldots\ \ldots., \boldsymbol{\Psi}_{\check{N}ap}) \end{matrix} \qquad (16)
$$

Here λ is the fingerprint database of the radio map. Order of the matrix is $\alpha * \beta$, where α is number of data points and β is dimensionality.

Similar procedure is applied for the creation of offline database of feature-scaling-based k-nearest neighbor (FS- kNN) method.

4.3 Online phase Query

In the online phase a query is put to the system having the values of the parameters used (RSS and Fs-kNN) and the system provides the estimated location by using matching technique, the value with the most similar match to the query is given as the estimated location and returned as coordinate points of the estimated location.

One thousand random test points are created in our simulation within the range of x ⤇1 to Lx and y ⤇ 1 to Ly where the matrix ℝ has the x coordinates of the generated test points and § has the y-coordinates of the test points. The distance from the APs and test points in the 2D space χ is measured in the same way by Euclidean distance.

$$d= \sqrt{(\mathbb{R}(i) - AP(x))^2 + (\S\,(j) - AP(y))^2} \qquad (17)$$

The RSS values are calculated for the test points as to be given as queries to the system using the same propagation model

$$\mathbb{Q} = [(\mathbb{R}i, \S i),\ \hat{\mathbf{S}}i = (\boldsymbol{\Psi}_1, \ldots \ldots, \boldsymbol{\Psi}_{\check{N}ap}) \qquad (18)$$

Order of the query matrix is 1000 by dimensionality

The two matrices λ and ℚ will be matched for location determination. For the matching purpose, we have used the K Nearest Neighbor (KNN) algorithm. KNN algorithm is the most widely used algorithm for fingerprint matching, the biggest advantage of using KNN is simplicity. The KNN chooses the fingerprints having the minimum Euclidean distance to the measurement given as a query. the value of K represents the number of nearest neighbors and then their value is averaged to give the coordinates of the estimated position. The value of K is differently used in the literature [12] [13] [14] [16] [19]. But usually the values of K are taken from 3 to 4, for large values of K the accuracy of the system degrades as outliers are also included as neighbors .weighed KNN or WKNN can also be used [19][20] for matching. The advantages of using WKNN is, it reduced the chances of including any outliers.

In classical KNN algorithm, the signal distance between the newly reported signal vector and fingerprint can be calculated using Euclidean distance.

$$\mathrm{d} = \sqrt{\sum_{i=1}^{k}(RSS_{k,i} - RSS_i)^2} \qquad (19)$$

In this equation (19) ($RSS_{k,i}$) fingerprint vector stored in the database and ($RSS_{i)}$ is newly reported signal vector that has to be estimate. After calculation of signal distances to each fingerprint using equation (10), all the finger print (i.e.. R*p*s) are sorted according to their signal distance similarly to the currently reported RSS vector in the increasing order. The positions (coordinates) of former K RPS are then weighted averaged and the estimated position of the Mobile location. However, the accuracy of KNN algorithm cannot be guaranteed all the time. Because the KNN fail in calculating the signal spaces of different RSS levels, as it cannot distinguish that, the different RSS values at different RSS level may not mean equal geometric distance in complex environment. To address this issue FS-KNN algorithm has been proposed that resolve this problem.

For KNN implementation and location determination following procedure was adopted.

KNN (λ, $\mathbb{Q}$)

$$\mathbb{Q} - \left(\mathbb{R}i, \S i\right) \Rightarrow [\hat{\mathrm{S}}i \; = (\boldsymbol{\Psi}\, 1, \ldots \; \ldots \; ., \boldsymbol{\Psi}\, \check{\mathrm{N}}\mathrm{ap}\right) = \omega \qquad (20)$$

$$(\mathbb{R}i, \S i) = \text{ Actual test Location} = \text{L} \tag{21}$$

$$D = \sqrt[1/2]{\sum(\omega(i) - \lambda})^2 \tag{22}$$

$$\text{Dsort } = argmin \sum_{i=1}^{1000} \sqrt[1/2]{\sum(\omega(i) - \lambda})^2 \tag{23}$$

$$Dsort = [dmin, \dots \dots dmax] \tag{24}$$

$$sort = indexes((Dsort) \tag{25}$$

$$Estimated\ position = \frac{\sum p(x), P(y)}{K} \tag{26}$$

$$Estimated\ positons\ (i = 1{:}\,1000) = E = p(xi)\, p(yi) \tag{27}$$

$$Error = \epsilon(i) = \sqrt{\sum E(i) - L(i)} \tag{28}$$

After the results from KNN and the position estimation tests performed for 1000 random queries, their Cumulative Density Function (CDF) of $Error = \epsilon$ was taken as a metric for precision of the system.

The values of parameters used in simulation:

Parameter	**Value**
Path loss exponent 'n'	3
Wall Attenuation Factor 'WAF'	2
Reference Distance 'd_0'	1
Power at d_0 'P(d_0)'	-30
Transmission Power 'P_{tx}'	10
No of Samples collected at RP 'samp'	10
'K' in KNN	5
Pixel value	2
Grid Size	2*2
No. of Position Queries	1000

4.4 RSS based FS Model

In this research, RSS-level-based scaling method as a new feature of scaling [40] is used for estimation of effective signal differences in computing the similarity among the different RSS value. Suppose d' represents the signal distance between an instantly reported RSS value and the RSS value (fingerprint) associated with the RP_k, then it can be calculated as follows:

$$d' = \sqrt{\sum_{i=1}^{k} (RSS_{k,i} - RSS_i)^2 \times \omega(RSS_i)} \qquad (29)$$

The scaling weight function $\omega(.)$ represents the amount of effective RSS distance to which the change in RSS with one unit at RSS level of RSSi and its values change with the change in value of actual RSS. The effective RSS distance returned by Equation (29) is best for describing the relationship between the actual RSS distance and the geometry distance.

The function $\omega(.)$ play an important role in FS model. It is difficult to calculate value of $\omega(.)$ in complex environment. To handle such situation the entire RSS space is divided into n equal intervals. When n=1, FS-KNN degenerates to classical KNN. RSS-level-scaling weight can be defined as:

$$\omega(x) = \sum_{j=1}^{n} \alpha_i x_i(x) \qquad (30)$$

Where $x$the actual RSS is value and $\omega(x)$ is the scaling function for signal difference at RSS value x. The key problem is how to properly tune the value of n and α_i. In practice, we can Adjusts the number of n interval as well as the interval coefficients $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ to achieve high accuracy of positioning. the granularity of scaling function $\omega(.)$ depend on the number of RSS n intervals .Smaller the granularity, better to represent the relationship among the Signal difference at different RSS level. Also, more number of RSS interval means large searching space in the implementation of SA (simulated annealing) and having a better location accuracy. In this research we consider the number of interval is small constant (e.g n=5). SA (simulated annealing algorithm) is used to tune the Coefficients . fig 4.1 gives a flowchart in FS-kNN for the tuning of coefficient .

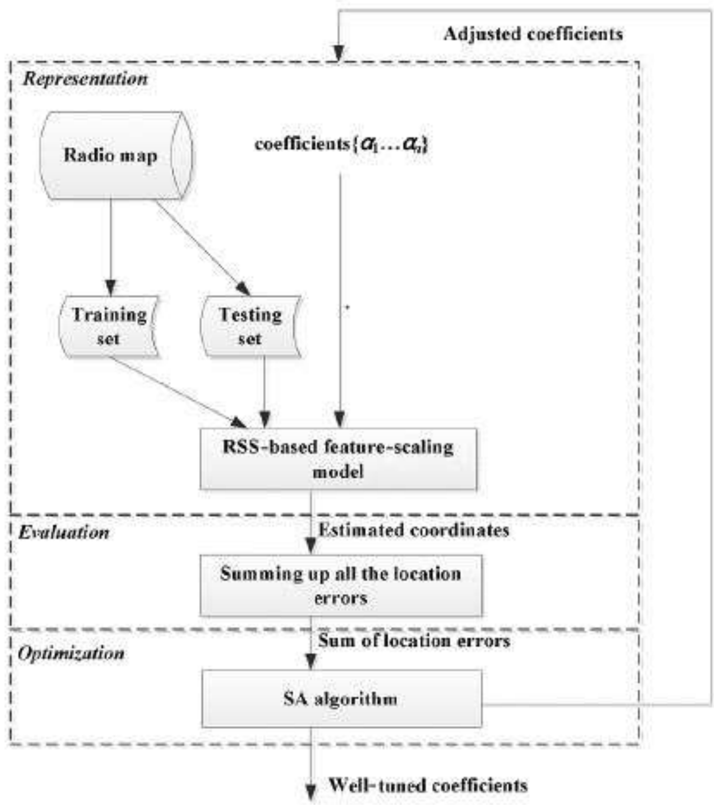

Figure.4.1: Flow chart for tuning coefficient in FS-KNN [40]

From figure.4.1, it can be seen that there are three process evolved in the flowchart 1) representation; 2) evaluation and 3) optimization, which are carried out iteratively for the tuning until certain predetermined condition(s) are satisfied. The output of the preceding component will be input the next component .in the optimization process some coefficient are changed randomly and then fed in to the next iteration. The iteration calculation is continue until the total number of iteration reaches a predetermine condition are satisfied.

4.4.1 Representation

*F*S-kNN can be represented by two main components: the radio map that has been calibrated during the offline phase and FS-based RSS model. To properly tune the weights for the FS model, here, the holdout method is used, which divides the radio map into two sets: 1) a training set and 2) testing set. The training set corresponds to the RPs (fingerprints) with known coordinates for localizing MS at unknown position. The testing set is used for iteratively evaluating the performance of adjusted coefficients obtained in each iteration. After obtaining a new set of coefficients, the testing set is fed into the FS model to produce estimated coordinate for each element in the set.

4.4.2 Evaluation

For each set of coefficients obtained, we shall evaluate its resulting performance by computing the sum of location errors, denoted by *cost*, as follows:

$$\text{Cost} = \sum_{i}^{m1} \sqrt{(xi - xi')^2 + (yi - yi')^2} \qquad (31)$$

where (*xi*, *yi*) is the actual coordinate of the *i*th element in the testing set, (*xi', yi'*) is its estimated coordinate, and *m*1 is the total number of elements in the testing set. It is worth noting that the so-called RPs in the testing set are actually considered as unknown positions during the evaluation process. The larger the sum, the higher the localization inaccuracy is. Obviously, the coefficients that make cost = 0 are the ideally optimal solution.

4.4.3 Optimization

SA is used to search for new coefficients resulting in better accuracy. SA uses a variable as the temperature for the cooling schedule, which will become lower and lower during iterative calculations. In each iteration, some of the coefficients are randomly changed. If the new cost is smaller, the new set of coefficients becomes the current coefficients. If not, the new coefficients can still become the current coefficients with certain probability *p*, which is calculated as follows:

$$p = e^{-\frac{\Delta\, cost}{temperature}}, \qquad \text{if}\ \Delta\ \text{cost} \geq 0 \tag{32}$$

where Δ cost represents the difference between the new cost and the optimal minimal cost. This process continues until the iteration number reaches a predetermined maximum number or certain predetermined conditions are met. In the process of coefficient tuning using SA, the maximum iteration number set to 35000. We have also tested other settings of maximum iteration number and find that the positioning error changes (decreases) insignificantly when the iteration number exceeds 35000.

4.5 Genetic Algorithm

In this research, an attempt has been made to improve indoor positioning using KNN, FSKNN algorithm by optimizing error distance to a minimum value using Genetic Algorithm. It is a subject of minimizing error distance of the estimated position versus actual position, which is subjected to constraints such as well-known accuracy of KNN and FSKNN classifier. Genetic algorithm is a self-adaptive procedure that solves problems through natural selection and evolutionary processes based on genetics. It works with a population of solutions and attempts to guide the search toward improvement, using the survival of the fittest principle.

Genetic algorithms differ from normal search procedures in the following ways

1. Genetic algorithms search from a set of points rather than a single point.
2. Genetic algorithms use objective functional information rather than derivatives or other secondary knowledge

3. Genetic Algorithms use probabilistic transition rules and not deterministic rules.

Genetic search technique start with an initial set of random solutions called population. Each individual in the population is called a chromosome and shows a solution to the problem. Chromosomes are symbols of strings. Although this is not necessary, it is usually a binary bit string. The elements in the chromosome are called genes. The chromosomes evolve successive iterations called generations. During each generation, the chromosomes are evaluated, using some measures of fitness

4.5.1 Methodology of Procedures

Assume that the distance between the estimated points (that are calculated using KNN algorithm and FS KNN algorithm) and the actual points, is selected to be minimum (equal zero). This assumption is achieved by the following steps:

- Assume a range of points (x, y) around each estimated point ($x_{estimated}$,$y_{estimated}$).
- This range follows the linear ratio around the value 1; where 1 represents the estimated point as following: **0.9 < Point Ratio(x,y) < 1.2 "as example"**
- The assumed linear point ratio can be expressed in the figure 4.2
- The assumed points inside the **ratio range** are the first generation that will be used by GA to find the nearest point **(min error)** to the actual point according to the objective function derived bellow.

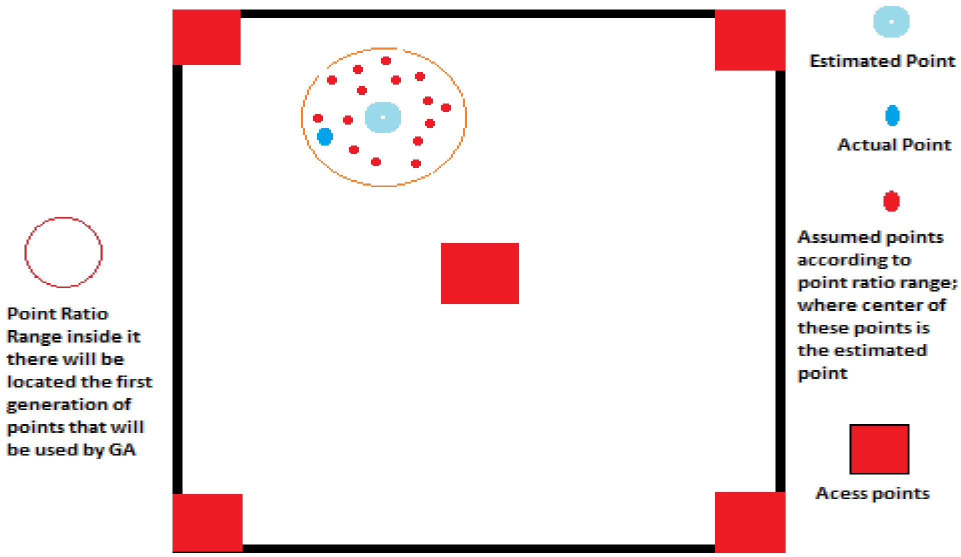

Figure 4.2: Genetic Algorithm

4.5.2 Derivation of Objective function

The distance between access point number 5 and estimated point using KNN method can be represented as:

$$\text{dist1} = \sqrt{(X_{\text{estimated by KNN}} - X_{\text{access point5}})^2 + (Y_{\text{estimated by KNN}} - Y_{\text{access point5}})^2} \quad (33)$$

After expansion

$$\text{dist1} \coloneqq \left(X^2_{\text{estimatedKNN}} - 2X_{\text{estimatedKNN}}\text{AccessP}_{5x} + \text{AccessP}^2_{5x} + Y^2_{\text{estimatedKNN}} - 2Y_{\text{estimatedKNN}}\text{AccessP}_{5y} + \text{AccessP}^2_{5y}\right)^{\frac{1}{2}} \quad (34)$$

Also the distance between the actual point and the access point number 5 can be represented as:

$$\text{dist2} = \sqrt{(X_{\text{actual point}} - X_{\text{access point5}})^2 + (Y_{\text{actual point}} - Y_{\text{access point5}})^2} \quad (35)$$

After expansion:

$$\text{dist2} \coloneqq (X^2_{\text{actual point}} - 2X_{\text{actual point}}\text{AccessP}_{5x} + \text{AccessP}^2_{5x} + Y^2_{\text{actual point}} - 2Y_{\text{actual point}}\text{AccessP}_{5y} + \text{AccessP}^2_{5y})^{1/2} \quad (36)$$

In order to get the most accurate prediction for the estimated point, it should be equal to the actual point; therefore the difference between **dist1** and **dist2** should be equal to **zero**:

Subtract Eq 34 to the power 2 from Eq 36 to the power 2:

$$0 = \mathrm{result} = (\mathrm{X_{estimated\ by\ KNN}} - \mathrm{X_{access\ point5}})^2 + (\mathrm{Y_{estimated\ by\ KNN}} - \mathrm{Y_{access\ point5}})^2\text{-}$$
$$(\mathrm{X_{actual\ point}} - \mathrm{X_{access\ point5}})^2 + (\mathrm{Y_{actual\ point}} - \mathrm{Y_{access\ point5}})^2 \qquad (37)$$

$$\begin{aligned}\mathrm{result} \coloneqq\ & \mathrm{X^2_{estimatedKNN}} - 2\mathrm{X_{estimatedKNN}AccessP_{5x}} + \mathrm{Y^2_{estimatedKNN}} - \\ & 2\mathrm{Y_{estimatedKNN}AccessP_{5y}} - \mathrm{X^2_{actual\ point}} + 2\mathrm{X_{actual\ point}AccessP_{5x}} - \\ & \mathrm{Y^2_{actual\ point}} - 2\mathrm{Y_{actual\ point}AccessP_{5y}}\end{aligned} \qquad (38)$$

By assuming that: $\mathrm{Y_{actual\ point}} = \frac{\mathrm{Y_{estimatedKNN}}}{\mathrm{X_{estimatedKNN}}}.\mathrm{X_{actual\ point}}$, and substituting in the **result** equation(38) , then divide the whole term by $X_{estimated\ by\ KNN}$^2 , then we get:

$$\begin{aligned}\mathrm{Equation1} \coloneqq\ & 1 - \frac{2\mathrm{AccessP_{5x}}}{\mathrm{X_{estimatedKNN}}} + \frac{\mathrm{Y^2_{estimatedKNN}}}{\mathrm{X^2_{estimatedKNN}}} - \frac{2\mathrm{Y_{estimatedKNN}AccessP_{5y}}}{\mathrm{X^2_{estimatedKNN}}} \\ & - \frac{\mathrm{X^2_{actual\ point}}}{\mathrm{X^2_{estimatedKNN}}} + \frac{2\mathrm{X_{actual\ point}AccessP_{5x}}}{\mathrm{X^2_{estimatedKNN}}} - \frac{\mathrm{Y^2_{estimatedKNN}X^2_{actual\ point}}}{\mathrm{X^4_{estimatedKNN}}} \\ & + \frac{2\mathrm{Y_{estimatedKNN}X_{actual\ point}AccessP_{5y}}}{\mathrm{X^3_{estimatedKNN}}}\end{aligned}$$
$$(39)$$

By assuming that: $X_{actual\ point} = \frac{X_{estimatedKNN}}{Y_{estimatedKNN}} \cdot Y_{actual\ point}$, and substituting in the **result** equation (38), then divide the whole term by $Y_{estimated\ by\ KNN}$^2 then we get:

$$\begin{aligned}\mathrm{Equation2} \coloneqq\ & \frac{\mathrm{X^2_{estimatedKNN}}}{\mathrm{Y^2_{estimatedKNN}}} - \frac{2\mathrm{X_{estimatedKNN}AccessP_{5x}}}{\mathrm{Y^2_{estimatedKNN}}} + 1 - \frac{2\mathrm{AccessP_{5y}}}{\mathrm{Y_{estimatedKNN}}} \\ & - \frac{\mathrm{Y^2_{actual\ point}X^2_{estimatedKNN}}}{\mathrm{Y^4_{estimatedKNN}}} + \frac{2\mathrm{X_{estimatedKNN}Y_{actual\ point}AccessP_{5x}}}{\mathrm{Y^3_{estimatedKNN}}} \\ & - \frac{\mathrm{Y^2_{actual\ point}}}{\mathrm{Y^2_{estimatedKNN}}} + \frac{2\mathrm{Y_{actual\ point}AccessP_{5y}}}{\mathrm{Y^2_{estimatedKNN}}}\end{aligned}$$
$$(40)$$

By assuming that: $X_{\text{actual point}} = \text{RationRange}_x * X_{\text{estimated KNN}}$, and substituting in the ***Equation1***, then we get:

$$\begin{aligned}\text{FirstRelation} \coloneqq 1 - \frac{2\text{AccessP}_{5x}}{X_{\text{estimatedKNN}}} + \frac{Y^2_{\text{estimatedKNN}}}{X^2_{\text{estimatedKNN}}} - \frac{2Y_{\text{estimatedKNN}}\text{AccessP}_{5y}}{X^2_{\text{estimatedKNN}}} \\ - \text{RatioRange}_x^2 + \frac{2\text{RatioRange}_x\text{AccessP}_{5x}}{X_{\text{estimatedKNN}}} - \frac{Y^2_{\text{estimatedKNN}}\text{RatioRange}_x^2}{X^2_{\text{estimatedKNN}}} \\ + \frac{2Y_{\text{estimateKNN}}\text{RatioRange}_x\text{AccessP}_{5y}}{X^2_{\text{estimatedKNN}}}\end{aligned} \tag{41}$$

By assuming that: $Y_{\text{actual point}} = \text{RatioRange}_y \cdot Y_{\text{estimatedKNN}}$, and substituting in the ***Equation2***, then we get:

$$\begin{aligned}\text{SecondRelation} : \\ = \frac{X^2_{\text{estimatedKNN}}}{Y^2_{\text{estimatedKNN}}} - \frac{2X_{\text{estimatedKNN}}\text{AccessP}_{5x}}{Y^2_{\text{estimatedKNN}}} + 1 - \frac{2\text{AccessP}_{5y}}{Y_{\text{estimatedKNN}}} \\ - \frac{\text{RatioRange}_y^2 X^2_{\text{estimatedKNN}}}{Y^2_{\text{estimatedKNN}}} + \frac{2X_{\text{estimatedKNN}}\text{RatioRange}_y\text{AccessP}_{5x}}{Y^2_{\text{estimatedKNN}}} \\ - \text{RatioRange}_y^2 + \frac{2\text{RatioRange}_y\text{AccessP}_{5y}}{Y_{\text{estimatedKNN}}}\end{aligned} \tag{42}$$

The used objective function will be the summation of the 2 relations **(FirstRelation and SecondRelation)**:

$$\begin{aligned}
\mathrm{GeneralRelation}: \\
= -\frac{1}{\mathrm{Y}^2_{\mathrm{estimatedKNN}}\mathrm{X}^2_{\mathrm{estimatedKNN}}}(&-2\mathrm{Y}^2_{\mathrm{estimatedKNN}}\mathrm{X}^2_{\mathrm{estimatedKNN}} \\
&+ 2\mathrm{X}_{\mathrm{estimatedKNN}}\mathrm{AccessP}_{5\mathrm{x}}\mathrm{Y}^2_{\mathrm{estimatedKNN}} - \mathrm{Y}^4_{\mathrm{estimatedKNN}} \\
&+ 2\mathrm{Y}^3_{\mathrm{estimatedKNN}}\mathrm{AccessP}_{5\mathrm{y}} + \mathrm{RatioRange}^2_{\mathrm{x}}\mathrm{Y}^2_{\mathrm{estimatedKNN}}\mathrm{X}^2_{\mathrm{estimatedKNN}} \\
&- 2\mathrm{RatioRange}_{\mathrm{x}}\mathrm{AccessP}_{5\mathrm{x}}\mathrm{Y}^2_{\mathrm{estimatedKNN}}\mathrm{X}_{\mathrm{estimatedKNN}} \\
&+ \mathrm{Y}^4_{\mathrm{estimatedKNN}}\mathrm{RatioRange}^2_{\mathrm{x}} - 2\mathrm{Y}^3_{\mathrm{estimatedKNN}}\mathrm{RatioRange}_{\mathrm{x}}\mathrm{AccessP}_{5\mathrm{y}} \\
&- \mathrm{X}^4_{\mathrm{estimatedKNN}} + 2\mathrm{X}^3_{\mathrm{estimatedKNN}}\mathrm{AccessP}_{5\mathrm{x}} \\
&+ 2\mathrm{Y}_{\mathrm{estimatedKNN}}\mathrm{AccessP}_{5\mathrm{y}}\mathrm{X}^2_{\mathrm{estimatedKNN}} + \mathrm{RatioRange}^2_{\mathrm{y}}\mathrm{X}^4_{\mathrm{estimatedKNN}} \\
&- 2\mathrm{X}^3_{\mathrm{estimatedKNN}}\mathrm{RatioRange}_{\mathrm{y}}\mathrm{AccessP}_{5\mathrm{y}} \\
&+ \mathrm{RatioRange}^2_{\mathrm{y}}\mathrm{Y}^2_{\mathrm{estimatedKNN}}\mathrm{X}^2_{\mathrm{estimatedKNN}} \\
&- 2\mathrm{RatioRange}_{\mathrm{y}}\mathrm{AccessP}_{5\mathrm{y}}\mathrm{Y}_{\mathrm{estimatedKNN}}\mathrm{X}^2_{\mathrm{estimatedKNN}})
\end{aligned} \tag{43}$$

Where the solution will be as following:

$$\begin{aligned}
\mathrm{X}_{\mathrm{ga,calculated}} &= \mathrm{RatioRange}_{\mathrm{x}} \cdot \mathrm{X}_{\mathrm{estimatedKNN}} \\
\mathrm{Y}_{\mathrm{ga,calculated}} &= \mathrm{RatioRange}_{\mathrm{y}} \cdot \mathrm{Y}_{\mathrm{estimatedKNN}}
\end{aligned} \tag{44}$$

If the selection process for selecting best individuals is not satisfied by GA process then the **GeneralRelation** can be modified as the following assumption:

$$\mathrm{RatioRange}^2_{\mathrm{x}} = \mathrm{RX}, \mathrm{RatioRange}^2_{\mathrm{y}} = \mathrm{RY} \tag{45}$$

Then the **GeneralRelation** will be modified as:

$$\text{Relation} := -\frac{1}{Y_{\text{estimatedKNN}}^{2} X_{\text{estimatedKNN}}^{2}} \Big(-2Y_{\text{estimatedKNN}}^{2} X_{\text{estimatedKNN}}^{2} + 2X_{\text{estimatedKNN}} \text{AccessP}_{5x} Y_{\text{estimatedKNN}}^{2} - Y_{\text{estimatedKNN}}^{4} + 2Y_{\text{estimatedKNN}}^{3} \text{AccessP}_{5y} + \text{RXY}_{\text{estimatedKNN}}^{2} X_{\text{estimatedKNN}}^{2} - 2\text{RatioRange}_{x} \text{AccessP}_{5x} Y_{\text{estimatedKNN}}^{2} X_{\text{estimatedKNN}} + Y_{\text{estimatedKNN}}^{4} \text{RX} - 2Y_{\text{estimatedKNN}}^{3} \text{RatioRange}_{x} \text{AccessP}_{5y} - X_{\text{estimatedKNN}}^{4} + 2X_{\text{estimatedKNN}}^{3} \text{AccessP}_{5x} + 2Y_{\text{estimatedKNN}} \text{AccessP}_{5y} X_{\text{estimatedKNN}}^{2} + \text{RYX}_{\text{estimatedKNN}}^{4} - 2X_{\text{estimatedKNN}}^{3} \text{RatioRange}_{y} \text{AccessP}_{5y} + \text{RYY}_{\text{estimatedKNN}}^{2} X_{\text{estimatedKNN}}^{2} - 2\text{RatioRange}_{y} \text{AccessP}_{5y} Y_{\text{estimatedKNN}} X_{\text{estimatedKNN}}^{2} \Big) \quad (46)$$

Where the solution of the objective function will be as following:

$$X_{\text{ga,calculated}} = \left(\left(\sqrt{\text{RX}}\right) + \text{RatioRange}_{x}\right)/2) \cdot X_{\text{estimatedKNN}}$$
$$Y_{\text{ga,calculated}} = \left(\left(\sqrt{\text{RY}}\right) + \text{RatioRange}_{y}\right)/2) \cdot Y_{\text{estimatedKNN}} \quad (47)$$

4.5.3 Genetic Algorithm Operation

According to the objective function derived above, the genetic algorithm can be used to find the min error distance that predicts the location of the actual points as following:

GA OPERATORS

The Genetic algorithm employs the following three operators:

a) Selection: After selecting the random points according to point ratio range **"**First Generation**"**, it is now to specify the selection method, by which the GA will select the best-estimated points that satisfy the objective function, to generate next generations Selection is one of the main operators in GAs, and relates directly to the Darwinian concept of the survival of the fittest. The main objective of the selection operators is to emphasize better solutions. This is achieved in two of the main steps of GA:

- **Selection of the new population**: A new population of candidate solutions is selected at the end of each generation to serve as the population of the next generation. The new population can be selected from only the offspring, or from both the parents and the offspring. The selection operator should ensure that good individuals do survive to the next generations.
- **Reproduction**: Reproduction is the process in which individual chromosomes are copied according to their fitness values, i.e., chromosomes with a higher value have a higher probability of contributing one or more offspring to the next generation. The fitness value (corresponds to value of the objective function) is the final arbiter of the chromosome's life or death. Reproduction producing offspring from selected parents by applying crossover and/or mutation operators.

There are many selection types to select the new population, for example Random selection, Proportional selection, Tournament selection and Rank-Based Selection. In this research, we used the Tournament selection method to selects a group of *nts* individuals randomly from the population, where $nts < ns$ (*ns* is the total number of individuals in the population). The performance of the selected *nts* individuals is compared and the best individual from this group is selected and returned by the operator.

b) Crossover: In crossover operation, members of the regenerated chromosomes are mated at random. In the first step, a pair of chromosomes from the regenerated pool is selected at random. Then the crossover site *x* is determined by generating a random number between 1 and *m-1* where *m* is the length of the chromosome. Then, two new chromosomes are formed by swapping all the genes between *x+1* and *m* inclusively.

Several crossover operators have been developed to compute the mask such as "Uniform crossover, Two-point crossover and One-point / Single point crossover. In our research, we focus "One-point / Single point crossover".

c) **Mutation:** The mutation operator arbitrarily alters one or more components of the selected structure, thus increasing the diversity of the population. Each position of each solution vector in the population undergoes a random change with a probability equal to the mutation rate. The aim of mutation is to introduce new genetic material into an existing individual; that is, to add diversity to the genetic characteristics of the population. Mutation is used in support of a crossover to ensure that the full range of allele is accessible to each gene. Mutation is applied at a certain probability, p to each gene of the offspring, to produce the mutated offspring.

4.5.4 Summary of GA steps

The improvement of indoor positioning using KNN algorithm by optimizing error distance to minimum its value using Genetic Algorithm, can be summarized in the following steps:

Step 1: After the estimated points are calculated using KNN or FSKNN Algorithm, Enter the GA and determine NVARS (number of variables in the objective function); it is initially taken to be equal 2 (x and y).

Step 2: Apply linear point ratio range as following:

- Assume a linear matrix of points for **x chromosome and y chromosome**;

Population Matrix = $\begin{bmatrix} \mathbf{0.8775} & \mathbf{1} \\ \boldsymbol{up\ to} & \boldsymbol{up\ to} \\ \mathbf{1} & \mathbf{1.115} \end{bmatrix}$ **; where** length of each column equals to population length **(contains 100 element).**

- This Population Matrix represents the first population **(first generation of chromosomes that will be entered the GA and will be treated by the GA operators)**

Step 3: Determine the selection method that will be applied by the GA. It is selected to be selection tournament method.

Step 4: Determine the crossover method that will be applied by the GA. It is selected to be cross over single point; where the population matrix will generate new chromosomes by single point crossover; for illustration, it will be like the following:

$$\begin{array}{cc} \mathbf{0.8775} & \mathbf{1} \\ \mathbf{0.88} & \mathbf{1.02} \\ \mathbf{0.9} & \mathbf{1.05} \\ \mathbf{0.92} & \mathbf{1.07} \\ \mathbf{0.94} & \mathbf{1.09} \\ \mathbf{1} & \mathbf{1.115} \end{array} \Longrightarrow \begin{array}{cc} \mathbf{1} & \mathbf{0.8775} \\ \mathbf{0.88} & \mathbf{1.02} \\ \mathbf{1.05} & \mathbf{0.9} \\ \mathbf{0.92} & \mathbf{1.07} \\ \mathbf{1.09} & \mathbf{0.94} \\ \mathbf{1} & \mathbf{1.115} \end{array}$$

Before Crossover   After Cross Over

Step 5: Determine the mutation method that will be applied by the GA. It is selected to be Gaussian mutation; where the population matrix will generate new chromosomes by using mutation; for illustration it will be like the following:

$$\begin{array}{cc} \mathbf{0.88} & \mathbf{0.8775} \\ \mathbf{1} & \mathbf{1.02} \\ \mathbf{1.05} & \mathbf{1.07} \\ \mathbf{0.92} & \mathbf{0.9} \\ \mathbf{1} & \mathbf{0.94} \\ \mathbf{1.09} & \mathbf{1.115} \end{array} \Longleftarrow \begin{array}{cc} \mathbf{1} & \mathbf{0.8775} \\ \mathbf{0.88} & \mathbf{1.02} \\ \mathbf{1.05} & \mathbf{0.9} \\ \mathbf{0.92} & \mathbf{1.07} \\ \mathbf{1.09} & \mathbf{0.94} \\ \mathbf{1} & \mathbf{1.115} \end{array}$$

After Mutation   Before Mutation

Step 6: GA takes the population matrix after mutation and enter its values into the objective function and find the fitness error of objective function, by this fitness error the selection method is used to select the suitable chromosomes (best individuals). This process is repeated 200 times (total number of generation) in each generation the 3 operators are applied to each chromosome until the GA selects the best individuals.

Step 7: GA selects the best individual and then the most accurate estimated points are calculated as following:

Let's say that the best chromosome is selected to be 1.105 for X_{ratio} and 0.97 for Y_{ratio}; then

$$X_{GA\ estimated} = X_{estimated\ by\ KNN\ or\ FSKNN} * X_{ratio}$$

$$Y_{GA\ estimated} = Y_{estimated\ by\ KNN\ or\ FSKNN} * Y_{ratio}$$

Chapter 5

RESULTS AND DISCUSSIONS

This chapter presents the simulation results carried out in MATLAB .Furthermore this chapter explain the simulation scenario and discussed the obtained results. The RSS based basic localization technique, the proposed technique tested in indoor environment, and results are compared. The number of Aps and its impact on improvement of results are also addressed.

5.1 Simulation result

Case A

The first scenario is mapped upon an area of dimension 100x100 (meters) in which there is a LOS distance between the APs and the RPs and the number of Aps used in this simulation are five and these five Aps are placed at the simulation area according to the coordinates points. The coordinates of the five Wi-Fi access points are shown in Table 1.

Table 1: Coordinates of 5 APs

AP #	X-coordinate	Y-coordinate
1	0	0
2	0	50
3	50	0
4	50	50
5	25	25

Radio map was created by dividing the total area i-e 10000 m^2 into grids by spacing of 2m in both x and y dimensions. See figure (5.1). The total 2500 reference points were generated and value of RSS are taken at each point across all APs and stored in the fingerprinting database. The 1000 random position were taken in inside coverage area and match with fingerprint database of 2500 points using our proposed method. Results are presented in CDF for both the existing RSS based fingerprinting technique and the proposed technique. The results are very encouraging, the CDF plot can be seen in figure (5.2-5.6) and results in Table 2.

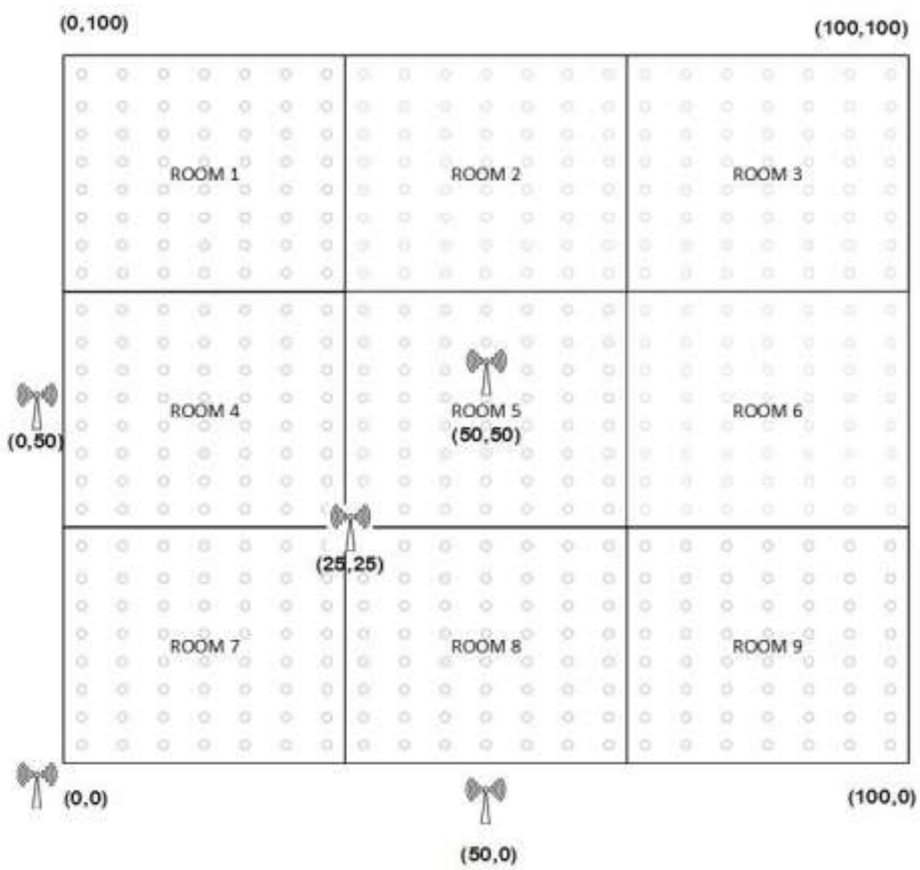

Figure 5.1: Room dimensions and Grids, APs placement also shown (Case A)

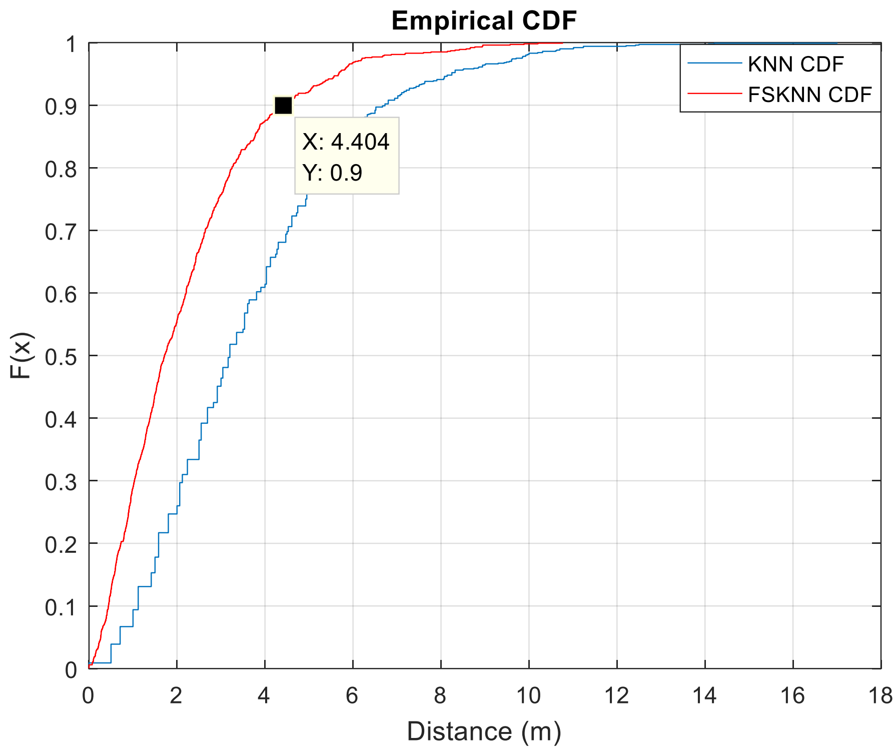

Fig 5.2 CDF Plot of KNN and FSKNN

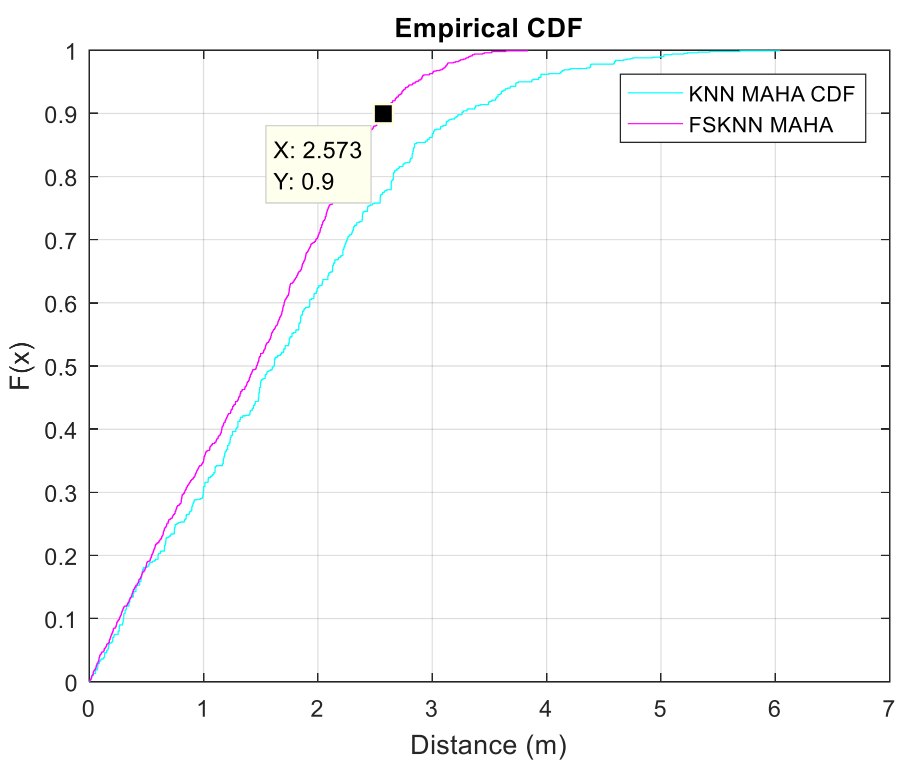

Fig 5.3 CDF Plot of KNN MAHA and FSKNN MAHA

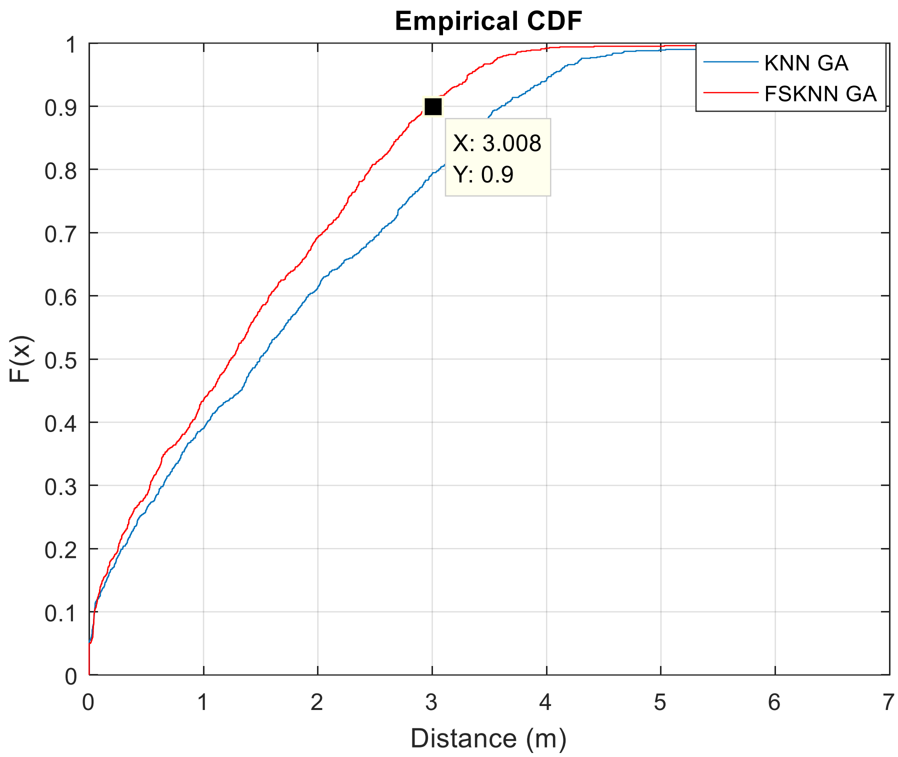

Fig 5.4CDF Plot of KNN GA and FSKNN GA

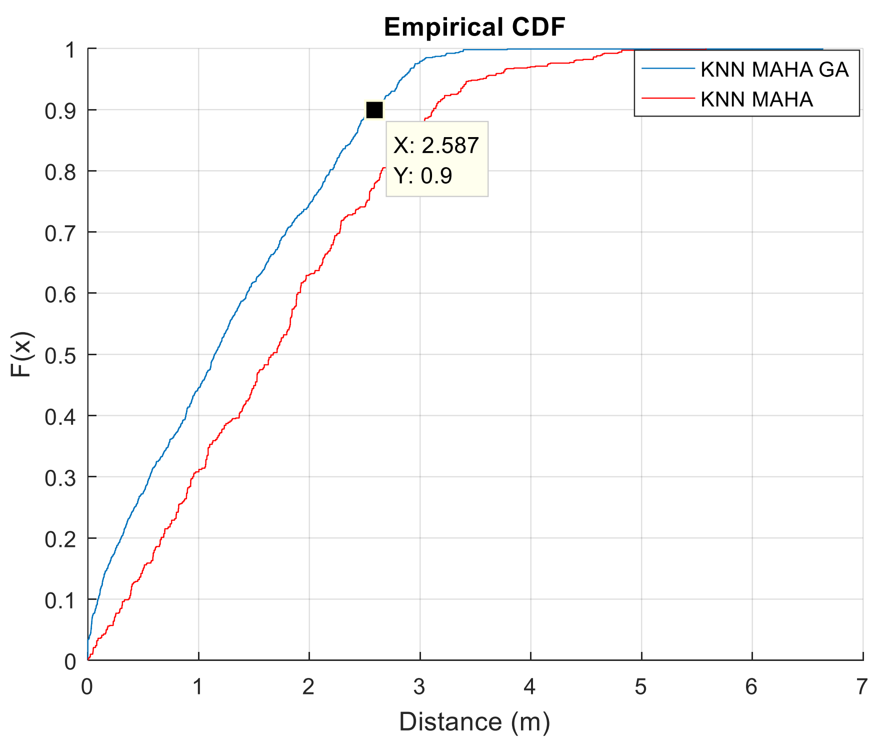

Fig 5.5 CDF Plot of KNN MAHA and KNN MAHA GA

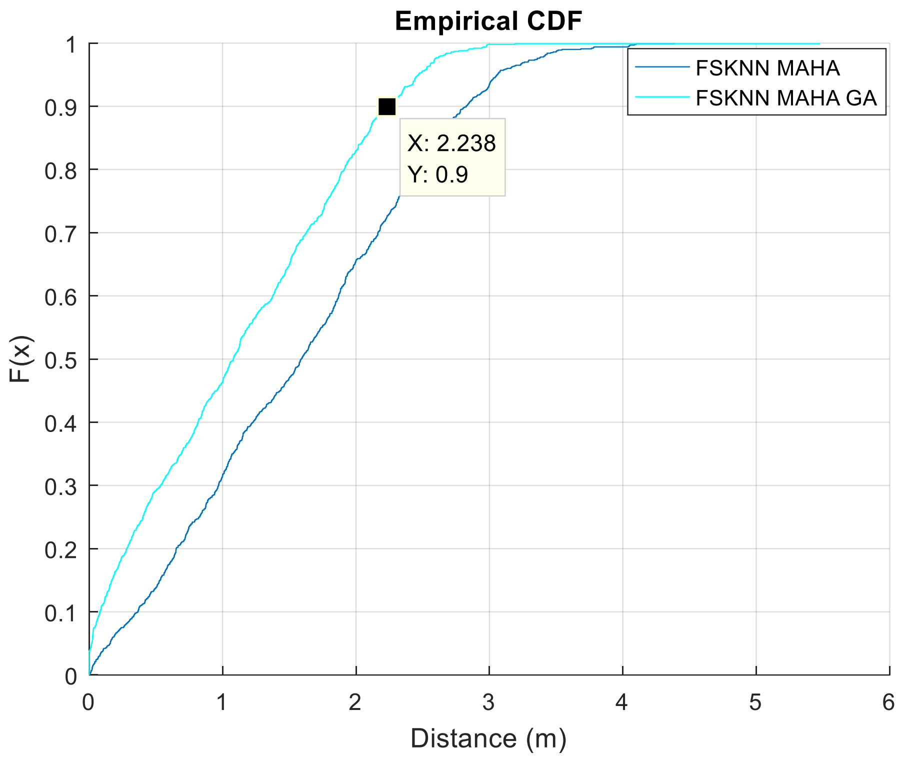

Fig 5.6 CDF Plot of FSKNN MAHA and FSKNN MAHA GA

Table 2: CDF Plot Error Comparison Case A

Technique	**CDF (error) at 0.9**
RSS based KNN	6.72
FSKNN	4.40
KNN GA	3.605
FSKNN GA	3.008
KNN MAHAL	3.26
FSKNN MAHAL	2.57
KNN MAHAL GA	2.587
FS KNN MAHAL GA	2.238

CASE B

In the second scenario, the area of simulation is similar to previous dimension i.e. 100×100m. However, the deployment of Aps are changed. The number of APs used in this Simulation are seven. The seven Aps are placed in the following coordinates points: AP1 (0, 0), AP2 (0, 50), AP3 (50, 50) and AP4 (50, 0), AP5 (25, 25), AP6 (50,25), AP7 (0, 100).). The spacing of grid point are same as in the previous case A. By increasing, the number of Aps the accuracy of our proposed technique is improved as compare to the existing RSS based technique. The CDF plot our proposed technique can be seen in figure.(5.7-5.12)

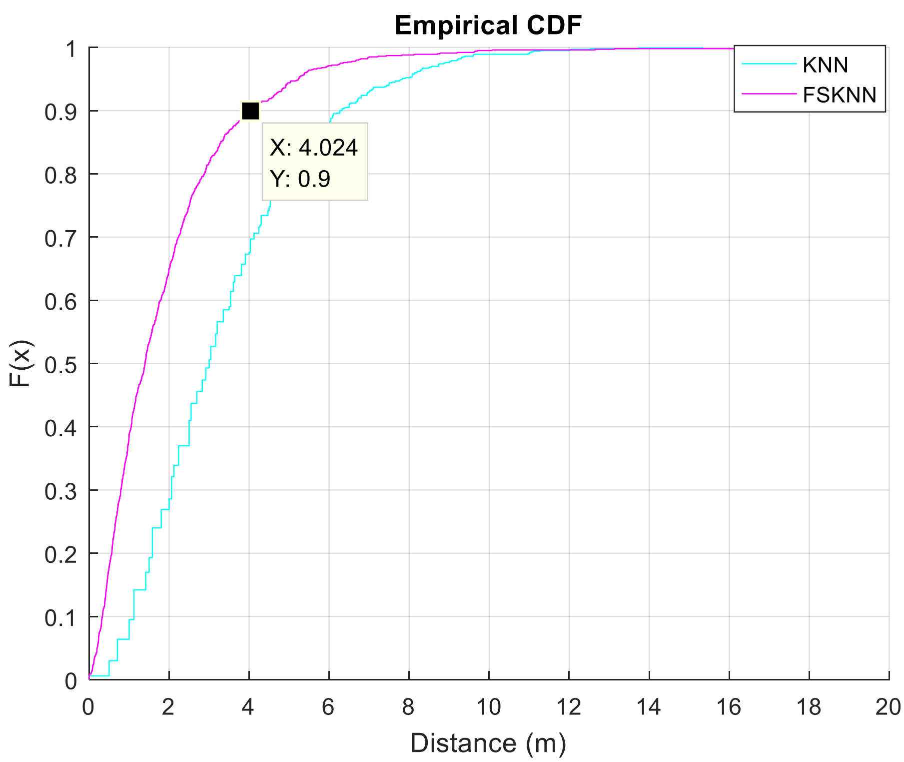

Fig 5.7 CDF Plot of KNN and FSKNN

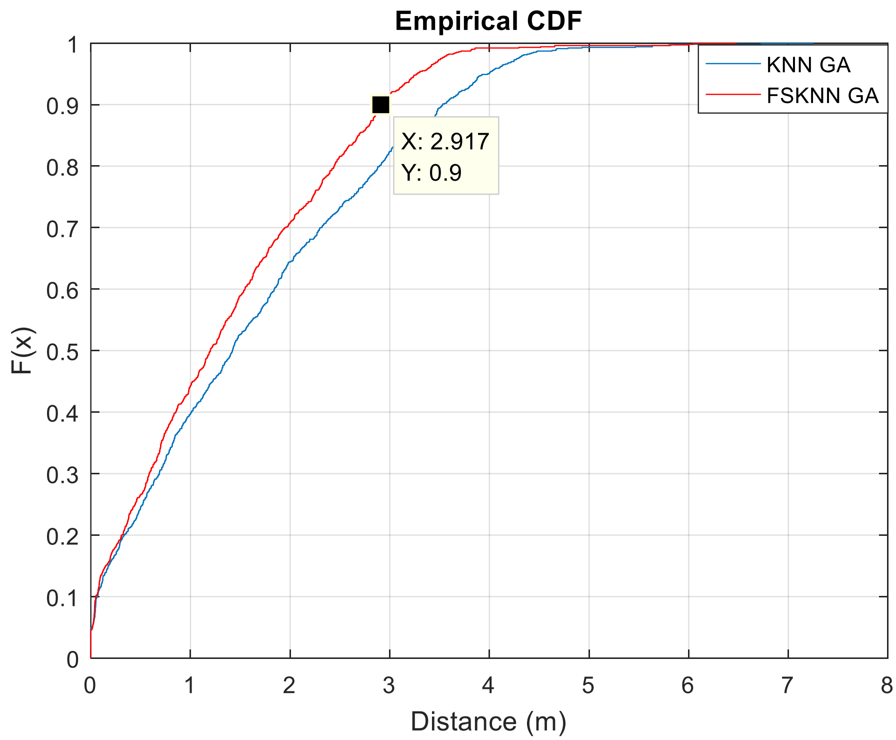

Fig 5.8 CDF Plot of KNN GA and FSKNN GA

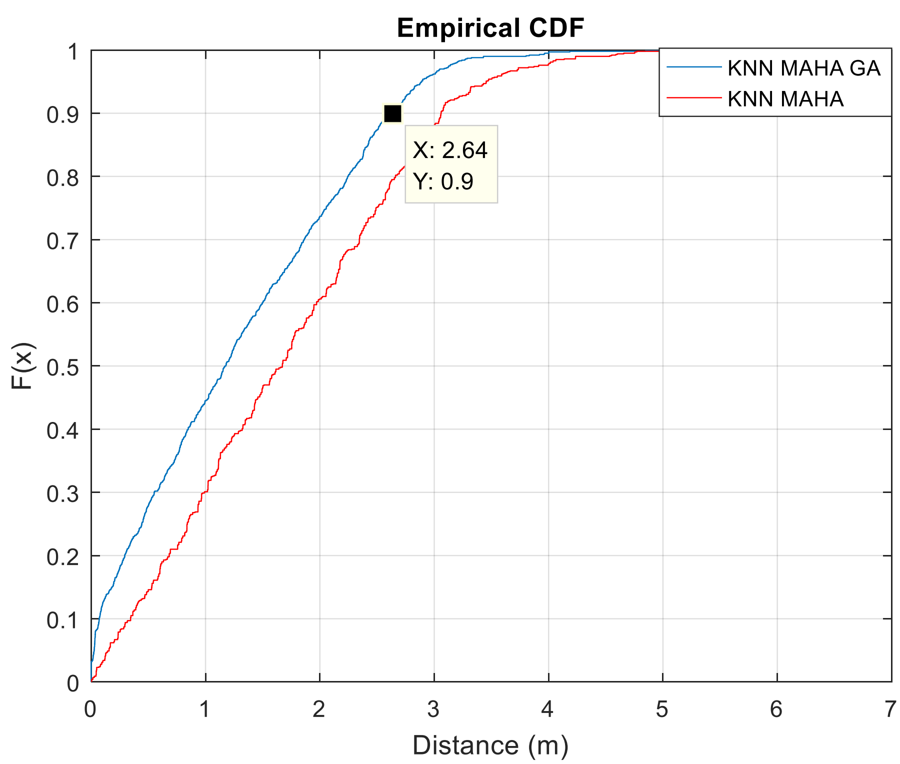

Fig 5.9 CDF Plot of KNN MAHA GA and FSKNN MAHA GA

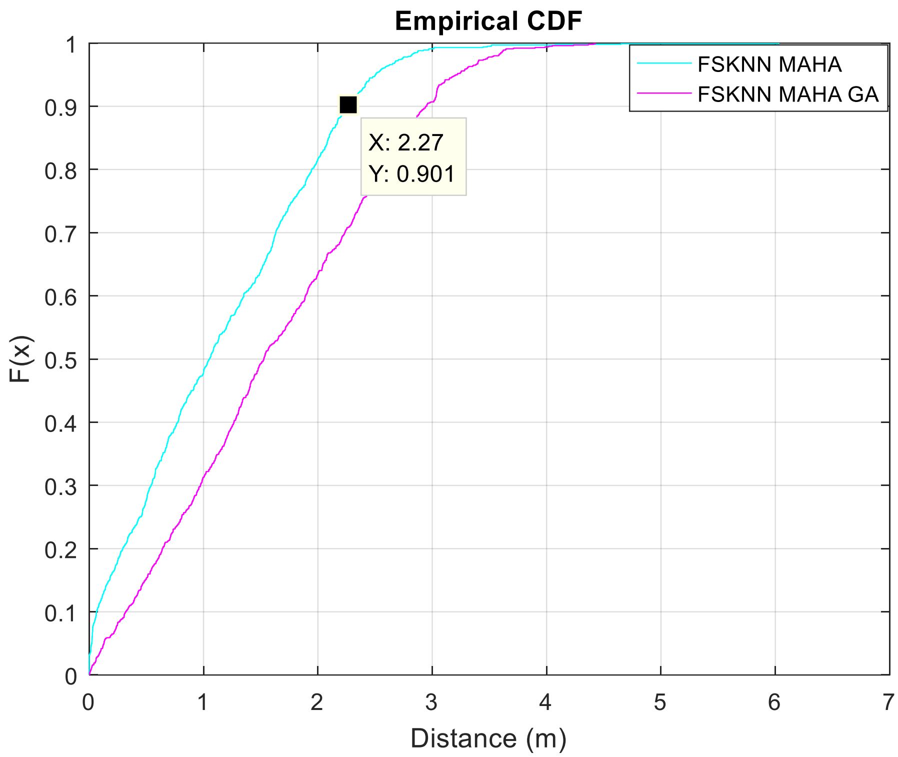

Fig 5.10 CDF Plot of FSKNN MAHA GA and FSKNN MAHA

Table 3: CDF Plot of Error Comparison Case B

Technique	**CDF (error) at 0.9**
RSS based KNN	6.625
FSKNN	4.024
KNN GA	3.531
FSKNN GA	2.917
KNN MAHAL	3.061
FSKNN MAHAL	2.952
KNN MAHAL GA	2.64
FS KNN MAHAL GA	2.27

It can be seen from the results that placement of APs has an impact on the accuracy of the system and it can be observed that increasing the number of APs can have a positive impact on the location accuracy.

Chapter 6

CONCLUSION

In this thesis, a fingerprinting technique has been proposed to improve the localization accuracy of commonly used WLAN localization systems. In our proposed technique, FS-KNN and Genetic algorithm are used along with the RSS values and shows a significant improvement in the overall results. The proposed technique does not require any extra infrastructure and is simple to understand. It makes the fingerprints unique and enriches them, which leads towards an improvement in the system's performance.

This research work also focuses on implement of FSKNN algorithm and Genetic algorithm. One of the big issue with KNN algorithm is that in calculating the signal distance of different RSS levels that consider the equal RSS distances and cannot be distinguished between the different RSS values at different RSS level that caused different physical distance in reality.

In this research, an attempt has been made to improve indoor positioning using KNN, FSKNN algorithm by optimizing error distance to a minimum value using Genetic Algorithm. It is a subject of minimizing error distance of the estimated position versus actual position and improve the localization of accuracy. As discussed in the results section that our proposed technique is consistent and provides better solution to the problem as compared to the systems based on RSS. The idea can be further investigated by other location dependent parameters for fingerprinting techniques and placement of AP to the development of RSS model and also relations depending on the characteristics of the signal. In future, it can be implemented having story of building of with multiple floor with different room dimension to provide better localization solutions for indoor environments.

Chapter 7

REFERENCES

[1] Liu, Hui, et al. "Survey of wireless indoor positioning techniques and systems." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.6 (2007): 1067-1080.

[2] Machaj, Juraj, Peter Brida, and Jozef Benikovsky. "Using GSM signals for fingerprint-based indoor positioning system." *ELEKTRO, 2014*. IEEE, 2014.

[3] Gundegard, David, et al. "Cellular positioning using fingerprinting based on observed time differences." *Smart Communications in Network Technologies (SaCoNeT), 2013 International Conference on*. Vol. 1. IEEE, 2013.

[4] Kasantikul, Kittipong, et al. "An enhanced technique for indoor navigation system based on WIFI-RSSI." *Ubiquitous and Future Networks (ICUFN), 2015 Seventh International Conference on*. IEEE, 2015.

[5] Kotanen, Antti, et al. "Experiments on local positioning with Bluetooth." *Information technology: Coding and computing [computers and communications], 2003. Proceedings. ITCC 2003. International conference on*. IEEE, 2003..

[6] Cong, Li, and Weihua Zhuang. "Hybrid TDOA/AOA mobile user location for wideband CDMA cellular systems." *IEEE Transactions on Wireless Communications* 1.3 (2002): 439-447.

[7] Wu, Kaishun, et al. "CSI-based indoor localization." *IEEE Transactions on Parallel and Distributed Systems* 24.7 (2013): 1300-1309.

[8] Guvenc, Ismail, and Chia-Chin Chong. "A survey on TOA based wireless localization and NLOS mitigation techniques." *IEEE Communications Surveys & Tutorials* 11.3 (2009).

[9] *R. S. a. J. S. N, "TIME OF ARRIVAL BASED LOCALISATION IN* WIRELESS *SENSOR NETWORKS: A LINEAR APPROACH," Signal & Image Processing : An International Journal (SIPIJ), vol. 4, no. 4, August 2013.*

[10] Khalaf-Allah, Mohamed. "Time of arrival (TOA)-based direct location method." *Radar Symposium (IRS), 2015 16th International*. IEEE, 2015.

[11] R. M. V. a. R. M. Buehrer, "Asynchronous Time-of-Arrival-Based Source Localisation," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, 2013.

[12] Bahl, Paramvir, and Venkata 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*. Vol. 2. Ieee, 2000.

[13] Khodayari, Shahrzad, Mina Maleki, and Elham Hamedi. "A RSS-based fingerprinting method for positioning based on historical data." *Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2010 International Symposium on*. IEEE, 2010.

[14] X. C. a. Y. D. Peerapong Torteeka, "Hybrid Technique for Indoor Positioning System based on Wi-Fi Received Signal Strength Indication," in *2014 International Conference on Indoor Positioning and Indoor Navigation*, 27th-30th October, 2014.

[15] J. M. a. P. Brida, "Impact of Wi-Fi Access Points on Performance of RBF Localisation Algorithm," in *ELEKTRO 2012*, May 2012.

[16] N. G. D. A. Ali Khalajmehrabadi, "Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges," *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, 2016.

[17] Kul, Gökhan, Tansel Özyer, and Bülent Tavli. "IEEE 802.11 WLAN based real time indoor positioning: Literature survey and experimental investigations." *Procedia Computer Science*34 (2014): 157-164.

[18] Fang, Shih-Hau, and Tsung-Nan Lin. "A novel access point placement approach for WLAN-Based Location systems." *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*.IEEE, 2010.

[19] Juraj MACHAJ, Peter BRIDA, "Using of GSM and Wi-Fi Signals for Indoor Positioning Based On Fingerprinting Algorithms", *INFORMATION AND COMMUNICATION TECHNOLOGIES AND SERVICES VOLUME: 13 NUMBER: 3, 2015 SEPTEMBER*

[20] Umair, Mir Yasir, Kopparapu Venkata Ramana, and Yang Dongkai. "An enhanced K-nearest neighbor algorithm for indoor positioning systems in a WLAN." *Computing, Communications and IT Applications Conference (ComComAp), 2014 IEEE*. IEEE, 2014.

[21] Jungmin So, Joo-Yub Lee, Cheal-Hwan Yoon and Hyunjae Park, "An Improved Location Estimation Method for Wi-Fi Fingerprint based Indoor Localization", International *Journal of Software Engineering and Its Applications Vol. 7, No. 3, May, 2013*

[22] Xizhe Wang, Jian Qiu*, Sheng Ye, Guojun Dai, "An Advanced Fingerprint-based Indoor Localization Scheme for WSNs", *2014 IEEE 9th Conference on Industrial Electronics and Applications (ICIEA)*

[23] Zengwei Zheng, Yuanyi Chen, Sinong Chen, Lin Sun, and Dan Chen, "BigLoc: A Two-Stage Positioning Method for Large Indoor Space", *International Journal of Distributed Sensor Networks Volume 2016, Article ID 1289013, 9 pages*

[24] Jukka Talvitie, Markku Renfors, and Elena Simona Lohan, "Distance-Based Interpolation and Extrapolation Methods for RSS-Based Localization With Indoor Wireless Signals", *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 64, NO. 4, APRIL 2015*

[25] Chi-Fu Hung, Chiu-Ching Tuan and Fong-Mao Jhuang," Improving Wi-Fi Indoor-Positioning Accuracy by Using AP Selection and Adaptive Pattern Matching", *Journal of Applied Science and Engineering, Vol. 19, No. 3, pp. 337_346 (2016)*

[26] Van Haute, Tom, et al. "Optimizing time-of-arrival localization solutions for challenging industrial environments." *IEEE Transactions on Industrial Informatics* 13.3 (2017): 1430-1439.

[27] Wang, Xuyu, et al. "CSI-based fingerprinting for indoor localization: A deep learning approach." *IEEE Transactions on Vehicular Technology* 66.1 (2017): 763-776.

[28] Dewing, Alexander, Robert Sun, and Xiaonan Tong. "Localization Through Wireless Access Point Channel State Information."

[29] Chen, Yingying, et al. "A practical approach to landmark deployment for indoor localization." *Sensor and Ad Hoc Communications and Networks, 2006. SECON'06. 2006 3rd Annual IEEE Communications Society on*. Vol. 1. IEEE, 2006.

[30] Laitinen, Elina, and Elena-Simona Lohan. "Are all the access points necessary in WLAN-based indoor positioning?." *Localization and GNSS (ICL-GNSS), 2015 International Conference on*. IEEE, 2015.

[31] Qiuyun Chen, Bang Wang, Xianjun Deng, Yijun Mo, and Laurence T. Yang, "Placement of Access Points for Indoor Wireless Coverage and Fingerprint-based Localization", *2013 IEEE International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing*

[32] Farkas, Károly, Árpád Huszák, and Győző Gódor. "Optimization of Wi-Fi access point placement forindoor localization." *Journal IIT (Informatics & IT Today)* 1.1 (2013): 28-33.

[33] Li, Dong, et al. "Measurement-Based Access Point Deployment Mechanism for Indoor Localization." *Global Communications Conference (GLOBECOM), 2015 IEEE*. IEEE, 2015.

[34] Tang, Shu, Lin Ma, and Yubin Xu. "A novel AP placement algorithm based on user distribution for indoor WLAN system." *China Communications* 13.10 (2016): 108-118.

[35] Baala, Oumaya, You Zheng, and Alexandre Caminada. "The impact of AP placement in WLAN-based indoor positioning system." *Networks, 2009. ICN'09. Eighth International Conference on*. IEEE, 2009.

[36] Tummala, Dinesh, and Robert G. Akl. *Indoor Propagation Modeling at 2.4 GHZ for IEEE 802.11 Networks*. Denton, TX: U of North Texas, 2005. Print.

[37] Fischer, Sven. "Observed time difference of arrival (OTDOA) positioning in 3GPP LTE." *Qualcomm White Pap* 1.1 (2014): 1-62.

[38] Suma S. Cherian and Ashok N. Rudrapatna, "LTE Location Technologies and Delivery Solutions" *Bell Labs Technical Journal 18(2), 175–194 (2013) © 2013 Alcatel-Lucent*

[39] Youssef, Moustafa, and Ashok Agrawala. "The Horus WLAN Location Determination System."*Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services - MobiSys '05* (2005)

[40] Li, D., Zhang, B. and Li, C., 2016. A Feature Scaling Based $ k $-Nearest Neighbor Algorithm for Indoor Positioning Systems. IEEE Internet of Things Journal, 3(4), pp.590-597.