POLITECNICO DI TORINO

Master degree course in Computer Engineering

Master Degree Thesis

Indoor localization based on Wi-Fi Probe Requests

Supervisor

Prof. GIOVANNI MALNATI

Candidate:

DERBEW ALEMAYEHU LISHANE

Mat. 202732

February 20

Summary

The Internet of Things (IoT) drives the evolution of the Internet and regarded as a great potential to improve quality of life for the surging number of elderly people significantly [9]. As Android operating system gains immense popularity and usage these days, it is a trend to get use of it for the wider access of Internet of Things (IoT) utility [9]. We are entering a new era of computing which the technology that many are calling the Internet of Things.

Global positioning system (GPS) does not provide generally a good positioning performance in an indoor location because of many reasons (Henniges, 2012). On the other hand, other alternatives such as the WI-FI technology has become recently in a popular use to provide indoor localization. That is due to many reasons, such as the wide spread of WI-FI infrastructure in the indoor environments and the low cost of this technology [12].

Saying that, the finding of low cost indoor localization tool or device lead us to design, implement this thesis as a generic solution, which bases on a low cost, tiny and portable microcontroller chip(Esp8266), by using the probe request in a promiscuous(sniffing) mode.

The work presented in this thesis mainly concerned about the study on the behavioral movement of smart devices users individually and jointly at a given time, using Wi-Fi controller integrated microcontroller chip. It report the existence of those users, based on point of reference(area) and time, aiming to provide the organized and tracked pattern of stations movement. It localize the frequent people and report to the system, that rely and need to make action which may promote the service offered by the unit to make decisions relying on this computed , processed data as input.

The main contribution of this thesis is to propose a generic indoor localization based solution for studying, analyzing smartphone, mobile users(stations) behavioral movement individually and by pairing, using Esp8266 microchip with full TCP/IP stack and microcontroller capability based

on Wi-F- probe request in promiscuous mode(Sniffing). The application is composed of different module, which can be divided in to three main components. Esp8266 client (station sniffer), Esp8266 server (instant sniffed data sender) and Main centralized server (received station identity, compute, apply algorithm and feed organized and relational data to a REST based web Service).

Acknowledgment

I would first like to thank my thesis supervisor Prof. Giovanni Malnati his continuous support whenever I had a question about my research. His patient guidance, excellent supervision and constant support was significant throughout my work. I also wish to express my gratitude to the stuff of Istituto Superiore Mario Boella (*ISMB*) who rendered support during the period of my project work.

Finally, I must express my very profound gratitude to my parents and dear friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Acronyms

IOT Internet of Things
RX/TX Transmit, Receive
Wi-Fi Wireless Fidelity
RFID Radio Frequency Identification
R&D Research and Development
1G first data transmission standards
2G Second Generation Technology
3G Third Generation Technology
4G Fourth Generation Technology
GPS Global Positioning System
UAV Unmanned Aerial Vehicle
FCC Federal Communication commission
GHZ Gigahertz
UWB Ultra Wide-Band
MHZ Megahertz
IEEE Institute of Electrical and Electronics Engineers
RSS Really Simple Syndication
LQI La Quinta Corporation
RSSI Received Signal Strength Indicator
VHF Very High Frequency
GSM Global System for Mobile (Communication)
WLAN Wireless Local Area Network
AP Access Point
SSID Service Set Identifier
MAC Media Access Control
HDK Hardware Development Kit
GPIO General-Purpose Input/output
UART Universal Asynchronous Receiver/ Transmitter
RAM Random Access Memory
ESP Encapsulating Security Control
DHCP Dynamic Host Configuration Protocol
DNS Domain Name System
SSL Secure Socket Layer

JSON JavaScript Object Notation
I/O Input/output
RST ReSet
VCC Virtual Channel Connection
GND Ground
IDE Integrated Development Environment
AMPDU Aggregated MAC Protocol Data Unit
LDPC Low-Density Parity-Check
WEP Wired Equivalent Privacy
MIC Minimum Inhibitory Concentration
CCMP Cipher Block Chaining Message
FCS Frame Check Sequence
USB Universal Serial Bus
USART Universal Synchronous/Asynchronous Receiver/Transmitter
TCP Transmission Control Protocol
UDP User Datagram Protocol
OUI Organizational Unique Identifier

HDK Hardware Development Kit

Contents

List of Figures

Chapter One

1. General introduction

1.1 Introduction

The Internet of Things is comprised of smart machines (device, sensor) interacting and communicating with each other or to other machines, environments, objects and infrastructures. As a result, huge amount of data is being generated, and that data is being processed into useful actions that can command and control things to make our lives much easier and safer and to reduce our impact on the environment [10].

We see the Internet of Things as billions of smart, interconnected "things" (a sort of "universal global neural network" in the cloud) that will encompass every aspect of our lives, and its creation is the intelligence that embedded processing provides [11]. The ability and creativity of this new era is limitless, with an amazing potential and efficiency to improve our lives. This thesis is an extensive reference to the possibilities, utility, applications and the evolution of the Internet of Things.

Probe Request frame is type of frame, which is transmitted by the client to scan for 802.11 networks in the area. The identifier of requested BSS network (SSID) along with the supported rates that the client station can communicate with, transmitted in the frame body of this request to the access points. A broadcast SSID used in case there is no specified or already defined network [12].

Wi-Fi is a major component for communication in mobile devices, for example, phone and tablet. Researchers take advantage of Wi-Fi signal to build indoor localization system, life pattern analysis, human activity recognition and so on [14].

The most interesting Wi-Fi packet to us in this case is the Probe Request Frame. Smartphones, laptops, and other devices, which are not currently connected to a Wi-Fi network send out this packet. Most Android and iPhone devices send out this request every 40 to 60 seconds, which makes using these to track the movement of people specifically useful. I should note that there is

no location information embedded into these packets. We just know, that if we received a probe request from a certain device, it is within a certain distance of the monitoring chip, A device that is not connected to any network will send out a probe request frame to not only the general public, but also targeting specific access points [13].

1.2 Statement of the problem

While the main The objective of this thesis is to study the behavioral movement of smart devices users individually and jointly at a given time using Wi-Fi controller integrated microcontroller chips and report the existence of those users based on point of reference and time aiming to solve the unorganized and untraced pattern of wireless network in a given area at specific place where the analysis is needed and localize the frequent people and report to system which can rely and need to make action which may promote the service offered by the unit or to make some decisions relying on this data as input.

Since Trace backing and real time based system are the leading technology in this 21 century, a lot can be said and done, this thesis mainly focuses on following, listening, learning and presenting users, behavioral, historical movement, basing the common device which entirely integrated to our day to day activity smartphone, tablets, laptops even desktop computer, The mechanism which brings this dummy devices into a very vital and key tools which can even base to be the analysis can be performed is the wireless Wi-Fi probe request, that can be traced by during authentication, data communication of our devices to the nearby internet feeder access point.

This second by second, minute by minute reaching up to hours and many days data collection, cascaded with a series of peers communication, transmission, studying and analyzing, algorithmically grouping pairs and a lot of techniques which is described in this thesis will directed us to time based analysis referring the aerial information will able many sectors to build up their systems with the need of their customer.

Hotels, Tourisms, Transportation system service providers, supermarkets and museums also the rest sectors who their aims are interested on users / customer based real time preference or historical and learned behavior may seek the solution of this thesis main work.

1.3 Main Contribution

The main contribution of this thesis is to propose a generic indoor localization based solution for studying, analyzing smartphone, mobile users(stations) behavioral movement individually and by pairing, using Esp8266 microchip with full TCP/IP stack and microcontroller capability based on Wi-F- probe request in promiscuous mode(Sniffing). The application is composed of different module, which can be divided in to three main components. Esp8266 client (station sniffer), Esp8266 server (instant sniffed data sender) and Main centralized server (received station identity, compute, apply algorithm and feed organized and relational data to a REST based web Service).

1.4 Thesis Outline

This paper is organized in different chapters the, in the next three chapters the Designed system is described and the final two chapter talks results with the sample tests and the future works.

Chapter 2:- continued in deep the value of internet of things in accordance with Indoor localization then moved in detail about Indoor localization meaning, introduction and specifically touch Wireless based indoor localization, its technology and significance, which is the main objective of the thesis. Finally concludes by stating the functionality and areas (sectors) which implements and benefit more.

Chapter 3:- directly talks in deep the used technology and implementation phase focusing on the design, setup, software, hardware, database used and their functionality, Next to that it proceed by defining how the implementation of the thesis act up on them briefly.

Chapter 4:- talks the rise (threat) of MAC Randomization up on Wi-Fi tracking process and attacks, which emerged to leverage this features.

Chapter 5:- sums up all and present each major steps of the entire implementation output on the collected sample data during the thesis work.

Chapter 6:- defined conclusion of the thesis, difficulties and limitation up on the implementation phase and finally future work of the thesis stated.

Chapter Two

2. State of the Art

2.1 The Value of the Internet of Things

The true value of the Internet of Things does not lay in the lights turning on when the car reaches the driveway, but rather the data that the connected devices collect about its users. Imagine a hospital with connected devices. The data collected from those devices outputs information on the status of patients and runs analytics on the various monitoring machine, helping the hospital to run as optimally as possible.

The collection of data from devices will allow consumers, businesses and even entire connected cities to run more efficiently. However, collecting large amounts of data presents challenges [15].

“Some of the challenges that still need to be figured out are partially around the algorithms that can process the data and give you something valuable out of it,” Galvez said. “What are you actually taking out of all this data you are collecting?” [15]

With the collection of information come major privacy and security concerns for consumers. Both Galvez and Jones agree that it’s up to the manufacturers of the products to ensure they are protecting user data.

“It’s up to the individual companies that are building these products to make sure they’re safeguarding their user’s information by protecting the data as best they can,” *Galvez said* [15].

Despite security concerns, data collection is a key component of the realization of the Internet of Things and is embedded in the end goal of the it.

“The goal would be for all these devices to talk to each other and people to have access to this information depending on the value they can device from it,” *Galvez said* [15].

When it comes to placing a number of the economy of the IoT, different experts have different predictions. Jones said some experts call it $1.2 trillion by 2025, while Cisco estimates the industry to reach $14.4 trillion in the same period. Jones simple estimated "the economic potential of this type of environment is something that is going to dwarf the Internet and mobile combined." The Internet of Things is not merely a step along the path to digital transformation; it is the driving force. By 2025, the Internet of Thing's economic impact could reach US$11 trillion, or 11% of global economic value, and by 2030 the Internet of Things could influence nearly the entire economy [15].

The heart of Internet of Things is smart devices (Sensors). They will come in every shape and size, from Nano chips and smart dust to gigantic machines. The number of connected things will grow exponentially from 15 billion in 2015 to 200 billion in 2020.

Meanwhile, while the Internet of Things seems to be just one technology, it actually inter-related other major technologies, such as cloud computing, data analytics, mobile, sensors, and machine-to-machine communications since the relevance of this all gathered and retrieved data are aiming some realistic objective.

The main and real value of the Internet of Things doesn't come from all the connections it creates but from the data it generates. With real-time data analytics, the Internet of Things becomes a live communications network for fostering insights and gradual improvements. It will also become the creation of Live Business, in which companies will be able to sense and respond to customers in the moment based on the analysis gained from those smart Information retrieved from the data.

2.2 New Interactions between People, Things and Machines

Algorithms that operate and control machines are becoming increasingly sophisticated and complex. We already have self-learning robots and self-driving cars. By setting up machine-to-machine communications and creating algorithms that enable drive things and machines to control each other, we will create a new level of automation that will dramatically change the way we interact, work, and collaborate in daily basis.

We will see many new communications between things, people and machines which today look like science fiction. Many of the things in our daily lives at home and work will interact with

each other, making us to depend and use them in new ways. In this connected world, having access to things like cars will become more important than owning them.

At the human-to-machine level, we will be able to gauge the status of machines, receive warnings when they need maintenance, and control their roles in production at any given moment. The IoT will also enable us to control robotic extensions of our bodies, such as replacements for lost or disabled limbs or suits to improve our strength and other capabilities. And humans and machines will work together as teams, communicating through the Internet of Things [16].

Using the Internet of Things, machines will coordinate each other and communicate with another machines to create large armies of automated system which is capable of acting together in swarms, as ants do in nature. Machines (devices) will also be able to monitor one another for potential problems and perform adjustment without the intervention of people.

While mobile made people a part of the digital revolution, the Internet of Things enables everything else to be part of that transformation. Because the Internet of Things will affect so many different aspects of the economy, its generated value will exceed that of mobile [16].

Mobile had a direct economic impact of $3 trillion on a global value chain in 2014, created 11 million jobs in the global value chain, and resulted in billions of R&D investments and startups. This transformation was the output of the huge investment from many players within telecommunications and IT, which developed the core infrastructures, mobile networks, technologies, apps, and devices [16].

Mobile started in 1990 with 2G, got more serious in 2000 with 3G, and finally kick-started the smartphone with 4G in 2010. With each technical leap, mobile's social and economic impact increased in areas such as healthcare, finance, and education [16].

The Internet may have connected many people in the world, but mobile gave them true global access. Mobile also allowed individuals and small and midsize enterprises to join a global economy that had long been dominated by big companies [16].

While behind mobile, the Internet of Things is following a similar path. It started as a concept around 2000, is now in its second wave of development, and is experiencing exponential growth.

Not only are IT and telecommunications companies investing in the Internet of Things so are players from outside the classical digital markets [16].

The Internet of Things won't remain behind mobile, however. It will spark a transformation that will exceed mobile's by connecting everything together. We expect nothing less than a reinvention of interactions, communications, and services on a global scale, creating new jobs and opportunities [16].

2.3 Introduction to Localization

Location based service is increasing important technologies which play a core role in modern life. Besides the application of mapping and navigation, location information can be also used for geometric based social network or other types of entertainment. The market size of Location based service is predicted to reach $3.8 billion by 2018 in North America alone. The localization scenarios of Location based service categorized into two main categories: outdoor localization and indoor localization [17].

For outdoor localization, GPS (global positioning system) is the de facto standard. GPS uses 24-satellite system that can provide global coverage and precision to the area of 1-5 meters. The GPS system is easy to scale and the cost of a transceiver GPS chip is low. However, GPS only works well in open area, since GPS signal can be blocked by building, thick forest and other types of physical obstacles. Hence, GPS does not work well in indoor environment [17].

Indoor location based service is not simply an extension of outdoor localization. It can work for specific scenario and improve the life quality as well as boost business market. Indoor location based service can be useful for navigation in large mall or complex yet unfamiliar place for users [17].

Some of the largest mall in United States can be 2,000,000 square feet and more than 300 stores [Wikipedia]. An indoor mapping and navigation system facilitates the localization of user position and help them find the best way to their desired store or nearby facilities like restroom, or diaper changing stations. Indoor localization can also be helpful for visually impaired person or robot navigating system in indoor environment [17].

The location information can tell them where they are and guide them through the building. Another very important scenario for indoor localization is fire or other emergency situation.

Indoor localization enabled devices, system can help the rescuers, or residents locate their positions and find the shortest way to escape from the building, where the thick, dark smoke blocks their field of vision. Unlike outdoor localization, indoor places might be small and easy to deploy extra infrastructure [17].

There are many researches and systems are proposed in the last decade. They are different in many aspects. Generally, they can be divided into several categories: Vision Based Indoor Localization Visual information can be collected and practiced in indoor navigation in many literatures. In the micro-flyer equipped with two camera, take pictures of the special texture on the wall. By analyzing the distortion of the captured texture, the system can infer its relative locations from the walls, which used to keep the micro flyer from crash [17].

The UAV (Unmanned Aerial Vehicle) shots laser beams to the surrounding. By capturing and analyzing the position of the laser points on the walls and ground, the UAV can predict its distance from ground and walls. However, image based localization will consume more computing resource (processing the image) and power. In robot navigation systems, it requires the robot to wonder around the floor for long time to narrow down its potential locations. What's more, the use of camera will increase the cost of the system, which disqualifies the scalability of the system [17].

Wireless Based Indoor Localization Unlike light, wireless wave can get through doors and walls and provide ubiquitous coverage of a building. Wireless based indoor localization uses features of received signal to infer the distance to known points and get the location of current point [17].

Most of current works in indoor localization uses wireless based indoor localization. The deployment of the system is easy and the existence of microwaves does not obstruct human activities in the building. The cost of wireless chips is much cheaper than cameras. The power and computing resource consumption also significantly less than vision based indoor localization [17]. Other Methods Besides vision based and wireless based indoor localization, there also many other ways for indoor localization [17]. By adopting a dead-reckon method, the location of current position, calculated as the previous location adding the movement [17].

Among recent literatures about indoor localization, wireless based indoor localization methods take up most proportion of them. In this paper, we are intended to investigate these literatures and summarize the technologies and methods used in wireless based indoor localization. The rest

of the paper is organized as follow. The second section gives an overview of indoor localization using wireless technologies, it briefs the challenge and evaluation criteria in wireless based indoor localization. The third section discusses the wireless technologies that used for indoor localization [17].

We will compare the advantages and disadvantages of each wireless technology as an indoor localization technology.

2.3.1 Overview of Wireless Based Indoor Localization [17].

The architecture of wireless based indoor localization system usually requires two parts, the beacon stations that emit the wireless signal and the user devices that receives the signal. The computation of localization can be resident in either parts. GPS can also be included as a wireless based localization technology. It uses wireless signals to communicate between satellites and the GPS devices. The 24 satellites are 24 beacon stations. The GPS devices calculate their locations based on received signals from satellites. However, indoor environment is quite different from outdoor environment. The propagation of wireless wave influenced by reflection, scattering, and diffraction [17].

The signal strength can be affected by multi path fading or shadow fading. In indoor environment, the walls, furniture's or walking people will change the propagation of the wireless wave and introduce variance to the wireless signal received by the user. For indoor localization, there are several criteria for evaluating a localization system [17].

1. *Precision*: Precision is similar to accuracy which indicates the correctness of the localization. However, accuracy means the mean of distance error, whereas precision can be seemed as the derivation of the distance error. The precision of different systems or methods can fluctuate from 5 m to 20 cm [17].
2. *Accuracy*: The accuracy is indicating the reliability of the system. It answers us, how likely will it give a right position? [17]
3. *Update interval*: Update interval indicates how frequently the location information updates and reflect the possible power consumption. A simple solution for reducing the power consumption of the localization system is to increase the update interval [17].

4. *Coverage*: means the area covered for accurate localization. Since different wireless technology has different distance, hence short-range wireless technology might need more devices to cover the same area [17].
5. *Computational cost*: is an important criterion for evaluation a localization method. More computational cost means more power consumption and more cost on user device [17].
6. *Infrastructure*: Whether it needs to build extra infrastructure for the deployment of the system or not is import to the budget cost. The reuse and maintenance of the infrastructure is also significant for the deployment. it also includes the user device cost [17].
7. *Offline Computing*: Some of the indoor localization methods require offline computing or site survey, which need labor-intensive work and more deployment time. This requirement increases the cost for deployment as well as maintenance cost. It needs to recalibrate every other interval to keep the accuracy [17].
8. *Localization time*: The time needed for localization for wireless based indoor localization varies for different methods. For methods that support localization for immobile object, this time can be very fast [17].

Wireless based indoor localization system is much cheaper comparing to other types of indoor localization technology. It can provide better coverage on the whole building or room. However, the multipath and shadow fading problems affect the accuracy of localization using wireless signal. While evaluating indoor localization system using wireless, we must consider both the efficiency and cost of the system [17].

2.3.2 Wireless Technologies for Indoor Localization [17]

The frequency less than 300 GHz electromagnetic spectrum called as radio spectrum. We have used the radio for wireless communication for long period of times (Centuries). Different radio frequency has been assigned to different usage. The use of spectrum regulated by Federal Communications Commission (FCC) [17]. There are many protocols and standards for wireless communication designed for different applications. The wireless technology used for indoor localization can be classified by the frequency it uses. Since the frequency of the wireless technology affects its abilities like coverage, wall penetration, and resistance to obstacles. In this paper, we classify them into 3 divisions [17]:

1. Long distance wireless technology
2. Middle distance wireless technology
3. Short distance technology. [17]

2.3.2.1 Short Distance Wireless Technology [17]

Bluetooth is a personal area network standard. It also uses 2.4 GHz and 5 GHz bands as WiFi does. Bluetooth is widely used for short distance communication like earphones, cell phones. Bluetooth concerns the power consumption; it uses a very low transmission power. So the coverage of Bluetooth is shorter than Wi-Fi and other WLAN technology. Hence, Bluetooth is not suit for localization for large area Precision is similar to accuracy, which indicates [17].

UWB (Ultra-Wide Band), unlike other technology, uses a sub-nanosecond radio pulse to transmit data in a wide range of bandwidth (normally greater than 500 MHz), Its transmits Precision is similar to accuracy which indicates ion can be regarded as background noise to other wireless technologies, hence in theory, it can use any spectrum without interfere with other users. It uses small transmission power -41.4dBm/MHz (which is limited by FCC) meaning the power consumption is low. Another advantage of UWB is its immune to multi-path problems [17].

RFID (Radio Frequency Identification) is a simple technology with a history of more than 50 years. It composes of two parts: tag and reader. The reader uses radio-frequency electromagnetic field to read the data in the tag and get the identification of the object the tag attached to. The tag can either have battery or not which makes it an active tag or a passive tag. The passive tag can be very cheap and have a long lifetime which is ideal for cost-sensitive scenario. However, the passive tag RFID suffers from both tag collision and reader collision problems. Tag collision happens when a reader reads more multiple tags, and reader collision happens when the coverage of two readers' overlaps and read the tag at the same time. The communication range of RFID is very short, around (1-2m), this increase the labor works for pre-deployment to cover the huge area [17].

To summarize all those the wireless technology used in literatures, different technology has different transmission range, different needs for dedicated infrastructure, and different power consumption [17].

2.3.2.2 Middle Distance Wireless Technology [17]

Wi-Fi is one of the most used wireless technologies. It follows a series of standards in IEEE 802.11. It uses two license-exempt bands: 2.4 GHz, and 5 GHz. Most buildings like super mall or office building have already deployed Wi-Fi hotspots that provide whole building coverage as network access point. And most commercial products, like phones, laptops and tablets, support Wi-Fi. That means the infrastructure cost and user device cost can be very low. Additionally, Wi-Fi based localization can be easily adopted by buildings and users. With all the advantages, Wi-Fi is a mainstream technology in literatures for indoor localization [17].

ZigBee is a specification based on IEEE 802.15.4 standard. It uses 868 MHz band in Europe, 915 MHz band in the USA and Australia, and 2.4 GHz in other regions. ZigBee is used for long distance transmission between devices in wireless mesh network. It has low cost, low data transfer rate, short latency time, comparing to Wi-Fi standards. In IEEE 802.15.4, standard Link Quality Indication (LQI) defined to indicate the quality of the link and used to derive Received Signal Strength (RSS). And there are integrated chips (CC2430/CC2431) been manufactured to get the RSSI, which makes the implementation of the system easier [17].

2.3.2.3 Long Distance Wireless Technologies [17]

Frequency Modulation used worldwide for regional radio broadcasts. In most regions, it uses 87.5 to 108.0 MHz radio spectrum. Using the VHF (very high frequency) which is less than the Wi-Fi and other modern wireless technology, FM is less distorted by interference, weather or other obstacles like walls (concrete). Since the ubiquity of FM, there is no need to build extra beacon infrastructure using FM for indoor localization [17].

FM receiver is cheap and has lower power consumption hence better battery life. However, the FM station is very far away and FM has large wave length (around 3m), which means that the signal strength of FM signal does not dramatically change in short distance. Hence FM works better for large area. Since different FM, stations use FDMA to share the spectrum, multiple channel signals used to reduce the variance or error introduced by single channel signal [17].

GSM/CDMA has been used in cellular network communication. The GMS/CMDA frequencies in different regions are different. Generally, it falls in 850MHz, 900MHz, 1800MHz, and 1900MHz bands. The GSM/CDMA network is already covered in most buildings; hence there is

no or less need for extra infrastructure. Unlike FM, GSM has a relatively small propagation distance in indoor environment. However, GSM/CDMA heavily patented, so it is hard to do modification or extensions based on GSM/CDMA, which limits the future development on it [17].

2.4 Wi-Fi-Based Indoor Localization [18]

One of the main usage of using Wi-Fi Positioning Systems is to locate the position of almost every Wi-Fi compatible device without installing extra software or manipulating the hardware. Beside this, in WLAN, line of sight is not required. Due to this advantage, Wi-Fi positioning systems have become the most widespread approach for indoor localization [118].

Most positioning systems based on WLAN (Wi-Fi) are available as commercial products as prototypes based on measurements on the received signal strength (RSS). Wi-Fi based positioning systems have several benefits [18].

1. In terms of cost effect, WLAN infrastructures implementation of position algorithms does not need any additional hardware as network interface cards measure signal strength values from all wireless access points in range of the receiver. Therefore, signals needed for positioning can be obtained directly from NICs available on most handheld computing devices. Due to the ubiquity of WLANs, this mode provides a particularly cost-effective solution for offering LBS in commercial and residential indoor environments [18].
2. WLAN positioning systems offer maintainability in two respects: first, no costly requirement of infrastructure and hardware and second the number of mobile devices subscribing to positioning services. Beside this, there are also certain WLAN limitations: signal attenuation of the static environment like wall, movement of furniture and doors [18].

With the advancement of technology in this 21st century, everyone is using smart phones and other smart electronic objects. With this, the use of Internet of Things is in the trend. So basically the Internet of things is the interconnection between these smart objects on Internet which helps to send and receive data. Here are some more definitions: Internet of things can be described as the interlinking of everyday objects on a network in which intelligence is used for global

connection. It is Internet of Things, which makes it possible to access remote areas only because of connection, with physical things on internet [18].

It is the internet of things which can connect anything with anything else in the world and is important to business and other enterprises. It facilitates exchange of data on global level with security of data. It acts as a bridge between real life time objects and their external environment or we can say physical world on the internet. The Internet of things will act as a medium to increase the efficiency of networks in the global world. It will also be the medium to keep everything open as every single real life object will be connected to the network. The purpose of IoT is to make every single daily life object network enabled and also describe a technique how internet can be applied to everyday objects [18].

The Internet of Things (IoT), also known as the Internet of objects or Cyber Physical Systems (CPS), can be described as relation between daily life objects on a network. It is a platform which connects various technologies with wired or wireless technologies. Thus, IoT helps to create an environment to share different things on the network in efficient manner. IoT allows access of objects which are at far off places and can also be controlled from other places and these two places are connected on the network and in this way connection between the physical and virtual world is increasing. This has advantage that it increases efficiency and reduces human intervention. This technology of IoT which contains sensors and actuators creates more cyber physical systems. This is the reason behind the concept of smart cities and smart homes. Thus IoT is the infrastructure of the information society.

2.4.1 Active Wi-Fi Tracking[19]

Active (non-passive) Wi-Fi tracking relies on active device participation by installing additional software on device for a specific AP. In 2005, proposed a device localization system called *Place Lab* using different kinds of radio beacons, including beacons from Wi-Fi APs, to overcome limitations of existing systems [19].

In particular, the iniquitousness of Wi-Fi systems allows for long coverage and easy deployment while simultaneously achieving fairly accurate localization results around 30 to 40 meters in urban areas. Similarly, also relying on a custom application installed on the device being localized, reported to have achieved very accurate (around 3 meters) real-time localization

results in their indoor experiment by using a path loss based estimation model. Instead of relying on client-side computation, analyzed Wi-Fi traces from APs to track any object equipped with a Wi-Fi tag [19].

They reported results of similar accuracy (*around* 4 meters indoor) compared to other, also analyze Wi-Fi traces from APs but their goal was to construct a mobility model focusing on movements of devices among popular regions [19].

2.4.2 Passive Wi-Fi Tracking [19]

In passive Wi-Fi tracking, a capturing device senses and tracks any Wi-Fi traffic within its range. They presented a system comprising of common, off-the-shelf Wi-Fi AP hardware that captures probe requests and implements several techniques to prompt devices for additional transmissions in order to obtain more valuable data. The collected data is then used to estimate the trajectory (*i.e.*, spatio-temporal path) of monitored devices. The authors propose a solution based on the Viterbi algorithm and Hidden Markov Model to overcome limitations of simple interpolation based approaches [19].

Although, they have employed passive Wi-Fi traffic capturing before for tracing movements of mobile users, their scenario was limited to periodic MAC address scans of APs for the purpose of device localization and thus did not include tracking unmolded devices. In a similar way, present a crowd sensing approach by leveraging commodity smartphones and exploiting the natural mobility of people to gather information (*e.g.*, bandwidth distribution) about the existing AP infrastructure in a specific area [19].

In the following, passive Wi-Fi tracking approaches are discussed that aim towards tracking unmodified mobile devices. Perform real-time pedestrian flow analyses in indoor and outdoor environments by collecting and investigating probe requests. Focus on classifying human presence into different activity patterns (*e.g.*, engaged or outside). Show that probe request traces can reveal insightful information about the social structure and socioeconomic status of device owners [19].

On large-scale datasets, graph-based models were used to derive relationship graphs and were combined with further features such as the owner language guessed from known service set

identifiers (SSIDs) or the device vendor inferred from commonly known MAC address prefixes [19].

2.4.3 Localization Meaning & Definition

Localization, as the word suggests its own meaning, is the process of making something local in an area. It restricts an object to a particular area. It makes a product adaptable to a specific locale or market. The purpose of localization is to create a product or an object for a particular market called target, without taking into linguistic, cultural or religion differences. Just as the software industry looks forward to create and develop the next big technological capability, so must the localization and language industry. If the inanimate objects that we use in our everyday lives are one day going to be smart, what impact will this have on the localization and translation industry? The cloud, social and mobile technologies are key growth areas and they are all areas that further enable global markets. Future software and technologies must increasingly speak to global markets in a local language to enhance user experience. Localization or tracking is considered a major problem in today's technology and has been under study in various fields including Global positioning system and other wireless networks.

Elements of Localization Process Translation: - which means converting data from one form to another is a part of localization process. Other elements area as follows:

- Making your product available to target market
- Making your product accessible to other markets by modifying content
- Modify design and layout according to translated text
- Making it more efficient for local requirements such as currency
- Displaying dates, addresses and phone numbers in proper local formats
- Taking into account rules and regulations of local areas One of essential components in the IoT is wireless sensor network, in which environmental data (e.g. temperature, humidity, and object movements) is collected and processed using hundreds of sensor nodes. To respond and react to this environmental data, location information which is collected by sensor nodes should be made available at the base station (datacentre, sensor fusion, access point). There are various actions in IoT like fire alarm, energy transfer and

emergency request are established on the data center, a way to identify the location information of all the nodes at the data center is of importance. In this approach, i.e. localization at datacenter, location information is sent to the datacenter after it has been collected by data sensor node. Using the obtained distance information, data center constructs a map of sensor nodes. For performing localization process at the data center, pair wise distance information between each sensor pair should be provided. It has been shown that if exact location of sensor nodes (also called anchor nodes) is provided, information regarding location can be found accurately.

- Major problem with the localization process is that may be the data center not have the enough information of the sensor nodes
- Moreover, it is difficult to recover the original Euclidean distance matrixD from a subset of its entries because for the unknown entries there are many completion options.

Performance Metric:- The performance in the area of localization is evaluated on the basis of following factors:

1. *Accuracy*: Accuracy or location error is one of the significant factors in positioning of localization system. It is defined as an error in distance between the calculated distance and exact device location.
2. *Responsiveness*: In general, responsiveness refers how quickly something reacts to a situation. In localization, responsiveness can be defined a show fast the device location can be updated.
3. *Coverage*: It is important in positioning system to determine the problem of network coverage area in a designated area. It is closely related to accuracy.
4. *Adaptiveness*: Adaptiveness as the name suggests is the ability to cope up with the environmental changes. Localization accuracy and performance can be affected with the environment. Thus an adaptive system is one which increases efficiency by making it adaptable to the environment. Need for repeated adjustments can be removed from adaptiveness.
5. *Scalability*: When system has to be operated in larger areas, its performance is measured in terms of Scalability. Lesser the scalability, lesser will be the performance.

6. *Cost and complexity*: The parameters that contribute to the cost of localization system are additional bandwidth, energy weight, lifetime, money etc. In addition to this it may include the charges of the installation and survey time during the period technology has been deployed.

While in principle two alternative methods to associate with an AP can be recognized, in advance AP-initiated Wi-Fi beacons and client device-initiated probe requests, Probe exploits the latter. In the first method, APs periodically announce (*e.g.*, every 100 ms) their presence by broadcasting beacon management frames, which contain network-information such as the supported data rates and the SSID (Service Set Identifier). To detect APs, client devices listen for beacons and reply with Wi-Fi association frames to initiate a connection. Within the second method, client devices actively discover APs by broadcasting WiFi probe requests on potentially multiple channels.

Probe requests contain information about the client (**Media Access Control (MAC) address**) and the preferred AP (**SSID**) with which the client device wishes to associate. Although this thesis focuses on probe requests, Probe is also able to support capturing any publicly receivable WiFi activities. Instead of dealing with highly sensitive encrypted Wi-Fi packets, the publicly broadcasted probe requests turned out to be sufficient to address the questions posed in the conducted case study. This case study demonstrates the capabilities of Probe and reveals interesting patterns with the following use cases:

U1 paired person tracking: for how long two people (pair) stay in a given time?

U2 Person tracking: to reproduce the daily, weekly …. routine of a person.

Traces of Wi-Fi activities have been captured and analyzed in research for many years. Many rely on active participation of the device being tracked or on traces being taken from APs. [23] present a first approach employing passive Wi-Fi tracking, thus allowing to capture Wi-Fi packets. This idea was pursued by further research and also lead to the development of tooling for Wi-Fi tracking studies. Collecting and analyzing Wi-Fi traffic raises questions regarding privacy of potential sensitive data.

2.5 Functionality of indoor localization

2.5.1 Advantages for merchants and restaurants

Merchants can send their potential customers tailored offers – for example picking those who have already been to similar shops or who are returning visitors. And for sure everyone is delighted by a discount of his favorite shop pushed directly on his smartphone.

2.5.2 Advantages for exhibitors

Of course exhibitors have the possibility to present themselves on the map of the area in the trade fair app, including pictures, contact data and description. Furthermore, they can use location based marketing in order to get the attention of the matching visitors. The analysis of visitor flows makes it possible to choose the best stand position.

2.5.3 Advantages for trade fair visitors

Indoor positioning in exhibition halls helps visitors to find the way to certain stands. The app can be personalized, which means that exactly those stands can be highlighted on a map which is interesting for a visitor. Intermodal transport is also possible let the app show you the way from your home to the stand you want to see you can even see arrival and departure times of environmentally friendly public transportation.

Chapter Three

3. Used Technologies and Implementation

3.1 What is Esp8266

The ESP866 is a tiny Wi-Fi chip that has the huge advantage regarding the ridiculous price point nearly from $5 - $6, And the best is that this chip also has a small processor onboard, so it can actually function in complete autonomy, without an additional Arduino board for example. Compare that to the cost of a traditional Arduino + Wi-Fi module solution (around $40), and you will immediately see why this chip is receiving so much interest these days [2].

Moreover, the main factor that makes this chip useable for this thesis is its Wi-Fi Module, which is a self-contained SOC with integrated TCP/IP protocol stack that can give any microcontroller access to your Wi-Fi network. The ESP8266 is capable of either hosting an application or offloading all Wi-Fi networking functions from another application processor.

The Esp8266 HDK includes the following the chip—ESP8266EX, the module—ESP-WROOM-02 and the development board—ESP-LAUNCHER.

ESP8266 was designed by the Chinese company Espressif, aiming for uses in Internet of Things (IoT) systems. ESP8266 is a complete Wi-Fi system on chip that incorporates a 32-bit processor, some RAM and depending on the vendor between 512KB and 4MB of flash memory. This allows the chip to either function as a wireless adapter that can extend other systems with Wi-Fi functionality, or as a standalone unit that can by itself execute simple applications.

Depending on the specific module variant (ESP-1 to ESP-12 at the time of this thesis) between 0 and 7 General Purpose Input/Output (GPIO) pins are available, in addition to Rx and Tx pins of the UART, making the module very suitable for IoT applications.

The Software Development Kit (SDK) provided by Espressif contains a lightweight implementation of a TCP/IP control stack for Wi-Fi communication. The modules house libraries for optional services such as Dynamic Host Configuration Protocol (DHCP), Domain Name System (DNS), JavaScript Object Notation (JSON) and Secure Socket Layer (SSL) libraries for Application Level programming. It incorporates 802.11 MAC extensions such as 802.11b/g/n/d/e/h/i/k/r that manage signal transmission, encapsulation, encryption, collision management and roaming functionality [3].

3.1.1 Technical Overview

The ESP8266 contains complete Wi-Fi networking solution, meaning it can be used to either host or offload Wi-Fi networking functions from any other processor. In the case of hosting an application, it boots up directly from an external flash, while the integrated cache improves the performance of the system. Alternatively, we can use it as a Wi-Fi adapter. It is one of the most integrated Wi-Fi chips' in the industry. Besides its' integrated antenna switch, RF balun, power amplifier and the likes, it also integrates an enhanced version of Tensilicas' L106 Diamond series 32-bit processor, with on-chip SRAM. It also often used with other applications through external sensors through its GPIO pins. The chip has been designed for mobile use, while aiming for the lowest power consumption possible.

Esp8266 power saving architecture operates in three modes: active mode, deep sleep mode and sleep mode. It consumes about 60 nanoA in deep sleep mode, and less than 1.0 mA or less than 0.5mA to stay connected to the access point. It can be programmed to wake up at any required interval, or when a specific condition is met. This feature allows to remain in low-power standby mode until Wi-Fi is needed.

3.1.2 Communication

The main benefits of the ESP8266 chip, being the smallest module it has the most limited I/O pins. At first, all of the pins used for programming.

The pins numbered from 1 - 8 and has the following name: - TX, CHPD, RST, VCC, GND, GIO2, GPIO0 and RX. Of these, VCC, GND, RST and CHPD are not I/O pins, but are necessary for the operation of the module. This leaves us with GPIO0, GPIO2, TX and RX as I/O pins, However even these have predefined functions. The GPIO pins determining what mode the module starts up and the TX/RX used to program the module, for Serial I/O and most notably used for debugging and testing the output. As mentioned above another way of interacting with the chip is directly through the Serial port.

The Arduino IDE supports this feature and allows setting the baud rate among other things. The form of communication is done through the AT instruction set. Often the chip has predefined commands, however user-defined commands supported though. Commands are sent by typing the AT prefix plus the command to be executed, Most often used commands are AT and AT+ RST to see if the board is working or to restart it.

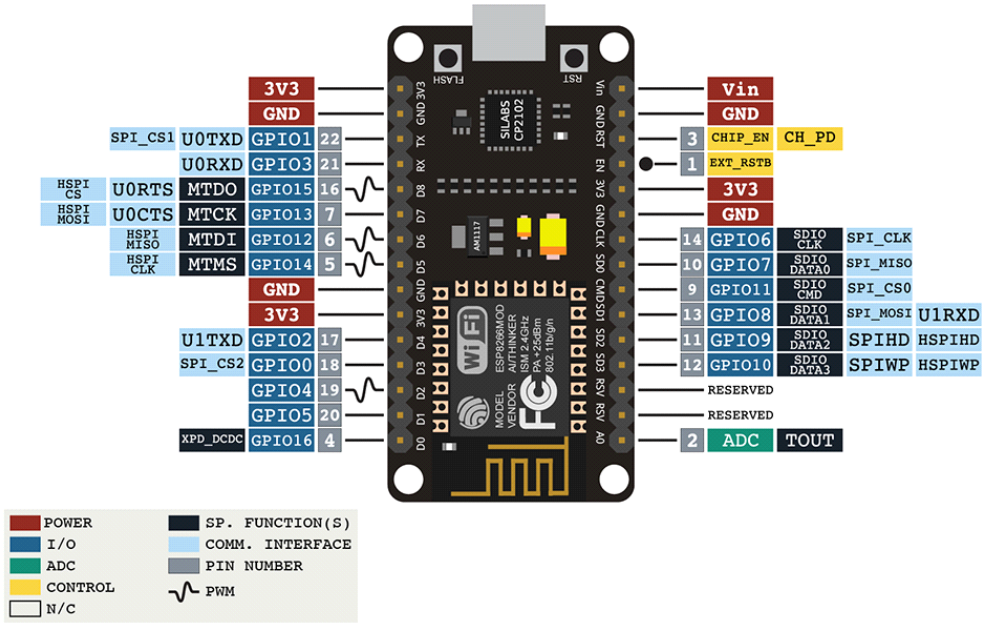


Figure 1:ESP8266 Hardware Descriptions

3.2 Why Esp8266

The ESP8266 module is an extremely cost effective module with a huge ever growing, community, with an easy integrated capability into space-constrained devices and Due to its small size, which is approximately 18mm x 20mm (ESP-WROOM-02) / 16 mm x 23 mm (ESP-WROOM-S2).

Module can operate in a low-power connectivity mode. For instance, if it operates in DTIM10, it consumes only 1.2mW while maintaining a Wi-Fi connection. The module also integrates an SPI flash of 16 Mbits, used for storing user programs, data and firmware

3.2.1 How it works

Before diving in in detail how it works and what activity with what consideration is performed during this thesis work, I will divide the main implementation has in to three,

- Collecting End user or Mobile users packet (sniffing mode)
- Two side Esp8266 serial communication (client server mode)
- Storing (database) and processing Received packet and learning, tracking user's movement.

The categorization allows us to illustrate and briefly reflect the functionality of the thesis with respect to the features offered by the Esp8266 module and the main objective of this project,

3.2.2 Collecting End user or Mobile users packet (sniffing mode)

Collecting the mac address of every mobile users handled in different mode, using the esp8266 module. but the most interesting and fascinating feature of this device and the thesis aim focus on capturing the user (Mobile user) identity which is localized using the mac address of the device in time of connecting through the wireless network in Promiscuous mode (sniffing mode).

Generally, any IEEE 802.11 packet, which sniffed using the Esp8266 module, have specific and generic type of structure and not all packet type sniffed too. Here is the list of IEEE 802.11 packet, which can be retrieved using Esp8266 board.

The following HT20 packets are support:

- 802.11b

- 802.11g
- 802.11n (from MCS0 to MCS7)
- AMPDU types of packets

The following not supported:

- HT40
- LDPC

Although ESP8266 cannot completely decipher these kinds of IEEE80211 packets completely, it can still obtain the length of these special packets. In summary, while in sniffer mode, ESP8266 can either capture completely the packets or obtain the length of the packet:

- Packets that ESP8266 can decipher completely; ESP8266 returns with the MAC address of the both side of communication and encryption type and the length of entire packet.
- Packets that ESP8266 can only partial decipher; ESP8266 returns with the length of packet structure.

RxControl and sniffer_buf used to represent these two kinds of packets. Structure Sniffer_buf contains structure RxControl.

```
struct RxControl {
    signed rssi:8; // signal intensity of packet
    unsigned rate:4;
    unsigned is_group:1;
    unsigned :1;
    unsigned sig_mode:2; // 0:is not 11n packet; non-0:is 11n packet;
    unsigned legacy_length:12; // if not 11n packet, shows length of packet.
    unsigned damatch0:1;
    unsigned damatch1:1;
    unsigned bssidmatch0:1;
    unsigned bssidmatch1:1;
    unsigned MCS:7; // if is 11n packet, shows themodulation // and code used (range from 0 to 76)
```

unsigned CWB:1; // if is 11n packet, shows if is HT40 packet or not

unsigned HT_length:16;// if is 11n packet, shows length of packet.

unsigned Smoothing:1;

unsigned Not_Sounding:1;

unsigned:1;

unsigned Aggregation:1;

unsigned STBC:2;

unsigned FEC_CODING:1; // if is 11n packet, shows if is LDPC packet or not.

unsigned SGI:1;

unsigned rxend_state:8;

unsigned ampdu_cnt:8;

unsigned channel:4; //which channel this packet in.

unsigned:12;

};

struct LenSeq {

u16 len; // length of packet

u16 seq; // serial number of packet, the high 12bits are serial

number, // low 14 bits are Fragment number (usually be 0)

u8 addr3[6]; // the third address in packet

};

struct sniffer_buf{

struct RxControl rx_ctrl;

u8 buf[36]; // head of ieee80211 packet

};

struct sniffer_buf2{

struct RxControl rx_ctrl;

u8 buf[112]; //may be 240, please refer to the real source code

u16 cnt;

u16 len; //length of packet

}; [4]

Note:

For the case of LEN == sizeof(struct RxControl), the methods to calculate the length of packet are as below:

• If sig_mode == 0, the length of packet is the legacy_length.

• Otherwise, the length of packet is in struct sniffer_buf and sniffer_buf2, and it is more reliable.

The callback function wifi_promiscuous_rx contains two parameters (buf and len).len shows the length of buf ,it can be: len = 128, len = X * 10, len = 12. [4]

LEN == 128 [4]

•buf contains structure sniffer_buf2: it is the management packet; it has 112 bytes of data.

•sniffer_buf2.cnt is 1.

•sniffer_buf2.len is the length of the management packet.

LEN == 12 [4]

•buf contains structure RxControl; but this structure is not reliable. It cannot show the MAC addresses of both communication sides, or the length of the packet header.

•It does not show the number or the length of the sub-packets of AMPDU packets.

•This structure contains length of the packet, RSSI and FEC_CODING

•RSSI and FEC_CODING used to judge whether the packets are from the same device.

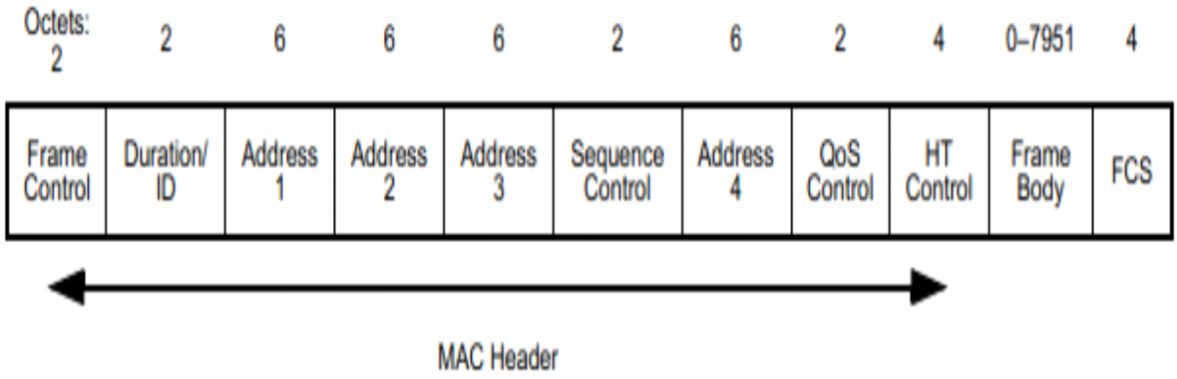


Figure 2:IEEE80211 Packet Format. [4]

The first 24 Bytes of MAC Header of data packet needed [4]:

- Address 4 field depends on From DS and To DS which is in Frame Control;
- QoS Control field depends on Subtype which is in Frame Control;
- HT Control field depends on Order Field which is in Frame Control;
- More details are found in IEEE Std 80211-2012.
- For WEP packets, MAC Header is followed by 4 Bytes IV and before FCS there are 4bytes ICV.
- For TKIP packet, MAC Header is followed by 4 Bytes IV and 4 bytes EIV, and before FCS there are 8 bytes MIC and 4 bytes ICV.
- For CCMP packet, MAC Header is followed by 8 Bytes CCMP header, and before FCS there are 8 bytes MIC

3.2.3 Scanning through all available channels or selected channels

IEEE 802.11 specifies 14 channels for low power Wi-Fi communication in the ISM (unlicensed) band, these channels are spaced 5MHz apart, except ch.14, which is spaced 12 MHz away from ch.13, In 802.11 b/g/n modes, the typical bandwidth requirement per channel is 20 MHz and a guard band of 2 MHz is added to that. Thus, there is a potential overlap in the operating frequency range when two transmitters operate in the same airspace.

The Demonstration, which performed for this thesis, sniffing the mobile users using the promiscuous mode went through two approach, depending on the available Wireless standard channel and both have its own pros and cons, let us have a quick reference on both methodology and which approach selected and for what reason.

3.2.4 Two side Esp8266 serial communication (client server mode)

After the completion of a successful testing using a standalone mode, which works in a selected area with a limited range coverage, the next aim of the thesis focused to distribute the given sensor (both client and servers Esp8266) in every needed area and standing a centralized server, which receive from every area of dedicated Servers.

In this section, we focus on how this client server mode implemented, in what fashion, means of connecting and the required hardware component for every activity. The client server paradigm help us to escalate a one origin (area) historical data analysis and move mental and behavioral study of a given mobile user. By putting the client (packet sniffer) module and server who receive the given packet from the device from the client and send to a given centralized server who study and analyze behavioral movement of the user singularly and with respect to another mobile users, in this thesis the scope is limited to mutual or pairs movement.

To implement this client server communication, we escalate the usage of a single Esp8266 module to two one which act as a client and the other same type of standalone sensor (same module but different functionality) bounded all together and putted together after we connected them with two pin RX-TX and TX-Rx fashion, as shown in the figure below.

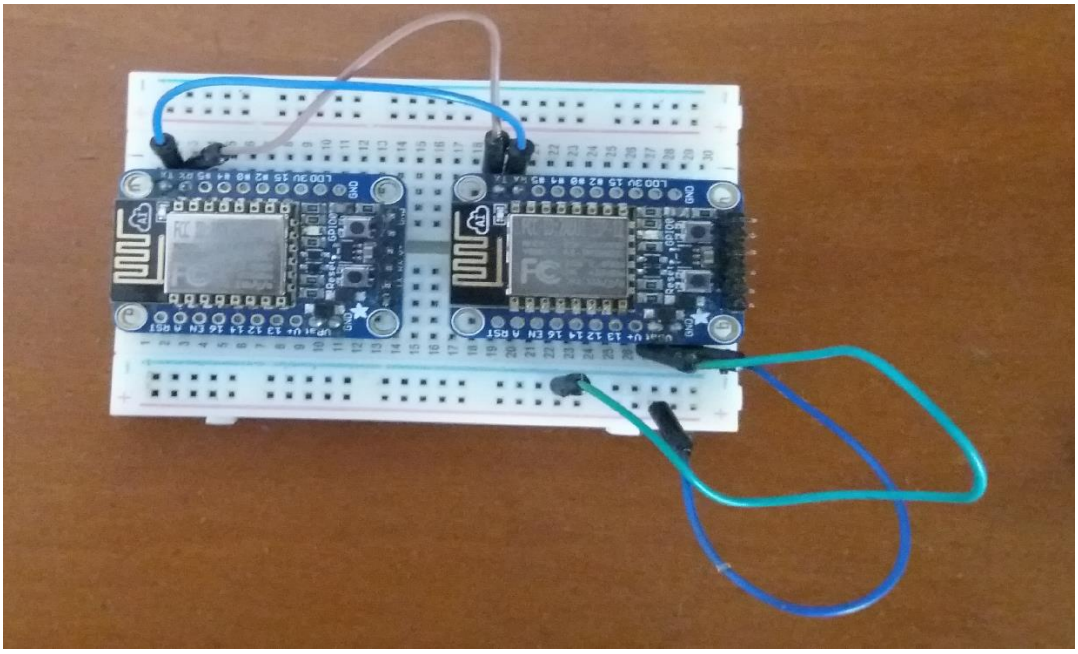


Figure 3: Esp8266 Serial communication design

ESP8266EX has two digital pins for power supply, Pin11 and Pin17. For digital power supply, there is no need to add additional filter capacitors. The operating voltage range of digital power supply pins are 1.8V ~ 3.3V.

3.4 Wi-Fi Probe Request

The Wi-Fi probe request service is a critical element of the IEEE 802.11 standard family. It allows a mobile device with wireless internet access to detect the known access point within its scanning region at regular intervals, only if the mobile users turn on the Wi-Fi function, It's called active discovery. Moreover, the modern smartphone also broadcasts probe requests regularly even when it has already connected to an AP. A broadcast probe request packet contains information of MAC address, operation channel, Encryption Mode (Encrypt Mode) and plaintext SSID of AP which has been connected previously (the other kind of probe request packet does not contain previous connected SSID information). Through capturing the probe requests with users' private information (the information in probe request is not encrypted), hacker realizes the network attacks. The TX/RX Pins are going to regular digital pins so the two ESP8266 via Software Serial libraries. Here is the basic needed serial communication sketch (program) between the two Esp8266 modules:

```
#include <SoftwareSerial.h>
SoftwareSerial ESPserial(2, 3); // RX | TX

void setup()
{
Serial.begin(115200); // communication with the host computer

// Start the software serial for communication with the ESP8266
ESPserial.begin(115200);

Serial.println("");
Serial.println("Remember to to set Both NL & CR in the serial monitor.");
Serial.println("Ready");
Serial.println("");
}

void loop()
{
// listen for communication from the ESP8266 and then write it to the serial monitor
if ( ESPserial.available() ) { Serial.write( ESPserial.read() ); }

// listen for user input and send it to the ESP8266
if ( Serial.available() ) { ESPserial.write( Serial.read() ); }
}
```

3.5 Client operation

The client Esp8266 chip intended to listen instantly in a real-time packet sniffed from the mobile user in promiscuous mode. unwrapping the Mac address (logical identification of a specific user) from the frame and send to the server using the two pin connection RX and TX, it uses the TX for sending the 48 bit Mac Address, TX :- transmitter RX :- Receiver.

Saying that the program which runs in the client Esp8266 needed to know in advance the SSID and Password of the access point which is the point of reference that the users is able to connect. Setting that feature allows the client to hear instantly any message between the user's mobile device and the access point, which provide the internet access.

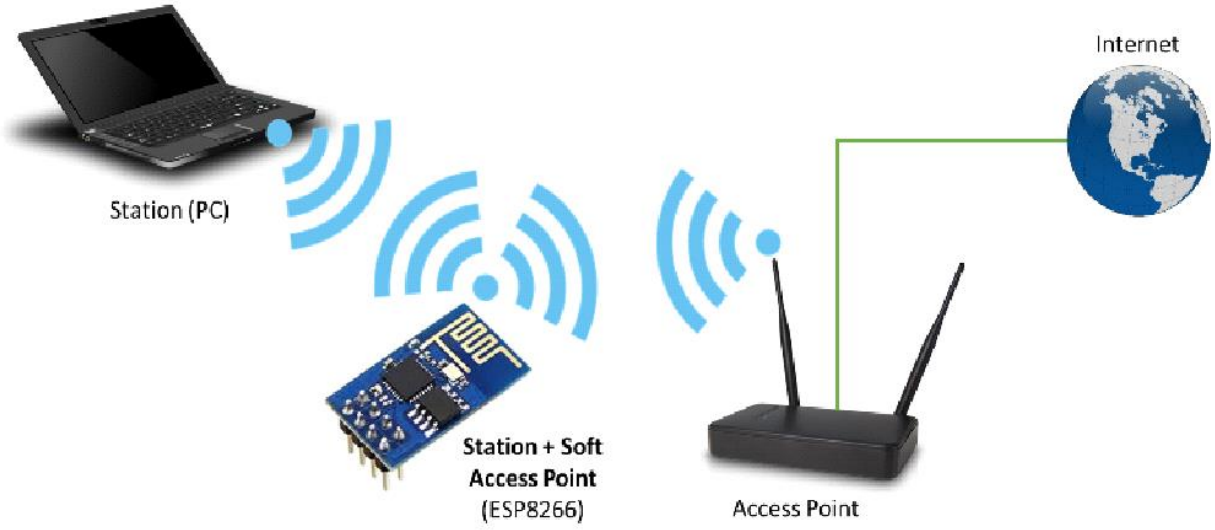


Figure 4: Esp8266 Client architecture setup. [5]

Note: - The callback function, which help the client to send the tracked user mac address comparing to the server Esp8266 socket speed connection to the Centralized server is much faster since its received more than 100 packet in a second to compensate this unlatching we put a delay of 0.2 second. I will reveal the problematic issue created during socket creation in the server operation portion of this thesis.

3.5.1 Generic use case Scenario, with basic sketch (program)

1. User make authentication and connected to the nearest access point.
2. In the time of authentication or data communication between the mobile user and the access point, the data sniffed by Esp8266 Client module.
3. The sniffed data packet, de encapsulated and the program runs under the client Esp8266 capture the mac address of the given user(identification)
4. Using the Tx Rx serial communication in between the server and Client Esp8266 module the client sends the **mac address** of a user, **id** of itself and the **timestamp** when the Esp8266 module captures the data.
5. This 4 steps of routine repeatedly executed by the client Esp8266 module for every mobile, smartphone, laptop and Desktop users.

3.6 Server operation

The server Esp8266 module function with the following main task:

- Make serial connection using TX, RX pin.
- Open a port and make http connection to the listening centralized server
- Send data with the predefined delay threshold

To understand better, we will illustrate the following sample code of basic http client, it works with standard http client mode but for this Thesis, it functions as a server to receive the data from the Esp8266 client module and send the aggregated and filtered data to the centralized server.

```
#include <ESP8266WiFi.h>

const char* ssid     = "your-ssid";

const char* password = "your-password";

const char* host = "data.sparkfun.com";

const char* streamId   = ".....";

const char* privateKey = ".....";


void setup() {
```

```
 Serial.begin(115200);
 delay(10);

 // We start by connecting to a Wi-Fi network

 Serial.println();
 Serial.println();
 Serial.print("Connecting to ");
 Serial.println(ssid);

 /* Explicitly set the ESP8266 to be a Wi-Fi-client, otherwise, it by default,
    would try to act as both a client and an access-point and could cause
    network-issues with your other Wi-Fi-devices on your Wi-Fi-network. */

 WiFi.mode(WIFI_STA);
 WiFi.begin(ssid, password);
  while (WiFi.status() != WL_CONNECTED) {
  delay(500);
  Serial.print(".");
 }
 Serial.println("");
 Serial.println("WiFi connected");
 Serial.println("IP address: ");
 Serial.println(WiFi.localIP());
}
```

```
int value = 0;
void loop() {
  delay(5000);
  ++value;
  Serial.print("connecting to ");
  Serial.println(host);
  // Use WiFiClient class to create TCP connections
  WiFiClient client;
  const int httpPort = 80;
  if (!client.connect(host, httpPort)) {
    Serial.println("connection failed");
    return;
  }
  // We now create a URI for the request
  String url = "/input/";
  url += streamId;
  url += "?private_key=";
  url += privateKey;
  url += "&value=";
  url += value;
  Serial.print("Requesting URL: ");
  Serial.println(url);
  // This will send the request to the server
  client.print(String("GET ") + url + " HTTP/1.1\r\n" +
               "Host: " + host + "\r\n" +
```

```
            "Connection: close\r\n\r\n");
unsigned long timeout = millis();
  while (client.available() == 0) {
    if (millis() - timeout > 5000) {
      Serial.println(">>> Client Timeout !");
      client.stop();
      return;
    }
  }

  while(client.available()){
    // Read all the lines of the reply from server and print them to Serial
    String line = client.readStringUntil('\r');
    Serial.print(line);
  }
  Serial.println();
  Serial.println("closing connection");
}
```

We must first establish a connection with the Wi-Fi network. The WiFi.begin class does this.

WiFi.begin(ssid, password);

The program continued only if a connection established

```
while (WiFi.status() != WL_CONNECTED) {
  delay(500);
  Serial.print(".");
```

```
}
```

Now that we have a Wi-Fi connection, we will send a message to the server at regular intervals. To do this, you must create a client object that will connect to the server.

WifiClient client;

```
if (!client.connect(host, port)) {
  Serial.println("connection failed");
  return;
}
```

A string is then constructed in which the data sent to the server is encoded. Here the time since the start of the ESP8266 as well as the IP address of the latter

```
String url = "/watchdog?command=watchdog&uptime=";
   url += String(millis());
   url += "&ip=";
   url += WiFi.localIP().toString();
```

All you have to do is send the message to the server

```
client.print(String("GET ") + url + " HTTP/1.1\r\n" +
             "Host: " + host + "\r\n" +
             "Connection: close\r\n\r\n");
```

Figure 5: ESP8266 Server Architecture setup

3.6.1 Generic Use Case Scenario

1. Before the arrival of the first data from the client Esp8266, the server Esp8266 make the following network setup, set up the username and password with the necessary wireless security, parameter to the pre known local area network that make hi reachable to the centralized server.
2. Checking the connectivity to the access point (internet access provider) if not repeat step 1 else jump to the next step.
3. Every incoming serial communication handled through the Tx, Rx pins communication between the client and server Esp8266 received by this module.
4. The data extracted and validity check occurred on the Mac Address, timestamp and Id of the client Esp8266 module.
5. Once the correctness of the data confirmed, a Tcp connection opened to the centralized server if and only if there is no opened session from the previous connection and finally send the data of a given user from one of the client area to the centralized server.
6. This all 5 steps are executed for each data coming from the allocated client Esp8266 module, who gathers the data from the mobile users who are active instantaneously.

3.7 Consideration taken to successfully communicate the client and server

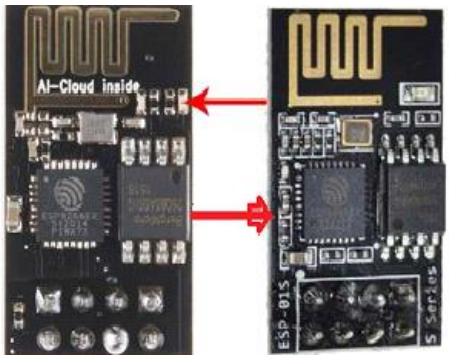


Figure 6: ESP8266 high level Client server communication

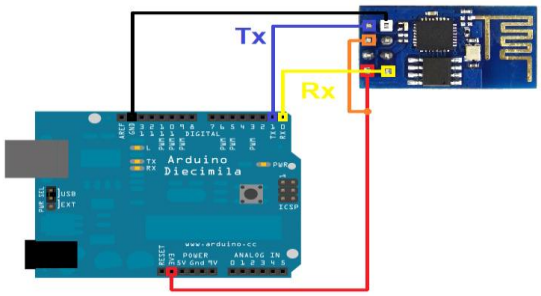


Figure 7: ESP8266 RX-TX Client server connection setup

3.7.1 Matching the baud rate

The baud rate specifies **how fast** data sent over a serial line. Usually expressed in units of bits-per-second (bps). If you invert the baud rate, you can find out just how long it takes to transmit a single bit. This value determines how long the transmitter holds a serial line high/low or at what period the receiving device samples its line.

Baud rates can be just about any value within reason. The only requirement is that both devices operate at the same rate. One of the more common baud rates, especially for simple stuff where speed is not critical, is **9600 bps**. Other "standard" baud are 1200, 2400, 4800, 19200, 38400, 57600, and 115200.

The higher a baud rate goes, the faster data is sent/received, but there are limits to how fast data can transferred. You usually do not see speeds exceeding 115200, which is fast for most microcontrollers. Get too high, and you'll begin to see errors on the receiving end, as clocks and sampling periods just can't keep up, so we set up both the client and server baud rate communication the same.

3.7.2 Making sure Rx, Tx connectivity and Serial Communication

For Level conversion from 3.3V to 5V we need only two components.

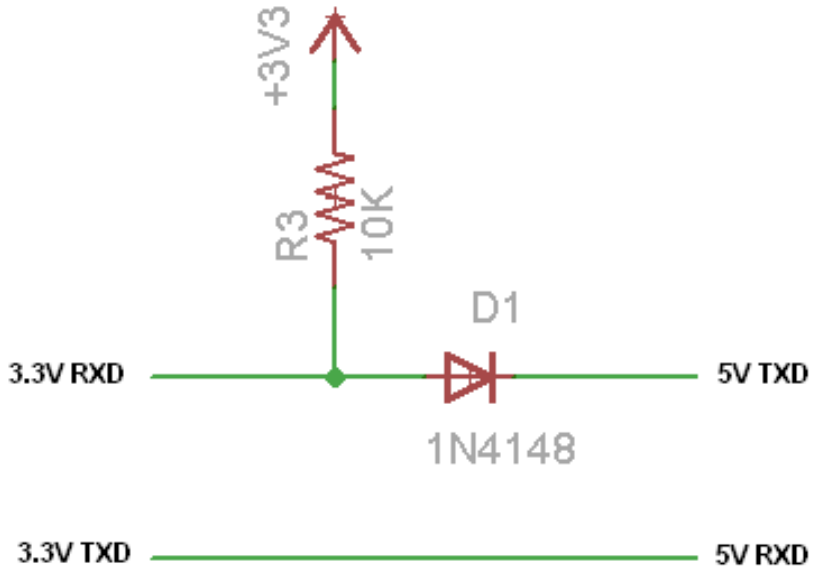


Figure 8: Logical Voltage Level Conversion

Assuming that you have already connected serial with your USB to Serial converter or using ESP Witty, Node MCU.

Serial communication on pins TX/RX uses TTL logic levels 3.3V. Do not connect these pins directly to an RS232 serial port; they operate at +/- 12V and can damage your ESP8266 board.

Serial used for communication between the Esp8266 board and a computer or other devices. All ESP boards have at least one serial port (also known as a UART or USART): Serial. It communicates on RX and TX.

3.7.3 Providing credential to the server Esp8266

In order to the server Esp8266 module to communicate to the centralized server, which receive from 1 or more Esp8266 server module each one of them needed to be connected to local network infrastructure to be reachable, means of network connectivity is handled through wireless network, SSID and Password of the wireless network is a key.

3.8 Means of communication

Once the sniffed user data reaches on the Esp8266 server from different area, it supposed to create a socket (statefull connection) to dedicated centralized server the following parameters

- Server Identification (the Id of the Esp8266 server)
- Timestamp
- Mac address

3.8.1 Server Identification

This parameter is the logical agreement stetted by the system to represent the local area Server, Means of knowhow on which location the user is moving around and the sniffed mac address sent together with the Id of theEsp8266 server.

This let us to track the exact movement of the user from time to time, knowing the timestamp also help us to narrate the user movement.

Timestamp

This parameter represents the real time presence of the user on specific place, having this and the Esp8266 Id information combining assure us the current exact user place and time his usage

Mac address

This parameter label to us the logical representation of a given user. To summarize, the above three parameter allow us the single data coming from the Esp8266 server answers the three W,

- Who (Mac Address),
- Where (Server Identification)
- When (Timestamp)

After successful socket on a given address and port, the communication start on in continues fashion from different sited area the necessary data given to the centralized dedicated server, which perform a routine, we will get back the operations carried by this Server later on.

3.9 Centralized Server data storage in Database and operations performed

This portion mainly focus on the operations, overall activities tasked on the Server, let us define the activities undergone and I will pass on them briefly in each lifecycle of those activities operation.

- Open a connection
- Assign a thread for every incoming data from the Esp8266 server
- Evaluating the data and insertion to database
- Job scheduler on analyzing historical data
- categorization algorithm

Apparently, except from the last activity (activity 5) for every incoming data from the esp8266 module (server) they passed through the first 4 activities.

Open a connection

A socket is a software endpoint that establishes bidirectional communication between a server program and one or more client programs. in this thesis the communication is one directional, from the client to the server, since the aim of the client is not to feed a data by the server but on committing the real time data gathered from the Esp8266 module (client).

The socket associates the server program with a specific hardware port on the machine where it runs so any client program anywhere in the network with a socket associated with that same port can communicate with the server program.

Now, you might ask what protocol you should use UDP or TCP? This depends on the client/server application you are writing, in short, TCP is useful for implementing network services such as remote login (rlogin, telnet) and file transfer (FTP), which require data of indefinite length to transfer. UDP is less complex and incurs fewer overheads. It often used in

implementing client/server applications in distributed systems built over local area networks, so since our need merely related to the benefit of the TCP I implemented on TCP protocol.

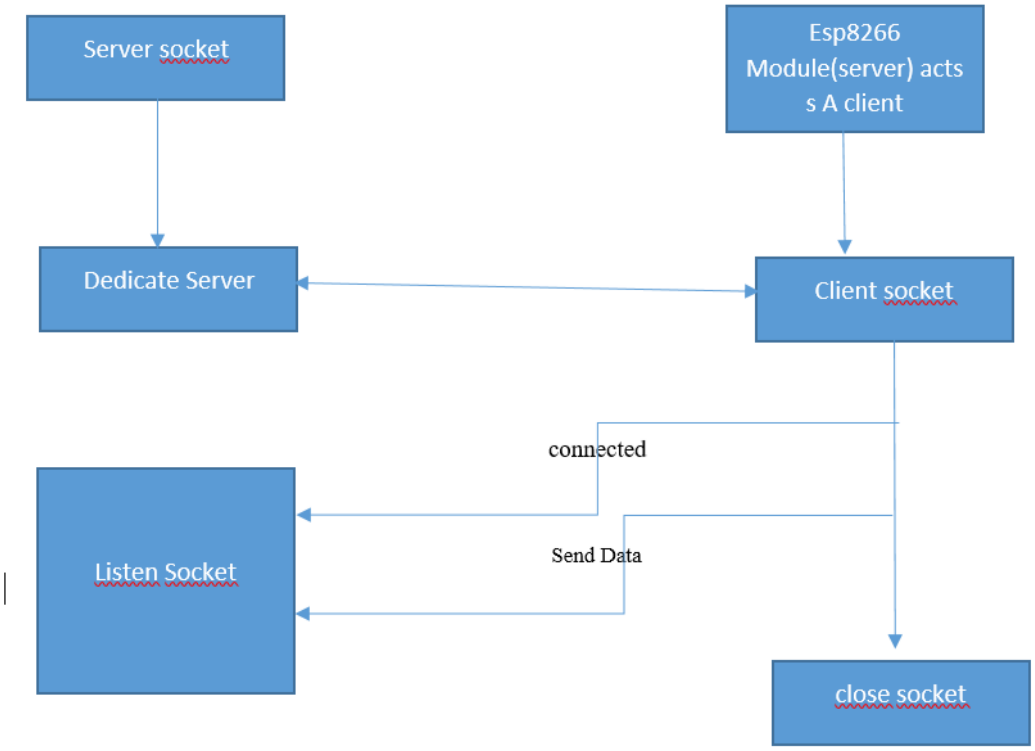


Figure 9: Client-server socket communication flow

3.9.1 Assign a thread for every incoming data from the Esp8266 server

Server program typically provides resources to a network of client programs. Client programs send requests to the server program, and the server program responds to the request.

One way to handle requests from more than one client is to make the server program multithreaded. A multi-threaded server creates a thread for each communication it accepts from a client which the dedicated Server do for every of incoming data from the Esp8266 (Server module). A thread is a sequence of instructions that run independently of the program and of any other threads.

Using threads, a multi-threaded server program can accept a connection from a client, start a thread for that communication, and continue listening for requests from other clients. Which

makes the server-to-server listen simultaneously while performing others task through the dedicated thread for the client request.

So the creation of threads and assignment of the routine goes like in the same routine for each and every data coming from the Esp8266‘s module, here below is the representation of diagram.

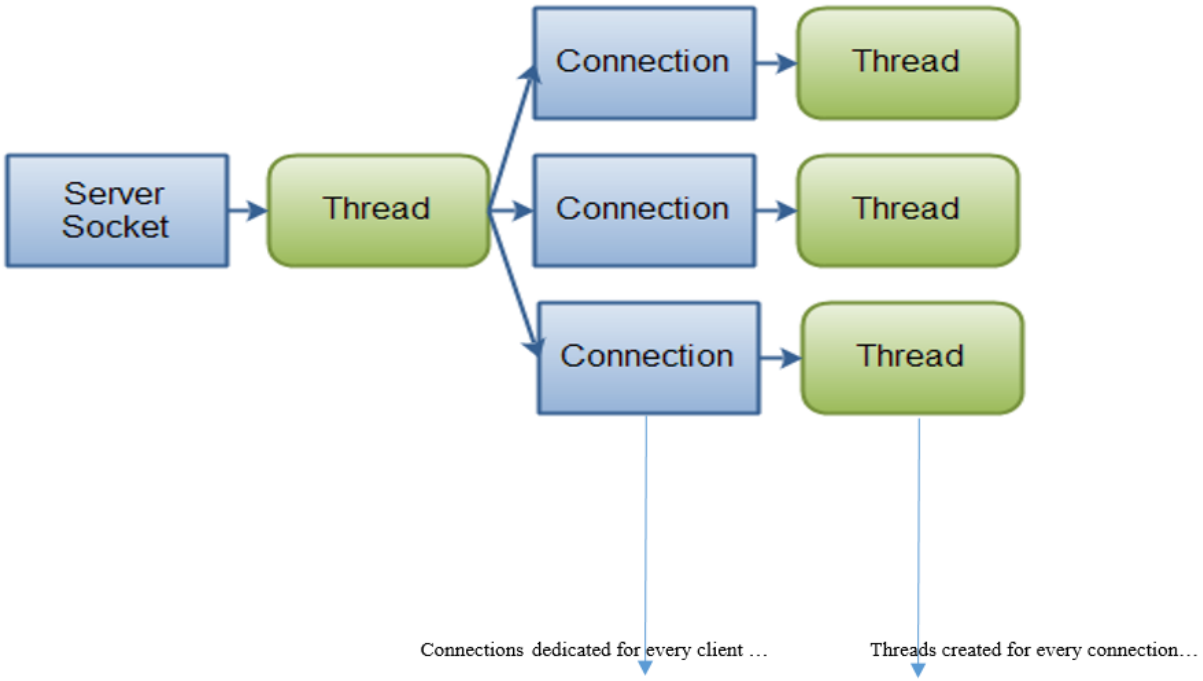


Figure 10: Thread creation and manipulation by centralized server

3.9.2 Evaluating the data and persisting to database

We have been looking the general overview of the communication setup and architecture from the Esp8266 server (which act as a client, when communication to the centralized server) to the dedicated Centralized server, once the data received the first operation handled is to check the validity of it, by

1. Comparing the mac address field which is a 24-bit of numbers
2. The timestamp whether it’s in the range of the past 10 minute at most
3. The validity of the Server Identification, which must have be in the list of areas Identifier.

Once this all checks performed and approved by the server, finally ends up the allocated threads to make a connection to the Centralize database and insert the data to it. For better understanding see below following diagram.

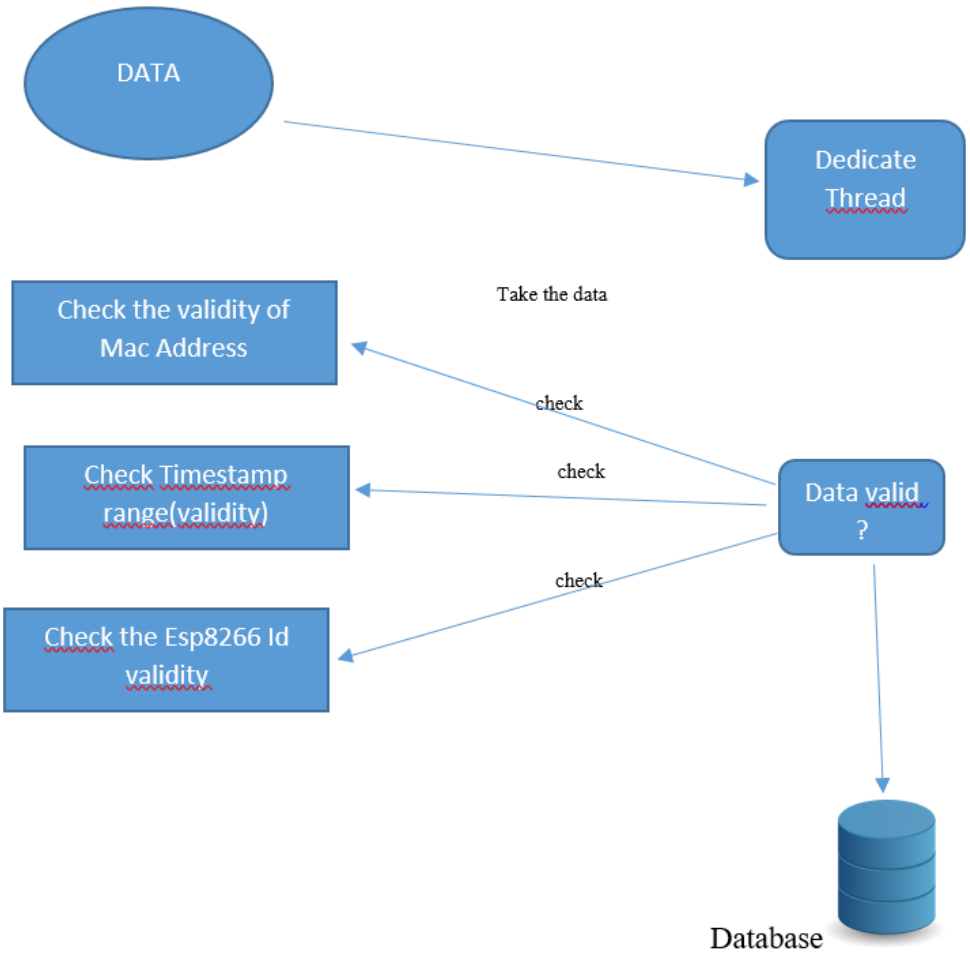


Figure 11: Packet validity verification

3.10 Job scheduler on analyzing historical data

Beside from the socket communication, the data retrieval, data processing and insertion to database. The main and smart functionality of the job scheduler is to scheduled execution, which aims for analyzing log history and running categorization algorithm, which is going to be present in the next section.

Analyzing log history

The objective of this operation is the main part of the thesis, which make crucial and vital operation by collecting and analyzing the event log (Smart device user's data). It collected from different area by many Esp8266 module (Server) to collect, parse and feed the Rest Service application program that expose the sniffed data in a meaningful, representational form and for further algorithmic operation, which we I will cover in the next section.

The data, which persisted to database by the centralized server, comes from different Esp8266 Server module (board). Assuming that, the scheduled job run every **5 minute.** This job aimed to create a neutral table which intended to store an aggregate data after running pair algorithm, and this data table in conjunction with the first table able to classify us the each and every single mobile users identifier (mac-address) is a station device (desktop) or moveable device (laptop, smartphone, tablet …)

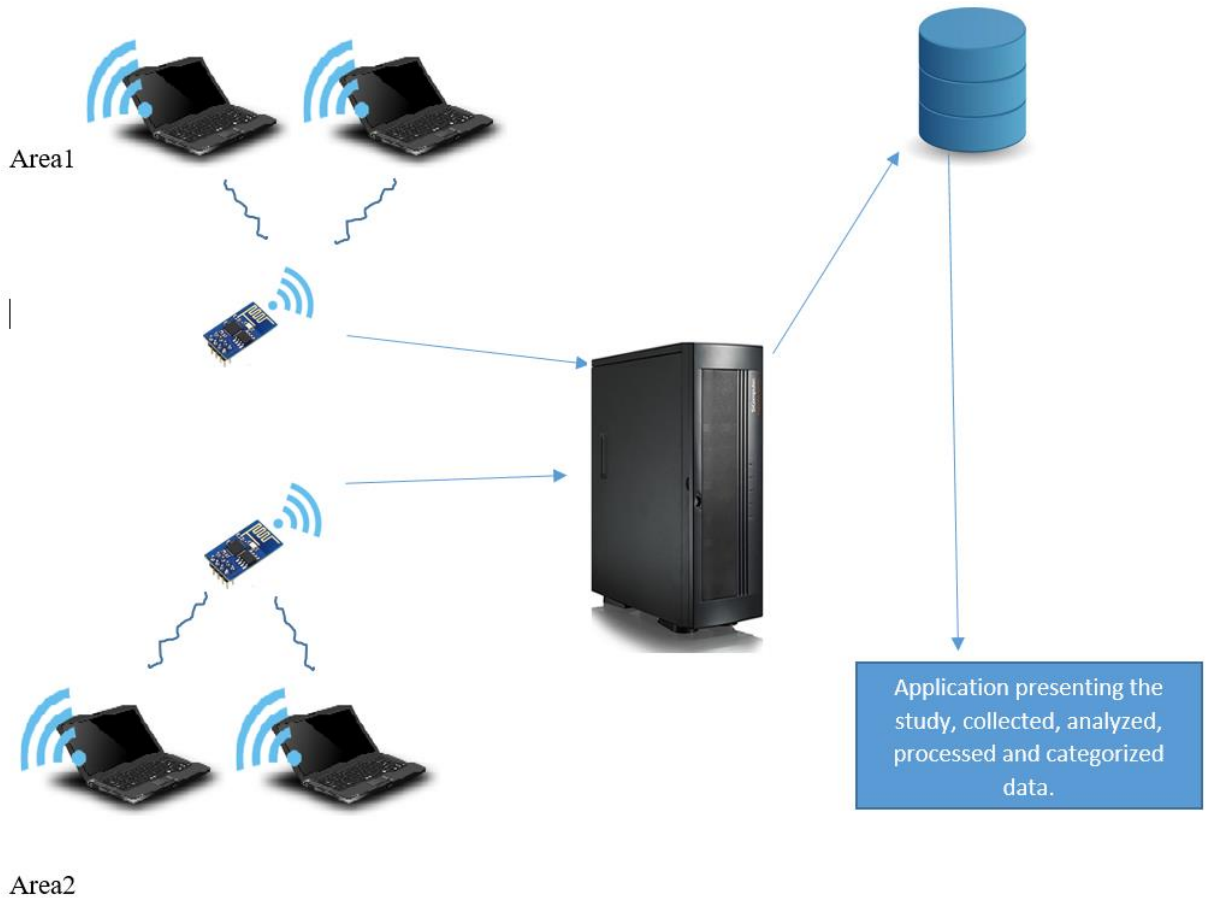


Figure 12: Centralized server data retrial setup

Differentiate mobile and station device

Knowing and differentiating users based on some characteristics is a key thing for this thesis. Since taking some control, actions and specific analysis is necessary on mobile device while station device is not the interest, the only way to exclude or differentiate this mixed data from both the mobile, smart phone device users from Desktop users achieved in this step.

Taking in control of this operation would have been better from the responding Esp8266 module (client nearby their area) but unfortunately due to the way Mac address organization.

MAC Address can only identify the manufacturer of the network controller chip. It encoded in the OUI (Organizational Unique Identifier) part of the MAC Address. Some manufacturers only make controller chips for smartphones, while some only make controller chips for computers. A few makes controller chips for both, but usually use very different model, so that each model has its own OUI (it's quite common that a manufacturer has more than one OUI).

Based on the OUI, you might be able to discern between MAC Addresses of smartphones and MAC Addresses of computers. Except in those cases where the manufacturer uses the same OUI for smartphone & computer controller chips.

So having in mind of this restriction it’s necessary to run some task in concatenation of grouping and learning the behavior of our users jointly and individually.

3.11 Characteristics of Desktop user’s uniqueness than the others

Based on the study and experiments of this thesis also assuming the general behavior of Desktop computers, logically it is easy to distinguish them from other users.

- Data origination from same source: interpreting the aggregated data collected by the centralized server, we can able easily to filter the same Mac address from different areas (at least from two areas) to falsify this characterization and stations, which always statically collected from the same origin, are countably a lot. This tell us from our study, this device is either peoples who eventually or occasionally sit on a given place or Desktop users, this can make us some percent assured but not 100%, we confirmed the differentiation in the second characteristics of the users.

- Time slot of internet or connectivity usage: Based on this two behavior I am able to group the needed mobile user's data from those Desktop computers, additionally I confirmed this work with 100% correctness by visiting their data from second origin (Esp8266 client area locator) source. The following picture describe well the real scenario of this identification.

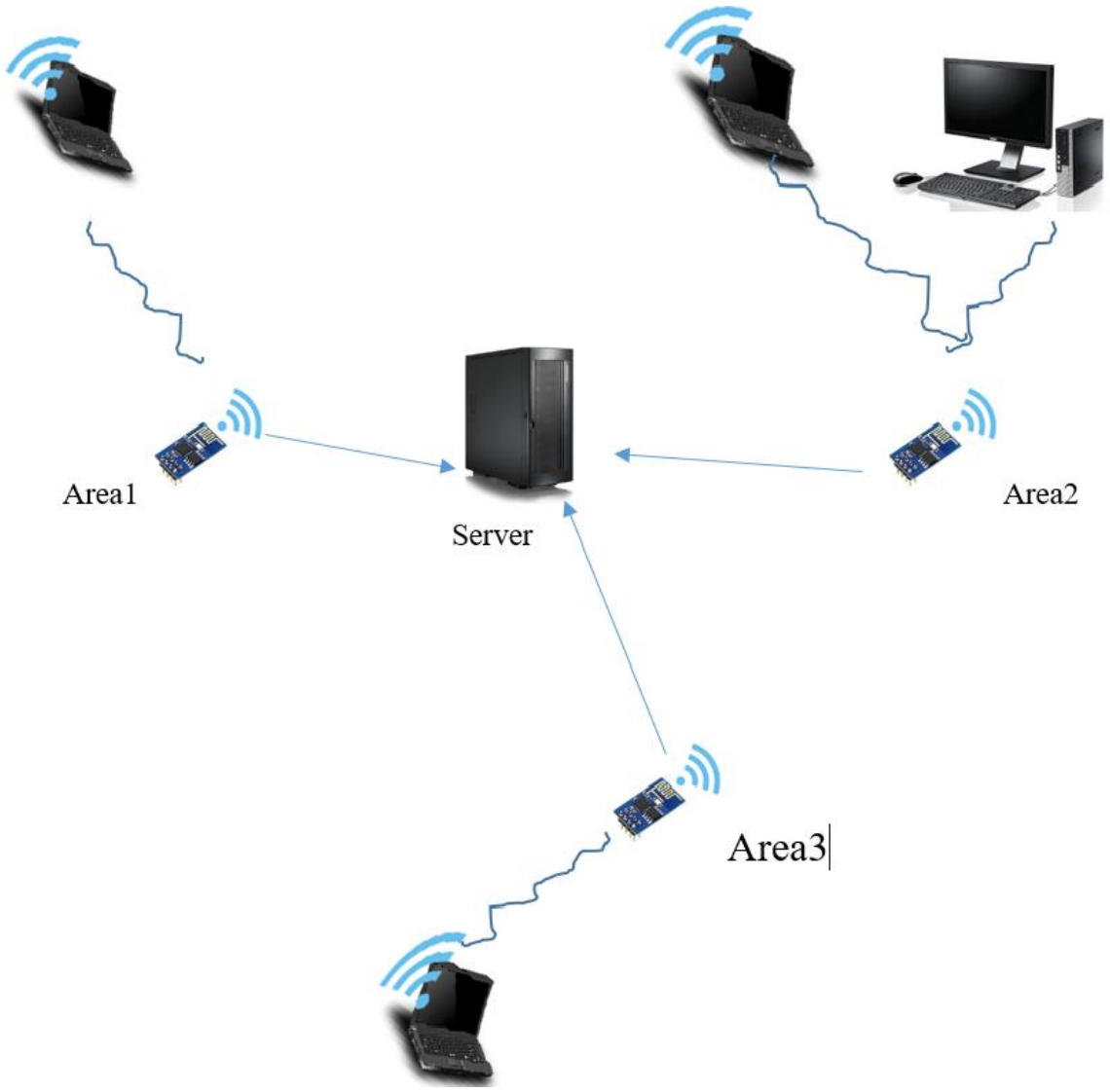


Figure 13: Differentiating desktop (Computer)-mobile scenario 1

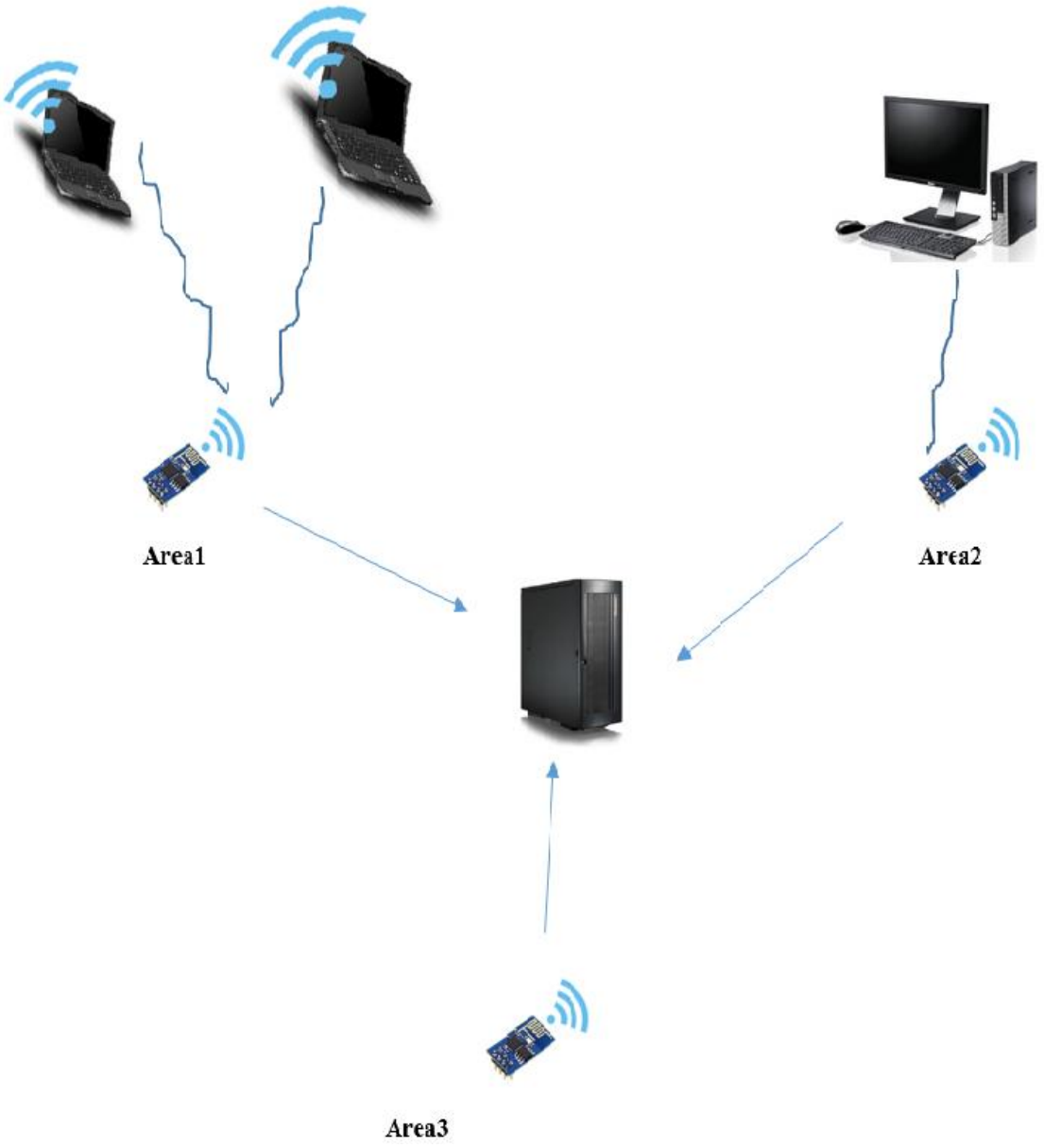


Figure 14: Differentiating desktop (Computer)-mobile scenario 2

This two diagram perfectly demonstrates the scenario I have been mentioned, as you can see from figure 1 in Area 2 putting out the parameter when (a time which the device tracked by Esp8266).

Scenario 1

Initially area 2 has two clients connected to the Esp8266 device since they are exchanging data to the given access point (Internet gateway) after some time t, the mobile device in this situation the

laptop. Users disconnected it session and moved out from that area presence, while the Desktop user statically remain there, fed up the data while surfing the internet.

Scenario 2

Initially Area 1 has 1 connected device to the Sniffer Esp8266 device,

Scenario 3

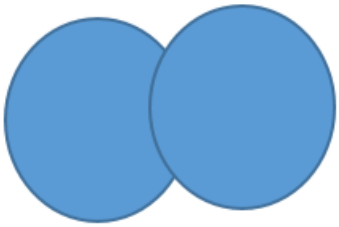
Assuming the previous two scenario, after a given time interval of repetition the initial status of both area identifier connection pools may increase or decrease, this change only occurred for the mobile devices, while the Desktop mac Address are rigidly stayed and recorded their identifier (Mac Address) from the same client(Esp8266).

3.12 Pairing Algorithm

After intensively recording the interested data which comes from dislocating area which this areas can be covered by one or more paired Esp8266 microcontroller chips, we are able to answer the 3 W , principal answers :-

- Who (the person identifier)
- Where (the data origins)
- When (the timestamp or time slice where the user's device recorded on)

This overall work will not answer this thesis objective; it is just like big piece of the slice missing from a given cake. once the scheduled job starts for the very first turn it starts its timer and reaching certain amount of duration, this unknown variable **Time "t"** reflect the duration the job can be executed without overlapping the next time scheduler job, it executes every registered mac address will be joined as a pair



Paired Mac addresses

Figure 15: Pairing Mac addresses

Since the analysis started from pairs study the first steps is to pair every users and propagate the data to newly table, then after for every pairs a counter started, which help us to answer the following crucial answers.

- *Who stay together?*
- *For how long they stayed together?*

Combining the two tables

Tabel1 the data, which hold the mac, **address**, the **timestamp**, client **Esp8266 Id**

Table 2 holds the **pairs data** and the **counter** (how long they stayed together)

The following questions will have answer.

Who stayed together?

For how long they stay together?

In which time those pairs meet?

In which area, time those pairs together?

In which time pairs those pairs appear mostly? etc

This type of interesting study behavioral analysis can be get answered, imagining couple of days analysis answers a lot which I can list it in this paper, what if this data collected for period of days, weeks, months, years…

This paper study main aim is to start studying the behavioral pattern of mobile users in different areas. It can be well-suited places to study and forecast based on the behavioral patter of the users study.

Hotel, Museums, Universities, colleges, Transportation solutionist, Supermarkets etc. A lot can benefited, in the next section I will try to describe in diagram the pair algorithm coupling, and time consideration.

Let us suppose in the time Range of 5 Minute, we collect this much new Users

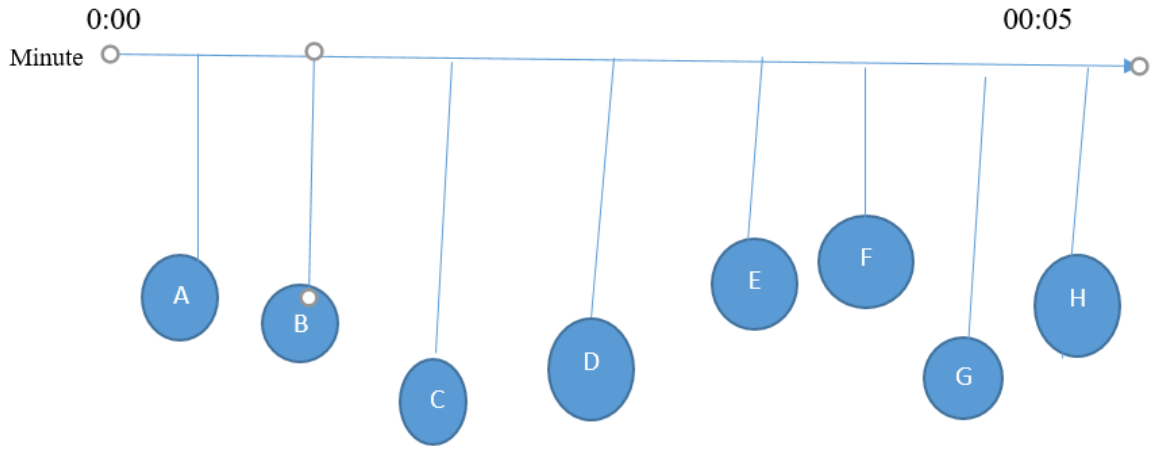


Figure 16: Sniffed Mac address in a given time slot

The algorithm start in take care of the very first x minute since the program start executing. The first execution will handle and gathered people in a given set of area at same time occurrence, and start pairing every of the persons mac-address from the first stable and start populating the second table with the paired String concatenating the mac addresses,

Figure 17: Mac address pairing computation mechanism

After each member of the users paired each other, the first round ends and wait for the next interval till the proceeding algorithm phase executed.

Starting from the first analysis, third parameter counter added on to magnify the time/duration of pairs or group of pairs staying together in a certain area, in a fixed or range of time range. As you can see all pairs counter value start from 1.

Even if the algorithm paired users, we can even dive to the analysis and make combinational analysis on group of pairs or even individual statistics may satisfy us more, that is up to the sector who are going to use this usage of the thesis.

Let's take second round and finalized the chapter, extending our example from the ending time of the above figure 00:05 till the next 5 minute 00:10 our members and their being paired structure will differ from the above, lets demonstrate that.

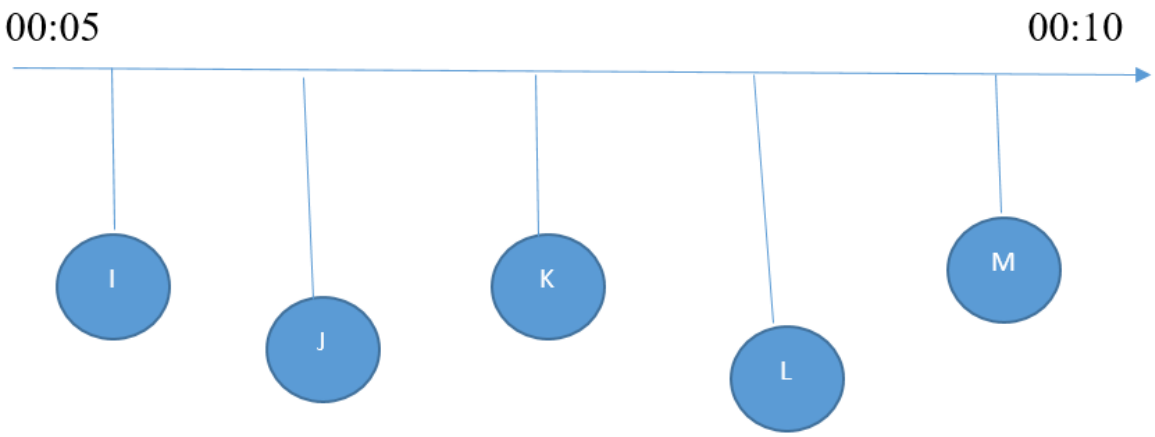


Figure 18: Round 2 sniffed Mac address in a given time slot

This figure shows us the previous users (users A-H) appear in a round due to many reasons that is not our interest but the pair algorithm start executing and structuring its users. The same as the first diagram showed us above. The next chapter deeply talks about the functionality and real scenario based introduction in many sectors, and we continue the **testing, result** and **limitations** and finalized our talk after we say a lot about the future work of this thesis.

Chapter Four

4. Wi-Fi Tracking and MAC Randomization

This chapter talks in detail the Wi-Fi tracking scenario in accordance with the pitfall of Mac address Randomization and current technologies and techniques, which used to treat the MAC Address randomization.

Wi-Fi has become a main ubiquitous technology and this days its integrated on the humans day to day handled devices like smartphones, wearable device and tablets … it is quiet frequent for smartphones and tablets because of their easy Internet usage thanks to the growing number of hotspots or wireless community networks.

As I mentioned briefly on chapter 3 the mechanism on how moveable device users are easily traced and a lot, the main goal of this chapter talks about MAC randomization, but before that we will roughly cover the steps and technological transitions or prevention method which used to eliminate the MAC Randomization technique.

Most smartphones and tablet users left open their Wi-Fi interface active and open knowingly or by habit which exposed them to be detected their presence and tracked their movement. This condition able to broadcast Wi-Fi packet, which even holds their unique Identifier to the nearest and available Access Point (AP), this packets are sent in clear text to the destination. The technique to detect and traced and analyzed the movement is named as Wi-Fi Tracking.

Passive Wi-Fi tracking lead to a huge privacy threat to the user mainly if those accumulated and unknowingly broadcasts data not properly anonymized. Saying this to overcome this type of issues and threats several countermeasures are proposed and applied, the mainly agreed and accepted widely is the birth of random identifier generator instead of continuously send the unique identifier in clear text.

The birth of MAC Randomization which try to solve the scrutiny of wireless users instead of sending a static and unique identifier it allows the device (Smartphones, tablets and laptops) to

introduce itself to the nearly available access point through different random identifier, however new research are demonstrated that Wi-Fi Tracking is still possible until this thesis is written.

Now let's briefly talks about MAC Randomization, how and in what way its preventing Wi-Fi tracking and finally we will cover the can be used to track Wi-Fi devices despite the use of MAC address randomization, including attacks based on the content and the timing of Wi-Fi frames.

4.1 Wi-Fi and Service Discovery

Any kind of Wi-Fi communication transmit their signal on Radio channels, and through this standardized protocol the key and vital thing is the way how the information is sent and received using service discovery mechanism. which able the wireless interface of the device to discover new access point or they can passively discover access point by listening the beacons they broadcast or it can also actively broadcast probe request from the nearby access point whom in response provide probe response to the wireless station.

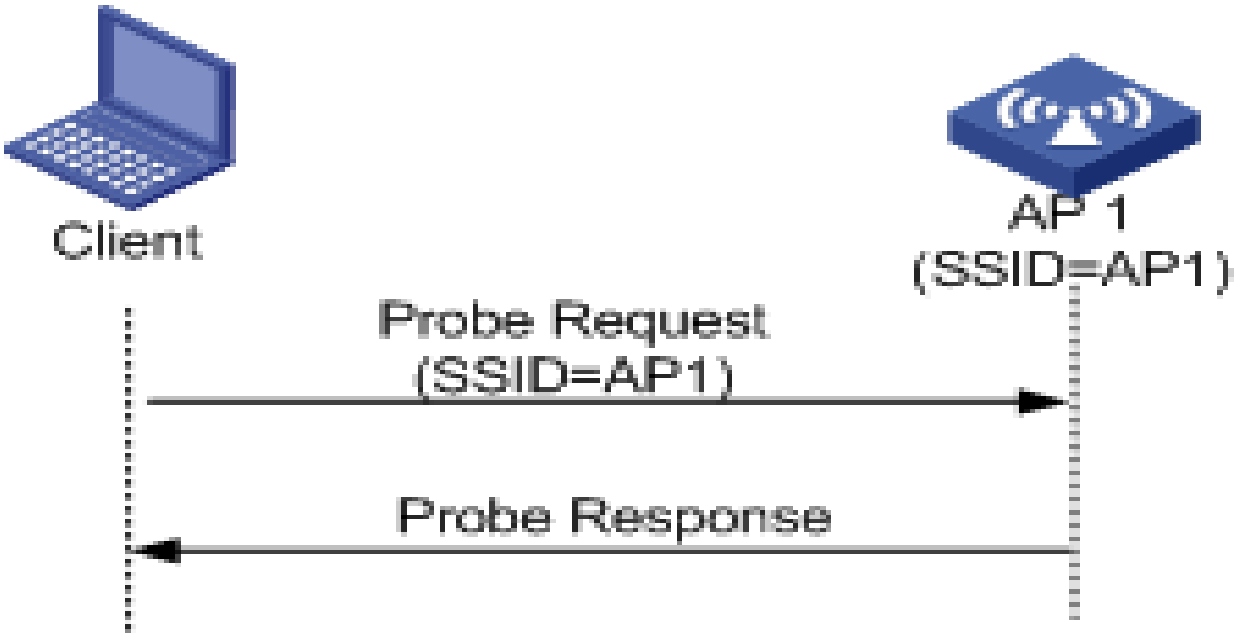


Figure 19: Active Scanning (broadcasting)

The Actively service discovery is mostly and widely used by Smartphones and tablets rather than the Passively service discovery mechanism , because of its less energy consumption makes it

favorable, so that Wi-Fi probe request is needed to be sent periodically for the device which are not connected to access point.

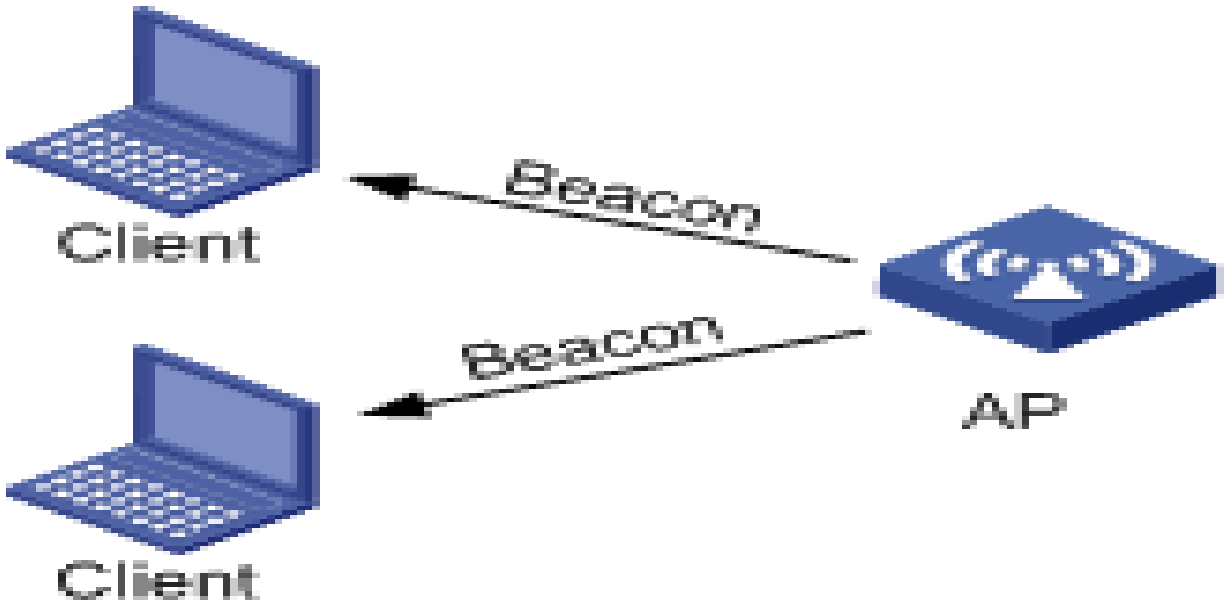


Figure 20: Passive Scanning (broadcasting)

4.2 Wi-Fi Based Physical Tracking

Including this thesis which main objective and Technological choice based on the Wi-Fi probe request to impose the targeted data retrieval mechanism, other studies and researches are based on the same issue Wi-Fi Based Physical Tracking.

In short and precise way physical tracking is a conversion of the physical world tracking happening in the digital worlds, specifically by listening and passively collecting the unique Identifier (Mac address) of a give wireless station. Third parties collect and investigate presence information (data) about a given object (Mobile device users) in unstoppable and limitless continuity.

Wireless devices identify access points within close proximity. Traditionally, devices perform active scanning where they broadcast probe request frames asking nearby APs to identify themselves and respond with 802.11 parameter information required for connection setup. These probe request frames require a source MAC address, but if an 802.11 device uses it's globally unique MAC Address then it is effectively, broadcasting its identity at all times to any wireless

receiver that is nearby. Wireless device users can then easily be tracked across temporal and spatial boundaries as their devices are transmitting with their unique identity. [8]

In case of Wi-Fi collection and gathering of data performed by a set of monitoring nodes, in this thesis context (Esp8266) deployed over an area of interest, which forward the collected information to a centralized server, refer chapter 3 for deep understanding.

Data collected by Wi Fi physical analytics generally obtained without explicit consent of the user. In addition, a number of Wi Fi trackers keep presence data in a raw or poorly anonymized format. Consumers exposed to a real privacy threat, which calls to technical solutions.

4.3 MAC Randomization

Now it is the time to introduce the proposal solution by *Gruteser*, to tackle the Wi-Fi tracking which exposed the security and a threat of any mobile users (smartphone, tablets …..).

Mac Randomization is a way to exchange a probe-request instead of using unique identifier (mac address) to use a temporary and periodically renewed unique identifier.

The first industrial stakeholder to apply this technology to prevent the Wi-Fi tracking and any kind of move mental, behavioral anonymization is The Apple Industry. They started to implement on the Mac Randomization. Sequentially the feature get started on Android 6.0, Linux iwlwifi driver and in windows 10, The Mac Randomization only adapted and only applied on Wi-Fi probe request when the wireless station is in scanning mode, however Windows extend this feature for association mode(in time of a device associated to the available wireless network).

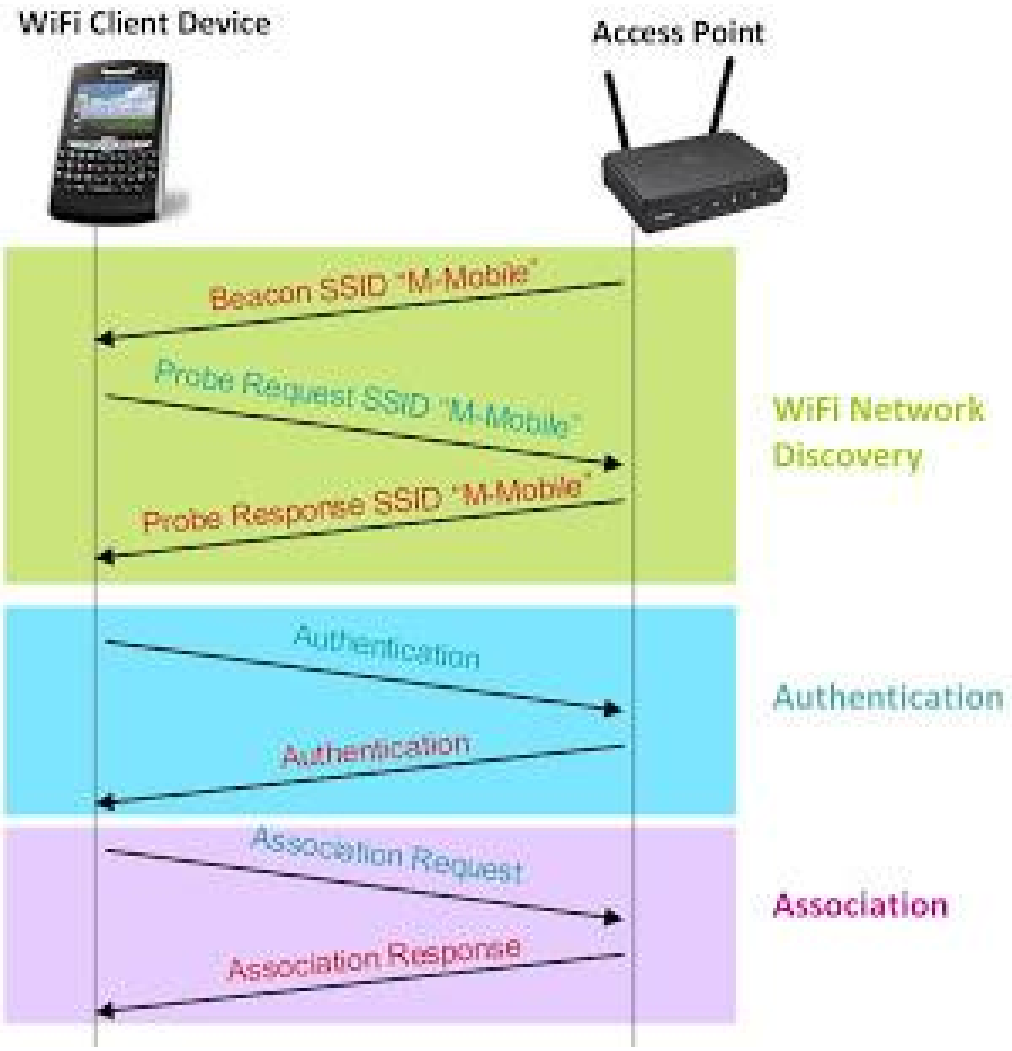


Figure 21: Wireless connection phase [7]

4.3.1 Pitfall or drawback of using MAC Randomization [8]

- Randomization techniques and schemes were easily identified from large collections of wireless traffic.
- Adoption rates for MAC randomization are low, particularly for Android devices.
- Passive and active techniques for determining true global identifiers is a trivial task due to flawed MAC randomization implementations, particularly for Android devices.
- The global MAC address was discoverable via a “control frame attack”. This allows tracking/surveillance for all known devices, irrespective of the OS, manufacturer and device type or randomization scheme. [8]

4.4 Content Based Attack

Mac Randomization on the first design and implementation feature provided by the Apple Industry emphasized only in the source field of the packet, which ensures during any probe request to change this source field, although Probe request also holds other field in both header and payload, which can leverage Wi-Fi tracking.

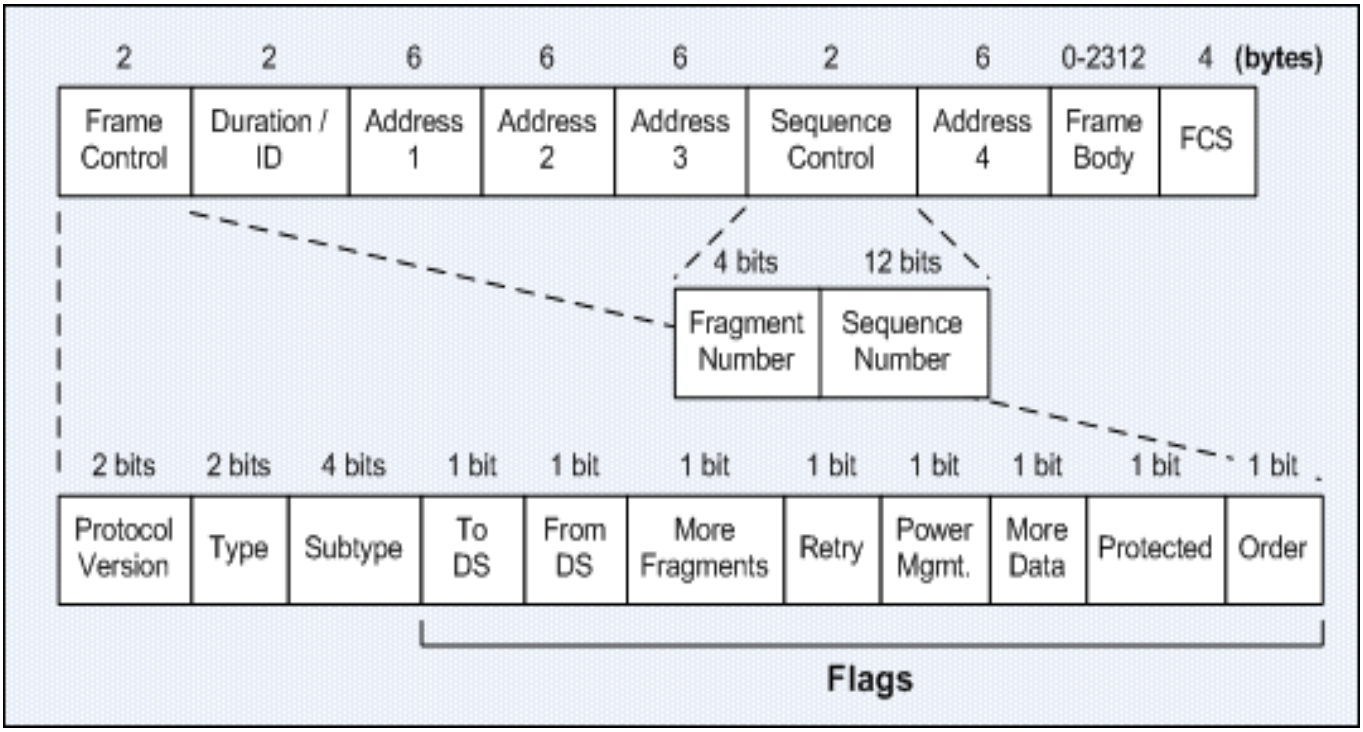


Figure 22: Sequence number field inside the frame [6]

This weakness of Mac Randomization feature identified by Freudiger, as his explanation the sequence number field of the packet not changed every time the new Unique Identifier generated and send by the Wi-Fi station to the nearby Access point. Having the consecutive sequence number of different randomly generated mac address still can be able to concatenate a random mac address according to the sequence number order and estimate a single mac address for those dynamically changed address.

Another issue, which can be predictable as that of a sequence field is identified on the physical layer, the scrambled seed used with OFDM used a predictable sequential number which let someone to be able link the randomly generated mac address (unique identifier)

4.5 Time Based Attack

This mechanism in the same way defeat the MAC Randomization using another resource, instead of basing on the frame content, which can be captured and de encapsulating the packet structure and finding a fields like sequence number is one way as we seen above.

Time based works or leverage the MAC Randomization using another resource than others, in time of Active service discovery mode Wi-Fi stations appeared or followed predictably and discovered temporary pattern.

Scanning an available AP performed using bursts, when stations scan and send probe-request in a given short period of time (400-500ms) by estimating the temporary distribution of mac address in a given bursts and in between two consecutive bursts it is possible to create a temporal fingerprint which allow us to separate and isolate a given device.

Studying, analyzing and some computation on bursts set (a group of burst, which hold the probability of similar randomly generated mac address in those groups of bursts)

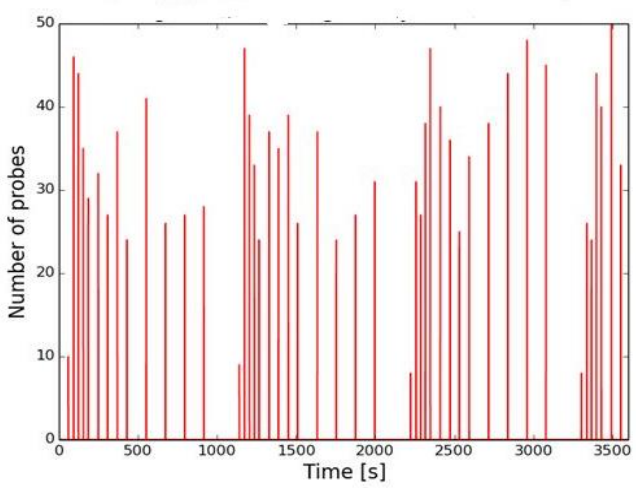


Figure 23: Regular bursts of Wi-Fi Probe Request [5]

4.5 Mac-Randomization on Thesis

If we talk this much about the new technological feature of MAC Randomization, Its time to Generalize about this thesis general aspect up on this technology.

Since the main goal of the thesis as briefly stated in Chapter 1 and Chapter 2 it directly related to the ideology of Wi-Fi tracking which is based on Wi-Fi probe request. The pitfall of the MAC Randomization will gone be the main concern of the implementation, However the Module(device) Esp8266 tackle it by somehow knowingly or unknowingly.

The rest part is out of this thesis scope, future consecutive part of the work will have deep look and solution too, considering until this thesis written there is no standardized implementation for MAC Randomization and the threats that leveraged the concept.

To summarize I will conclude this chapter by saying how I tried to tackle MAC Randomization partially using this Wireless controller integrated device(Chip).

Esp8266 works in promiscuous mode, which capture the packet by sniffing (scanning) stations, which send Wi-Fi probe request, authenticate and associate during the wireless connection phase beside windows 10 the rest smartphone devices will not change the unique mac address after the probe request. Windows 10 device apply the in different fashion by periodically changing the unique identifier even after the Association mode, this will let us to not accurately track the movement or behavior of those users.

To summarize using the Esp8266 full Wireless integrate kit, it's possible to keep listening the station and beacon communication in the all feature, although HT40 and LDPC packet format is not supported, for better look I do recommend to have a brief look of Chapter 3(Implementation and Technology).

Here below is the Mobile type and their supported MAC Randomization feature.

Chapter Five

5. Results and Expected Outputs

This chapter applies the approach introduce in chapter 3 and present the expected output gathered and taken out during the testing phase the implementation phase. To briefly understand and summarized the key part of each module of the entire application, the outputted data briefly described how and what to get.

It begins by presenting firstly the Esp8266 client output, then we proceed furtherly and discus the Esp8266 (Client – Server Serial communication) output and finally have a look the Rest based web Service which exposed the collected data from different Esp8266 Server(area represented), aggregated and computed data from the centralized server.

Before jumping in to the result, the sampled data which is captured by the Esp8266 module during the Wi-Fi tracking phase is taken from the Polytechnic Di Torino, laboratory Room and no risky and threat is performed by this thesis objective and goal.

5.1 Esp8266 Client

This portion of the application gives the output, mainly the sniffed stations Unique Identifier (Mac-Address), as specified in chapter 3, we analyses two scenario for Wi-Fi Tracking, here below it states the result in two form,

Scanning through Specified Channel and Scanning through Specified Channel

5.1.1 Scanning through Specified Channel

Client	Client Mac-Address	SSID	Beacon Mac Address	Channel
CLIENT:	44aaf5e4bb1b works with: [	FASTWEB-3mU6kJ]	01005e7ffffa	6
CLIENT:	44c34664abc9 works with:	8416f9ea05c6	8416f9ea05c6	0
CLIENT:	e02a82e141c6 works with: [	Vodafone-WiFi]	bc15ace9e66d	1
CLIENT:	5460094a061a works with: [	FASTWEB-BACCI]	00603b32e883	6
CLIENT:	90356efd99b0 works with: [	Vodafone-33461553]	01005e7ffffa	9
CLIENT:	0019fb751a5a works with: [	BaraNet]	01005e7ffffa	11
CLIENT:	6459f84f9240 works with: [	Vodafone-33789148]	01005e7ffffa	13
CLIENT:	a88195dd9d5a works with: [	Vodafone-33789148]	ffffffffffff	13
CLIENT:	00603b32e883 works with: [	FASTWEB-BACCI]	9801a7c682e3	6
CLIENT:	a018288389c1 works with: [	Vodafone2.4GHz-33784626]	01005e0000fb	11
CLIENT:	6459f84e77a0 works with: [	Vodafone2.4GHz-33784626]	01005e7ffffa	11
CLIENT:	9cd35bbcd1dd works with: [	Vodafone-33461553]	90356efd99b0	9
CLIENT:	0c8ffff58311 works with: [	TIM-84781271]	ffffffffffff	6
CLIENT:	40d28a67d169 works with: [	TNCAP9C1B5A]	a4b1e99c1b59	1
CLIENT:	e8d1lbc26b2b works with:	e8d1lbc26b2d	01005e0000fb	0
CLIENT:	6459f8425240 works with: [	Vodafone-33731874]	01005e0000fb	13
CLIENT:	48437c9ebdc7 works with: [	FASTWEB-3mU6kJ]	01005e0000fb	6
CLIENT:	78810238d9d0 works with: [	TIM-84781271]	333300000001	6
CLIENT:	9c9726b90874 works with: [	BaraNet]	01005e7ffffa	11
CLIENT:	6459f83b8480 works with: [	Vodafone-33704006]	01005e7ffffa	10
CLIENT:	b0702dd0768f works with: [	InfostradaWiFi-1UAZ7J]	01005e7ffffa	6
CLIENT:	78a5dd0461d0 works with: [	TNCAP9C1B5A]	ffffffffffff	1
CLIENT:	74811437f8e3 works with:	001cb3ae9d59	01005e0000fb	0
CLIENT:	b8e856088e44 works with: [	FASTWEB-1-ZF4wCXqfWXwY]	3333000000fb	1
CLIENT:	b8e85666da8b works with: [	TIM-18404477]	3333000000fb	1
CLIENT:	0026d91c0142 works with: [	atlantis]	01005e7ffffa	9
CLIENT:	e0aa96a87f10 works with: [	Vodafone-33731874]	ffffffffffff	13
CLIENT:	9c9726c4ab78 works with: [	FASTWEB-1-C4AB79]	33330000000c	1
CLIENT:	64126982bff3 works with: [	alice]	01005e7ffffa	6
CLIENT:	ec107b5b8057 works with: [	FASTWEB-1-uhexi432rWk0]	84261531e6f6	1
CLIENT:	9c4fdac9d4c7 works with: [	FASTWEB-1-kNK35r0ehews]	333300000002	1
CLIENT:	a491bl18d47c works with: [	TIM-18404477]	ffffffffffff	1
CLIENT:	a0d79517e89e works with: [	TIM-18404477]	3333000000fb	1
CLIENT:	8c7967983a72 works with: [	Vodafone-77562728]	333300000016	1

Figure 24: Result of scanning through specified channel

As the name implies of, this way of scanning (Wi-Fi Tracking) listen probe request only on those selected for this sample we used channel 1,6,9,11,13.

The main drawback of this mode is, it only listen (sniff) beacons, which implicitly define to work on the selected channel, for other channel, it stay quiet and no Tracking performed. Its main advantage is it perform well while we identify our service provider access point and only want to

Track, analyze and compute the pairing algorithm on that specified access point and its supported channel, see out chapter 3 for brief comparison of those two scenario in detail.

5.1.2 Scanning through Available Channel

In reverse to Scanning through Specified Channel, this mode allows the Esp8266 board to stay in promiscuous mode and sniff mobile users regardless of any channel selection. As we can see from the output neither the SSID nor the Channel not stated out, because that would not the main interest of the application. This is the raw data (Mac-address), which instantly, sent to the Esp8266 Server board (module). See out chapter 3 for brief comparison of those two scenario in detail.

```
CLIENT: c4ea1db8e4dc
CLIENT: 44aaf5e4bb1b
CLIENT: 44c34664abc9
CLIENT: e02a82e141c6
CLIENT: 5460094a061a
CLIENT: 90356efd99b0
CLIENT: 0019fb751a5a
CLIENT: 6459f84f9240
CLIENT: a88195dd9d5a
CLIENT: 00603b32e883
CLIENT: a018288389c1
CLIENT: 6459f84e77a0
CLIENT: 9cd35bbcd1dd
CLIENT: 0c8ffff58311
CLIENT: 40d28a67d169
CLIENT: e8d11bc26b2b
CLIENT: 6459f8425240
CLIENT: 48437c9ebdc7
CLIENT: 78810238d9d0
CLIENT: 9c9726b90874

CLIENT: 6459f83b8480
CLIENT: b0702dd0768f
CLIENT: 78a5dd0461d0
CLIENT: 74811437f8e3
CLIENT: b8e856088e44
CLIENT: b8e85666da8b
CLIENT: 0026d91c0142
CLIENT: e0aa96a87f10
CLIENT: 9c9726c4ab78
CLIENT: 64126982bff3
CLIENT: ec107b5b8057
CLIENT: 9c4fdac9d4c7
CLIENT: a491b118d47c
CLIENT: a0d79517e89e
CLIENT: 8c7967983a72

CLIENT: d8c7717ef832
CLIENT: 38f8898e2d7e
CLIENT: f41ba18fc28b
CLIENT: 682737599209
```

Figure 25: Result of scanning through available Channel

5.2 Computational result by centralized Server

In this part, we will examine different parameter(input) based data retrieval mechanism, which are helpful to aggregate and present the output in a meaningful format instead of visualizing the raw and meaningless collection of mac-addresses, which are collected and applied paired algorithm in chapter 3.

This do not well present until we apply some queried parameter for better description of the 3 w (who, when and where).

5.2.1 Date and area parameterized result

Here below we start to explore some of the main queried search based on the three optional

parameters: -

Date (when)

Mac-Adress (Who)

It works by searching the combination of this three meaningful data. If one the misses it queries by two, which provided, if only one of them provided it query up on the specified parameter, if none inputted all the captured analysis of areas, which contain the captured address regardless of time duration will be presented.

In this sample example from, to keyword limit the time duration, and Area specifically bound in a given area, from the Get request of the URI.

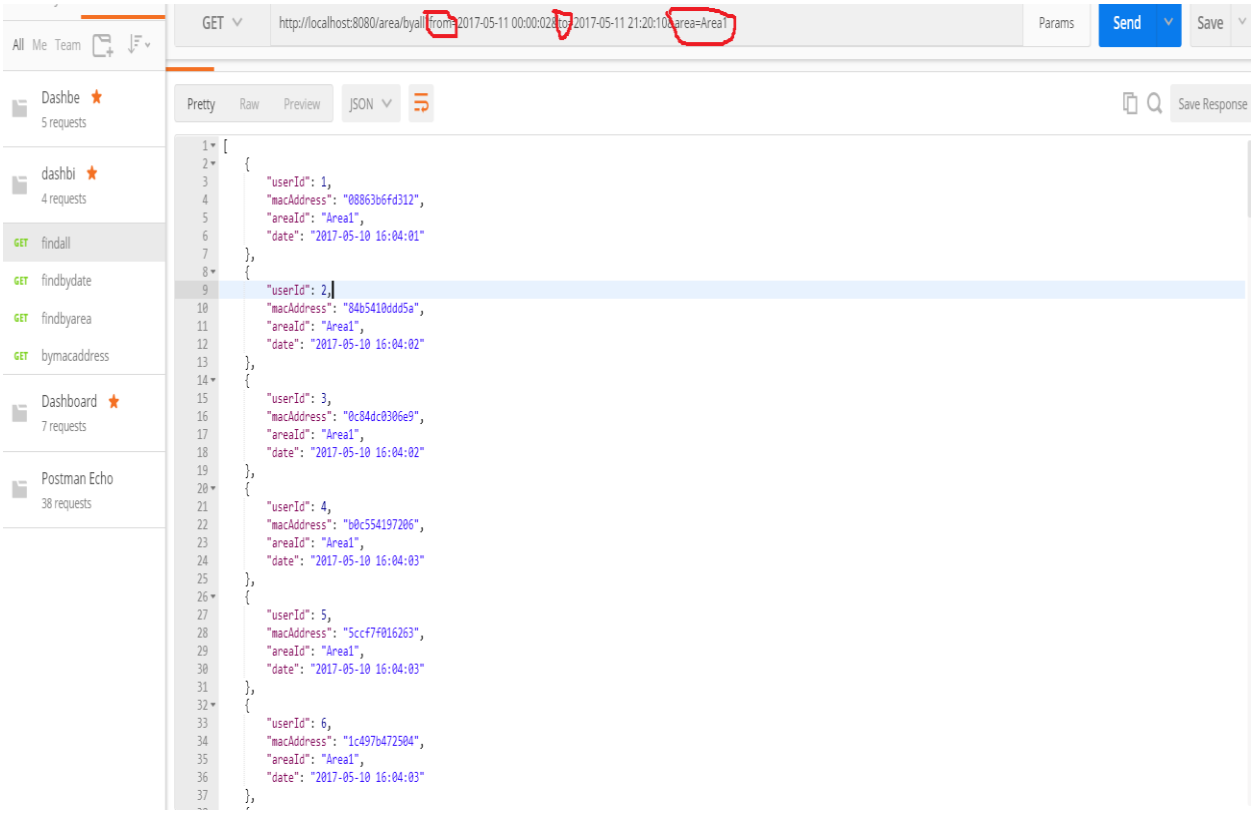


Figure 26: Result of filter by Time, Mac-Address and Area, Time, Mac-Address and Area as a filter input

5.2.2 Mac address based result

This search exclude the date and area involvement up on the filter and specifically provide a single users (station) information in all areas, without and restriction of date. This helps us to e and track a single user movement study, and behavior without overriding the date where he appeared and the time of appearance As the result this picture provide a sample output of Mac-Address:- **c4301d90f1b** output.

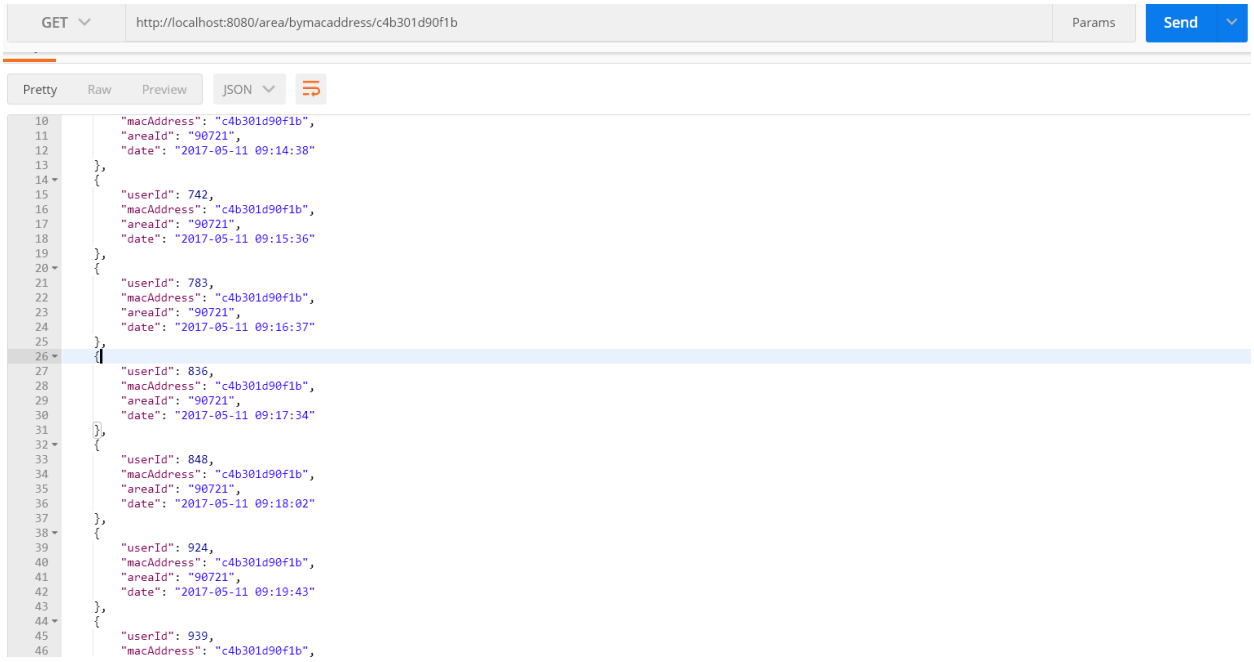


Figure 27: Result of filter by specific mac address

5.2.3 Area based result

This is one of the most important filter based search which applied a specific area as an input without considering the two **who** and **when** metadata information.

It enable any sector, which can be benefited from this solution by providing an Information to get the ability to control its users and have access on one of specific demarcated area according to the area identifier Esp8266.

Here below is the demonstrated search during this Thesis implementation, on two of the area

Area1 and **Area2**

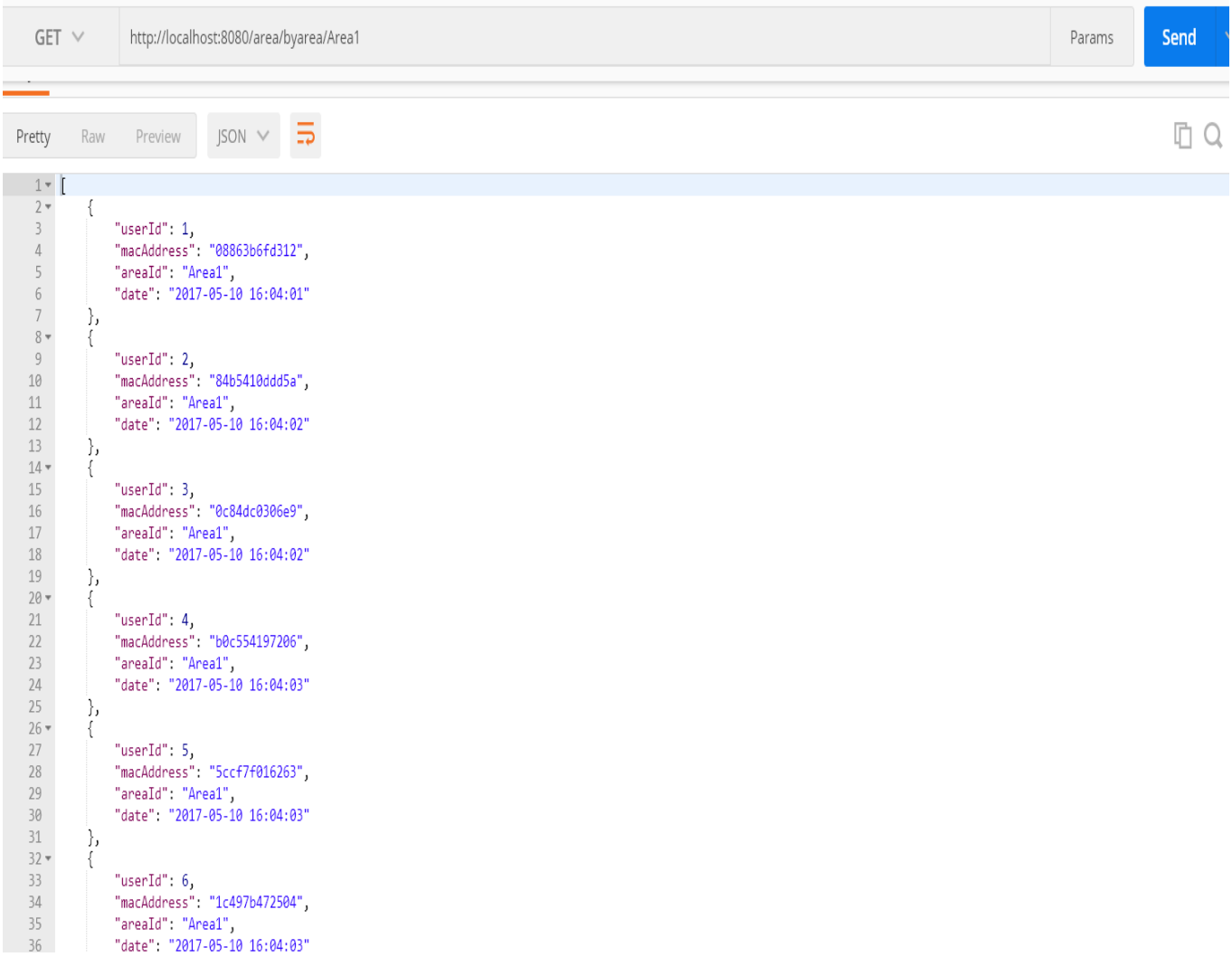


Figure 28: Result of filter by Area, Area1 as a filter input

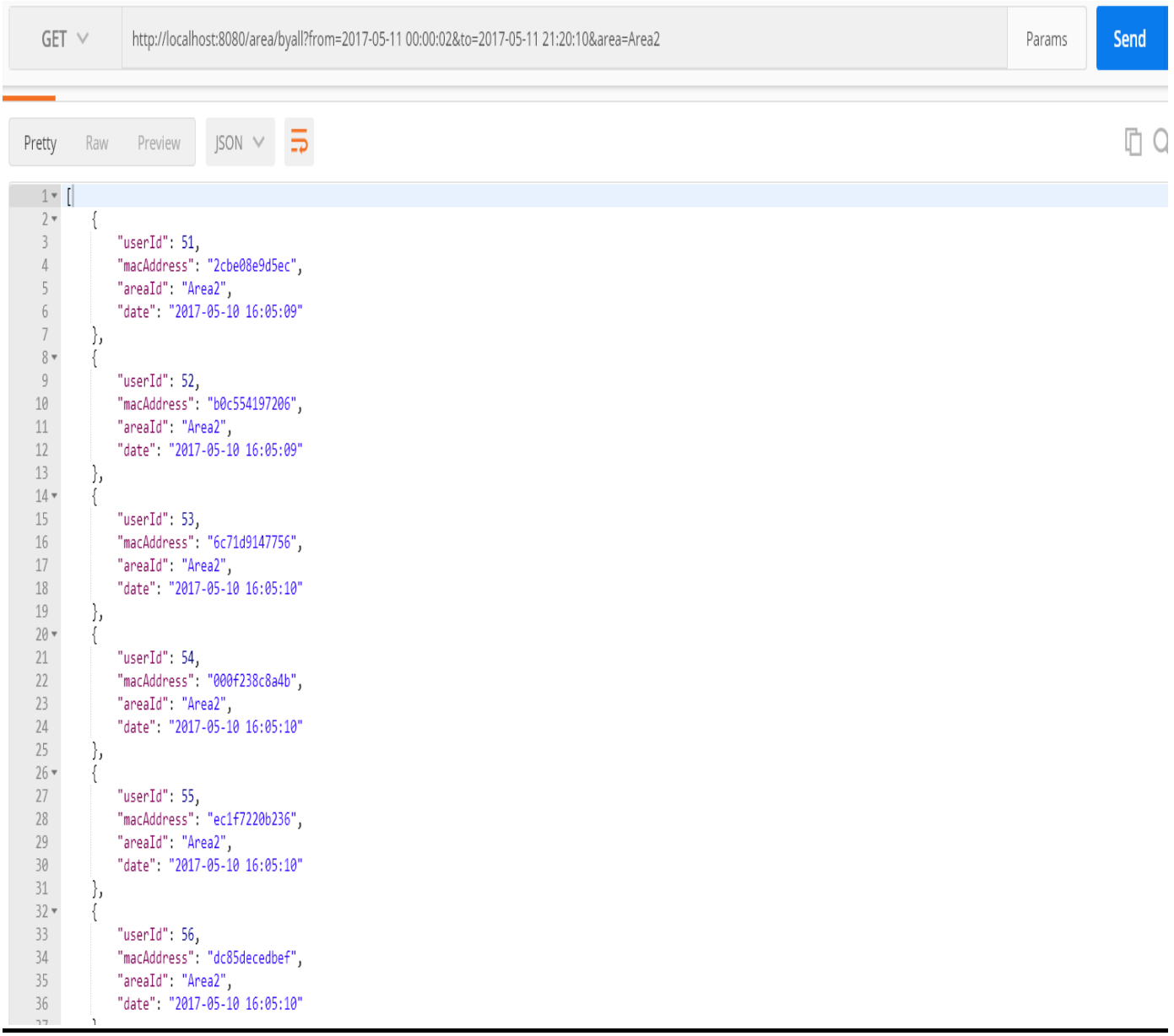


Figure 29: Result of filter by Area, Area2 as a filter input

5.4 Paired algorithm Result

User1	User2	counter
000e8fb18427	8c2daa537651	5
000e8fb18427	8c3ae3633f6a	6
000e8fb18427	8c705a100c68	5
000e8fb18427	9cb6d0d313c7	3
000e8fb18427	a00bbae43ad2	4
000e8fb18427	ac220b1d5561	7
000e8fb18427	ac220b8ea132	3
000e8fb18427	ac220b8ea133	1
000e8fb18427	acbc32a4709f	7
000e8fb18427	acbc32d61871	5
000e8fb18427	bc4cc41880f0	3
000e8fb18427	c0335e336589	6
000e8fb18427	c4b301d90f1b	7
000e8fb18427	c69a02299cfb	1
000e8fb18427	c8b373369221	7
000e8fb18427	c8b373369223	4
000e8fb18427	d83062621c29	6
000e8fb18427	dcd916642c1d	4
000e8fb18427	e02a82e141c6	2
000e8fb18427	e06995b44477	6
000e8fb18427	e0f84726b338	5
000e8fb18427	e8150e0dad09	2
000e8fb18427	e8508b6c5ec9	4
000e8fb18427	e8b1fcce23f4	2
000e8fb18427	f05b7bb530ea	2
000e8fb18427	f895c7f59cf7	5
00155d081301	441ca82aa68d	1
00155d081301	54271e762187	2
00155d081301	5ccf7f016263	2
00155d081301	80ee73607204	1
00155d081301	acbc32d61871	1
00155d081301	c8b373369223	2
00206bdbfde2	0008cace3cd3	3

Figure 30: Pairing algorithm, result for specific time duration

Chapter Six

6. Conclusion, Limitation and Future works

This thesis basing Wi-Fi probe request try to trace out (in Wi-Fi tracking trend) and analyze the human movement and behavior using Esp8266 micro controller integrated module(chip), this chapter summaries the contributions of this work, highlight its limitations and give directions for future work.

6.1 Conclusion

This thesis first defined the need for Indoors localization based on Wi-Fi probe request. Then warm up by introducing a general overview on the meaning and role of IOT in current technological contributions in Chapter 1.

Chapter 2 continued in deep the value of internet of things in accordance with Indoor localization then moved in detail about Indoor localization meaning, introduction and specifically touch Wireless based indoor localization its technology and its significance, which is the main objective of the thesis. Finally concludes by stating the functionality and areas (sectors) which implements and benefit more.

Chapter 3 directly talks in deep the used technology and implementation phase focusing on the design, setup, software, hardware, database used and their functionality, Next to that it proceed by defining how the implementation of the thesis act up on them briefly.

Chapter 4 talks the rise (threat) of MAC Randomization up on Wi-Fi tracking process and attacks, which emerged to leverage this features.

Chapter 5 sums up all and present each major steps of the entire implementation output on the collected sample data during the thesis work.

6.2 Limitations

The most important scenarios in which the contribution related to current indoor localization based service using wireless technology mainly relying on Wi-Fi probe request also in this thesis is likely to be limited are The emergence of the newly technological feature in most smartphones. MAC Randomization, which able the station (Mobile users) Mac Address to be changed during the time Probe Request and Probe Response.

Last but not the least the shortage of enough internal Memory storage of Esp8266 Microcontroller chip caused the needed and expected real-time data communication between the two Esp8266 modules (client- server) to follow a stored and forward scenario. The necessity of frequent and load of TCP server connection, which is established to the centralized server by Esp8266 server (board), hold a huge amount of space when new User data received from the Esp8266 client.

6.3 Future Work

Using the Wi-Fi probe request for indoor localization based solution or any Probe request based technology adding a new feature by smartphone stakeholder. MAC Randomization applied in almost all smartphone Industries to be not interfere the scrutiny and privacy of station (mobile users), as stated briefly in **Chapter 4**.

One of The future or continuity of this thesis can be based on this new Technology to overcome the threats, which raised to leverage its functionality of MAC Randomization. Moreover, by Deeping the business logic from the current proposed solution, which is generic to a specific sector as specified and briefed in **Chapter 2**, can be in Hospital, Transportation, and Museums we can exploit the major impact of this handy and flexible Esp8266 microchip, which hold full TCP/IP stack and microcontroller.

Finally yet importantly improving and extending the filtering mechanism, which enable us to consume the meaningful data as described in **Chapter 5** can be another sub area of the future work.

Bibliography

[1] Joel Scheuner, Alessandro De Carli *et al.* *Probr - A Generic and PassiveWiFi Tracking System,* Zurich 8050, Switzerland

[2] Marco Schwartz , Retrieved from <*https://openhomeautomation.net/getting-started-esp8266*> 2015, March 3

[3] Alexander Pukhanov , *Wi-Fi Extension for Drought Early-Warning Detection System Components*, Linköping December 10, 2015

[4] ESP8266 RTOS SDK Programming Guide, Retrieved from <*https://www.espressif.com/sites/default/files/20aesp8266_rtos_sdk_programming_guide_en_v1.4.0.pdf*>

[5] Julien Freudiger, PARC (A Xerox Company), *How Talkative is your Mobile Device? An experimental Study of Wi-Fi Probe Requests*

[6] Yap Chin Hoong*, IEEE 802.11 Frame Types, <http://www.itcertnotes.com/2011/05/ieee-80211-frame-types.html > , May, 2011*

[7] Airegis*, Deadlock in WiFi Networks*, *< http://wifidot11.blogspot.it/2010/10/deadlock-in-wifi-network.html*> , October 31, 2010

[8] Torjunkie, *MAC Address Randomization: Not as Random as You Think*, April 25, 2017

[9] Yu Wang, UPPSALA UNIVERSITET, *Intelligent Medicine System Prototype of The Internet of Things,* August 2012

[10] Freescale, the Internet of Things, *What the Internet of Things (IoT) Needs to Become a Reality, < https://eu.mouser.com/applications/iot-reality/>*

[11] Freescale, freescale.com / arm.com*, What the Internet of Things (IoT) Needs to Become a Reality,* Document Number: INTOTHNGSWP REV 2 May 2014 < *https://www.nxp.com/docs/en/white-paper/INTOTHNGSWP.pdf>*

[12] Raafat Hantoush, NOVA Information Management School, *Evaluating WI-FI indoor positioning approaches In a real world environment,* October 2016

[13] Nik Harris, Tracking People & Devices with WiFi**,** < http://nikharris.com/tracking-people/>**,** JANUARY 22, 2015

[14] Lin Sun, Sinong et al*, Mobile Device Passive Localization Based on IEEE 802.11 Probe Request Frames,* June 19, 2017 < *https://www.hindawi.com/journals/misy/2017/7821585/* >

[15] Hannah Becker, *What is the Internet of Things and Why is it Important?* , July 10, 2013
[16] Kai Goerlich , *The Importance of the Internet of Things*

The value of the Internet of Things won't be in the connections; it will be in the exponential amounts of data generated by those connections and processes. May 9, 2016

[17] Junjie Liu, f Prof. Raj Jain ,, *Survey of Wireless Based Indoor Localization Technologies*, April 30, 2014

[18]

[19] Joel Scheuner, Alessandro De Carli, Genc Mazlami, Sebastian Stephan, Dominik Sch¨oni, Thomas Bocek, and Burkhard Stiller , *Probr – A Generic and Passive WiFi Tracking System University of Zurich, Department of Informatics IFI, Zurich 8050, Switzerland,*