5-2021

Time of Flight and Fingerprinting Based Methods for Wireless Rogue Device Detection

Daniel Chege
Montclair State University

Abstract

Existing network detection techniques rely on SSIDs, network patterns or MAC addresses of genuine wireless devices to identify malicious attacks on the network. However, these device characteristics can be manipulated posing a security threat to information integrity, lowering detection accuracy, and weakening device protection. This research study focuses on empirical analysis to elaborate the relationship between received signal strength (RSSI) and distance; investigates methods to detect rogue devices and access points on Wi-Fi networks using network traffic analysis and fingerprint identification methods. In this paper, we conducted three experiments to evaluate the performance of RSSI and clock skews as features to detect rogue devices for indoor and outdoor locations. Results from the experiments suggest different devices connected to the same access point can be detected ($p < 0.05$) using RSSI values. However, the magnitude of the difference was not consistent as devices were placed further from the same access point. Therefore, an optimal distance for maximizing the detection rate requires further examination. The random forest classifier provided the best performance with a mean accuracy of 79% across all distances. Our experiment on clock skew shows improved accuracy in using beacon timestamps to detect rogue APs on the network.

Keywords*:* DDoS (Distributed Denial of Service), Traffic detection, Fingerprinting, Evil Twin attack, RSSI (Received Signal Strength Indicator), Clock skew

MONTCLAIR STATE UNIVERSITY

Time of Flight and Fingerprinting Based Methods
for Wireless Rogue Device Detection

by

Daniel Chege

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

May 2021

College of Science and Mathematics

Department of Computer Science

Thesis Committee.

[redacted]

Dr. Christopher Leberknight
Thesis Sponsor

[redacted]

Dr. Boxiang Dong
Committee Member

[redacted]

Dr. Bharath Samanthula
Committee Member

TIME OF FLIGHT AND FINGERPRINTING BASED METHODS FOR WIRELESS ROGUE DEVICE DETECTION

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

by

Daniel Chege

Montclair State University

Montclair, NJ

2021

Contents

List of Figures

List of Tables

Chapter One

Introduction

The continuous evolution of wireless networks technology and ubiquitous use of mobile devices has fueled the spread of malicious activities on the internet. The adoption of wireless sensor networks has progressively become an enticing focus for vindictive assaults. Wi-Fi technology evolution and deployment as a low-cost infrastructure provide great opportunity for indoor localization (Sun, 2014).Wi-Fi localization systems incorporates 802.11 protocol for position accuracy and plays a pivotal role in determining receiver location (Zhang, 2019). Besides, most wireless devices rely on these positioning features for optimum performance and accuracy. Wi-Fi localization uses 2.4GHz and exploits signal path loss propagation due to Wi-Fi signal variations to determine how close the receiver is to a certain AP. Continuous radio waves attenuation with reference to inverse-square law indicates the distance can be estimated based on the transmitted and received signal strengths relationship

As stipulated in (Chen Y. T., 2007), wireless network openness has propelled cybercrimes such as IP spoofing, identity theft and DDoS attacks. Additionally, wireless localization over the years has attracted substantial research exertion due to the increased ease to deploy and interoperability of wireless fingerprinting localization (Yiu, 2017). IP spoofing attacks have continuously posed an incurable threat to the internet as recent attacks on most popular network infrastructure and websites prints the damaging effects of these attacks (Duan, 2008). Besides, IP spoofing still is a contributive ingredient to identity theft and makes it harder to isolate network attacks from legitimate traffic.

This survey elucidates recent rising advancements with the concentration towards IP spoofing and its effect to information security, privacy, and identity theft. This paper gives an in-depth analysis on existing literature in mitigating the effect on Infrastructure based DDoS attacks. Continuous adoption of wireless sensors has increased indoor localization feasibility in Wireless Local Area Network (WLAN) (Feng, 2010). On the contrary, spoofing attacks compromise the legitimacy of users on the network and pose a serious threat by masquerading as legit nodes on the network. These advanced characteristics affect information integrity as well as cause havoc on network systems by injecting traffic attacks such as evil twin access point attacks.

Although traditional approaches proposed in (Chen Y. T., 2007) to address spoofing attacks, cryptographic authentication has proven to offer additional infrastructural overhead and computational power linked with cumbersome key management issues. Wireless localization requires low-cost hardware and maintenance through dependency on the existing network infrastructure such as IEEE 802.11 protocol. Wireless fingerprinting relies on the RSS values for effective positioning on the WLAN. Moreover, these sensors on the network use node distance between each other to determine their actual position (Kaemarungsi, 2004).

In this study, machine learning algorithms were used for classification to determine centroid distance, evaluate classification accuracy, and determine relationship between the actual and estimated distance of the nodes and access points in the signal space. Additionally, this research study focuses on RSSI as signal input to detect rogue clients on the network and evaluates beacon timestamps performance in detecting rogue access points on the signal space. This paper is organized as follows; Chapter 2 provides in-depth analysis on previous research work in detecting rogue Wi-Fi clients on the network and rogue APs on the signal space. In Chapter 3, we describe machine learning techniques used in classification performance as well as methods used

to detect and identify rogue Wi-Fi clients with reference to RSSI and clock skew. Chapter 4 examines and interprets experiment results through determining RSSI reliability and clock skew effect in detecting rogue clients and APs. Lastly, Chapter 5 gives in-depth analysis on the research study done and proposes future exploration of the research study.

Chapter Two

Related work

The rapid rise and adoption of IoT technologies has precipitated a large volume of research to explore DDoS attacks, network identity theft, and their effect in the smart home environment. The rising number of interconnected devices in the IT sector has led to unprecedented increase in digital disruption. Besides, linked IoT devices have largely contributed to the widespread DDoS attacks due to low security features integration. Furthermore, vast user information stored on the cloud has powered the increase in DDoS attacks (Nawir, 2016). Likewise, spoofing attacks can cause a variety of security breaches such as rogue access point (AP) masquerading as legit access point. While there is a significant amount of research contributions with respect to the smart IoT environment, the ensuing literature review will focus on prior studies that are most closely related to our work.

The deployment of cryptography as native security approach to cope with the alarming rates of DDoS attacks in IoT devices has yielded partial effectiveness. Besides, cryptographic schemes are predisposed to node compromise and due to its infrastructural and computational overhead; its partially functional (Chen Y. T., 2007). Moreover, wireless devices have limited power thus makes it impossible to deploy authentication and key management. These predicaments have proven to require additional human resource to manage the devices on the network layer.

(Jana, 2009) proposed two different methods for estimating APs clock skews. First, they used linear programming approach to estimate clock skew deviation by calculating difference in the upper-bound time offsets and beacon arrival time at the fingerprinting node. Secondly, the other method was based on identifying a line with the least square distance from all time offsets. Furthermore, (Jana, 2009) proposed a more heuristic approach to differentiate frames sent by fake APs and Real APs by evaluating beacon timestamps and frames transfer rate.

(Zander, 2008) hypothesized synchronized sampling that reduces quantization error and improves clock skew accuracy by up to two magnitudes on low-resolution timestamps and one magnitude on high-resolution timestamps respectively. As indicated in (Krishna, 2009), calculating relative clock drifts can be achieved through examining two-way beacon frame TOA location tracking.

Numerous research studies have conflicting arguments whether RSSI is a good Wi-Fi localization candidate or not. (Sadowski, 2018) suggests indoor localization is hindered by several factors such as noise, environment and manufacturers' hardware capabilities thus affecting position accuracy. Furthermore, (Sadowski, 2018) highlights Indoor Wi-Fi localization offers better discriminative characteristics in node fingerprinting since it requires low cost to operate. Alternately, (Navarro, 2010) emphasizes RSSI fingerprinting is not a good Wi-Fi localization candidate. The objective of the research study is to investigate RSSI reliability in device fingerprinting and explore how clock skew can be used in detecting rogue access points.

IP Spoofing and Mac Filtering

(Li, 2006) proposed light-weight MAC (medium access control) security feature through forge-resistant analyzing relationships from the incoming packets on the spoofed networks. This approach introduces security within the MAC address as a fingerprint to identify multiple fake devices using the same MAC address. As discussed in (Park, 2001), route-based packet filtering was used to address proactive spoofed IP packets from reaching to their destination as well as to identify reactive spoofed IP traceback flows. Moreover, DPF (route-based distributed packet filtering) uses routing information to determine if the packet has a genuine source or destination address. (Cota-Ruiz, 2013) proposed WSN localization technique which estimates sensor node coordinates based on local spatial constraints by providing constant updates with reference to adjacent nodes within the communication spectrum. (Nuo, 2012) proposed a cluster based WSN localization algorithm to reduce measurement errors on multi-hop nodes and aimed at improving nodes localization accuracy.

Although (Park, 2001)proposed route-based packet filtering, routers can be flawed with rogue IP addresses as well as fake MAC address resulting to security breach. Adversaries can bypass the strict rule set on the routers through masquerade and cause havoc on the entire network infrastructure. Alternatively, these strict rules can discard legit frames or devices as the maximum threshold number of users is reached resulting to DDoS attack. Furthermore, this paper proposes a method for detecting spoofing attacks using RSSI (Received Strength Signal Index) as a Wi-Fi localization technique.

Wi-Fi Localization Techniques

Fingerprinting and Traffic Detection – RSSI localization technique relies on the signal strength to create a road map from various nodes readings from the network environment.

Basically, fingerprint maps are reference points at predetermined points coupled with various signal strength. Fingerprint mapping includes all measurements from different positions and their corresponding received signal strength (Pei, 2017). Moreover, clock skew can be used for device fingerprinting through monitoring and analyzing beacon/probe frames (Jana, 2009). Besides, beacon timestamps have substantial advantages in identifying rogue APs on the signal space. Time synchronization function (TSF) has a higher transfer rate (10 to 100 frames per second) considering TCP clock synchronization latency and provides accurate beacon timestamp compared to TCP timestamp delay (Jana, 2009).

Signal based – Wireless localization systems takes signal strength measures form different access points to determine the connected device distance from all access points.

Signal based localization makes use of signal power level in reversed approach for AP to identify multiple signals emitted by multiple devices on the network (Chen Y. &., 2002).

Angle of arrival (AoA)- This technique uses angle measurements of the RF signal from the device. Besides, AoA triangulation doesn't incorporate distance to signal but it also relies on angles to determine the position of an object. Furthermore, AoA uses signal strength and difference of signal arrival to determine the best angle of arrival (Wielandt, 2017). For effective accuracy, two or beacons are required for location estimation. Although AoA technique seems effective, there are limitations since AoA requires additional antennas resulting to high implementation cost (Farid, 2013).

Time of Flight (TOF) – measures the time taken for a signal to travel from the object to the reference point. TOF deploys time synchronization between when the signal was sent by the device and the exact time the signal is received at the reference point and also incorporating signal speed (Anjum, 2020). Although, TOF based systems are widely replacing signal to noise ratio

(SNR) based systems; they have drawbacks in implementation since they have inherent challenges in obtaining accurate measurements. Additionally, 802.11 protocol requires single frequency band to operate and TDOA requires better time synchronization for better accuracy.

Even though, RSSI localization is widely explored to determine the node positioning on the network, there exist lots of unsolved research problems. For instance, signal strength errors are a common phenomenon ranging from unknown distance from the access point (Hyo, 2009). Besides, most localization systems depend on wireless device transmission power with reference to signal strength to determine node positioning. Since the propagation of signal is affected by reflection, diffraction and scattering, there's a substantial impact on measurement accuracy. As indicated in (Bekcibasi, 2014) RSSI is affected by some factors such as environmental and device errors that result to localization error and measurement inaccuracy.

Problem Statement

There have been many solutions proposed to improve existing limitations with common standards across various layers in the IoT stack. Yet there remain challenges with attacks on different protocols within the network layer. The focus of this research is to propose solutions for a subset of network attack vectors or techniques used to create a distributed denial of service (DDOS). Specifically, this study examines how to prevent network nodes from gaining unauthorized access to an IoT network by exploiting IP spoofing or MAC address hijacking and SSID manipulation. There are two approaches to identify these types of attacks such as traffic detection and fingerprint identification (Tang, 2017).

Research Questions

Traffic Detection. Examines differences in traffic patterns. Received Signal Strength (RSS) will be used to identify the location of nodes in the wireless network.

First, this research study aims to identify the optimal distance between valid and rogue Wi-Fi clients connected to the same access point. While prior research suggests received signal strength (RSS) can be used to detect rogue devices (Chen, 2007); the distance between Wi-Fi clients and access points can significantly impact detection accuracy. This research study assumes rogue Wi-Fi clients have hijacked mac addresses from legitimate nodes.

RQ1: What is the optimal distance to differentiate between rogue and valid devices on a Wi-Fi network using received signal strengths (RSS)?

The null hypothesis is:

H1: There is no statistical difference between devices that are 5 meters from the access point compared to devices that are 10 meters from the same access point.

Secondly, to perform near real-time detection of rogue Wi-Fi clients, computational overhead must be minimized. One way this can be achieved is by investigating the tradeoff between the detection accuracy and the amount of RSS data required. The more data required to detect the rogue clients the more computational resources required to perform the analysis. If calculations are to be done on IoT devices the limitation with computational resources is further amplified.

RQ2: What is the trade-off between the amount of RSS data and detection accuracy?

The null hypothesis is:

H2: There is no statistical difference in detection accuracy for same devices with 50% fewer data points connected to the same access point.

RQ2 will be evaluated using different machine learning algorithms to identify the impact on detection accuracy by decreasing the sample size of the data. The complete sample size collected for each individual device is 297 data points

Fingerprint identification. There will be a significant difference between temporal network characteristics of rogue devices compared to authorized devices. For example, the duration field in a MAC address can be used to identify the chipset for a device (Cache J., 2006). In instances where hardware of rogue devices matches valid nodes on the network, the nodes' location or proximity within the network can be used to detect the rogue device.

While RQ1 and RQ2 are directed toward helping network operators to detect rouge devices connected to their network, the next set of research questions are aimed at addressing client-side vulnerabilities. For example, if a Wi-Fi user is connected to a rogue access point no safeguards exist that can alert the user. Therefore, this part of the research aims to detect rogue access points also known as "evil twins". Furthermore, this research hypothesizes that at proximity RSS values between devices may not provide enough variance to detect rogue devices. For example, an attacker may place a rogue access point near a valid access point to mask any detectable differences in received signal strength. In such situations, alternatives discrimination characteristics must be used. One characteristic that has shown to aid in the detection of rogue devices is clock skew. As discussed in (Arackaparambil, 2010) the clock skew is the temporal difference in the hardware clocks between the access points and the Wi-Fi client.

Thirdly, this research study seeks to investigate if clock skews can provide enough discrimination quality to detect rogue devices (Wi-Fi clients and access points).

RQ3: Do clock skews provide higher discrimination to detect rogue Wi-Fi clients compared to received signal strength? The null hypothesis is:

H3: There is no statistical difference in clock skews between rogue and valid Wi-Fi clients.

RQ4: Does distance impact the discrimination quality to detect rogue access points? The null hypothesis is:

H4: There is no statistical difference in clock skews between rogue and valid access points.

For this research its assumed that IoT Wi-Fi clients will have the capability to compute, maintain and compare historical clock skew data with new data from the access point. Since IoT devices have limited processing power understanding the amount of data required to accurately detect rouge access points is a critical component for designing security controls.

Chapter Three

Methodology

This section describes the experimental environment along with the hardware components that were used for RSS fingerprinting. We conducted practical experiment to investigate whether RSS can be used as a candidate for Wireless Sensor Network (WSN) localization. In our localization model, we used Weka to process the measurements taken from the experiment. As discussed in (Adewumi, 2013), RSSI mean can be used to determine the position of nodes in the signal space at given distance. As indicated in (Kaur, 2015), machine learning classification techniques can be used for prediction purposes and performance evaluation.

Hardware used in the experiments consist of a Verizon FIOS Actiontec MI424WR router that served at the valid access point; Dell M3800 notebook with an Intel® Dual Band Wireless-AC 7260 as our valid device and CanaKit Raspberry Pi 4 as our rogue device with a constant 5 meters incremental distance from 5 meters to 20 meters from the access point. We collected 297 RSS readings from both wireless nodes with varying distance considered. Based on a statistical power analysis in Figure 3.1, 297 samples are required to detect a large effect size.

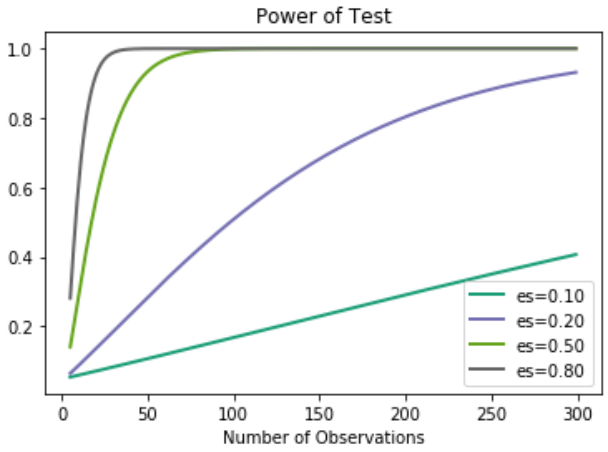
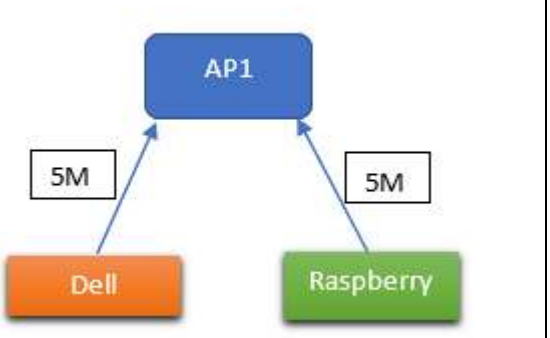


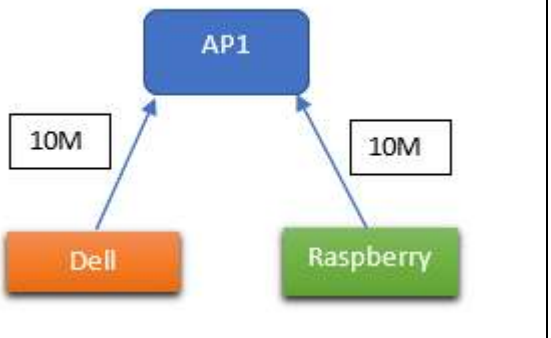
Figure 3.1 Power analysis to determine sample size to detect large effect

Experiment I – Rogue Wi-Fi Clients (RQ1: Distance)

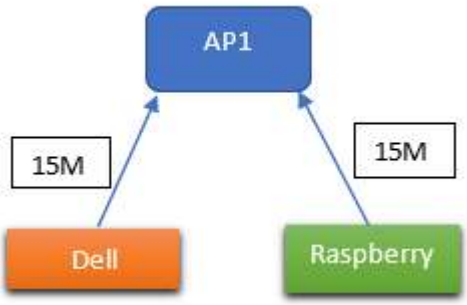
In this experiment, we deployed two wireless nodes: Dell as the valid node and Raspberry Pi4 as the rogue node on the signal space and Verizon FIOS Actiontec as our AP (access point). The two nodes were positioned 5M apart from the AP and with a 5m incremental distance from the AP. Additionally, 297 RSS readings were recorded for analysis purposes and evaluation.



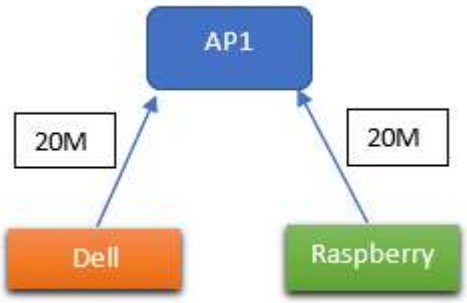
(a) Dell and Raspberry Pi4 @ AP1



(b) Dell and Raspberry Pi4 @ AP1



(c) Dell and Raspberry Pi4 @ AP1



(d) Dell and Raspberry Pi4 @ AP1

Figure 3.2. Experimental setup

A Two-Sample Mean T-Test is used to test hypothesis, H1.

Experiment II – Amount of Data

Next, the amount of data required is examined. Since data collection can be a time-consuming process that can delay the detection of rogue devices, RQ2 investigates detection accuracy with 50% of the original data. The original dataset contains 594 data points for each distance (i.e., 297 data points were collected for each device (Dell and Pi) at 5, 10, 15 and 20 meters from the access point) 148 data points were removed for each device at each distance to create the 50% reduction. The modified dataset contains 298 data points. A total of 148 data points was used for each device. Minimizing the amount of data collected will minimize the amount of time to perform near-real time analysis. The optimal algorithm identified in experiment I will also be used in this experiment to examine classification performance with reduced data.

Experiment III – Clock Skew

Clock skews are the inherent tiny drifts in the clocks of hardware devices due to variations in the manufacturing process (Arackaparambil, 2010). (Kasera, 2008) showed that the clock skews of wireless devices remain consistent over time. In their research that observed that, due to the essentially zero latency and the availability of a high frequency stream of high precision beacon

timestamps, the process of measuring clock skews became more accurate and effective in wireless networks.

The objective of this experiment is to empirically validate prior research by investigating the accuracy of using clock skews to discriminate between valid and rogue access points. To accomplish this objective, two timestamps are collected: (1) the beacon timestamp, which is the amount of time, in microseconds, that an access point has been active and (2) the packet arrival time, which is the time the operating system receives a packet and will rely on the kernel to give it a valid timestamp. The kernel will get the timestamp from either the network interface driver or the networking stack. The packet arrival time is supplied as seconds since January 1, 1970, 00:00:00 UTC (also known as UNIX time or Epoch time). Timestamps will be collected from access points at two distinct time periods. Raspberry Pi's will serve as the Wi-Fi client which receives access point broadcasts of beacon frames. Linear regression and statistical tests will be conducted to examine statistical significance and effect size.

Chapter Four

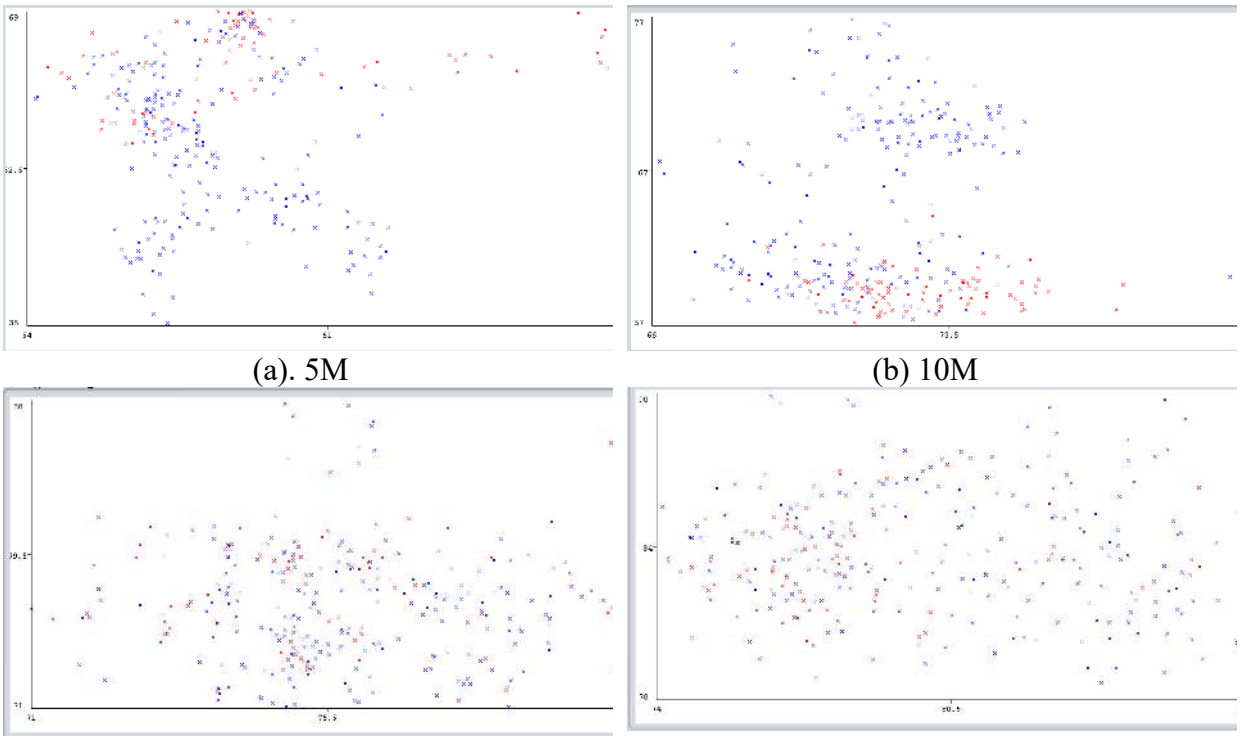
Results and Analysis

Experiment I- Rogue Wi-Fi Clients (RQ1: Distance)

In this experiment, we have examined how RSS values change with reference to different distances. RSS readings were taken with a 5-meter incremental distance with Dell and Raspberry connected to AP1. Furthermore, we performed a statistical t-test to determine if RSS can be used for wireless localization.

K-Means Clustering

In this section, we used RSS readings collected from Experiment I to compute and determine cluster points distributions on the signal space using K-Means clustering algorithm.



(a). 5M (b) 10M

(c) 15M (d) 20M

Figure 4.1. Dell cluster points vs Rasp cluster points at distances between 5 to 20 meters (x =Dell, x = Pi4)

The results were plotted from 5m-20m where distance served as discriminative feature to determine cluster points positioning as shown in Figure 4.1. The difference in node positioning can be visualized at 10M (Figure 4.1(b)) and the change in cluster pattern can be detected compared cluster points positioning as shown in Figure 4.1 compared to both nodes connected to the same access point at 5M (Figure 4.1(a)). The LQI (Link Quality Index) is largely affected by the decrease in TX power and increase in distance between the node and the access point. Weak AP signals affects the ability to detect rogue clients since the nodes are recording in almost similar RSS value range.

Since the k-means results are not conclusive, the data was further examined with a t-test. Results suggest a difference between devices can be detected based on their received signal strength. As indicated in Table 4.1, p-values calculated in the T-Test suggest a difference in RSSI distributions between difference devices connected to the same access point from the same

distance. From the experiment, we can clearly pinpoint the p-value is less than 0.05; hence we can conclude that is likely a difference in RSS values between rogue and valid nodes. Moreover, since the p-value is lower than 0.05, we reject the null hypothesis.

Distance (meters)	No. of samples	RSSI Mean		P-value
		Dell	Pi 4	
5	297	58.99	57.58	2.90286E-02
10	297	71.94	63.84	1.02045E-70
15	297	75.65	76.70	1.77791E-06
20	297	79.99	83.74	5.01165E-39

Table 4. 1 Corresponding P-values and RSSI mean of the nodes connected to AP1

Based on the p-value results in Table 4.1 it is evident that the strongest difference is detected at 10 meters. In statistical analysis, p value and effects are correlated in that the p value reports if the effect exists while as the effect size shows substantive significance of the results. For instance, any observed discrepancy is presumed to be explained by sampling unpredictability if the P value is greater than the chosen alpha level (e.g.,.05). A statistical test with a large enough sample will almost always show a substantial difference, unless there is no effect at all, in which case the effect size is exactly zero. The P value is considered indecisive because it depends on the sample size. The main result of the calculation may mean that only large samples were used. As a result, the magnitude of the effect is calculated to test the effect of distance on signal strength. Results provided in Table 4.2. suggest the largest effect sizes are observed at 10 and 20 meters. A larger effect size suggests there is a larger difference in the distribution of RSS values between the rogue device (raspberry Pi) and the legitimate device (Dell laptop).

5 meters	10 meters	mean (5-10m)	15 meters	20 meters	mean (15-20m)
0.116024	0.693934605	0.4049793025	-0.19854501	-0.547907971	0.37

Table 4. 2 Effect size (N=297)

While a statistical difference and large effect size (>0.5) suggest signal strength may be used to detect rogue Wi-Fi client's; further examination is required to see how this difference

actually performs as a feature and which machine learning algorithm provides the best accuracy. To further validate the statistical results from experiment, three machine learning algorithms are evaluated to investigate the detection accuracy. The following popular machine learning algorithms are used to examine the classification performance for detecting rogue Wi-Fi devices using RSS values: naïve Bayes, multilayer perceptron, and random forest. Tables 4.3 - 4.6 contain details regarding detection accuracy for discriminating between a raspberry Pi (rogue device) and a Dell laptop (legitimate device) at different distances from the same access point.

Distance (d = 5 meters)					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Naive Bayes	69%	0.729	0.69	0.677	0.797
Multilayer Perceptron	63%	0.656	0.633	0.619	0.632
Random Forest	83%	0.834	0.827	0.826	0.919

Table 4. 3 Classifier performance for detecting a rogue device (raspberry Pi) at 5 meters

Distance (d = 10 meters)					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Naive Bayes	82%	0.851	0.816	0.812	0.891
Multilayer Perceptron	80%	0.814	0.798	0.795	0.881
Random Forest	91%	0.907	0.906	0.906	0.96

Table 4. 4 Classifier performance for detecting a rogue device (raspberry Pi) at 10 meters

Distance (d=15 meters)					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Naive Bayes	62%	0.633	0.621	0.613	0.687
Multilayer Perceptron	56%	0.564	0.564	0.564	0.616
Random Forest	65%	0.651	0.65	0.649	0.729

Table 4. 5 Classifier performance for detecting a rogue device (raspberry Pi) at 15 meters

Distance (d=20 meters)					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Naive Bayes	75%	0.759	0.751	0.749	0.787
Multilayer Perceptron	76%	0.805	0.763	0.754	0.789
Random Forest	76%	0.804	0.763	0.754	0.783

Table 4. 6 Classifier performance for detecting a rogue device (raspberry Pi) at 20 meters

At distances ranging from 5 to 20 meters, results in the tables 4.3-4.6 consistently indicate the random forest classification algorithm outperforms both Naïve Bayes and Multilayer Perceptron. The best classification accuracy of 91% was observed at 10 meters. However, even though a large effect size was detected at 20 meters the detection accuracy is significantly low compared to the large effect size detected at 10 meters (78% vs 91%). Another observation presented in Figure 4.2 is that the classification rate drops when the distance is increased from 10 meters to 15 meters beyond 10 meters, but then slightly increases from 15 meters to 20 meters.

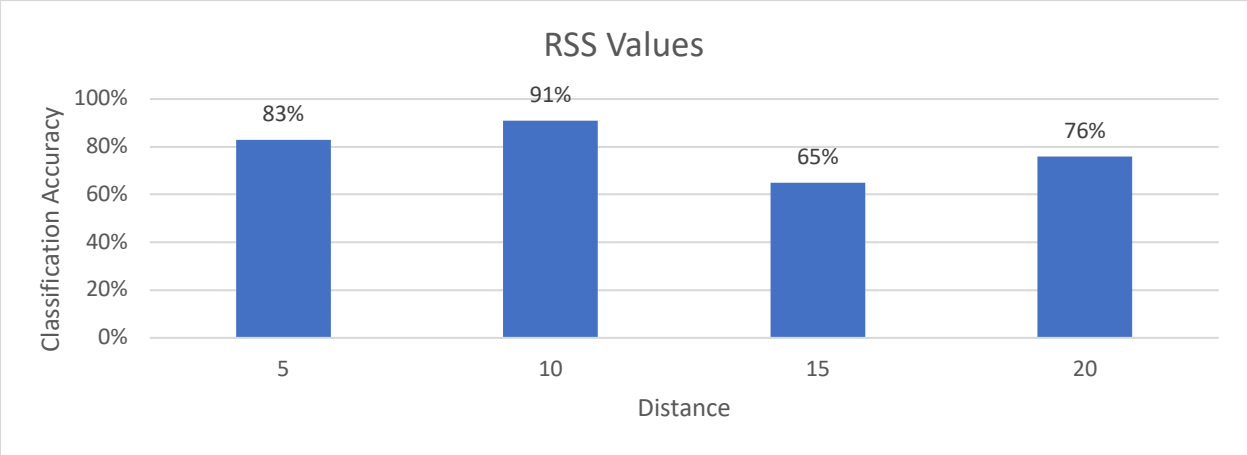


Figure 4.2. Random Forest classification rate for RSS values at distances between 5 to 20 meters

One assumption is that the noticeable differences between classification rates below 10 meters compared to classification rates above 10 meters many signal a demarcation point or

threshold value that suggests the demarcation for using RSS values for detecting rogue devices in Wi-Fi networks. To further investigate this assumption, the average classification rates above and below 10 meters is presented in Figure 4.3.

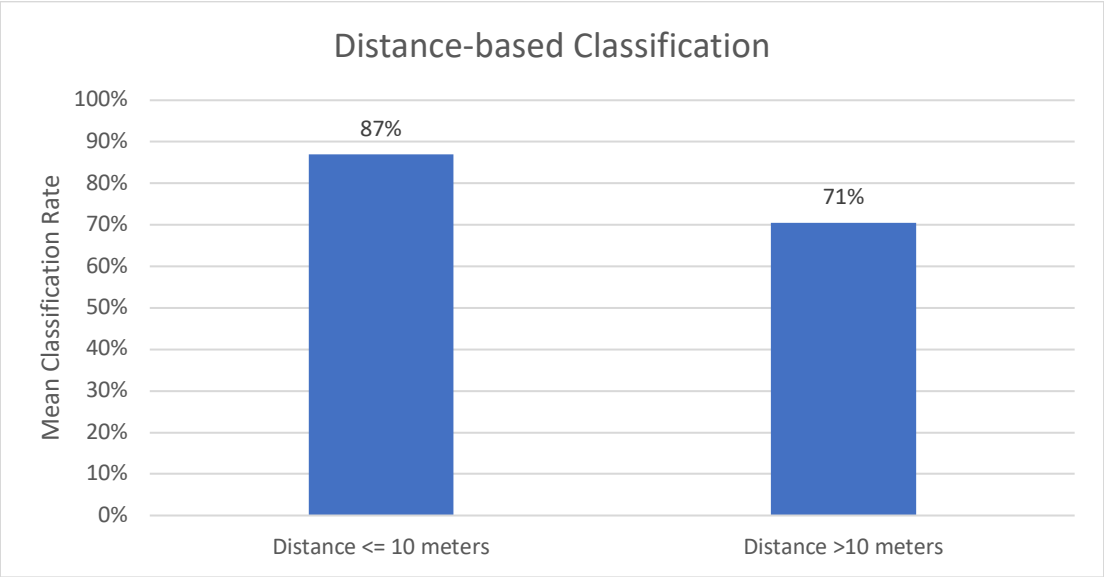


Figure 4.3 Mean random forest classification rate below and above 10 meters

It can be observed that detecting rogue devices using received signal strength values at distances greater than 10 meters provides lower discrimination quality compared to devices that are 10 meters or less from a given access point.

Therefore, the potential for using RSS may be proven most useful under certain attack scenarios where an adversary has the capability to place a rogue device in the same proximity as a legitimate user. However, in certain situations the adversary may not have the ability to place a rogue device at the exact same distance as a legitimate user. The most challenging scenario, based on the previous results, is if the adversary is within the 10-meter range. The hypothesis for RQ1 specifically examines this case. H_1 states that:

There will be no statistical difference in received signal strength between devices that are 5 meters from an access point compared to devices that are 10 meters from the same access point. Since, the random forest algorithm outperformed the other classification methods in experiment I it will also be used to address H1. The Raspberry Pi is a low cost highly portable device hence in subsequent analyses it has been selected as the device for the adversary. In an attack scenario considered the adversary does not have access to be within the same distance to the access point but we assume she/he does have the ability to be within proximity. We examine the case where the adversary is 10 meters from the access point and the legitimate user is within 5 meters from the same access point. RSS data are collected from the Pi at 10 meters and the Dell at 5 meters from the same access point. Results are presented in Table 4.7.

d= Dell at 5 vs. Pi at 10 meters					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Random Forest	85%	0.849	0.848	0.848	0.918

Table 4. 7 Classification accuracy for Dell at 5 meters and Pi4 at 10 meters from an AP

An 85% classification accuracy was achieved with the random forest algorithm. Therefore, at close distances less devices than or equal 10 meters from an access point received signal strength can help to detect rogue devices.

Experiment II- Rogue Wi-Fi Clients (RQ2: Data reduction)

Unlike Experiment I which investigated the classification performance for detecting different Wi-Fi devices connected to the same access point at the same distance, the second experiment analyzes classification performance for different devices connected to the same access point at different distances with reduce data. Placing a constraint on the amount of data required for analysis can provide near real-time analysis. Results from experiment II demonstrate that the

accuracy for discriminating between wireless devices using received signal strength (RSS) does not negatively impact performance.

d= Dell at 5 vs. Pi at 10 meters					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Random Forest	87%	0.894	0.866	0.863	0.92

Table 4. 8 Modified dataset with 50% reduction in data

As can be seen in Table 4.8 an 87% classification accuracy was achieved with a 50% reduction in data (148 data points for each device vs 297 data points for each devices). To examine this in greater detail the original dataset was used with 20% of the data used for training and the remainder for the test data. Results are provided in Table 4.9.

d= Dell at 5 vs. Pi at 10 meters					
	Accuracy	**Precision**	**Recall**	**F**	**ROC Area**
Random Forest	86%	0.857	0.857	0.857	0.928

Table 4. 9 Original dataset with 20% used for training and the remainder used for test

Again, it can be observed that the received signal strength data contains enough discrimination quality even with significantly less data. This underscores the value for using RSS data as an attribute to detect rogue devices that may have access to be within proximity to an access point as legitimate users.

However, results in Figure 4.1 show that the p-values after 10 meters goes down at 15 meters and then up again at 20 meters. This oscillating pattern may suggest a more stable method is required that is not impacted by distance or latency. Consequently, alternate methods for detecting rogue devices at longer ranges from an access point are required.

One alternate method is to examine the differences in clock skews between Wi-Fi devices. Prior research has shown that the clock skews of wireless devices remain consistent over time (Jana, 2009). Clocks skews will be examined between the rogue (raspberry Pi) and legitimate device (Dell laptop) connected to the same access point. One access point is within the same indoor location as the rogue and legitimate device. This access point has an SSID = 53BL2 and is within close range to the Wi-Fi devices. The other access points are located outdoors from the Wi-Fi clients. The distance is unknown but is at least 20 meters away from the Wi-Fi client devices. Figures 4.4 – 4.7 illustrate the differences in clock offsets. In Figure 4.4 all devices (Wi-Fi clients and access point are located within 5-10 meters. In Figure 4.5-4.7, Wi-Fi clients (raspberry Pi and Dell laptop) are within 5-10 meters from each other, and the access points are located more than 20 meters from the Wi-Fi clients.

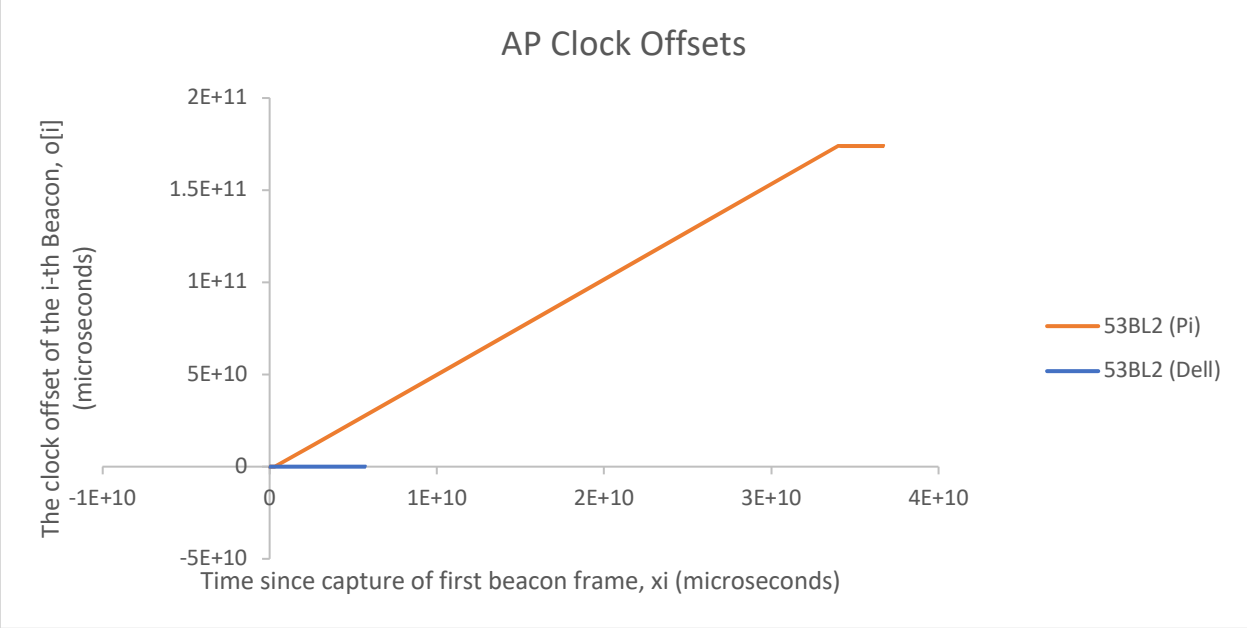


Figure 4.4 Clock offset Dell vs Pi connect to access point with SSID 53BL2

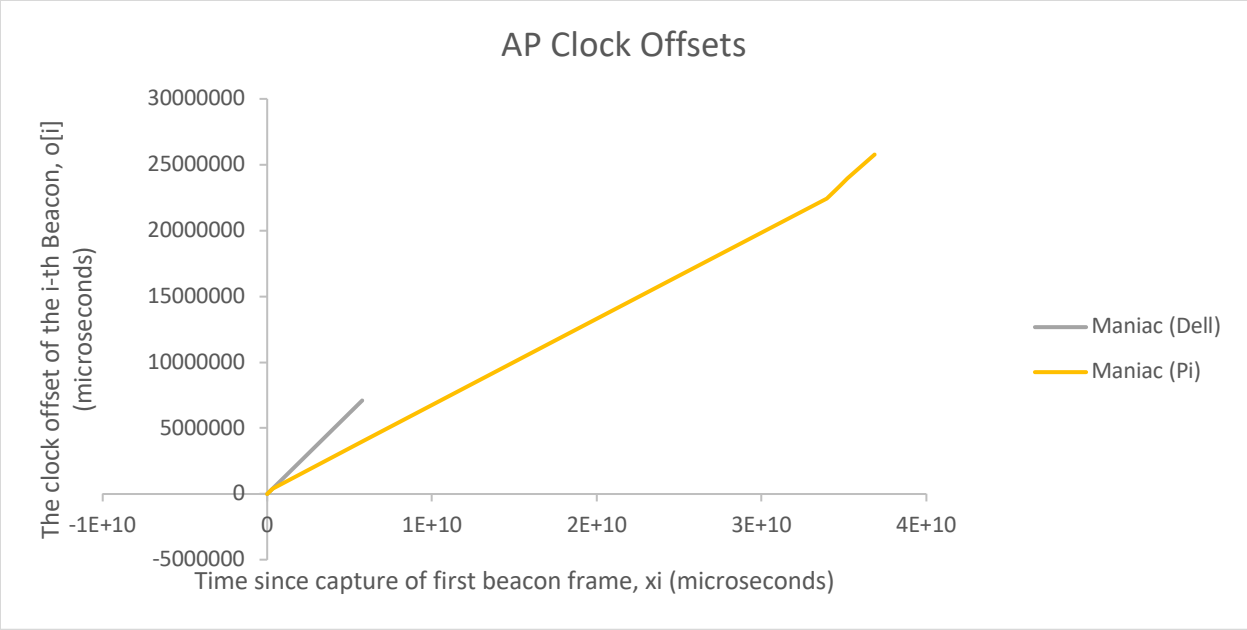


Figure 4.5 Clock offset Dell vs Pi connect to access point with SSID Maniac

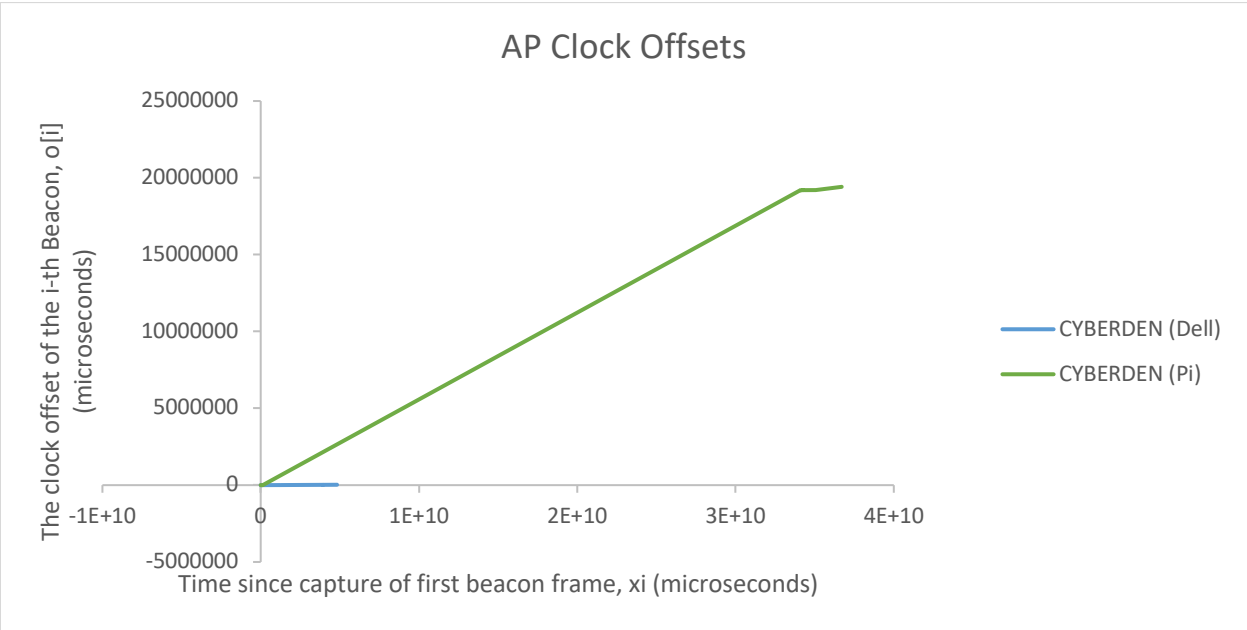


Figure 4.6 Clock offset Dell vs Pi connect to access point with SSID CYBERDEN

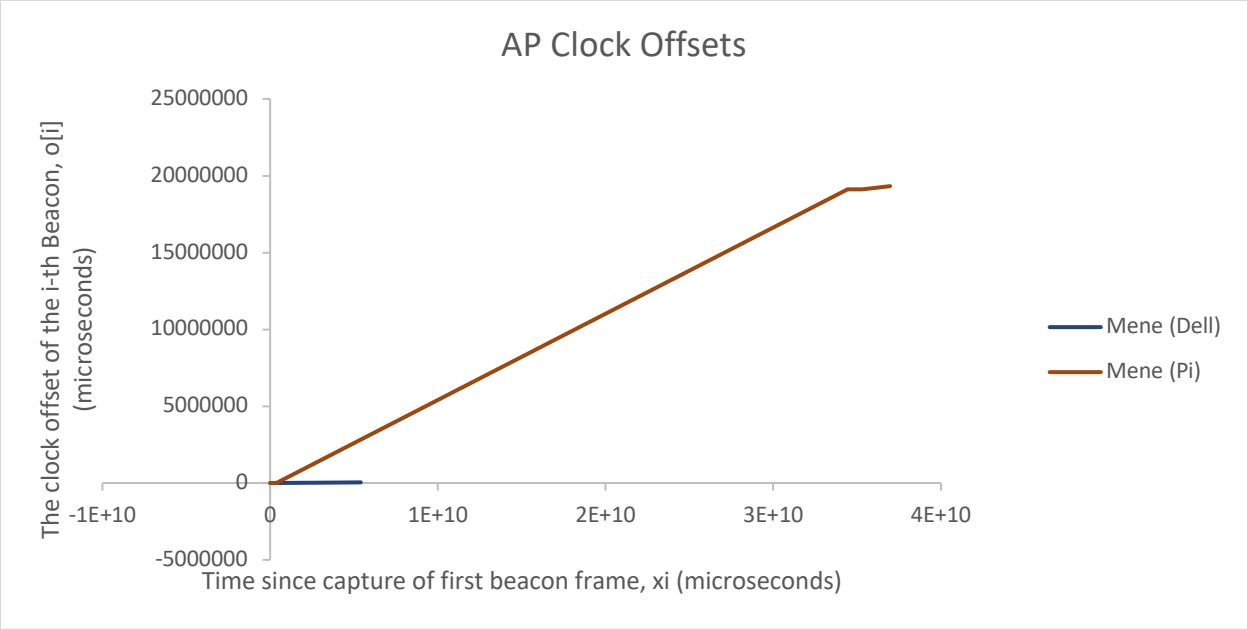


Figure 4.7 Clock offset Dell vs Pi connect to access point with SSID Mene

Visual inspection of the Figures 4.4-4.7 suggests a widespread or separation between clock offsets for the rogue (raspberry Pi) and legitimate node (Dell laptop). To validate the results p-values and effect sizes were computed in Table 4.10 and 4.11.

		Raspberry Pi			
		53BL2	Maniac	Cyberden	Mene
Dell	53BL2	9.3E-265			
	Maniac		0		
	Cyberden			8.7E-09	
	Mene				0.057417

	N
53BL2	3342
Maniac	9962
Cyberden	26
Mene	25

Table 4. 10 p-values and sample sizes

Results in Table 4.10 indicate clock skews is an effective metric for detecting rogue devices. In every case regardless of the distance between the Wi-Fi clients (rogue = Raspberry Pi

and legitimate = Dell laptop) the access points results are statistically significant. While all results in Table 4.10 are statistically significant, the sample sizes (N) for the Cyberden and Mene access points are too small to detect a large or medium effect and therefore, these results should be interpreted with caution. Further analysis is required. The effect sizes for 53BL2 and Maniac are presented in Table 4.11.

	53BL2 (Pi)	Maniac (Pi)
53BL2 (Dell)	-0.422913	
Maniac (Dell)		0.732330567

Table 4. 11 Effect size based on Clock Skews

It can be observed from Table 4.11 that the largest effect size was detected between the rogue (Pi) and legitimate device (Dell) when connected to the Maniac access point. The distance between this access point and the Wi-Fi clients is significantly further away compared to the 53BL2 access point. This result suggests strong support for using clock skews over received signal strength values for large distances between Wi-Fi clients and access points. Specifically, large effect sizes can be detected between rogue and legitimate Wi-Fi clients connected at distances greater than 20 meters from an access point. Further investigation is required to examine the impact of the sample sizes. However, in both cases sample sizes are large enough to detect a large effect with a statistical power of 0.8. Experiment III investigates using clock skews for detecting rogue access points also known as “evil twins.

Experiment III- Clock Skew

Access points (AP’s) broadcast beacon frames to alert any Wi-Fi client that they are willing to accept a connection. Beacon frames were broadcast to the Raspberry Pi Wi-Fi client during two different time periods: November 2020 and March 2021. Data for all access points (AP’s) within

range of the Raspberry Pi was collected for one minute. Since the beacon frames and packet timestamps are in different units (microseconds vs Epoch time in seconds) the first packet is converted to time 0. The difference of each successive packet is computed as the difference between the previous packet time. Figure 4.8 presents the clock offsets of four different access points obtained from a Raspberry Pi Wi-Fi client during November 2020. Results demonstrate that after 15 seconds (15,000,000 microseconds) clock offsets appear to represent unique linear functions. This suggests clock offsets provide good discrimination quality after 15 seconds.

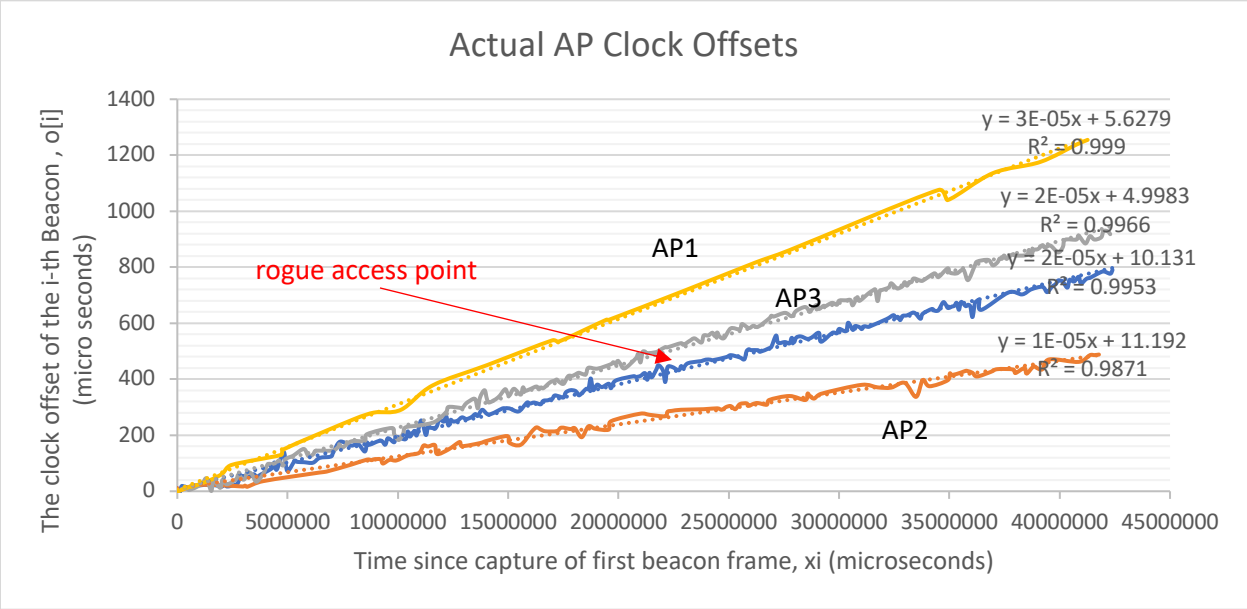


Figure 4.8 Access point (AP) Actual Clock Offsets (November 2020)

To further examine the predictive quality of clock offsets a linear regression was performed with 75% of the data for training and 25% of the data for testing.

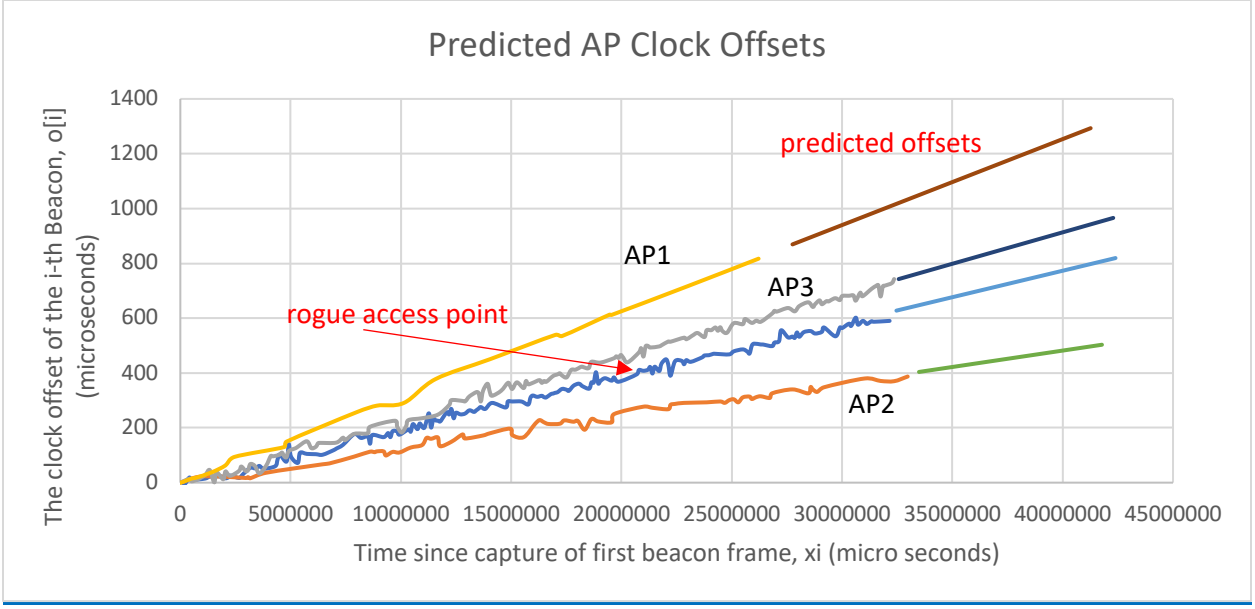


Figure 4.9 Access point (AP) Predicted Clock Offsets (November 2020)

As can be observed in Figure 4.9 the predicted clock offsets align with their respective linear functions.

To further verify results obtained in Figure 4.8 the experiment was conducted using a different Raspberry Pi Wi-Fi client several months later in March 2021. Results are presented in Figure 4.10.

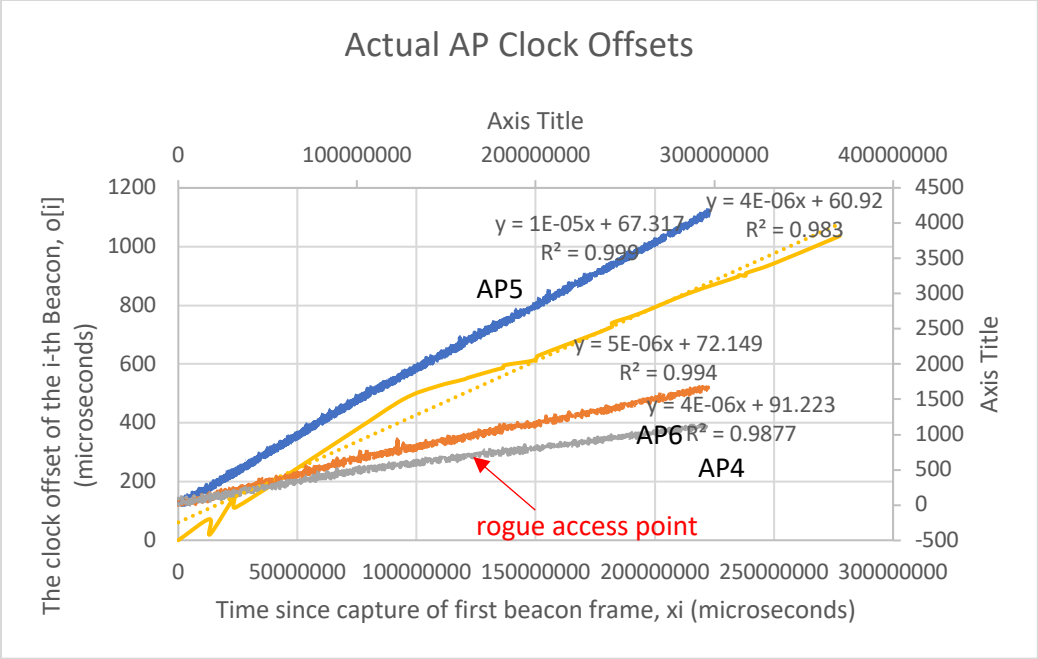


Figure 4.10 Access point (AP) Actual Clock Offsets (March 2021)

Like the same experiment conducted in November 2020 there is a clear separation between the different access points using clock offsets as a feature. A clean separation appears after about 7 seconds (7,000,000 microseconds) from the arrival of the first beacon frame.

To validate the differences observed in Figure 4.5 and Figure 4.6 statistical t-test were performed with access points that had at least 97 beacon frames (i.e. n = 97 data points). Access points that had fewer than 97 beacon frames were excluded from the analysis. To perform the statistical tests, we made sure all datasets were balanced. AP2 only had 97 data points so to make the data balanced only 97 data points were chosen. A power analysis with 97 data points is only able to detect moderate effect size of 0.5 with a power of 80%. Consequently, one access point was removed for November (AP1) and one from March 2021 (AP6) since they had fewer than 97 beacon frames. Results from the t-test for data collected in November 2020 and March 2021 are

presented in Table 4.12 and Table 4.14 and respective effect sizes are provided in Table 4.13 and Table 4.15.

	Rogue AP	**AP2**	**AP3**
Rogue AP		7.37214E-08	0.029861
AP2	7.37214E-08		7.44E-39
AP3	0.029861298	7.44E-39	

Table 4. 12 t-test p-values Clock Skews (November 2020)

	AP2	**AP3**
Rogue AP	0.404892	-0.16452093

Table 4. 13 Clock skew effect size (November 2020)

It can be observed that the differences in clock skew distributions between the rogue access point and other access points is statistically significant ($p < 0.05$). In addition, the effect size for AP2 is higher than AP3, but in both cases only a moderate (AP2) to small effect (AP3) are detected.

	Rogue AP	**AP4**	**AP5**
Rogue AP		0.007381	1.43E-07
AP4	0.007381		6.80E-08
AP5	1.43E-07	6.80E-08	

Table 4. 14 t-test p-values Clock Skews (March 2021)

	AP4	**AP5**
Rogue AP	-0.09718	-0.0744608

Table 4. 15 Clock skew effect size (March 2021)

During the March 2021 collection period higher results were observed. The differences in clock skew distributions between the rogue access point and other access points is statistically significant ($p < 0.05$), and the effect size for AP4 is higher than AP5. A large effect was detected for both access points. The combined results of p-value and effect sizes for both time periods

provides support for using clock skews to detect rogue access points. The main difference between the two time periods is that large effect sizes were detected during the second time (March 2021) compared to the first time (November 2020) with the same sample sizes of only 97 beacon frames. Future research will investigate the possible causes for the differences in effect sizes.

Chapter Five

Conclusion and Future Exploration

In this paper, we have exhaustively analyzed Wi-Fi localization based on RSSI and packet-based time synchronization on the IEEE 802.11 protocol on highlighting the impact on network effectiveness and performance. This study investigates the RSSI pattern through RSSI distribution in the signal space. RSS based positioning scheme was used to test for reliability and accuracy of using RSSI as a determinant feature to identify rogue Wi-Fi clients. Wi-Fi localization under these conditions remains a fundamental platform to explore reflecting on the application of RSSI in identifying the positioning of the rogue AP on the signal space.

The main attraction to RSSI as a metric is that the measurement and calculation is simple, and less tasking compared to other localization metrics. Increased frequency has cultivated the need to focus on wireless sensor networks due to their recent adoption and applicability on a range of devices in the IoT environments. Network architectural designs over the years have faced the challenge in providing clock synchronization on the sensor network since there are many applications that process large chunks of data on the networks. This paper explored the effectiveness of beacon time synchronization on identifying the presence of rogue APs on the signal space. When a device wants to send a packet from the AP to the node, the device broadcast

its beacon frames to announce its presence on wireless network. A summary of results from this research is provided in Table 5.1.

Hypothesis	Description	Result
H_1	There is no statistical difference between devices that are 5 meters from the access point compared to devices that are 10 meters from the same access point.	**Supported.** While there was some volatility in signal strength as the distance between devices and the access point was increased from 5-20 meters, the mean effect size and was larger when the devices were less than 10 meters from the access point than when the devices were more than 15 meters from the access point (Table 4.2: 0.41 for d <10m and 0.37 for d>15m). The same was observed with the classification rates (Figure 4.3: 87% for d <10m and 71% for d>15m). At proximity when one device was 5m from the access point and the other devices was placed at 10m from the access point an 85% accuracy was achieved (Table 4.7).
H_2	There is no statistical difference in detection accuracy for same devices with 50% fewer data points connected to the same access point.	A significant decrease is accuracy was not observed with a 50% reduction in data. This suggests a small amount of data collection is required facilitating near real-time analysis. An 86% detection accuracy was achieved (Table 4.9).
H_3	There is no statistical difference in clock skews between rogue and valid Wi-Fi clients.	**Supported.** At a distance of at least 20m from the access points a large effect size was detected using clock skews (Table 4.11). Considering results obtained for H1, this result suggests clock skews provide better discrimination quality compared to signal strength at distances greater than 20 meters.
H_4	There is no statistical difference in clock skews between rogue and valid access points.	**Supported.** In 2 out of the 3 test cases a large effect was detected between the rouge and valid access points. This was observed when the distance between the rogue and valid access points was 20m or more. This is consistent with results obtained for H3 and addresses limitations with using signal strength for distances more than 15m as observed with results from H1. Further research is required to understand the one exception when a small effect was detected.

Table 5. 1 Summary of results

The presence of rogue APs on the signal space has contributed to identity theft, service unavailability and has jeopardized information integrity. Besides, the adversaries have widely exposed the CIAD metrics resulting to user manipulation and distributed denial of service (DDOS).

Future exploration

Our approach focused on the hardware-based traffic detection and model implemented on physical and datalink of the network layered model. We propose more diverse fingerprint identification technique which will explore significant difference between temporal network characteristics of rogue devices compared to authorized devices. For example, the duration field in a MAC address can be used to identify the chipset for a device (Cache, 2006). In instances where hardware of rogue devices matches valid nodes on the network, the nodes' location or proximity within the network can be used to detect the rogue device.

Moreover, synchronous timestamping approach offers a valid technique to detect and identify rogue APs. The difference in clock skew signals presence of rogue APs and can be implemented into two layered forms; software-based and infrastructural-based (hardware-based). This paper solely focused on the infrastructural-based section of the research study. Likewise, we suggest studies to inter-connect the software based and hardware-based time synchronization to minimize the overhead when there is a large pool of devices broadcasting their beacon frames.

References

Adewumi, O. G. (2013). RSSI based indoor and outdoor distance estimation for localization in WSN., (pp. 1534-1539).

Anjum, M. K. (2020). RSSI Fingerprinting-based Localization Using Machine Learning in LoRa Networks.

Arackaparambil, C. B. (2010). On the reliability of wireless fingerprinting using clock skews. *In Proceedings of the third ACM conference on Wireless network securit*, (pp. 169-174).

Bekcibasi, U. &. (2014). Increasing RSSI localization accuracy with distance reference anchor in wireless sensor networks. *Acta Polytechnica Hungarica*.

Cache, J. (2006). Fingerprinting 802.11 implementations via statistical analysis of the duration field. Uninformed. org, 5.

Chen, Y. &. (2002). Signal strength based indoor geolocation. *International Conference on Communications*, (pp. 436-439).

Chen, Y. T. (2007). Detecting and localizing wireless spoofing attacks. *In 2007 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, (pp. 193-202).

Cota-Ruiz, J. R.-P. (2013). A distributed localization algorithm for wireless sensor networks based on the solutions of spatially-constrained local problems. *Sensors Journal*, 2181-2191.

Duan, Z. X. (2008). Controlling IP Spoofing through Interdomain Packet Filters., *5*, pp. 22-36.

Farid, Z. N. (2013). Recent advances in wireless indoor localization techniques and system. *Journal of Computer Networks and Communications*.

Feng, C. A. (2010). Compressive sensing based positioning using RSS of WLAN access points., (pp. 1-9).

Hyo, A. S. (2009). Environmental-Adaptive RSSI-Based Indoor Localization,. *ransactions on Automation Science and Engineering*, 626-633.

Jana, S. &. (2009). On fast and accurate detection of unauthorized wireless access points using clock skews., (pp. 449-462).

Kaemarungsi, K. &. (2004). Properties of indoor received signal strength for WLAN location fingerprinting. *In The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services,*, (pp. 14-23).

Kasera, S. J. (2008). On Fast and Accurate Detection of Unauthorized Wireless Access Points using Clock Skews. *Mobile Computing and Networking.*

Kaur, P. S. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. 500-508.

Krishna, R. &. (2009). Calculating relative clock drifts using IEEE 802.11 beacons., (pp. 1-4).

Li, Q. &. (2006). Light-weight detection of spoofing attacks in wireless networks. *International Conference on Mobile Ad Hoc and Sensor Systems*, (pp. 845-851).

Navarro, E. P. (2010). Wi-fi localization using RSSI fingerprinting .

Nawir, M. A. (2016). Internet of Things (IoT): Taxonomy of security attacks., (pp. 321-326).

Nuo, W. Q.-L. (2012). Three-dimensional localization algorithm of wireless sensor networks base on particle swarm optimization. *The Journal of China Universities of Posts and Telecommunications*, 7-12.

Park, K. &. (2001). On the effectiveness of route-based packet filtering for distributed DoS attack prevention in power-law internets. *ACM SIGCOMM computer communication review*, (pp. 15-26).

Pei, L. L. (2017). Evaluation of fingerprinting-based WiFi indoor localization coexisted with Bluetooth. *The Journal of Global Positioning Systems*.

Sadowski, S. &. (2018). Rssi-based indoor localization with the internet of things.

Sun, Y. L. (2014). WiFi signal strength-based robot indoor localization., (pp. 250-256).

Tang, Z. Z. (2017). Exploiting wireless received signal strength indicators to detect evil-twin attacks in smart homes. *Mobile Information Systems*.

Wielandt, S. &. (2017). Indoor multipath assisted angle of arrival localization. *sensors*, 2522.

Yiu, S. D.-C. (2017). Wireless RSSI fingerprinting localization., (pp. 131,234-244).

Zander, S. &. (2008). An Improved Clock-skew Measurement Technique for Revealing Hidden Services., (pp. 211-226).

Zhang, Y. D. (2019). Zhang, Yao, Zhongliang Deng, and Yuhui Gao. "Angle of Arrival Passive Location Algorithm Based on Proximal Policy Optimization., (p. 1558).