Privacy Leaks from Wi-Fi Probing[1]

Pejman Najafi, Andreas Georgiou, Dina Shachneva and

Ioannis Vlavianos

University College London

March 2014

[1]This report is submitted as part requirement for the MSc Information Security course. It is substantially the result of our own work except where explicitly indicated in the text. The report may be freely copied and distributed provided the source is explicitly acknowledged.

ABSTRACT

In a modern era where smartphone use has exponentially increased, we investigate the amount of private information an adversary can extract by looking at the active service discover in Wifi where a wireless station broadcasts the list of its preferred wireless networks, without user's consent or knowledge.

This report describes a range of different techniques which violate users' privacy by using Wi-Fi fingerprints emitted from devices. One of the main points of the report include the relationship discovery that was first implemented by Cunche, et al and explains how to infer relationships between users through their SSIDs sets. Other attacks can be mounted that will reveal not only a social link between users but the actual path that a device follows in real time. These techniques are currently used for other purposes as well, such as commercial use and public safety. In addition, a set of countermeasures are proposed for these problems and discuss their effectiveness for each specific attack.

Table of Contents

CHAPTER 1 - INTRODUCTION

During the last decade the number of portable Wi-Fi devices such as smartphones, tablets, and notebooks have increased dramatically. Nowadays, the majority of users constantly carry their mobile devices with them not only at workplace but also at home.During the last decade handheld devices became powerful enough so that is more convenient to execute more and more tasks using them than a personal computer. The Internet is no longerconsidered as a luxury anymore but anecessityforour everyday life.In this regard more that 60% of the cell owners use their smartphone to access the internet [1]

All mobile devices connect to the internet either using 3G or Wi-Fi technology. Wi-Fi offers not only high speeds and reliability but also more affordable internet access. For this reasons, it has become the most preferable way to provide internet access for medium range connectivity.In this regard many recent scientific papers have investigated how much information can be obtained from user interaction with a Wi-Fi network. Particularly observing the fields that are transmitted in plaintext such as *probe requests*

In order to connect to the network a mobile phone is sending the probes requests that contain the SSID name of previously associated networks. This mode is called Active Discovery Mode (ADM). Wi-Fi supports encryption and authentication standards (e.g. WPA2) to ensure that the transmitted data between a client and an Access Point are safe from eavesdroppers. However, before we reach the phase of credentials exchanging there is the phase of Network Discovery where a Wi-Fi enabled client discovers and contacts the AP for the first time. In this phase there are several messages exchanged between the two parties (e.g. network discovery probe requests) that are transmitted necessarily in plaintext. These packets (also called frames) include in their headers, among others, the Media Access Control (MAC) address of the Wi-Fi enabled device. This address serves as a unique identifier for the said device. Based on this two features: unique identifier MAC and probes it seems feasible to identify users with a high probability.This fact enables the identification of a user by its device's MAC address. Practically, the device serves as an interface to reach the user behind and the MAC address of the device becomes as a nickname of the user. As a result, the user becomes vulnerable to what we would call fingerprinting attacks. Fingerprinting is called "the process by which a device, it's driver or the OS a machine is running can be uniquely identified by its externally observable characteristics" [2]. Therefore an adversary can extract information about someone previous history location or even track him down in small proximity.

The rest of present paper is structured into three chapters. In this regard, Chapter 2 provides the background information, covering the basic definition used throughout the rest of the paper; chapter 3 illustrates several issues regarding the active service discovery that could be considered as user's privacy violation. In this regard an entity is capable of leveraging information by observing the probe request transmission; next, chapter 4 proposes and discuses several approaches which are considered as a countermeasure against the problems discussed in chapter 3; and finally, chapter 5 comprises the conclusion

CHAPTER 2 - BACKGROUND INFORMATION

This section introduces the basic definitions and features as describe in the IEEE 802.11 standard for the wireless local area network (WLAN) communications.

2.1 IEEE 802.11 Frames

The Wi-Fi packets also known as frames are divided into three categories: Management frames, Control frames and Data frames. Management frames in 802.11 include association, authentication and probe requests between APs and stations. Similarly Control frames are used to accommodate Data frames. They coordinate the stations' communications and by using commands like Request to send (RTS), Clear to send (CTS) and Acknowledgement (ACK). The last category contains the actual data of the communication and it can be either encrypted or in plaintext.

2.2 MAC Address

The MAC (Media Access Control) address is a 48-bit long unique identifier of a network interface. The first 24 bits represent the manufacture of the network interface. The MAC address of the transmitter is included in the header field of all management and data frames. Although the content is could be encrypted the header is never encrypted and therefore it is always broadcasted in plaintext. The device is not only advertising its presence for eavesdropping equipment but also enables adversaries to uniquely track its location.

2.3 Preferred Network List

The Preferred Network List (PNL) also called Configured Network List (CNL [3]) is a list of the networks the device has been previously associated with. The OS of the device is responsible for keeping this list. The list contains the SSIDs of those networks and its security configuration but it does not store the BSSID (often the MAC address of the AP's wireless adaptor). Depending on the vendor and the operating system of the client device, the request probes sent by the device in the Active Discovery Mode contain the SSIDs of the networks stored in the PNL.

2.4 Network Discovery Modes

The IEEE 802.11 protocol supports only two modes of AP discovery, passive and active.

2.4.1 Passive mode

In the Passive Discovery Mode (PDM), the Wi-Fi client passively monitors the spectrum for AP's advertising packets called beacons. These beacons are periodically broadcasted and are practically advertising to the nearby clients their existence and that they are ready to serve clients.

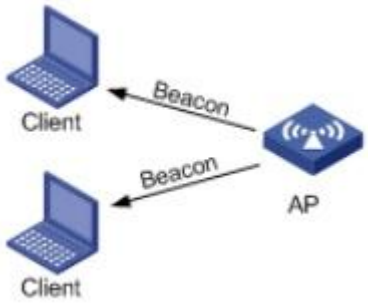

Figure 2.1 - Passive Network Discovery mode [4]

2.4.2 Active mode

In the Active Discovery Mode (ADM), the Wi-Fi client actively broadcasts discovery packets called Request Probes to the spectrum and receives responses from the available access points containing their SSIDs (Service Set Identifiers). There are two types of ADM, as illustrated at figure 2.2.a, the client sends out a named-specific SSID probe request where only the APs with that SSID nameresponds. In the second one figure 2.2.b, the client sends out a broadcast message with the SSID field left null and waits for a reply from any alive APs.

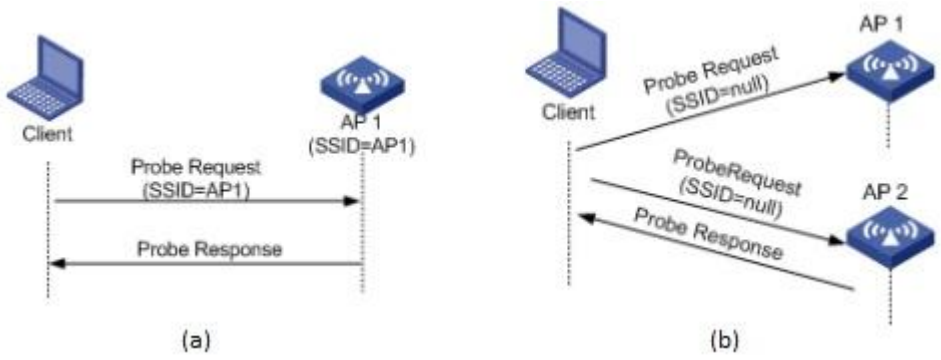

Figure 2.2 - Active Network Discovery mode [4]

This process is repeated on each channel until the client associates with an AP. In addition, active network discovery is used in the case of hidden networks where the AP does not broadcast any beacons. Due to the fact that the active discovery mode allows much faster AP association than the passive mode, most Wi-Fi clients are configured to actively search for networks by default.

CHAPTER 3 - PRIVACY ISSUES

In this chapter, we will present and analyse several privacy issues regarding Wi-Fi network active discovery mode. Furthermore we illustrate how much information can be obtained by monitoring the Wi-Fi traffic, in particular the probe request which is sent in plain text.

3.1 Infer Social Link

Since the majority of private WLAN network are encrypted and password protected, users who are capable of accessing the same private network are most likely to know each other.Wi-Fi fingerprints (i.e. list of SSIDs probed by a specific device) can be used to infer a social link between the device holders (possible relationship between the users of the device). This can be achieved by measuring the similarity between the Wi-Fi fingerprints of the devices. This is known as Record Linkage (techniques to find records that may belong to the same entity in two or more datasets).

The similarity depends on: (1) The size of the intersection between two fingerprints; and (2) Rarity of the intersection, since common SSIDs such as eduroam or BTOpenzone are more likely to be sharedby the majority of fingerprints. Therefore a weight must be assigned to each SSID which is inversely proportional to its frequency of being probed.

The similarity between two fingerprints can be measured using different similaritymetrics such as *Jaccard Index, IDF similarity (TF-IDF),* and *Adamicsimilarity*which are typically used in record linkage-related problems. In this regard, these techniques can be used to measure the similarity and classify fingerprint pairs as linked, if the similarity value is above a given threshold or non-linked otherwise.

Cunche, et al. in [5] evaluated several similarity metrics and compared their performance. In this regard the authors collected the Wi-Fi fingerprints from a set of volunteers, along with the fingerprints of individuals with whom they maintain a strong social relationship and categorised device pairs in two groups: *Linked* and *Non-Linked*as a ground truth evidence. Next, the fingerprint pairs on both Linked and Non-Linked sets, were separated and shuffled and considered as one set. Finally the *cosine-idf, jaccard, adamic,* and *Psim-3*(modified version of *adamic*) similarity metrics were usedas classifiers to compute fingerprint pairs similarity for each potential couples and classify the pair as linked or non-linked based on several thresholds.

The similarity metrics are calculated as follows

$$Jaccard\ (X,Y) = \left(\frac{|X \cap Y|}{|X \cup Y|}\right)$$

$$Cosine - IDF\ (X,Y) = \left(\frac{\sum_{x \in X\ \cap\ Y}\ IDF_x^2}{\sqrt{\sum_{x \in X}\ IDF_x^2}\sqrt{\sum_{y \in Y}\ IDF_y^2}}\right) \quad , \quad IDF_i = \log\frac{1}{F_i}$$

$$Psim - 3\ \ (X,Y) = \sum_{z\ \in X\ \cap\ Y} \frac{1}{f_z^3}$$

Where *X* and *Y* are fingerprints containing a list of SSIDs.IDF_i is the inverse document frequency of the term *i*in the considered corpus. $f_z$is the frequency of the element *z*. The results of *jaccard* and *cosie-idf*range from 0 (not linked) to 1 (linked), and Psim-3 ranges from 0 to a maximum value depending on the considered corpus.

They observed that *cosine-idf*and *Psim-3* metrics had the best performance compared to the other metrics, the term "performance" was measured using the false positive rate (FPR) - proportion of all devices that do not correspond to an underlying link, but they are wrongly reported as positive by the test; and the true positive rate (TPR) which is the proportion of linked devices that have been rightly reported as such by the test.

$$FPR = \frac{n_{FP}}{n_{FP} + n_{TN}} \qquad TPR = \frac{n_{TP}}{n_{TP} + n_{FN}}$$

Where n_{TP} is the number of true positives (devices or pairs that are correctly identified as linked). n_{TN}isthenumber of true negatives. n_{FP} refers to the number of false positives (number of device pairs being wrongly identified to be linked). n_{FN}numberfalse negatives.

The best performance is when the TPR is at its highest value while the FPR is at its lowest value. Figure 1 showsthe performance evaluation of the classifiers with several thresholds carried out by Cunche, et al. [5].

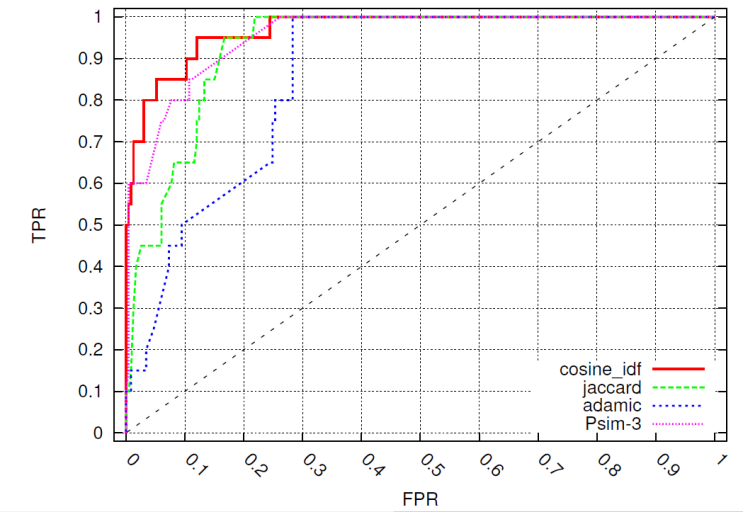

Fig. 3.1. ROC curve of the classifier [5].

One of the principal reasons for a better performance of *cosine-idf* and *Psim-3* similarity compared to Jaccard is based on the fact that, Jaccard index considers similarity as only the ratio of the intersection's cardinal by the union's cardinal, in other words it does not consider the rarity of the intersection, therefore causing a less efficient result. Whereas the *cosine-idf*and*Psim-3* take into consideration not only the size of the intersection but also its rarity.

According to their study [5] the value of threshold has a direct impact on the FPR and TPR. In this regard, a lower threshold causes an increase in the true positive rate and also an increase in the false positive rate. In other words increasing the threshold results in a more aggressive classification that correctly classifies only a few of the actually linked devices (true positives), and wrongly reports a small proportion of devices as being linked (false positive). The choice of the threshold depends on the application of the classifier (e.g. favouring the increase in n_{TP} at the cost of increasing n_{FP} or favouring the lower n_{TP} to reducing the n_{FP}). However the paper suggests that the set of thresholdsthat cause the FPR to be less than 0.1 provide the most optimal performance. For instance for the Psim-3 metric, a threshold of $t = 1.5710^{-5}$ leads to TPR=0.8 and FPR=0.0772, will limit the number of *false positives*. The figure below represents their finding for *cosine-idf* and *Psim-3* classifier tested on different thresholds [5].

	Thresh.	Nb. detected pairs	TPR[4]	FPR[4]
Cosine-IDF	0.188	215 384	0.5	0
	0.117	502 102	0.8	0.0300
	0.0153	3 934 564	1	0.2446
Psim-3	0.162	408	0.1	0
	$1.57\ 10^{-5}$	88 476	0.8	0.0772
	$1.19\ 10^{-8}$	3 812 524	1	0.2618

Fig. 3.2. Selected threshold and associated number of pairs detected. And corresponding value of TPR and FPR [5].

Cheng, et al. [6] proposed that when collecting the Wi-Fi probe requests, recording the *location* and the *time-stamp* in addition to the MAC address and the list of SSIDs, will allow to further improve the device linkage and classification by measuring:*Location proximity* and *Spatio-temporal co-occurrence probability.*

MAC address	SSID	Location	Timeslot
a1:b2:c3:d4:e5:f6	attwifi	starbucks	1pm-2pm
a1:b2:c3:d4:e5:f6	hello	starbucks	1pm-2pm
a1:b2:c3:d4:e5:f6	lisa's network	Bldg1	3pm-4pm

Fig. 3.3. Example of Wi-Fi probe request details collected from a user [6]

3.1.1 Location proximity

Geographic location proximity can be used to further improve the classification and explore more possible relationships. Consider two students from the same university who know each other, however each access their own department APs. Since their device Wi-Fi-fingerprints will not have common SSIDs, classifying only based on similarity of their SSID lists will indicate no relationship between the two students. Whereas adding classifiers to also consider location-based similarity will allow this relationship discovery.

It is possible to map SSID names with a geographic location. This can be done using either an online AP mapping databases such as WiGLE [11] or self-created mapping database by activities such as

Wardriving (a person who maps Wi-Fi networks). In this regard whenever two users probe for two different SSIDs from the same geo-location, they are mapped to the same cluster (possibly a building). This information can be used to better classify the devices with their corresponding fingerprints as linked or not linked. In this regard the classifier computes the *cosine-idf*similarity between two sets of geo-locationsmappingto classify two devices as Linked or Non-Linked based on a threshold value.

$$Similarity\ (d_1, d_2) = Cosine\ \big(M(X), M(\ Y)\big) = \left(\frac{\sum_{x\in\ M(X)\cap M(Y)} IDF_x^2}{\sqrt{\sum_{x\in M(X)} IDF_x^2}\sqrt{\sum_{y\in M(Y)} IDF_y^2}}\right)$$

$$IDF = \frac{1}{f}$$

Where *M* is the function that maps SSIDs into their geographic location. *X*and*Y*correspond to the fingerprints (the SSID lists) of device 1 (d_1) and device 2 (d_2). IDF is the weight of a *geo-location* which is inversely proportional to its frequency

Cheng, et al. [6] discussed that it is possible to detect more than 30% more relationships by computing the potential relationship of a pair of devices that their SSIDs can be mapped to a same location.

3.1.2 Spatio-temporal co-occurrence probability

Spatio-temporal co-occurrence is the probability of two users to be in the same place at the same time. The frequency and duration of the meeting of two users, can be a good indicator of a possible relationship or its strength. This is based on the fact that two persons with strong relationship (excluding online relationships) are expected to meet and share more time together compared to two unrelated individuals.

It is possible to construct a user spatio-temporal profile using a matrix where the columns represent location, the rows represent timeslots and each entry is the show up frequency for that specific user (MAC address). Measuring the *Spatio-temporal similarity* of two users profile can also be used as a complementary factor for relationship discovery.

Inferring social link can also be applied to scenario; where given a device MAC address and its fingerprints, we would like to identify potential linked devices with the target-device. For instance in criminal investigation where the MAC address and fingerprints of a potential suspect is collected, it is possible to identify other individuals which may have a social relationship with the suspect or targeted advertising

3.2 Wi-Fi Tracking

The fact that probes are broadcasted by wireless devices contain the device's MAC address it make it possible for an adversary not only to infer social links between people but actually track down a user's location in real time. Wi-Fi tracking is not a theoretical attack, on the contrast it is already used for commercial purposes. Passive Wi-Fi monitoring prototypes are used from shopping malls and museums to generate people analytics and log users' behaviour [3]. User profiling was even used by Renew Solutions during a smart-recycle scheme before the City of London halted the project due to privacy concerns [7]. RS had installed smart bins in several areas around London. Smart recycles bins were able to track pedestrians' MAC addresses and display targeted advertisement according to the user's recycling habits.

Below three Wi-Fi tracking strategies are analysed that show how Wi-Fi probes can leak information that allow down devices that eventually link to individuals exposing further concerns about mobile's users privacy. All attacks are using as unique identifier the wireless fingerprints of their devices. This does not necessarily mean that is only the MAC address but sometimes is the combination of characteristics and observations that make it unique like Operating System version, Firmware version and Signal strength.

3.2.1 Stalker Attack

As the author of the paper "*Targeted tracking of individual using Wi-Fi"* named it Stalker attack as the implementation requires actual stalking of the victim. It is based on the idea of continuously monitoring the wireless traffic by physically following the victim and log all the Wi-Fi traffic during that period. After the target is located, all the Wi-Fi traffic is monitored and logged for further identification. You can then simply identify the MAC address that is persistent in the logs for the entire amount of time during stalking.

In the case where the MAC address of the high value individual is already known, the attacker can also use "booby trap". A monitor that scans all channels and can alert when a detection of that specific MAC address occurs. This kind of mechanism it might be useful only for targeting high profile individuals.

The team has carried out a series of test on two of well-known and commonly used mobile devices, Samsung Galaxy SII and Apple iPhone 4S in order to identify the maximum distance that any frames can be received. Frames from iPhone 4S received from a distance of 30 metres and for Galaxy SII approximately 100 metres away.

Explicitly the application of the attack described above is fairly easy to be implemented but the main issue is not practicality but stealthiness of the adversary. As the attacker tries to stay in the transmission range by maintaining physical proximity to the target he identity can be exposed.

3.2.2 Beacon Replay Attack

The Beacon Replay attack was first presented in *"I know your MAC Address:Targeted tracking of individual using Wi-Fi"* paper and used the combination of Work/Home locations as a unique identifier. The author was inspired by the research work done by PhillipeGolle and Kurt Partidge "On

the anonymity of home/work location pairs". Golle and Partidge paper presented an analysis of the probability of tracking down people by assuming that a person who works in a certain area is highly unlikely to live in the same geographical location with another person working there.

The attack described can be divided into two main phases, first collect a set of SSIDs near the target's home and then replay at its workplace to identify that person. The first phase, *AP-Fingerprint*, was implemented by trying to harvest a large set of *personally identifying wireless networks (PIWN)* using a simple Wi-Fi network fingerprinter software. The tool will store information about the Wi-Fi networks detected to be used later in the second phase. The following figure 5.1 illustrates the attack.

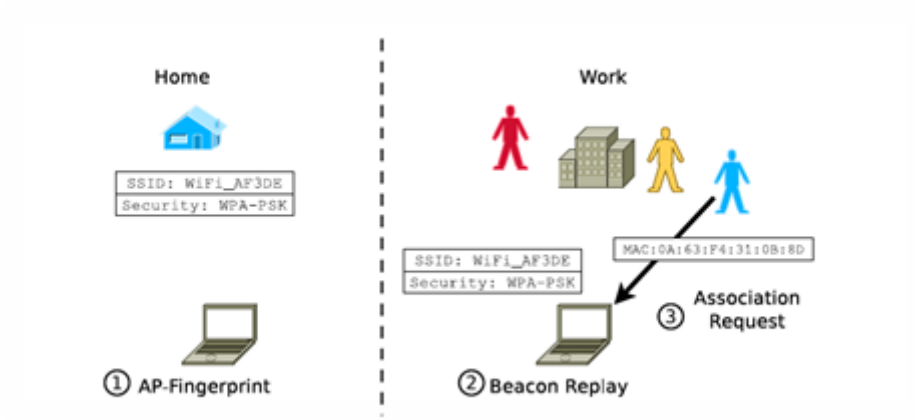

Figure 3.4 - Description of the attack phases

At the second part of the attack, *Beacon Replay*, the attacker creates fake APs in order to trigger reaction from the victim's mobile device. In simple words the list created during the first phase is replayed back using *Aircrack-ng* and *Base-ng* software. If the beacon replayer is in the range of the targeted device, the system will record the MAC addresses of any devices that will send association requests, try to connect to the rogue AP. Therefore creating a valid link between home and work locations of the mobile user thus creating privacy concerns.

The accuracy of this technique could be significantly improved if the signal strength is reduced since the transmission range will be decreased and the amount of APs at home location will be significantly less. On the other hand this will subsequently imply that the adversary will have to be closer to the victim which similarly with *Stalker Attack* raises stealthiness issues.

3.2.3 Estimating Smartphone Trajectories

Successful strategy relies on the efforts of tracking the path that an individual follows by calculating the most probable trajectory of his smartphone. A team of two researchers from the University of Illinois in their paper *"Tracking Unmodified Smartphones Using Wi-Fi Monitors"* expose a new invasive attack using low cost equipment and tools.

Similar with the previous attacks analysed in the previous sections the main unique identifier was based on the MAC address of the mobile device. They developed a system of sensors that scan all channels for Wi-Fi fingerprints and then passed to a central server for further processing. The system is clearly illustrated at Figure 5.2. Then by feeding the data to a selection of algorithms and

combinedwith some GPS ground truths they managed to model the movement of different station in the area which the experiment took place.

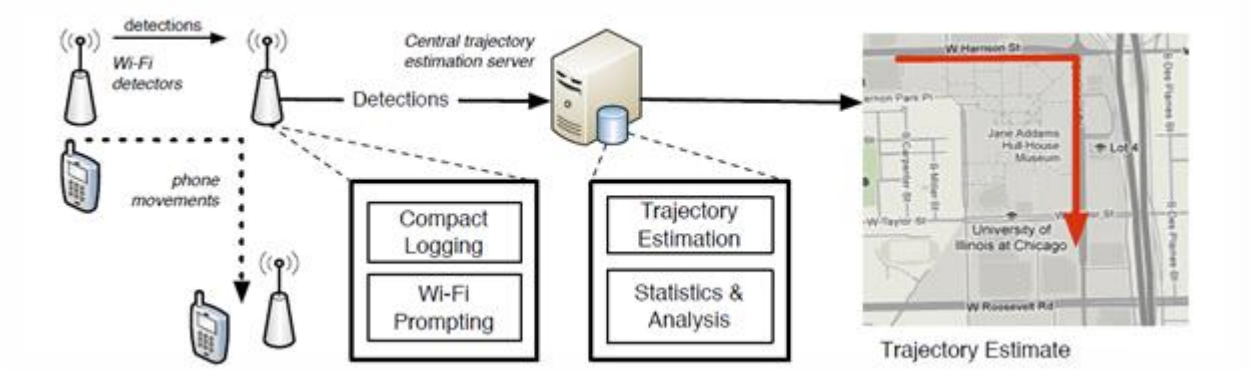

Figure 3.5 - Description of the Wi-Fi tracking system developed

The Naive approach is the simple approach whenever a hit is recorder you retain the record and the compared with the other sensors to match the overlaps and recreate a possible path. This solution is obviously inaccurate for a number of reasons. The signal strength fluctuates so much that a motionless device can been seen as moving around. Therefore even the use of Received Signal Strength (RSS) indicator does not provide reliable reading due to the dynamic shape of urban terrain.

In this scientific experiments security researchers took the project one step further. They applied a sequence of statistical algorithms to construct a model that will describe as accurate as possible the probabilistic path of the wireless devices in the area. Their main assumptions is that people either by walking or driving they use roads or walk paths. So they applied a spatial network, road map, to restrict the possible locations that a device could be. They also took into account any road limitations like turn restrictions and speed limits

Each observation from a monitor is recorded as state in time. Then it was modelled using hidden Makarov model (HHM). The data from sensors are used as an input to the model to calculate the probability of transition from one state to another.

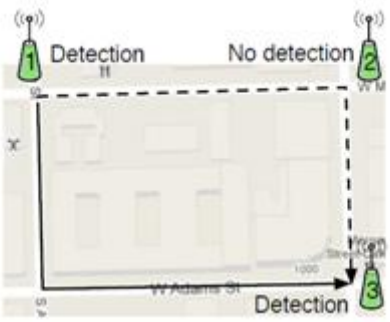

Figure 3.6 - Example of path probability

The Viterbi algorithm was then used to create a trajectory. The Viterbi algorithm was used to choose the path with the highest probability through a selection of Makarov states. The algorithm is applied after receiving no further transmission. In simple words the algorithm will decide which is the most probable path the user has taken considering the previous states that he has visited. As seen in Figure 5.3.

Three main strategies where used to increase the packet flow and the number of targeted Wi-Fi devices. This aims to minimise the error percentage and maximise the packet transmissions from each phone.

Broadcast Popular SSIDs

Network providers usually have many Wi-Fi hotspots to accommodate fast and reliable internet access for their customers but on the same time decrease the workload from the mobile broadband. A clever way to exploit this fact is by advertising popular SSIDs names like *eduroam, O2 Wifi*and*VirginMediaWiFi*. In this way they raise the possibilities to detect more devices if any of the users has previously connected to these Wi-Fi networks since a device by responding to the broadcast request will reveal it's MAC Address.

Create Fake APs

As it was already covered before, some phones when out of range will send probe requests to connect to Wi-Fi networks that are listed in their CNL list. During experiment they were constantly scanning for station's probe request and record the SSIDs information included. Then by spoofing the SSID networks that previously logged they have managed to increase the traffic between sensors and the stations. However one factor for a successful connection is the encryption scheme used by the network and this bit of information is not broadcasted in the probes. If the emulated network does not match with the one that the device previously had connected then no connection will be attempted. To maximise the number of connections they created multiple fake networks with the same SSID but different security protocols.

Send RTS/CTS messages

RTS/CTS stands for Request to Send/Clear to send is a duple of common commands found in the 802.11 wireless protocol [12]. Every time a station receives an RTS request it should respond with a CTS packet. The problem with the particular technique is that the station will not broadcast its MAC address in the CTS response instead the station will use the RTS address to respond. They overcame this issue by issuing a different MAC address for each RTS request in order to track down the device.

Their evaluation process included GPS data collected from mobile devices of several volunteers has verified the success of the attack. Therefore it was proven that using statistical algorithms you can provide an approximation of 70 metres from the real trajectory of the users traced [12].

3.3 Other issues

As seen previously the Wi-Fi specification itself is vulnerable to a series of attacks and techniques that compromise the privacy of the users. However there are other issue that could potentially allow an attacker leverage more information. [2] For example, difference in the physical properties between various signals can help identify radio-frequency based devices. [2] Another example, relevant to the fingerprinting analysis of paper [2] is the use of clock skews caused by small deviations between the clock oscillators (circuits that produces pulses to pace the system of a machine) of different machines. This metric has been found to create a suitable fingerprint to distinguish wired physical machines. [8] These skews can be observed from the network: the fingerprinting of the device is based on the approximate clock skew between the target and the observer by collecting timestamps from the TCP header of the packets sent from the target. [8] All of the aforementioned characteristics can serve to create different fingerprints to attempt to distinguish devices and possibly identify the users behind them.

CHAPTER 4 - DISCUSSION

The order in which the *probe* requests are sent from the device depends on the operating system and in some cases the device vendors, for instance some OS sends in the order of preferred SSID, whereas others send based on the most recent to oldest. This could introduce difficulties which could limit the amount of probes captured for the analysis, since the device stops sending probes once is connected to an AP. On the other hand, some could argue the order could increase the accuracy in some analysis, for instance increasing the confidences in inferring a social link since it is also possible to use the latest accessed APs as additional classifier to improve the results

4.1 Countermeasures

Throughout this paper we described several issues which violation the user's privacy, such smartphone trajectory or inferring user relationship. This section will focus on proposing and analysing few countermeasures which could eliminate or limit the privacy violation.

4.1.1 Active service discovery mode restrictions

One of the most obvious solutions that can be implemented in the smartphones is to disable or reduce the frequency in which probe requests are broadcasted. This method seems to be reasonable due to the fact that; it can solve the problem with smartphones constantly advertising its presence. However, there are several issues with this measure. First, implementing this method might cause a performance penalty which seems to be an undesirable feature for a smartphone phone. The second problem is that, this measure might prevent the device from associating with APs with hidden SSIDs. The most important issue, however, is that disabling active probe is not effective against all attacks(e.g. stalker attack). Furthermore reducing the frequency in which the probe requests are sent, can only postpone the problem. In this regard an attacker only need to stay long enough within the person's perimeter to capture the probes.

4.1.2 Blindprobe requests (broadcast)

As mentioned in the background section in active mode discovery, it is possible for a device to sends out a broadcast message with the SSID field left empty to get a broadcast beacon from the alive AP. This could be an effective technique countering smartphone trajectory attack or social link detection. In this regard an adversary would not be able to obtain the list of networks the user's device trusts and therefore the attacker will have difficulties to carry out his attack. The implementation of such a modification will require changes in the software for both sides which can lead to significant delays of the adoption of such method. Some operating systems like Android or IOS (new versions) are already using this modification. However, the old versions of different Android systems are still vulnerable and can be compromised. On the other hand this technique is ineffective against other type of attacks such as stalker attack. That is due to the fact that, even though the smartphone is no longer

broadcasting its SSID it is still actively advertising itself and can be uniquely identified by its MAC address.

4.1.3 MAC address spoofing

MAC address spoofing is another approach to counter the privacy issues covered in this paper. Changing the MAC address to a random value requires the installation and authorisation of a third party mobile application. There are several issues that might appear regarding the implementation of such solution. The first concern is that performing that action will require root address that will maybe introduce other security issues. Furthermore other techniques could be still used to fingerprint mobile devices and distinguish spoofed MAC addresses. For instance as suggested in [9], the timing pattern of probe messages can be used to identify the wireless interface's driver, and this information can be used to check if manufacturing vendor suggested by the MAC address matches the founding.

4.1.4 Geofencing

The last privacy preserving technique that needs to be mentioned is geofencing. The main idea behind it is that user should enable its Wi-Fi mode only in trusted areas e.g. work or at home. This method can eliminate some potential attackers that might be capturing in the tube or in other public places, but it is not so effective against the stalker attack were the adversary is targeting and following a specific person. There is also the possibility of an inside attacker in the working place which makes geofencing completely irrelevant to use.

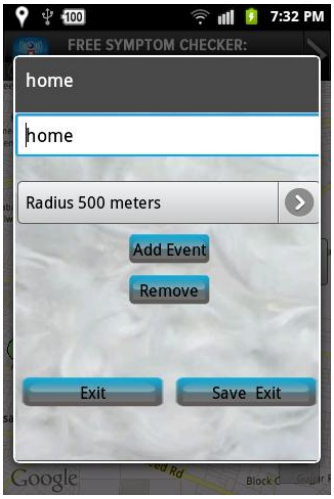

Figure 4.1 - "Geofence Free Tracker" is a free GeoFencing Android Application [10]

4.1.5 IEEE802.11 protocol modification

Another type of solution which can be deployed is related to the 802.11 protocol. Encrypting or obfuscating all identifiers is one of the modifications that can be applied to improve the security of smartphones. Implementing such techniques which require significant changes in the protocol might cause multiple difficulties not only for the APs but also for the devices. An example of this can be the backwards-compatibility with the older systems.

Taking into account all security measures that were listed above it is important to mention that there is not an ideal solution which can fix all existing vulnerabilities in the Wi-Fi protocol. Using different measures together seems to have the most promising effects on the system. Such techniques as MAC spoofing and sending probes with blank SSID field can be implemented at the same time for achieving the better result. However, there is no guarantee that introducing some new application on MAC spoofing will not trigger additional issues that have not been discovered yet.

CHAPTER 5 - CONCLUSION

As we have seen throughout the analysis, there are multiple privacy-related concerns and threats that are associated with the specification of the 802.11 family of protocols. These threats are as we have seen severe and range from disclosure of the identity of the user-target to tracking its movements, identify its working and living addresses etc. What is even more alarming is that most of these attacks do not require any high-cost specific equipment or knowledge but can me mounted with open-source tools (e.g. Airmon-ng, Wireshark) and off-the-shelf hardware (e.g. USB wireless interfaces).

In general, the family of privacy-related attacks we have described is based in one realistic assumption: first that the users constantly carry with them Wi-Fi enabled devices (i.e. smart-phones). Nowadays, users tend to be so attached to their smart-phones that the phone's MAC address tends to be to MAC address of the user in a sense. Furthermore, the dataset matching attacks are based on one additional, realistic as we have seen, assumption: that people that are socially connected (e.g. family, friends, co-workers) tend to share one or several unique network SSIDs.

The attacks we have described can cover a broad range of applications: individual targeting (e.g. stalker attack, physical-location) or crowd targeting (e.g. identifying relations between individuals by their CNL).

As we have seen in Section 3.2, the attacks we have discussed can be used for malicious as well as commercial purposes. However, although the techniques applied by these institutions for commercial purposes are of non-intrusive purpose, they raise some ethical concerns: whether the users are aware that they are being tracked and if they willing in order to shop to have their MAC address tracked.

In terms of countermeasures, there is some progress being done by the vendors but this progress as we have seen covers only a part of these attacks. Furthermore, even though in theory these security issues have been dealt with, in practice the users are still vulnerable as their devices are not upgraded.

As we have showed throughout this report, the techniques of privacy-related attacks we analyzed are usually very easy to mount, very cheap and have been occurring in the users' everyday life for malicious or even commercial purposes. The amount of information a willing adversary can extract for a target-user is substantial and therefore these security gaps should be dealt with seriously. There has been progress in this area of security by the technology vendors but still the attacks remain potent; the users as we have seen are still vulnerable from some of the attacks. We eagerly expect more research to be done in this field that will result in the mitigation of the attacks, either by adjusting the Wi-Fi protocol to cover the gaps or by reinventing it.

REFERENCES

[1] Pew Research, “Cell Phone Activities 2013” URL: http://www.pewinternet.org/2013/09/19/cell-phone-activities-2013/ [Accessed 22/03/2014]

[2] Desmond, Loh Chin Choong, et al. "Identifying unique devices through wireless fingerprinting." *Proceedings of the first ACM conference on Wireless network security*. ACM, 2008.

[3] Cunche, Mathieu. "I know your MAC Address: Targeted tracking of individual using Wi-Fi." *Journal of Computer Virology and Hacking Techniques* (2013): 1-9.

[4] Sidiropoulos, Nikos, M. Mioduszewski, P. Oljasz, and Edwin Schaap. "Open Wifi SSID Broadcast vulnerability." (2012).

[5] M. Cunche, M. A. Kaafar, and R. Boreli, “I know who you will meet this evening! linking wireless devices using wi-fi probe requests,” in *13th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2012.

[6] Cheng, Ningning, et al. "Inferring user relationship from hidden information in WLANS." MILITARY COMMUNICATIONS CONFERENCE, 2012-MILCOM 2012. IEEE, 2012.

[7] BBC news “City of Lonwon calls halt to smartphone tracking bins” URL: http://www.bbc.co.uk/news/technology-23665490. [Accessed 25/03/2014]

[8] Tadayoshi Kohno, Andre Broido, K.C. Claffy. Remote Physical Device Fingerprinting. In Proceedings of the 2005 IEEE Symposium on Security and Privacy (SP 2005), Washington, DC, USA.

[9] Jason Franklin, Damon McCoy, Parisa Tabriz, Vicentiu Neagoe, Jamie Van Randwyk, and Douglas Sicker. Passive data link layer 802.11 wireless device driver fingerprinting. In Proceedings of the 15th conference on USENIX Security Symposium - Volume 15, Berkeley, CA, USA, 2006. USENIX Association.

[10] https://play.google.com/store/apps/details?id=sh.geofence.tracker.tasker

[11] WiGLE, Wireless Geogrphic Logging Engine. http://wigle.net/

[12] Musa, A. B. M., and Jakob Eriksson. "Tracking unmodified smartphones using wi-fi monitors." *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 2012