

IEEE802.11 Access Point's Service Set Identifier (SSID) for Localization and Tracking

Mohammad Z. Masoud^{1,*}, Yousef Jaradat¹ and Mohammad Alia²

¹Electrical Engineering Department, Al-Zaytoonah University of Jordan, Amman, 11733, Jordan

²Computer Science Department, Al-Zaytoonah University of Jordan, Amman, 11733, Jordan

*Corresponding Author: Mohammad Z. Masoud. Email: m.zakaria@zuj.edu.jo

Received: 21 September 2021; Accepted: 01 December 2021

Abstract: IEEE802.11, known as WiFi has proliferated in the last decade. It can be found in smartphones, laptops, smart TVs and surveillance cameras. This popularity has revealed many issues in health, data privacy and security. In this work, a WiFi measurement study has been conducted in Amman, the capital city of Jordan. An Android App has been written to harvest WiFi information of the transmitted frames of any surrounding Access points (APs). More than 240,000 APs information has been harvested in this work. The harvested data have been analyzed to find statistics of WiFi devices in this city. Moreover, three power distribution models have been derived from the data for three different areas, closed, open and hybrid areas. In addition, the collected data revealed that the SSID can be leveraged as a landmark for the access points (APs). To this end, SSIDtrack algorithm is proposed to track shoppers/walkers in closed areas, such as malls to find their walking route utilizing only the SSID information collected from the surrounding area. The algorithm has been tested in two different malls that consist of four different floors. The accuracy recorded for the algorithm acceded 95%.

Keywords: WiFi; IEEE802.11; MiFi; power model; SSID; BSSID; SSIDtrack

1 Introduction

Smart devices have proliferated in the past decade. Smartness is defined as devices ability to communicate and exchange data and information [1]. Nowadays, Internet allows data sharing among millions of devices across the globe. Internet converted the smartness of the devices dream into reality. However, sharing the Internet connection among multiple devices may increases the subscribers cost and complexity. IEEE 802.11, commercially known as WiFi, emerged to tackle these issues. This technology allowed subscribers to share one Internet connection among their devices in houses, cars, work and even in the street. Smart devices are required to be equipped with a wireless network interface card (WNIC) to join the smartness era. However, the proliferation of WiFi technology may create passive and active opportunity in one hand and other hazards in the health care field [2,3], interference and data privacy [4] on the other hand. WiFi is a wireless standard that operates in the scientific and



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

medical (ISM) band 2.4 GHz and in the 5 GHz band. Each WiFi access point is identified by its service set identifier (SSID) which is the primary name associated with the network that allows other devices to join or participate in the network. Moreover, WiFi access points are identified by their unique MAC address known as basic service set identifier (BSSID). However, since this address is unique, it may infer private information and can be tracked [5]. Many algorithms have been proposed and adopted to tackle the uniqueness privacy issue of the BSSID, such as, MAC randomization [4] that have been implemented in different operating systems, such as Android, IOS and Windows [6]. However, SSID has been used as a safe variable since it is not unique. The passive SSID Hidden SSID algorithm is the only implemented algorithm that treated the SSID as an issue that may reduce the hacking attempts on any access point. Moreover, many algorithms have been proposed to detect these hidden SSIDs [7].

In this work, a measurement study has been conducted to show the amount of information that can be revealed from SSID of WiFi networks. An Android App has been written to harvest APs around the user every two seconds. The recorded data consists of BSSID, SSID and timestamp only. We have recorded the signal strength also; however, it has not been utilized in this work. Three field experiments have been conducted. The first one in open urban area, the second in the closed areas buildings and the third in a hybrid location which consists of open areas and buildings. More than 240k APs have been harvested in these field experiments. The harvested data have been analyzed according to three main folds. The first fold focuses on finding WiFi statistics in each different area and to find the popularity of certain technologies, such as WiFi direct and MiFi portable devices. In the second fold, distribution models have been derived for each area utilizing maximum likelihood and different popular distribution functions, such as, Gamma, power, Pearson and normal. The third fold focuses on proposing SSIDtrack tracking algorithm that depends on WiFi SSIDs and malls maps to find paths of shoppers/walkers in such closed areas. This algorithm is easy to implement and it has many advantages for customers and for business owners. For example, this algorithm may be utilized to find the hot locations in malls for advertising and to rent stores. It can be utilized to find missing children in malls if they carry smartphones. The applications of this algorithm are limitless and it does not require complex computation of signal strength or GPS information. Our main contributions in this work are summarized as follow:

- A measurement study of WiFi networks in Amman, the capital city of Jordan. We have collected more than 240,000 WiFi access point in this measurement study. We are the first to make this harvesting study in one of the Middle East and North Africa (MENA) countries to show the popularity of WiFi in this part of the world.
- Developing power distribution models for WiFi networks in three areas. These models can be used in wireless health research, data distortion mitigation, and noisy signal countermeasures. Furthermore, the power distribution models have been estimated using various techniques. The estimated models demonstrate the popularity of the WiFi signal. Furthermore, they show an estimate of how many signals may surround a user in any location, whether it is a close area, an open area, or a hybrid area. We are, to the best of our knowledge, the first to try to estimate these models.
- Proposing and deploying SSIDtrack algorithm which utilizes time stamp and WiFi SSID to find the location of a user. The algorithm is easy to implement since it does not require the implementation of any new sensors or devices in the areas. The algorithm is implemented as a smartphone App. The App can be distributed among the customers. Subsequently, the harvested data from these applications can be leveraged for advertisement and statistical studies of the customers' behaviors. Furthermore, shop owners can use this information to locate the best location in a mall to lease a shop.

- The proposed algorithm tracks the patterns of shoppers and walkers by using the SSIDs of mall stores. To our knowledge, we are the first to use SSID in tracking, tracing, and localization. Python scripts were used to create the algorithm. This algorithm is used to track people in a restricted area in order to determine their walking path. It places no emphasis on localization accuracy because the algorithm used WiFi signals from APs located in mall shops without knowing the location of these devices.

The rest of this paper is organized as follows: the next section overviews the background of MAC address randomization and the related works that have been conducted in the areas of WiFi tracking and WiFi measurement studies. Section 3 describes SSIDtrack algorithm and its phases. Section 4 overviews the experiment and the analyzed results. Finally, we conclude this paper in Section 5.

2 Related Works

The related work section is divided mainly into three main parts. The first part is the background of the MAC address randomization. The second part summarizes the WiFi measurement studies that have been conducted and the final part overviews the indoor localization mechanism.

2.1 MAC Address Randomization

Paper MAC address randomization is a technique that has been proposed to tackle the security issue of tracking unique MAC addresses. In this algorithm the seventh bit of the first byte of the organization unique identifier (OUI) part of the MAC address is set. This bit is called the unique universal bit. Moreover, a random number is leveraged for the rest of the MAC address. The eighth bit of the first byte of the MAC address has been reserved for multicasting and unicasting. It has been reported [6] that some companies may purchase new OUI addresses from IEEE to utilize them for MAC address randomization. However, in our dataset we have recorded many MAC addresses OUIs that have not been listed in the IEEE OUI list.

2.2 WiFi Measurement Studies

Studying WiFi networks have proliferated in the past few years. Many measurement studies have been conducted to gain more insights of these networks and their abilities to drive the future of Internet sharing. In [8], a measurement study has been conducted for a campus network of a university. A smartphone has been utilized with WiFitrack App to harvest the information of all APs surrounding the users. Different statistics have been shown. Moreover, the harvested data has shown the WiFi networks are deployed heavily in campus networks. In [9], the authors conducted another WiFi measurement study of a dense campus network. The authors analyzed the setup phase of these network and the RSSI values. The authors reported that the connection time of almost 80% of the APs is less than 10 s. Moreover, they reported a low connection setup value. In [6], the authors conducted a WiFi measurement study to gain insights of MAC address randomization algorithm and its usage. Moreover, they have shown different attacking methods that may breach the MAC address randomization. These measurement studies defer from this study in four main points; first the area or scope of the measurement. We have conducted the measurement in three main areas. We called these areas open, close and hybrid area. Second, we aimed to generate power models of the received signals. Third, we have measured the popularity of MiFi devices and WiFi hotspots. Finally, we attempted to analyze the SSIDs of the APs to infer information that can be used for localization in closed or indoor areas.

2.3 WiFi Localization

Indoor localization has attracted researchers for many years. Many survey papers have been written to gather and compare different techniques and algorithms [10]. However, Angle of Arrival (AoA), Time of Flight (ToF), Return Time of Flight (RTOF), Received Signal Strength Indicator (RSSI) dominated in this area. Many algorithms have been written to utilize RSSI for indoor localization [11–13]. Nevertheless, these algorithms are complex and require massive signal filtration methods [14]. In [15], the author surveyed the limitation and issues that encounter WiFi measurement in inferring pedestrian behavior, such as, lack of coordination and limitation of RSSI. All of indoor localization algorithms attempted to enhance the localization error over the algorithm complexity. In [16], the authors utilized fuzzy logic Type-2 with Bluetooth Low Energy (BLE) to localize and track peoples in closed areas. The experiment utilized 6 points deployed in a room of approximately 4 m × 4 m size. The recorded accuracy has an error of less than 0.4 m. In [17], the authors utilized the BLE with fuzzy logic and a smartphone to locate people in a closed area. Different number of BLE AP has been deployed in a room and different number of algorithms has been compared. The author proposed a model to localize people according to the room size, number of bacons and their locations. An error of less than 0.5 m has been recording according to the room diminutions. [Tab. 1](#) summarizes different methods and algorithms proposed in the area of indoor localizations.

Table 1: Summary of Indoor localization proposed algorithms

Method	Equipment's	Environment	Error
[18]	BLE	Room (52.2 m × 12.5 m)	0.41 m
[17]	BLE	Rooms 5 m ² , 15 m ² and 25 m ²	from 0.1 to 0.5 m
SALMA [19]	UWB	Room (4 m × 6 m)	Less than 0.2 m
[20]	GPS, UWB, MARG	Mixed area	3.2 m
[21]	WiFi, PDR	780 m ² and 1200 m ²	3.5 m
[22]	[WiFi, PDR	university floor	Less than 2.4 m
[23]	BLE	Office (1200 m ²)	1.6 m
[24]	Ultra-wide	7.35 m × 5.41 m and 50 cm × 50 cm	7.43 cm
[25]	Wi-Fi	A floor in a collage	0.5 m
[26]	WiFi	Meeting room 9 m × 7.7 m	0.87 m

In this work, we proposed SSIDtrack algorithm to track the pedestrian in indoor area. However, we did not emphasize on the accuracy of the location as other algorithms. Our main contribution of this work is to find the track of pedestrian or people in hybrid areas without any deployment of new sensors or components. The proposed algorithm is a viable algorithm that requires a smartphone application and a map of the closed area to map the harvested data on the downloaded map. No new sensors are required; this is way it is very hard to find the error or the accuracy of the method. We attempted to find the path with a simple viable algorithm. Moreover, our algorithm utilizes the SSID of the APs rather than complex information, such as, AoA and RSSI. For the best of our knowledge, we are the first to utilize SSID for indoor localization.

3 SSIDTRACK Tracking Algorithm

SSIDtrack is a simple weighted tracking algorithm. The algorithm depends on SSID of APs and does not require signal strength computation or any three static locations for any trajectory calculations. The algorithm requires two inputs; mall maps and the WiFi APs scanned lists. The malls' maps are easy to get from their websites. The maps of the three malls have been downloaded and the shops names have been extracted from the maps. A graph has been constructed for all the shops in each floor in each mall. For example, four graphs have been constructed for a mall with four different floors. A python script has been written to crawl the website of each mall. However, for each mall, the script has to be modified since each mall's website is different from the others.

The second input of the algorithm is the AP lists that have been recorded from the WiFi tracking application that have been written for the first part of this work. Each time the App scans for APs, the BSSID, SSID, RSSI and the timestamp are recorded. The timestamp is an important component in this algorithm. This value will allow us to count the numbers of occurrence of any SSID in the harvested lists. The algorithm has two phases, initialization phase and the tracking phase. In the initialization phase, the weight of each node is initialized to zero and the timestamp array is initialized to zeros. Another data cleaning step takes place in this phase. The APs with default configurations and without names are deleted from the list. This may reduce the time required for the tracking algorithm. The last initialization step is to convert the names found in the APs list to the real shops name. To implement this, we wrote a python spelling correction script that attempts to match shops names with SSIDs from the list.

In the tracking phase, the APs list is scanned from up to down. If a shop name is found in the list, it is popped out and the weight of the node is increased by one. The time stamp is added to the timestamp list of the node. The scanning algorithm stops when the algorithm reaches the last SSID. The rest SSIDs in the list cannot be used since they may belong to hotspots or MiFi devices.

After this step, the weights of all nodes are averaged. The average is utilized as a threshold value. If the weight of any node is less than the average value, the weight is set to zero. This step is required to delete the occurrence of some nodes that we received their signals with small RSSI. Finally, the timestamp array for each node is cleaned. Alg.1 shows the pseudo code of the algorithm.

An example of the algorithm is shown in [Fig. 1](#). In the figure, one second has been utilized as the scanning time of the application. This time is 10 ms in the real application. However, to understand the example 3 s have been utilized. The averaging process occurs for the scanning processes in this example. The time is important to show which list occurs before the other. Moreover, the occurrence of an AP in the list in the same position belongs to its signal strength. In the three lists recorded in the left, four APs have been recorded. The three lists are merged in the list in the right list. The SSID APs' names in the three lists have been converted to the real shops names utilizing the names correction algorithm written in python. Subsequently, if an AP has occurred more than one time, its weight is incremented. The APs without names for example number four in the list; will be deleted from the final list. Finally, the APs are sorted according to their weight. This mean in the last 3 s, the customer is between shops 1 and 2. Moreover, he is moving to reach shop 3. The algorithm works in real time, the scanner will scan every 10 ms to obtain new lists and the track of the user will be shown according to the AP recorded weight.

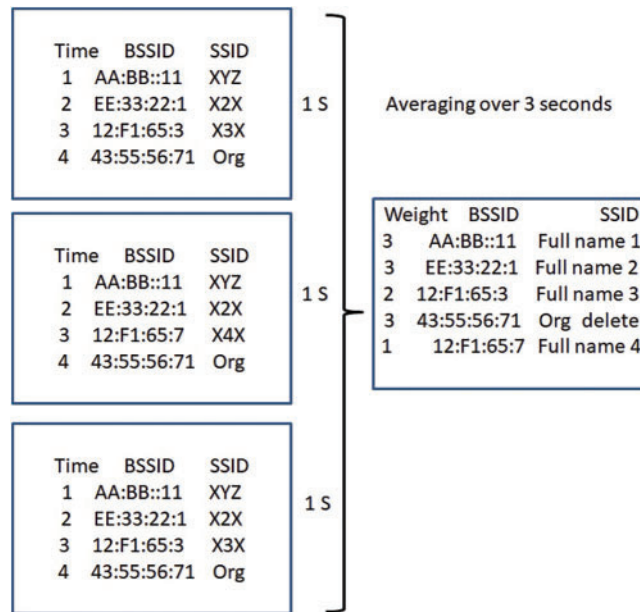


Figure 1: Example of the algorithm

Algorithm 1: SSIDTrack

Result: Location of the user according to the SSID of the surrounding APs

LAP: WiFi AP List, the list has timestamp, SSID, BSSID, RSSID;

LAP(T): Timestamp of the current list;

MAP(x): The map of the x floor of the area;

W: A list of weights for all APs;

TSL: List of timestamps when APs has been scanned

TH: Threshold value of scanning speed

Search-SP: Error spelling algorithm to find the best name for any SSID;

Search: Algorithm to search the map for the name of the shop;

While true do

LAP = SCAN();

For *n* **in** *LAP* **do**

if *TSL*[*n*] = 0 **or** *TSL*[*n*] ≥ *LAB*[*AP*, timestamp] + *TH* **then**

W[*n*] = *W*[*n*] + 1;

end

if *time* ≥ *TH* **then**

Temp = ∑ *W*/*n*

for *W* **do**

W[*i*] = *W*[*i*]/*temp*;

name = Search-SP(*AP*[*i*]);

if *W*[*i*] ≤ 1 **then**

delete(*AP*);

else

temp2 = Search(*map*, name)

(Continued)

Algorithm 1: Continued

```

        if  $W[i] \geq 1$  then
            Map(temp2) = high;
        end
        if  $W[i] = 1$  then
            Map(temp2) = mid;
        end
    end
end
end
end
end
end
end
end

```

4 Experiments

To collect WiFi data, an Android App has been written to harvest surrounding WiFi access points. The App scans for new APs every 1 s. We attempted to use this speed to collect different power levels of the surrounding APs. The App records a list of access points, time stamp and received signal strength in dbm. The recorded time stamp is required to count the number of APs signals at any moment. The App has been installed on Huawei p10 lite smartphone with Android version 8. The experiment took place over a week between August the first, 2019 and August the eighth, 2019. The App has been utilized to harvest the APs data in three different locations; closed, open areas and hybrid areas. In the open area, we selected a driving test in the neighborhood area. For the closed areas, three malls have been selected. Finally, for a hybrid area we selected the university. The harvested files have been filtered and analyzed utilizing Python 2.27 scripts. The following subsections describe the steps, data and the results of each area.

4.1 The Open Area

In this part, the written App has been installed on a Huawei p10 lite smartphone. The smartphone has been placed in a car that travels with a max speed of 20 Km/H. the car traveled in the streets of a neighborhood, named “Tareq”, in Amman, the capital city of Jordan. Approximately 20687 APs have been scanned and recorded in 3 days between the first to the fourth of August 2019. The statistics of the recorded data is shown in [Tab. 2](#). Five main observations or inferred information are drawn from this data.

Table 2: Harvested Data Statistics

Variable	Open area	Closed area	Hybrid
Number of Aps	20687	1326	191
Number of hidden SSIDs	447 (2.16%)	55 (4.1%)	4(2%)
Number of named SSIDs	8550 (41.33%)	795 (60.07%)	90(47.1%)
Number of MiFi devices	3346 (16.2%)	209 (15.7%)	34(17.8)
Number of hotspots	2371 (11.4%)	134 (10.1%)	38(19.89)%
Standstill APs	5111 (24.9%)	58 (4.4%)	5(2.6)%

(Continued)

Table 2: Continued

Variable	Open area	Closed area	Hybrid
Direct Wifi devices	41 (2%)	21 (1.6%)	4(2%)
Wireless printers	769 (3.8%)	32 (2.4%)	16(8.3)%
5 GHz devices	707 (3.4%)	40 (2.9%)	3(1.6)%
2.4 GHz devices	19980 (96.5%)	1286 (97.09%)	188(98.4%)

First, the popularity of portable WiFi AP, which can be divided into MiFi and hotspots are approximately 28% of the harvested default named APs. In other words, this number is more than 52% of the AP with the default configurations. We could not count this number in the named APs since the utilized names cover this property. This means that body area network (BAN) and personal area networks (PAN) can be connected to the Internet in an easy way in the future. Moreover, Internet of Everything (IoE) devices can be designed with short range network connections that can be harvested from gateway connected to WiFi access points. However, health concerns are emerged since the transmitter is very close to the human, for example inside his/her pocket.

Second, the popularity of wireless printers; these printers comprise about 3.8% of the harvested default named data. More than 90% of these printers are HP printers. The series and version also can be shown in the harvested data. In addition, the direct WiFi access technology comprises 2% of the harvested data. This technology is utilized to connect to WiFi TVs, stereos or even WiFi cameras. However, keeping the default configuration of this technology is a massive security threat. We hope that the manufactures of these devices technologies reduce this risk by disabling these technologies until configuration is set.

Third, some Internet service providers (ISPs) utilized a default name that reveals the type of the last mile connection. For example, we counted 429 FttH connections and another 353 4G connections. Such information may infer private information of the subscribers.

Fourth, smartphone brands popularity, hotspots with default names may infer the brand name of the smartphones or tablets. For example we found that Huawei smartphones are the most popular smartphones in the neighborhood. Moreover, we found that Android phones comprise about 85% of the smartphones.

Finally, the name of the ISPs; this information has been recorded smoothly in the stand still and MiFi devices. Such information may help ISP to find their popularity in different location to tailor their services to fit the subscribers in different locations. A simple drive test may show such information without the requirement of subscribers' surveys. It is worth mentioning that the harvested data shows that WiFi 2.4 GHz is the most popular technology with more than 96% compared to the 5 GHz band.

4.2 The Closed Area

The same smartphone utilized in the first part has been carried while shopping in three different malls; "Makkah", "Taj" and "Barakeh" malls in Amman, Jordan. The shopping experiment took place in two different days 5th August, 2019 and 6th August 2019. All floors of the malls have been covered. The harvested data have been cleaned and analyzed utilizing the same Python scripts. [Tab. 2](#) shows the statistics of the harvested data from all malls.

From [Tab. 3](#) we can observe that the named APs have been increased over the open area data. The malls shops attempt to change the SSID of their shops to their real names for advertisement purposes. Over than 60% of the harvested APs have names. Most of the remaining harvested APs are MiFi and hotspots as observed. These access points belong to the works in the malls or the shoppers. We have extracted the names of shops from the harvested data and wrote a Python script to match them with the real names of the shops that have been collected from the malls websites. The python script has been tuned for spilling mistakes and short names. [Tab. 3](#) shows the percentage of shops found with their SSID names in the data. [Tab. 3](#) also shows that shops names may be utilized as marks for the localization issues inside these closed massive areas. These percentages motivated us to propose our tracking algorithm passed on SSIDs.

Table 3: Named Shops percentages in the harvested data

Mall name	Percentages
Makkeh	90%
Taj	86%
Barakeh	92%

4.3 Campus Scenario

In the third scenario, Our App has been installed on a Huawei Mate smartphone and the phone took a tour in the campus of Al-Zaytoonah University of Jordan. All the buildings and all the streets have been covert except the building of the university presidency. [Tab. 2](#) shows the collected data.

Three main points can be observed. First, the number of mobile WiFi APs ‘MiFi or hotspots’ is massive with more than 71.2% of the default configuration APs of this category. Second, the percentage of named AP is approximately the same as the open area. This shows that most of the subscribers keep the default configuration and they changed only for offices or shops. Finally, the number of harvested APs is massive for a small Area, such as, the university. The following sections show the health concerns of such numbers.

4.4 Power Distribution Models

The Massive number of harvested WiFi AP in our dataset motivated the study of the received power distributions models. These models are important as a benchmark for health studies, WiFi noise and IoT devices. The harvested data has been divided into the above three scenarios and three different distribution models have been driven. Maximum Likelihood (MLH) analysis has been utilized to estimate the distribution models parameters.

Four main distribution models, namely, Normal, Lognormal, Pearson type 3 and Gamma, have been tested in each scenario and Kolmogorov Smirnov statistical test [18] have been utilized to select the best fit between the four models.

4.4.1 Open Area Distribution Model

After deploying the four main distribution models on the open area dataset and MLH utilized to fit their parameters, two distributions obtained a high p-value over 0.05 for Kolmogorov Smirnov statistical test; Pearson type III and Gamma Distribution. We will utilize Pearson type III since its

test result exceeds the Gamma distribution. Eq. (1) shows Pearson type III distribution PDF. We can observe that this distribution requires three main parameters to be fitted; standard deviation ‘ α ’, the ‘skew’ value of the distribution and ‘loc’ value. The values are required to calculate Eqs. (2)–(4). Finally, the values calculated from Eqs. (2)–(4) and utilized to extract the value of Pearson distribution.

$$y = \frac{\beta}{\gamma(\alpha)} (\beta(x - Z))^{\alpha-1} e^{-\beta(x-Z)} \quad (1)$$

where $\gamma(\alpha)$ is the Gamma Function and

$$\beta = \frac{2}{skew * \sigma} \quad (2)$$

$$\alpha = (\alpha * \beta)^2 \quad (3)$$

$$Z = \frac{loc - \alpha}{\beta} \quad (4)$$

The Values of these three values estimated utilizing MLH are (Skew = 1.2638, Loc = -75.67799, $\alpha = 6.97$) and the p-value for the Kolmogorov Smirnov statistical test is 0.32269. Fig. 2 shows the histogram of the open area harvested data and the fitted model.

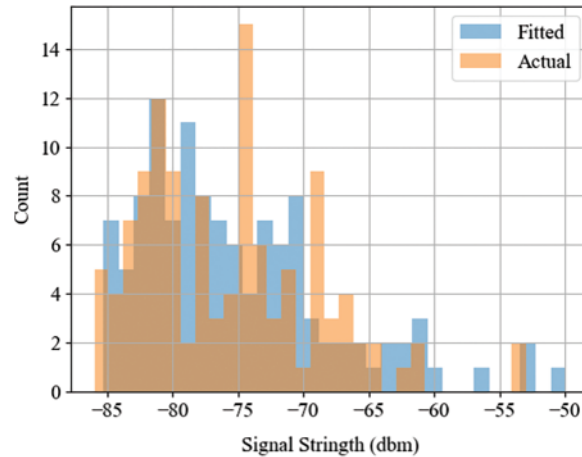


Figure 2: Histogram of the original data and fitted data

4.4.2 Closed Area Models

We wrote two different distribution models for only two malls, “Makah” and “Taj” since the number of APs harvested from these malls exceeds the third mall. Subsequently, we compared the two estimated models to show the similarities. For “Makkah” mall dataset, Normal distribution has dominated. Eq. (5) shows the PDF of this distribution.

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

The distribution has two main parameters to be fitted using MLH; μ and σ^2 . The values for these parameters are ($\mu = -63.7535$, $\sigma^2 = 13.37499$). The p-value of the statistical test is 0.1482. It worth

mentioning that the second distribution that obtained the second highest p-value is the Pearson type III distribution. Fig. 3 shows the histogram of the data and the fitted model. In the other hand, after fitting the dataset of “Taj” mall, Normal distribution also dominated. The values of its parameters are ($\mu = -67.29, \sigma^2 = 13.488968$) and the obtained p-value is (0.2885). We can observe the similarity of the fitted parameters between both models. This shows the Normal distribution can be utilized to model the received WiFi signal strength indoor. It also worth mentioning that the second distribution that obtained the second highest p-value is the Pearson type III distribution with p-value of (0.2874262). Fig. 4 shows the histogram of the data and the fitted model.

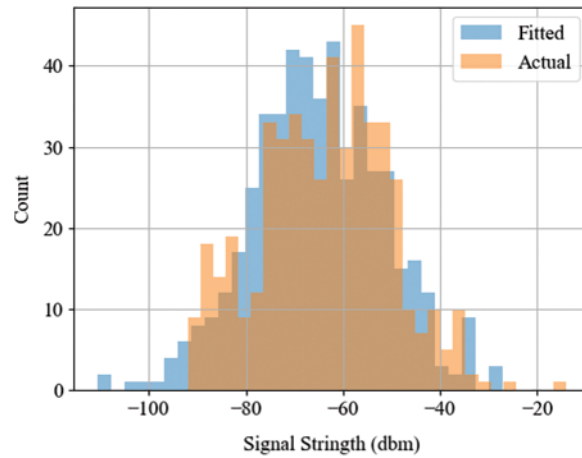


Figure 3: The histogram of “Makkah” mall dataset with the fitted model

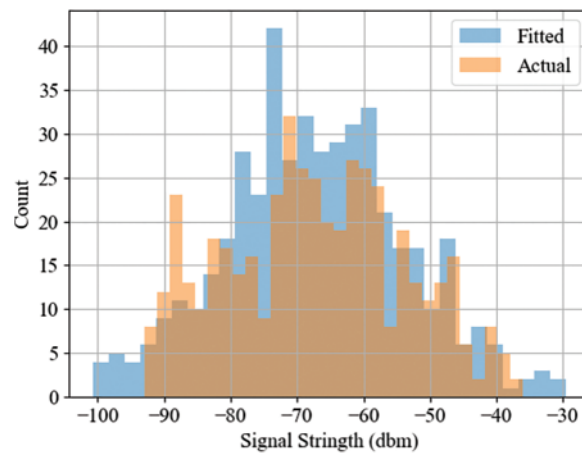


Figure 4: The histogram of taj mall dataset with the fitted model

4.4.3 Hybrid Area Models

The harvested dataset from the University has been utilized for the last model. Pearson type III recorded the highest p-value in Kolmogorov Smirnov statistical test. The parameters values are (1.1307948, $-73.3947198, 8.65325$) and the p-value is (0.29). It worth mentioning that, the second distribution model is the Gamma model with p-value of (0.29). Fig. 5 shows the histogram with the fitted model.

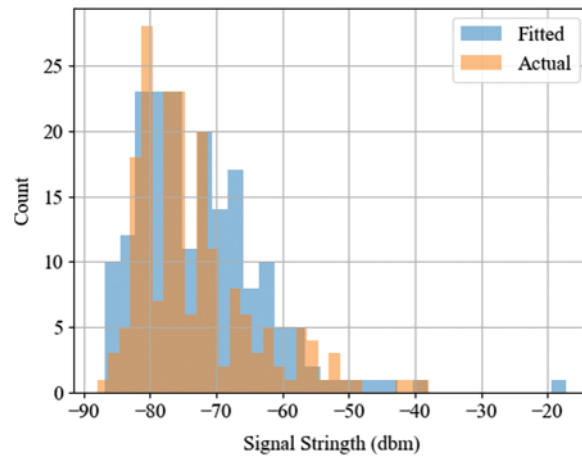


Figure 5: The histogram of university dataset with the fitted model

4.5 Number of APs Surrounding Users

This is was the reason to write our own App for this experiment. Each time the App scans for APs, it adds a time stamp. Using this time stamp we could count the number of access points surrounding the user at any given time. This point is important in two folds. First, popularity of WiFi and its ability to drive the new era of IoT devices. Second, the health concerns of the impact of received signals on our health. We divided as the data for three scenarios as above, open, close and hybrid. Fig. 6 shows the histogram of the number of received APs in the open area. We can observe that the number on average is 8. This means that at any given time in the urban area eight different APs can be received from any device carried by the user. Moreover, this means that our bodies absorb the power of eight different APs at any given time if we do not carry a hotspot or a MiFi device. This number increases in close areas as shown in Fig. 7 for the data harvested from the malls. The number of APs surrounding us in the mall at any given time on an average is 14. Finally, the hybrid area is shown in Fig. 8. This number is approximate 10 in the university. These numbers requires more investigation for health concerns. Do not forget that these numbers increase with the addition of more students since MiFi will increases. If all of this power is absorbed by our bodies, what about the aggregate value with smartphones signals, Bluetooth smartwatches and earphones?

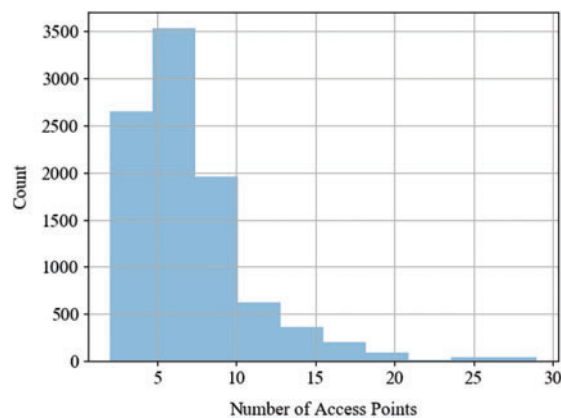


Figure 6: Number of APs around us in open area

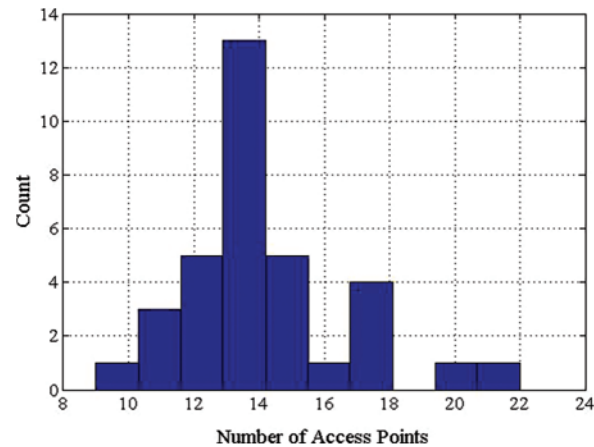


Figure 7: Number of APs around us in close area

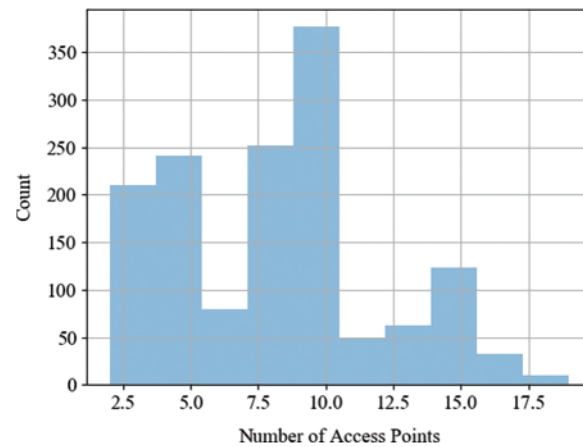


Figure 8: Number of APs around us in university

4.6 Study of MAC Addresses

To track devices utilizing WiFi technology, SSID or BSSID may be utilized. BSSID or the APs MAC have the issue of randomization [6]. This algorithm has been proposed and implemented to reduce the tracking risk for smartphone users. However, many devices have implemented such a technology. On the other hand, the hidden SSID has been implemented to reduce the security breach of WiFi APs. We have observed from the statistics that the hidden SSID technology is not popular in the dataset harvested in the three scenarios. However, what about the randomized MAC addresses? To study their impact, we have wrote a python script to extract the first part of the MAC address of the harvested APs. Subsequently, check the MAC address against the MAC address OUI list downloaded from IEEE. The list consists of 27118 companies and their OUI section. Subsequently, if the MAC address is found in the list, we record the company and country of origin to show the dominated companies and countries in the area of network interface cards (NIC) designing and manufacturing. Finally, if the MAC address not found in the list, we complimented the seventh bit of its OUI and we check against the list again. We have added all the MAC addresses harvested from all areas in one dataset. The dataset consists of 240k APs MAC addresses. The dataset has been harvested over three

months, from May to August 2019, from different areas in Amman city. [Tab. 4](#) shows the popular companies recorded from the harvested data. We can observe that Huawei owns more than 32% of all the WNIC manufactured and run in Amman. We also observed that Cisco has only 3% of the running WNICs.

Table 4: Companies percentages in the dataset that occurred more than 2%

Mall name	Percentages
Huawei 32.23	Huawei 32.23
Tp-link 8.58	Tp-link 8.58
Sagemcom 8.57	Sagemcom 8.57
Samsung 3.8	Samsung 3.8
Cisco 3.0	Cisco 3.0
ZTE 2.93	ZTE 2.93
Gemtek 2.86	Gemtek 2.86
Aruba 2.49	Aruba 2.49
D-Link 2.08	D-Link 2.08
Ruckus 2.08	Ruckus 2.08

From the harvested list, it has been found that, 9390 MAC addresses, 3.9%, of the harvested MAC addresses leverage the randomization technology. Moreover, 165 companies have been recorded in the dataset. 32 companies '18.9%' have recorded in the dataset with only the randomization technology. In other words, these 32 companies utilized only the MAC randomization. However, they count for less than 5% of the recorded MACs. A full list of these companies can be downloaded from our website (<http://bayadata.net>). Moreover, 0.9% '2160' MAC addresses are utilized and they have not been recorded for any company [Tab. 5](#) shows these MAC addresses.

Table 5: Unknown MACs

MAC (OUI) percentage	MAC (OUI) percentage
00:B0:C6	0.21%
04:8D:39	0.04%
00:4F:81	0.02%
34:0A:91	0.02%

[Fig. 9](#) shows the popular countries of manufacturing WNICs in the dataset. We can observe China has more than 51% of the WNICs recorded in the dataset and US has less than 13%.

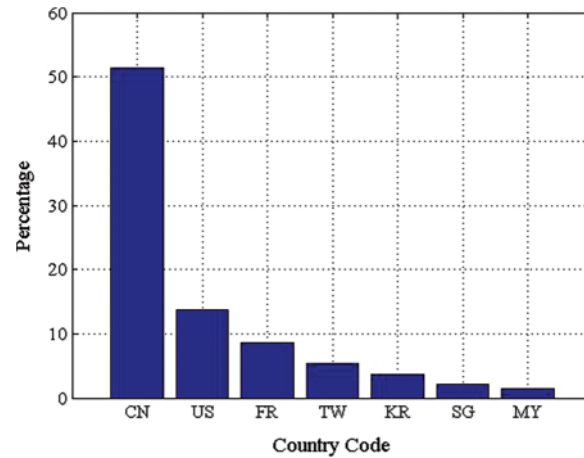


Figure 9: WINCs Countries of origins

4.7 SSIDtrack Algorithm

To study the accuracy of the tracking algorithm, we walked in the two malls in each floor we recorded on paper the name of shops that we came cross and the name of the shops that we entered. Moreover, we have marked our walking on the map of each floor. Subsequently, we recorded the harvested lists of APs in two different files for each mall. We fed the recorded lists to the SSIDtrack algorithm and we have colored the shops with three colors according to their appearance in the output weighted file of the algorithm. Fig. 10 shows the output map of one floor of “Taj” mall. The red color has the highest weight, the green has the second highest weights and finally the orange color has the smallest weight. This color will be deleted from the output since its weight values are very small and can be neglected. This color is the color of the AP recorded with weak RSSI value. Their weight is too small to be considered in the walking path. The black arrow shows our real path in the mall. We can observe that the path has been recorded from our algorithm. Moreover, the timestamp in the algorithm shows when we have passed beside certain shops. In addition, if we add more colors to the figure we can easily find which shops we have entered and which we only passed by from their weights. Fig. 11 shows the accuracy of the track that has been drawn for each floor in the first mall. The accuracy has been recorded according the appearance of all SSIDs of the shops in the track and their weights. Fig. 12 shows the same results for Mecca mall. To obtain the tracking or localization accuracy in cm, the location of each WiFi access point has to be recorded or obtained from the shops in the mall. If the location of these access points is recorded, it is easy to calculate the accuracy in cm. However, it was hard to obtain these data from the shops in one hand and to measure our real location to these points on the other hands. Nevertheless, we focused on tracking the movement of the customers which is more important to the shop owners since it gain insights of where to locate any advertisements or where to lease a new shop or booth. In addition, the corridors in the malls are narrowed in width ‘less than 3 meters’ this means that the error of tracking a pedestrian will be less than 3 m.

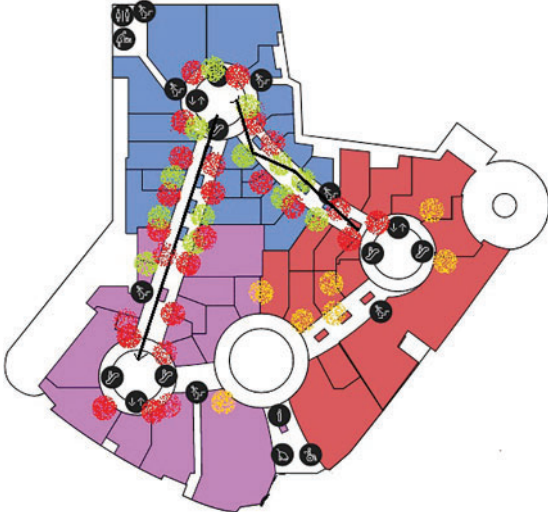


Figure 10: The recorded track and SSIDtrack output map

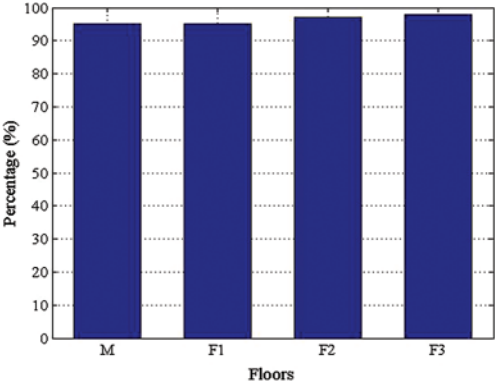


Figure 11: “Taj” mall tracks accuracy

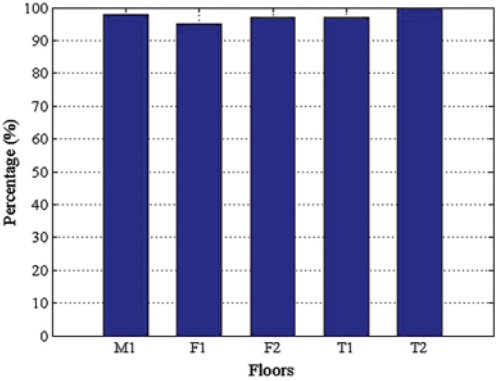


Figure 12: The mecca mall tracks accuracy

5 Conclusion

In this work, APs have been utilized as passive sensors for three main contributions. First, power models of the APs in three different scenarios have been driven. These models could be the foundation of different WiFi studies in the area of health, noise and interference. We have shown that Pearson III dominated in open areas and in hybrid campus areas. However, normal distribution models the power distribution in closed areas, such as malls. Second, an algorithm has been proposed, named, SSIDtrack to track pedestrian and shoppers' activities in malls. The algorithm is simple and easy to implement. Our experimental results have shown that the accuracy of the algorithm exceeded 95% in two different malls. This algorithm may be utilized for advertising purposes since it shows the most popular locations in malls. Moreover, it can be used to track children in very closed areas. The algorithm requires only the SSIDs of the surrounding APs without calculating AoA or RSSI. Finally, we have shown statistics of APs massive numbers in Amman city and we have shown the popularity of MiFi devices and hotspot technologies. These statistics shows that the popularity of WiFi will play an important role in the future of Internet of everything (IoE).

Funding Statement: This work is funded by Al-Zaytoonah University of Jordan under project name "miniature distributed architecture for massive data processing" with the grant number 15/12/2019-2020.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. Masoud, Y. Jaradat, A. Manasrah and I. Jannoud, "Sensors of smart devices in the internet of everything (IoE) era: Big opportunities and massive doubts," *Journal of Sensors*, vol. 2019, no. 1, pp. 122, 2019.
- [2] C. Sage and D. O. Carpenter, "Public health implications of wireless technologies," *Pathophysiology*, vol. 16, no. 2-3, pp. 233-246, 2009.
- [3] K. Foster and J. Moulder, "Wi-Fi and health: Review of current status of research," *Health Physics*, vol. 105, no. 6, pp. 561-575, 2013.
- [4] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown *et al.*, "A study of MAC address randomization in mobile devices and when it fails," *Proceedings on Privacy Enhancing Technologies*, vol. 2017, no. 4, pp. 365-383, 2017.
- [5] M. Cunche, "I know your MAC address: Targeted tracking of individual using Wi-Fi," *Journal of Computer Virology and Hacking Techniques*, vol. 10, no. 4, pp. 219-227, 2014.
- [6] M. Vanhoef, C. Matte, M. Cunche, L. S. Cardoso and F. Piessens, "Why MAC address randomization is not enough: An analysis of Wi-Fi network discovery mechanisms," in *Proc. of the 11th ACM on Asia Conf. on Computer and Communications Security*, Google, USA, pp. 413-424, 2016.
- [7] S. Xie, X. Zhu and J. Shen, "Method and system for automatically adapting to Wi-Fi network with hidden SSID," US Patent 10,117,169, 2018.
- [8] C. Zhang, X. Hei, Y. Fan and L. Xiao, "Dissecting campus WiFi connections in an empirical view," *Journal of Computers*, vol. 30, no. 1, pp. 64-74, 2019.
- [9] C. Zhang, X. Hei and B. Bensaou, "A measurement study of campus WiFi networks using WiFiTracer," in *Cyber-Physical Systems: Architecture, Security and Application*, Switzerland: Springer, pp. 19-42, 2019.
- [10] F. Zafari, A. Gkelias and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2568-2599, 2019.
- [11] Z. Yang, Z. Zhou and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1-32, 2013.

- [12] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach *et al.*, “Practical robust localization over large-scale 802.11 wireless networks,” in *Proc. of the 10th Annual Int. Conf. on Mobile Computing and Networking*, Philadelphia, PA, USA, pp. 70–84, 2004.
- [13] P. Kumar, L. Reddy and S. Varma, “Distance measurement and error estimation scheme for RSSI based localization in wireless sensor networks,” in *Fifth Int. Conf. on Wireless Communication and Sensor Networks (WCSN)*, Allahabad, India: IEEE, pp. 1–4, 2009.
- [14] J. Xiao, K. Wu, Y. Yi, L. Wang and L. M. Ni, “Pilot: Passive device-free indoor localization using channel state information,” in *IEEE 33rd Int. Conf. on Distributed Computing Systems*, Philadelphia, PA, USA: IEEE, pp. 236–245, 2013.
- [15] A. Petre, C. Chilipirea, M. Baratchi, C. Dobre and M. van Steen, “WiFi tracking of pedestrian behavior,” in *Smart Sensors Networks*, Amsterdam, Netherland: Elsevier, vol. 2017, no. 1, pp. 309–337, 2017.
- [16] X. Dang, X. Si, Z. Hao and Y. Huang, “A novel passive indoor localization method by fusion CSI amplitude and phase information,” *Sensors*, vol. 19, no. 4, pp. 875, 2019.
- [17] F. Orujov, R. Maskeliunas, R. Damasevicius, W. Wei and Y. Li, “Smartphone based intelligent indoor positioning using fuzzy logic,” *Future Generation Computer Systems*, vol. 89, no. 1, pp. 335–348, 2018.
- [18] B. Al-Madani, F. Orujov, R. Maskeliunas, R. Damasevicius and A. Venckauskas, “Fuzzy logic type-2 based wireless indoor localization system for navigation of visually impaired people in buildings,” *Sensors*, vol. 19, no. 9, pp. 2114, 2019.
- [19] B. Groswindhager, M. Rath, J. Kulmer, M. S. Bakr, C. A. Boano, K. Witrissal *et al.*, “Salma: UWB-based single-anchor localization system using multipath assistance,” in *Proc. of the 16th ACM Conf. on Embedded Networked Sensor Systems*, New York, NY, United States, pp. 132–144, 2018.
- [20] K. Zhang, C. Shen, Q. Zhou, H. Wang, Q. Gao *et al.*, “A combined GPS UWB and MARG locationing algorithm for indoor and outdoor mixed scenario,” *Cluster Computing*, vol. 22, no. 3, pp. 5965–5974, 2019.
- [21] W. Xu, L. Liu, S. Zlatanova, W. Penard and Q. Xiong, “A pedestrian tracking algorithm using grid-based indoor model,” *Automation in Construction*, vol. 92, no. 2, pp. 173–187, 2018.
- [22] K. Nguyen-Huu, K. Lee and S. W. Lee, “An indoor positioning system using pedestrian dead reckoning with WiFi and map-matching aided,” in *2017 Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN)*, Sapporo, Japan: IEEE, pp. 1–8, 2017.
- [23] M. Patel, A. Girgensohn and J. Biehl, “Fusing map information with a probabilistic sensor model for indoor localization using RF beacons,” in *2018 Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN)*, Nantes, France: IEEE, pp. 1–8, 2018.
- [24] T. Arsan, “Accurate indoor positioning with ultra-wide band sensors,” in *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 2, pp. 22, 2020.
- [25] M. Elbes, E. Almaita, T. Alrawashdeh, T. Kanan and S. AlZu’bi, “An indoor localization approach based on deep learning for indoor location-based services,” in *2019 IEEE Jordan Int. Joint Conf. on Electrical Engineering and Information Technology (JEEIT)*, Amman, Jordan: IEEE, pp. 437–441, 2019.
- [26] Z. Tian, W. Yang, Y. Jin, L. Xie and Z. Huang, “MFPL: Multi-frequency phase difference combination based device-free localization,” *Computers, Materials & Continua (CMC)*, vol. 62, no. 2, pp. 861–876, 2020.