

SVEUČILIŠTE U SPLITU
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I
BRODOGRADNJE

POSLIJEDIPLOMSKI DOKTORSKI STUDIJ ELEKTROTEHNIKE I
INFORMACIJSKE TEHNOLOGIJE

KVALIFIKACIJSKI ISPIT

LOCATION PRIVACY AND WIFI NETWORKS

Ante Dagelić

Split, Rujan 2019.

Contents

- 1 Introduction** **1**

- 2 WiFi connection initialization protocols** **3**
 - 2.1 WiFi networks and standards 3
 - 2.2 Passive Service Discovery 6
 - 2.3 Active Service Discovery with Broadcast 7
 - 2.4 Active Service Discovery 9

- 3 WiFi localization techniques** **11**
 - 3.1 Current localization 11
 - 3.2 Dynamic localization 17
 - 3.3 Previous whereabouts and analytics 21

- 4 Current and future work** **25**

- 5 Conclusion** **29**

- BIBLIOGRAPHY** **31**

- Labels** **36**
 - Abstract 37

1. Introduction

One of the most challenging problems in today's IoT mobile era is location privacy. Different mobile devices, such as smartphones, tablets, wearable gadgets and more, constantly collect different information from its surrounding area. That information can be then used to develop different technologies such as WiFi based positioning systems [1]. Other examples are radio signals emitted by devices connected to different mobile carrier networks where you determine one's location using cell tower trilateration [2]. Data collected this way has a lot of usages on the market, not only being a huge threat to the actual security of the user being followed, but also as a big data research tool, indoor tracking, localization or as a marketing tool delivering targeted services based on location or other profiled data for a particular user.

Carrying a wireless connected device (e.g. a smartphone) makes the user act as a portable beacon. Data transmitted over the wireless channel can be monitored by third parties. Even when applying the latest encryption algorithms to the transmitted message content, the data monitored is still relevant and can threaten your location privacy. This threat is furthermore increased by today's omnipresence of WiFi networks. The authors of [3] show an increasing number of WiFi enabled devices and total traffic transmitted over WiFi networks, as can be seen from Figure 1.1.

There have been a number of papers discussing the potential threats to user privacy

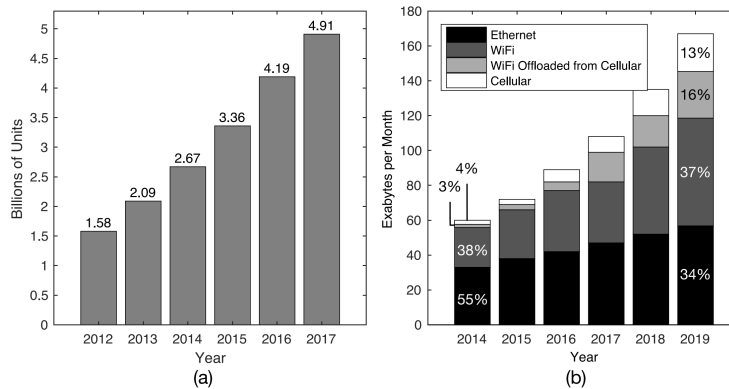


Figure 1.1. (a) Increase in the number of WiFi-enabled devices shipped during 2012 to 2017. (b) Traffic growth of WiFi, cellular and Ethernet networks during 2014 to 2019 [3]

upon collecting different management frames from mobile devices (beacon frames containing an access points service set identifier SSID) collected and used in Google services [4]. This technique is used by most of the devices in order to discover nearby access points to connect to. Aggregating and managing these probe requests can lead us to finding out the users Preferred Network List - PNL which contains the list of all the previous SSIDs his/hers device had been connected to in the past. This data can then be used to figure out the users previous whereabouts, since a large number of SSID's can be geologically located (e.g. using WIGLE service [5]). Additionally different social relations, or habits can also be discovered [6, 7, 8].

Mobile device manufacturers and the leading standardization institutions have of course recognized location privacy as a major issue, and in terms of e.g. WiFi, developed certain modifications to the connection establishing protocols. This is called passive scanning technique, which is slower and less user-friendly than active scanning techniques. Based on that experience, they have developed a hybrid of the two, also in the active group, which does not leak the PNL and still establishes quick connections. This mode is used in most of the todays mobile devices and is the proposed solution for hiding the contents of ones PNL.

In this paper we provide an overview of existing scientific work on location privacy within WiFi networks. In Chapter 2 we overview WiFi authentication protocols, which are the basis of most localization vulnerabilities. In Chapter 3 we categorize the existing work on WiFi location privacy into three categories: current localization, dynamic localization and previous whereabouts, followed by Chapter 4 where we briefly cover our current and future work. Finally, we conclude in Chapter 5.

2. WiFi connection initialization protocols

In this chapter we will briefly overview WiFi as a technology, primarily focusing on the first step in WiFi communication - connection initialization which is the basis of our scientific work. We start by defining WiFi networks and their operation and then move to different service discovery protocols used for connection initialization. Please note that inadequately configured WiFi networks contain a wide range of security vulnerabilities as given in [9]. However, in our work we focus on location privacy vulnerabilities which are to the best of our knowledge exploited the most by targeting the WiFi connection initialization protocols, thus a more detailed technical overview is only given for the connection initialization protocols.

2.1. WiFi networks and standards

WiFi is a commonly used as a name for a family of radio technologies used for wireless local area networking based on IEEE 802.11 standards. Multiple parts of IEEE 802.11 protocols are used and it is intended for the technology to work seamlessly with wired protocol Ethernet [10]. Using wired connections, multiple devices can exchange network packets either directly or through a series of network switches as can be seen in Figure 2.1. Such networks can be upgraded to support the wireless transmission of packets through the usage of electromagnetic waves which can greatly increase the simpleness and ease of setup for the clients, as can be seen in Figure 2.2.

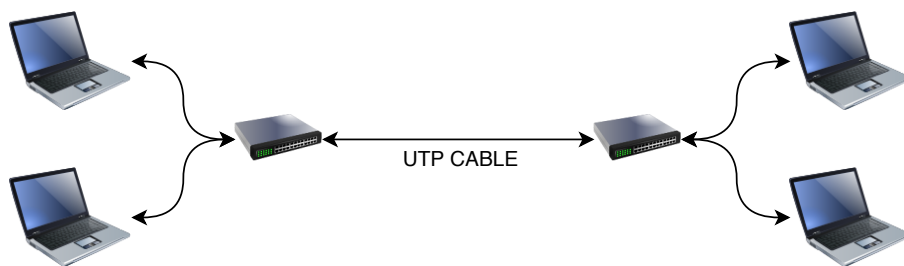


Figure 2.1. Classical wired networks

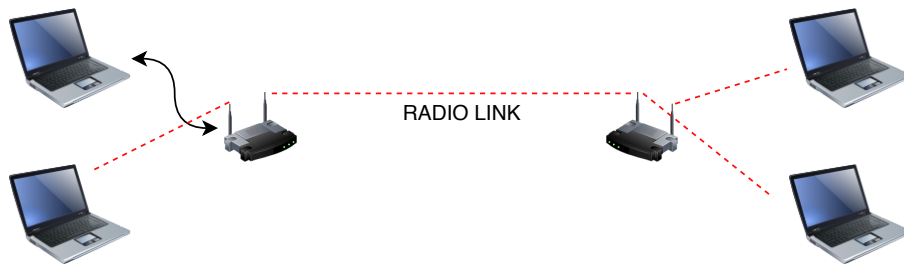


Figure 2.2. Wireless Local Area Networking - WLAN

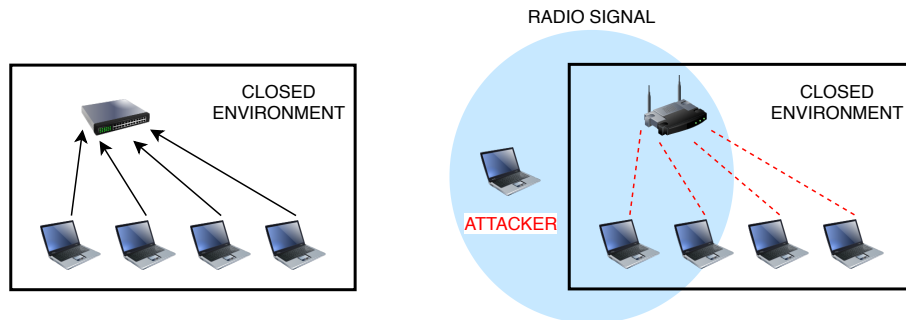


Figure 2.3. Wired vs Wireless attacker advantage

WiFi clients range from notebooks, smartphones, tablets, smartwatches to various IoT sensing devices, internet connected house appliances, fitness trackers and many others. Although the WiFi is a great technology bringing many improvements for the users, in security terms things are not so optimistic. Contrary to the wired networks where the attacker can compromise the victim only by having a wired connection to the network, WiFi and other wireless networking technologies provide the attacker with the option to do malicious actions by simply being in the range of the WiFi Access Point or users WiFi enabled device, which is a simpler feat as depicted in Figure 2.3.

IEEE 802.11 standards

IEEE 802.11 specifies the Media Access Control (MAC) and physical layer (PHY) protocols for wireless local area network (WLAN) communication [11]. It is a part of IEEE 802 set of protocols defined for local area networks. IEEE 802.11 includes communication in various frequencies, out of which the most used ones in WiFi networks are 2.4GHz and 5GHz, with the former having a higher usage percentage [12].

2.400-2.500 GHz Industrial Scientific and Medical (ISM) radio band spectrum is used by 802.11b, 802.11g, and 802.11n-2.4. 802.11a, 802.11n and 802.11ac use the more heavily regulated 4.915–5.825 GHz band. Both of the commonly called 2.4GHz and 5GHz WiFi spectrum are divided into multiple channels each with certain frequency and bandwidth. In Figure 2.4 a graphical representation of the 2.4GHz spectrum is given. There is a total of 14

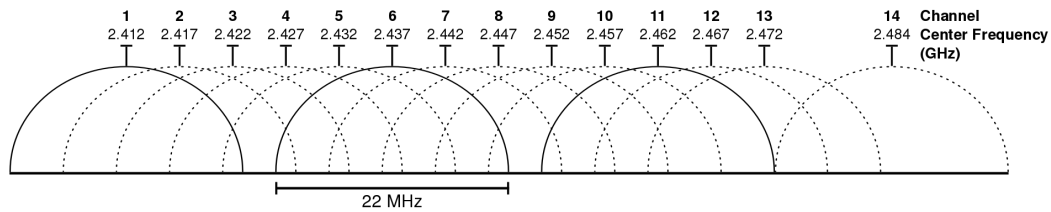


Figure 2.4. Graphical representation of WiFi channels in the 2.4GHz band [11]

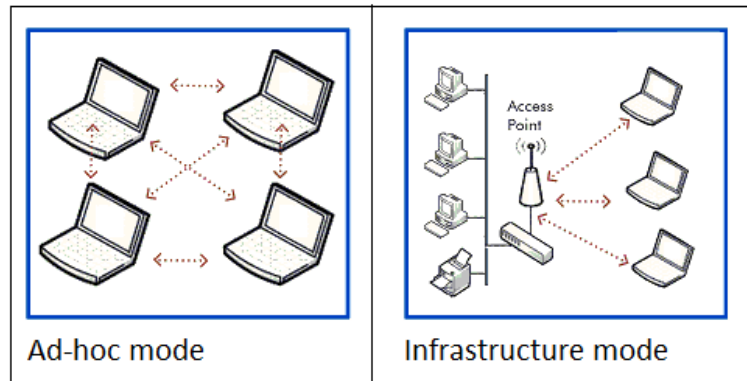


Figure 2.5. Ad hoc mode vs Infrastructure mode [14]

channels, with a bandwidth of 22MHz, where the higher numbered channels sometimes occur with additional restrictions based on a domain. As can be seen from Figure 2.4, the channels are overlapping, and interference is possible which can result in lower signal quality. Different regulatory domains allow the networking equipment to use only certain channels (i.e. channel #14 is not available in North America, but it is in Europe).

For our work, more details about channel operations and collisions are not important, just the fact that only two devices in range of each other can communicate on one channel at a time. The same way, an attacker using one WiFi card can interfere in some matter with only one channel at a time.

Ad hoc and Infrastructure mode

WiFi networks are generally used in one of two modes Ad hoc and Infrastructure mode [13]. In Ad hoc mode, WiFi networks communicate with each other directly using their WiFi cards, and no middle device is required. Each device's WiFi card needs to be configured to use the same Service Set Identifier (SSID) - name of the WiFi network, as well as use the same WiFi Channel. Ad hoc network is simple to setup and is usually used when no Access Points are available and a wireless exchange of data is required.

In infrastructure mode, WiFi devices are connected to a wireless Access Point which

is the central point switching the data traffic from one device to another. A graphical representation of both Ad hoc and Infrastructure mode WiFi networks are given in Figure 2.5. WiFi networks in infrastructure mode are an extension of a bigger wired network and provide WiFi users with the benefit of using the wired infrastructure often including an access to Internet gateway. The Access Point is given an SSID and allows the devices to exchange data through it on a predefined channel. Wireless Access Points are usually either pre-configured to work on a free channel, or have the option to scan the channel occupancy in a certain time range and pick the optimal channel themselves. WiFi networks setup in infrastructure mode are the preferred setup ranging from industry and events to home networks.

There are three main ways a WiFi enabled device can connect to an Access Point: using Active Service Discovery, using Active Service Discovery with Broadcast or by using Passive Service Discovery. The former two are similar in approach, but provide different advantages for an attacker, whereas the latter uses a different approach. In the following sections we will focus on how each service discovery works and which advantages the attacker can gain.

2.2. Passive Service Discovery

WiFi enabled devices using passive service discovery scan for Access Points in their vicinity, and if an Access Point the device has previously connected to has been discovered, the device will initiate the Authorization process as show in Figure 2.6. Due to energy savings, the device can not constantly scan for neighboring Access Points, so it uses intervals of various length (depending on current device activities). The same way, an Access Point can not constantly transmit the Beacons towards WiFi enabled devices. All of this in most cases makes Passive Service Discovery the slowest of the three connection protocols presented in our work.

Security-wise, as will be seen in Chapter 3, passive service discovery is the most secure of the three, considering the device is not actively transmitting any data the attacker can use to disclose private information about the user. This means if the attacker would just simply passively monitor the WiFi traffic, no data would be picked up from the victims device. However, as we will present later, active attacks can still be performed, mainly using the Karma attack [15, 16] where the attacker mounts a fake Access Point which constantly transmits Beacon packets often changing its Service Set Identifier and trying to provoke the device to attempt Authentication.

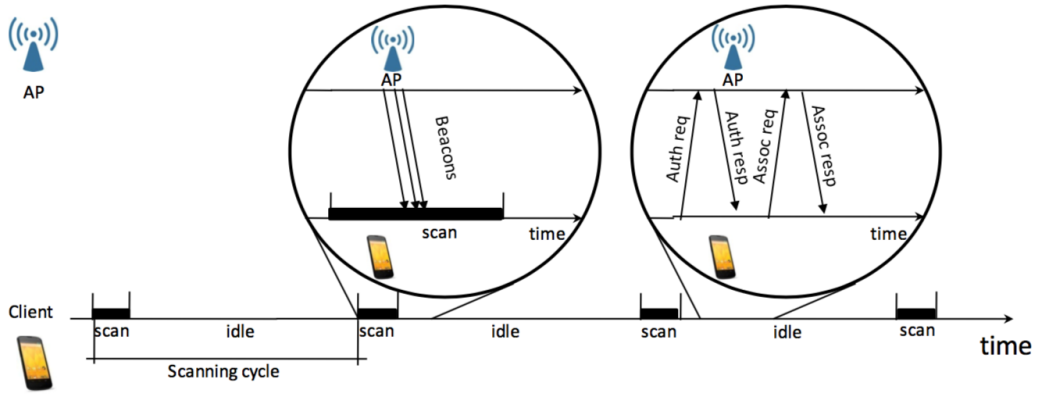


Figure 2.6. WiFi enabled device connection process in Passive Service Discovery

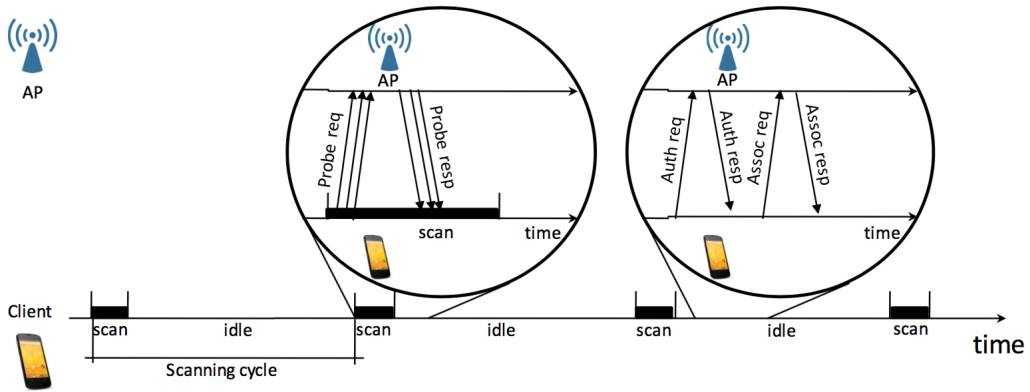


Figure 2.7. WiFi enabled device connection process in Active Service Discovery

2.3. Active Service Discovery with Broadcast

Unlike the passive service discovery, devices using Active Service Discovery are actively trying to connect to Access Points in their vicinity. The WiFi enabled device will, once again in certain periods, transmit Probe Request packets waiting for the neighboring Access Points to respond with Probe Response packet [17]. If the device has previously connected to the Access Point, the Authentication will begin, as shown in Figure 2.7. The authors of [18] and [19] show the ratio of scanning intervals in active service discovery for various device manufacturers and operating systems. A CDF graph about average scanning intervals has been given in Figure 2.8

The contents of the Probe Request packet sent by the device is shown in Figure 2.9. When active service discovery is used with Broadcast packets, the Probe Request packet sent by the device will hold an empty SSID field within the Frame Body block (containing “Broadcast” string).

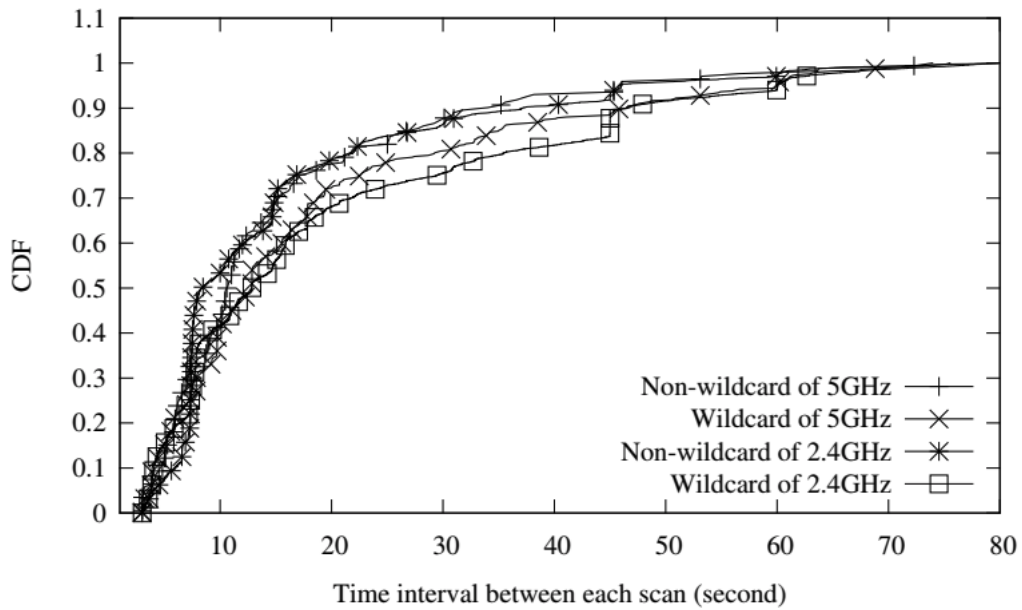


Figure 2.8. Time interval between each scan for Active Service Discovery with Broadcast (Wildcard) and Active Service Discovery (Non-wildcard) [19]

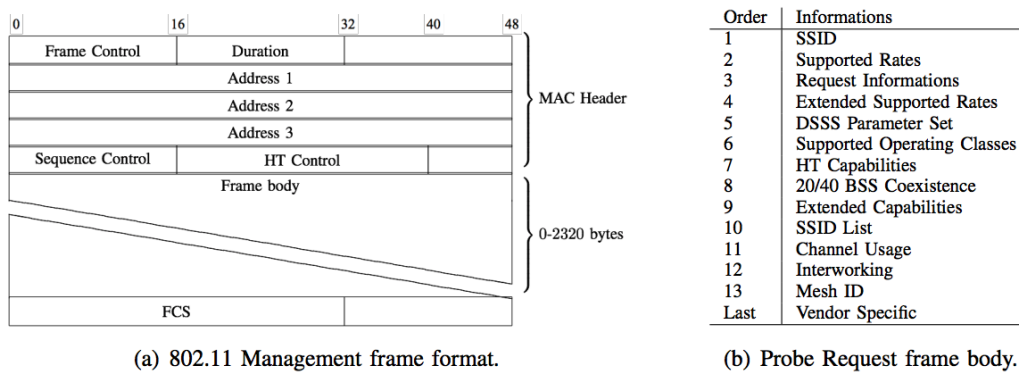


Figure 2.9. Management frame format and body of a probe request frame in the 802.11 protocol [20]

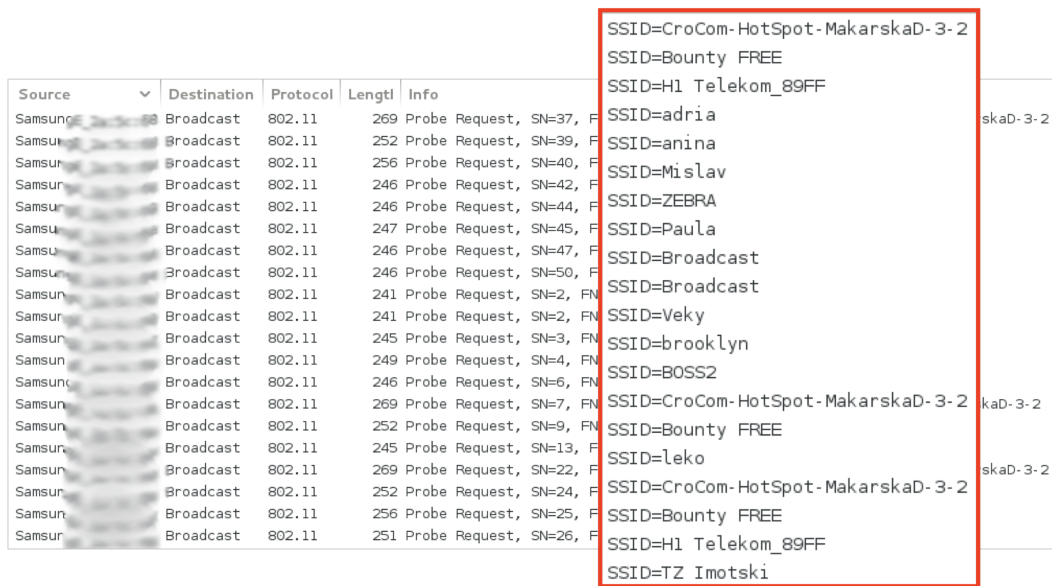


Figure 2.10. Example trace of Probe Request packets in Active Service Discovery [21]

Security-wise, active service discovery with broadcast can mainly be exploited by the attacker in order to assert the presence of the victims device. By using a number of strategically positioned sensing devices which are monitoring WiFi traffic, the attacker can conclude at what time a certain device has been on a certain place, thus violating users location privacy. Even though such attack requires a certain amount of preparation (guessing the area a potential victim will visit) its implications are still significant, as the victim does not know s/he is being monitored. More on the actual implications and techniques is given in Chapter 3.

2.4. Active Service Discovery

WiFi enabled devices using Active Service Discovery are using a similar approach as Active Service Discovery presented in Subsection 2.3 and Figure 2.7. However, unlike in the case of Broadcast usage where each Probe Request packet contains an empty SSID in the frame body as presented in Figure 2.9, in Active Service Discovery the WiFi device will send out the SSID of an Access Point it has previously connected to. The device will send out multiple probe request packets in one scanning interval, each containing a different Access Point's SSID.

An attacker can compromise the security of the WiFi user using Active Service Discovery by simply monitoring the WiFi channel for Probe Request packets and reveal the WiFi user's Preferred Network List (PNL) - the list of user's previously used Access Points. This can be done due to the flaw in WiFi connection protocol which sends the SSID string in clear. An example trace of collected Probe Request packets is given in Figure 2.10.

Although finding out the Access Points an user has connected to before may look like a weak security implication, researchers have proven otherwise. There is a wide range of different implications on user's location privacy such as geographically matching the SSIDs and revealing where the user has previously been, using the similarities in two user's PNL to conclude their social relations, determining user's sociological profile and many other. In the following chapter we will present the related work in this field by categorizing the existing work in 3 categories: current localization, dynamic localization and previous whereabouts.

3. WiFi localization techniques

In this chapter we will overview the current research on location privacy related topics within WiFi networks. We have divided the scientific publications into three categories: current localization, dynamic localization and previous whereabouts and analytics. Most publications can be categorized this way, where as some publications will be mentioned multiple times as their work fits two or even all categories. Please note that the entire scientific field is much larger than work overview in this chapter, WiFi in general has been researched a lot in the security and privacy field, as well as other wireless networks and various data sources in location privacy field. However we are focusing only on the intersection of WiFi networks and location privacy.

3.1. Current localization

By the term *Current localization* we categorize all the work concerning positioning and direction of movement estimations. Current Received Signal Strength Indicator (RSSI) is often used to calculate the position of a transmitting WiFi network device, as well as the time required for the signal to reach the device. WiFi enabled devices transmit packets in the WiFi bandwidth, and such packets can be monitored by the attacker as can be seen from Figure 3.2. The contents of the packet are not even important when attempting to position a user, only each packets signal strength. Even if the user is currently not connected to a WiFi network, and even is the device is not being used at the moment, it is still transmitting data on the WiFi channel, as the authors of [22] have concluded. This fact further more contributes to the ability to perform passive positioning, as even a pocketed smartphone device can be positioned.

By using multiple monitoring devices set through an area, the attacker can use the signal strength received by each sensing device to triangulate the physical position of a person as depicted in Figure 3.1. Even though WiFi positioning might seam like a simple simple setup, it is hardly so. The signal strength of a transmitting device varies depending on the path loss and indoor spaces are often filled with various obstacles, each having its own contribution to signal strength variation. Therefore, using a simple triangulation will not give valid results due to high amount of reflection, diffraction and scattering as well as differences in WiFi Chipsets

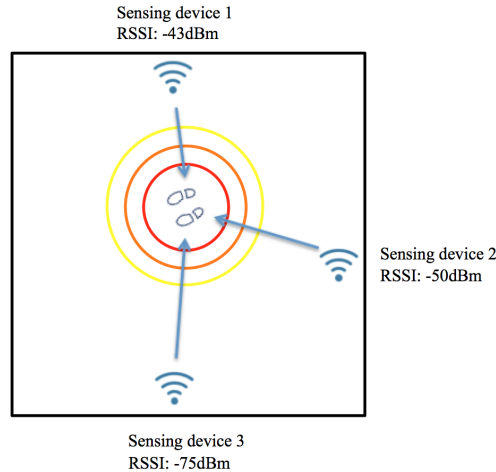


Figure 3.1. Using RSSI to triangulate the physical position of a person

```

▼ 802.11 radio information
  PHY type: 802.11g (6)
  Short preamble: False
  Proprietary mode: None (0)
  Data rate: 1.0 Mb/s
  Channel: 5
  Frequency: 2432MHz
  Signal strength (dBm): -51dBm
  Noise level (dBm): -98dBm
  TSF timestamp: 328268717
  ► [Duration: 2952µs]

```

Figure 3.2. Example trace of 802.11 radio information presented in Wireshark Tool [24], RSSI highlighted

[23]. There are better solutions for indoor tracking at the market, however they require the users to wear some type of a tag which is being tracked, so researchers have invested time to improve WiFi positioning accuracy.

The authors of [25] have developed a practical path loss model for indoor WiFi positioning enhancement, achieving the mean error in the distance estimation of 2.3m and 2.9m for line of sight and non line of sight environments, respectively. The authors consider the real life situation where the environment often does not have an unobstructed LOS (Line of sight) path between the sensing devices and the device being tracked. Many moving and still objects will greatly contribute to signal strength variation, so the signal strength is not dependent on distance alone. In [25] the Hata Okumara model [26] for signal propagation has been used, as given in Eq 3.1.

$$\log d = \frac{1}{10n} (P_{TX} - P_{RX} + G_{TX} + G_{RX} - X_\lambda + 20 \log \lambda - 20 \log(4\pi)) \quad (3.1)$$

Where in Eq 3.1, d represents the estimated distance between the transmitter and receiver. P_{TX} (dBm) and P_{RX} represent the transmitted power level and measurer power level respectively. Similarly, G_{TX} (dBi) is the antennae gain of the transmitter and G_{RX} the antennae gain of the receiver. λ (m) denotes the wavelength, n is the measure of the obstacles influence and X_λ is a normal random variable with a standard deviation of λ .

In their work, the authors of [25] conclude that for the middle 802.11b channel the frequency of 2442 MHz results in $\lambda = 0.12m$ and the standard deviation X_λ is in the range of 3 dB to 20 dB depending on the building construction. The measure of obstacle influence n is 2 for free space, and for obstructed paths it is in between 4 and 5. The authors conclude that a prototype should be developed to estimate n and to check if the antenna parameters of both transmitter and receiver are given correctly in the antennae documentation. In Figure 3.3 the results of Line Of Sight distance measuring using the Hata Okumara model on WiFi networks is given. To account for overestimations at lower distances and underestimation at higher distances as well as non LOS conditions, the authors use a multi model approach with different n values for lower and higher distances and achieve the mean error in the distance of 2.3m for LOS and 2.9m for non LOS contitions.

Local principal gradient direction is used to improve the WiFi indoor positioning in [27]. The authors have developed a mesh of indoor sensing calibration devices each with its own principal gradient direction. Then, distance correlation is used in order to discover nearest calibration points for the device being positioned. Finally, weighted squared Euclidean distance between the nearest calibration point is used to estimate the position of the device.

Mapping the target area and estimating the RSSI has been a strategy of choice in [28]. Recurrent Neural Network (RNN) and Long Short Term Memory (LSMT) is used as a deep learning technique in order to position the WiFi device. The authors conclude to have achieved a 99.7% accuracy when predicting which floor the device belongs to, with distance errors in the 2.5m to 2.7m range.

The authors conclude that adding multiple layers in the RNN and LSTM algorithm provides little improvement, thought increasing the required training and testing time. Similarly, in [29] the authors train the system with RSSI readings beforehand and use the data to perform real time positioning. Deep learning has also been used in [30] to estimate movement of human body based on WiFi channel state information.

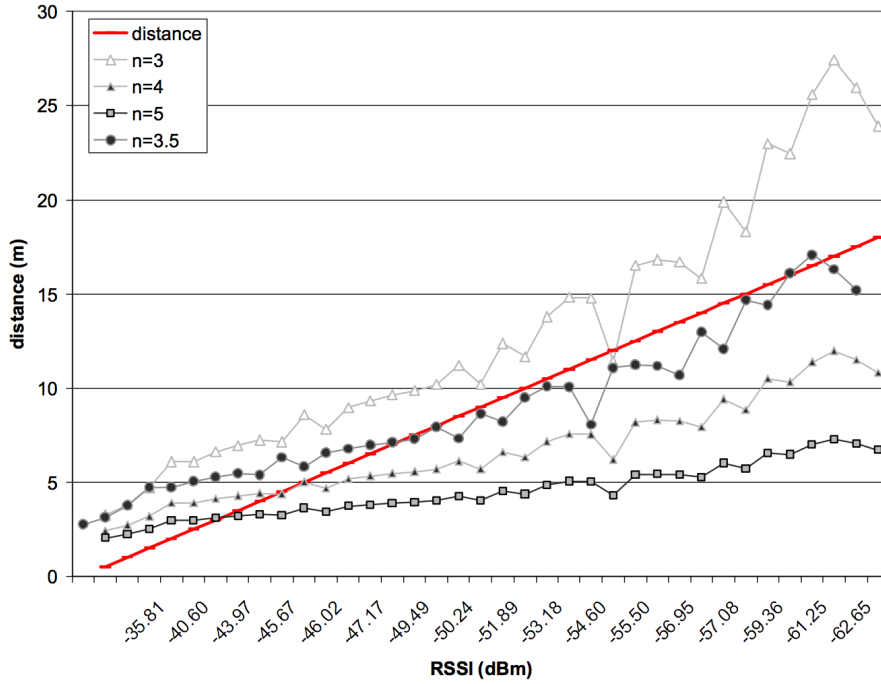


Figure 3.3. Choosing the parameters of Hata-Okumara model to measure the distance between WiFi transmitter and receiver in LOS conditions [25]

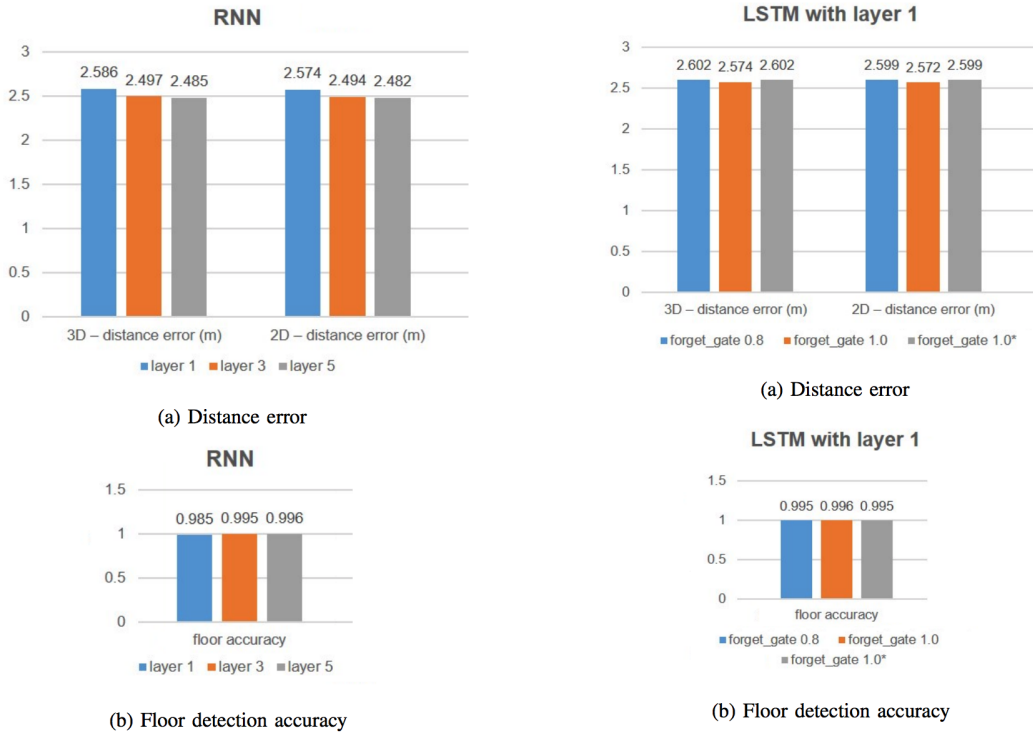


Figure 3.4. Distance error and floor detection accuracy by using RNN (left) and LSTM(right) [28]

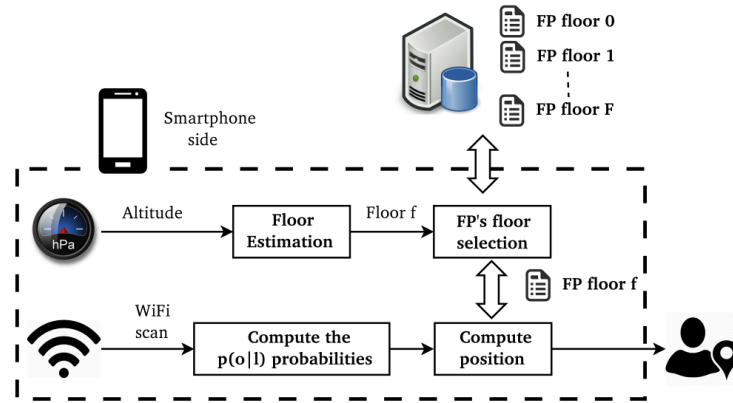


Figure 3.5. The logical scheme of the proposed 3D indoor algorithm using WiFi data and barometer information [31]

Additional information can also be included in order to further increase the success rate of WiFi indoor positioning. An interesting approach has been made by the authors of [31] where barometer information has been used along side data gathered by 5 Access Points in order to position the WiFi user in 3D space. The authors have conducted the research at University of Genoa and University of Bologna and were able to accurately localize the user with an error below 1.2m. A new 3D indoor localization algorithm has been introduced, whose logical scheme is given in Figure 3.5 and results in Figure 3.6. The conclusion is that the consideration of barometer information impacts the positioning success by up to 22% when more than 5 APs are used.

Drones carrying WiFi monitoring devices have been proposed to perform search and rescue operations in [32]. The authors have proven that such a sensing device can reliably locate a WiFi device on distances up to 200m as can be seen from Figure 3.7.

A number of other papers are present on the topic of WiFi based positioning. Weighted kernel assisted Bayes algorithm is used for localization based on received probe request packets has been used in [33]. In [34], the authors are using the data gathered from inertial sensor to correct WiFi indoor positioning results, using a variety of approaches including the Closest neighbor algorithm, Kalman filter, Particle filter and others. Using inertial sensor data has proven to decrease the error by up to 33%. Similarly, in [35] gyroscope, accelerometer and magnetometer are used to determine user's heading based on extended Kalman filter and that data is fed into WiFi positioning system, also achieving an increase in positioning success. Various signal propagation and area mapping algorithm are used in order to decrease the error in positioning, including the already mentioned Euclidean distance [36], using different signal propagation models for different regions [37] as well as Dempster Shafer fusion theory in [38]. In [39] the path and the direction of a user is modeled based on traces of WiFi packets, where

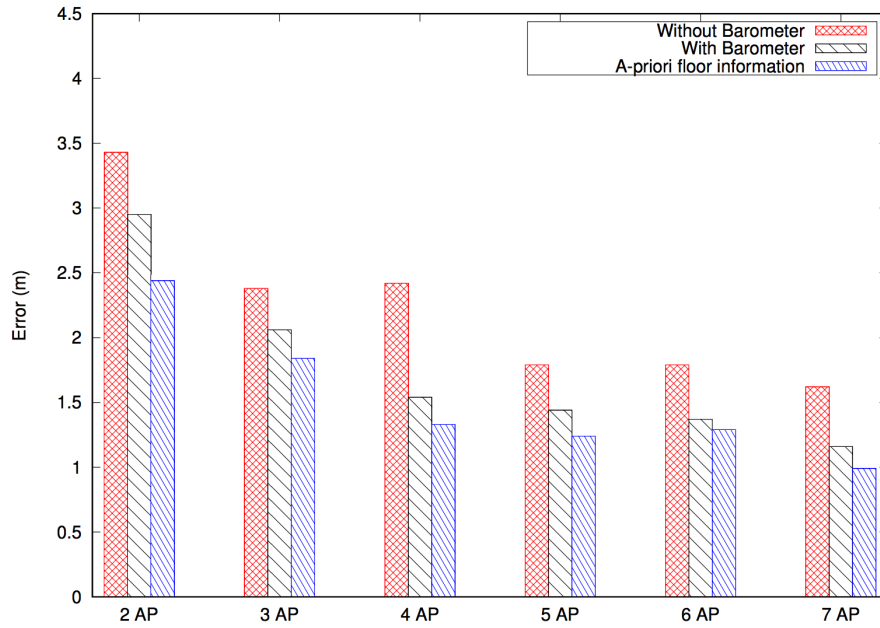


Figure 3.6. The positioning error for University of Genoa case-study of indoor positioning using WiFi and barometer data [31]

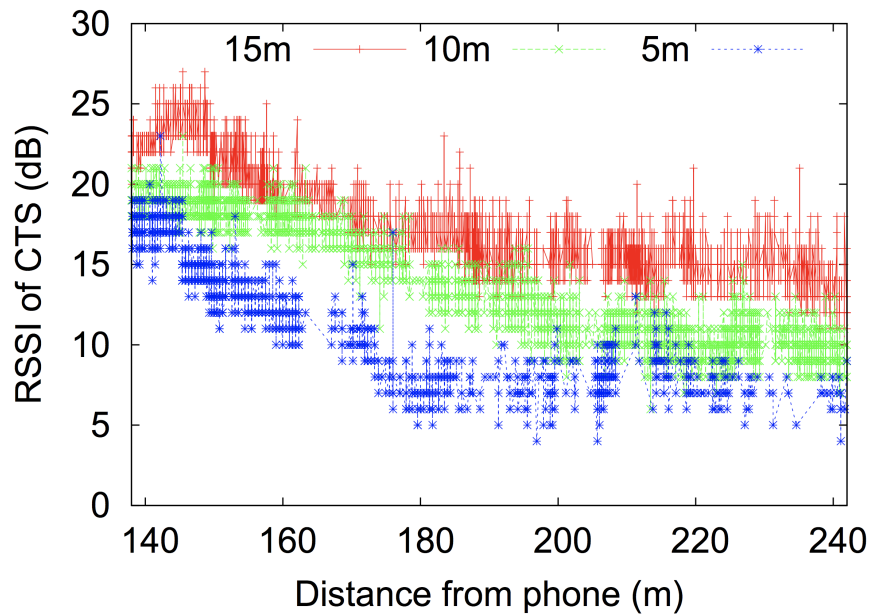


Figure 3.7. Plot of RSSI from Samsung smartphone to the search drone [32]

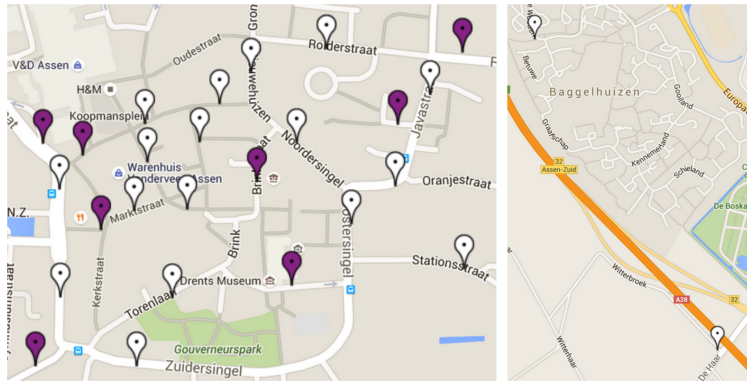


Figure 3.8. Placement of scanners in city center of Assen, Netherlands (left) and festival camp site (right) [40]

the performance of the method has been evaluated inside the room, outside the room, and in outdoor scenarios, as well as three kinds of walking paths, for example, horizontal, vertical, and slanted, are tested.

Even though indoor WiFi positioning is a very popular topic among researchers and a serious location privacy breach, in our work we plan on focusing more on: WiFi dynamic localization and WiFi based previous whereabouts, covered in the following sections.

3.2. Dynamic localization

By the term *Dynamic localization*, we categorize work which focuses on long term user tracking, rather than pinpointing his/hers exact indoor / outdoor position. The base idea for an attacker attempting to track a user is similar to Current localization (positioning) covered in Section 3.1 where the attacker is monitoring the packets on WiFi channel in order to figure out the presence of a certain device. Recall, WiFi devices will transmit packets even if they are not connected to a WiFi Access Point, and even if the device is not currently in use. That is an even more important fact for dynamic tracking than current localization, as the exact location is not important in dynamic tracking meaning that the transmission occurring some seconds before or later does not change the attacks success.

MAC address is a part of every WiFi packet, as well as the management probes earlier presented in Figure 2.9. As a MAC address uniquely distinguishes each WiFi enabled device, it is the basis for dynamic tracking of users. Setting up a sensor network over a certain area (e.g. a shopping mall, music festival, city landmarks) can provide the attacker with the ability to track users movements. Such data can be used in many different scenarios such as crowd control, marketing, general analytics about the popularity of various locations, using big data to predict

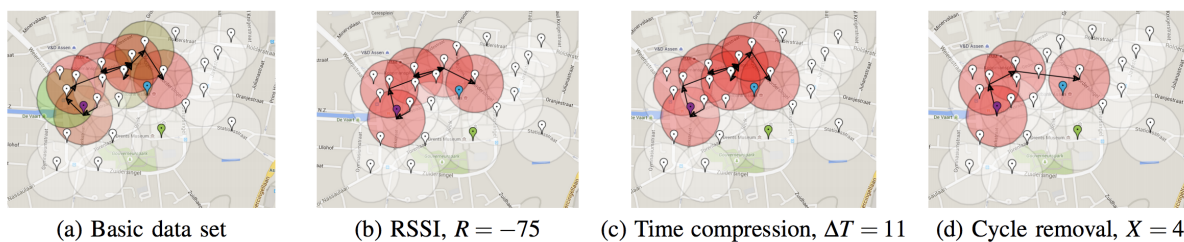


Figure 3.9. 10-minute path made by a device (the circles are 100m visual guides, they do not represent the coverage radius) [40]

users next location or even intersecting user traces with other dataset in order perform datasets matching.

Downtown area of a Dutch city has been covered with 27 WiFi sensors with the intent to monitor the movement of a 3-day music festival attended by 130,000 people in [40]. In Figure 3.8 the layout of the sensors is given. The authors have gathered a total of 15,135,611 detections of 248,192 devices over a 13-day monitoring period. To counter signal strength indicator (RSSI) variance, as well as multi sensing device detections, user moving in circles or multiple detections within the same period, the authors have proposed a data cleanup algorithm. An example trace of a single user is given in Figure 3.9. This is a nice example of how a WiFi enabled smartphone user's movement can be tracked, without the user ever knowing about the location privacy attack.

Another interesting example of dynamic tracking of users is given in [41]. The authors argue that the current facilities planning relies foremost on manual observations or coarse unverified assumptions and therefore do not properly scale or provide realistic data to inform facility planning. A new method is proposed to inform building facility planning by using WiFi traces in large building complexes. Similarly to [40], in [41] the authors work towards removing noise from the monitored WiFi packets using temporal and spatial features, and ultimately provide methods of quantification of area densities and user flows. They also attempt to classify user behavior (splitting users into roles, e.g. visitor, hospitalized or employee).

A spatio-temporal visualization tool has been built on top of [41] to enable building planners to inspect and explore extracted WiFi information to their gains, as has been depicted in Figure 3.10. More than a billion individual WiFi measurements have been recorded from 18,000 different devices. Similarly, in [42] a 1 year experiment has been conducted based on data obtained from 9 WiFi sensors which concluded how dynamic tracking can be used for optimizing the management and operation of the school, as well as for other similar infrastructures and, in general, for other kind of applications which require not very accurate people flow monitoring at low cost.

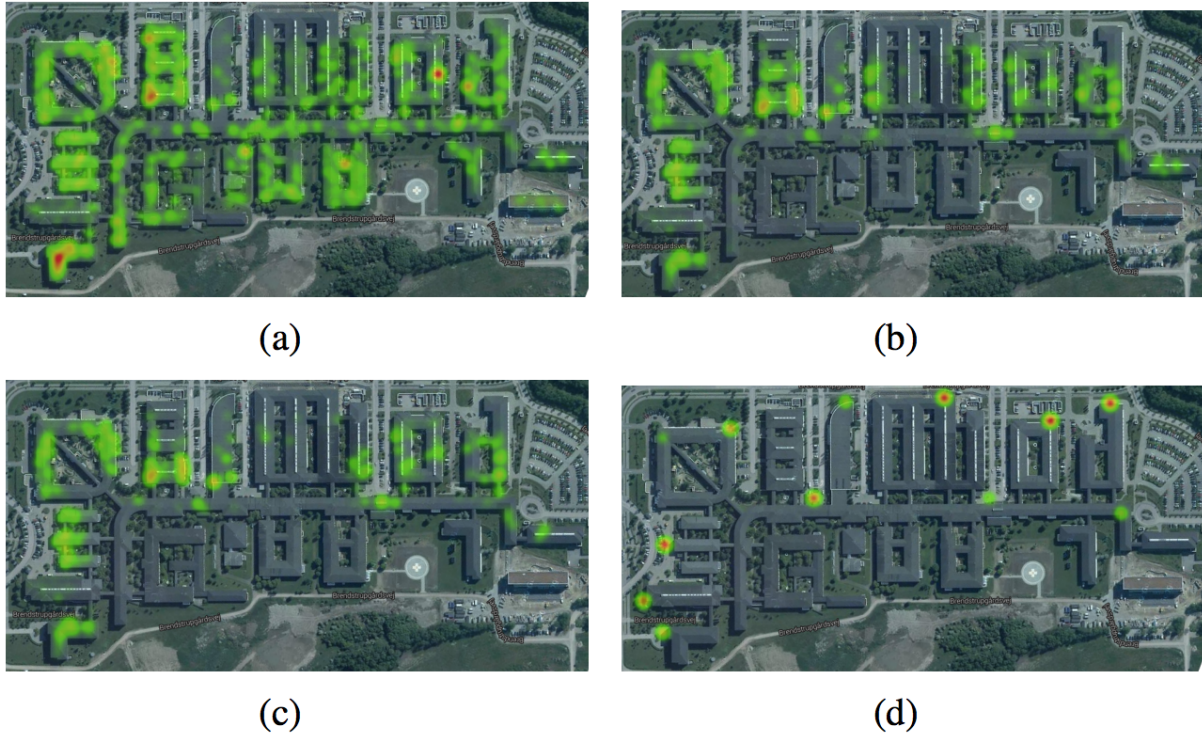


Figure 3.10. (a) Heat-maps representing all device positions; (b) Only inside to outside movements (leaving the building); (c) Positions of filtered exits; (d) Estimated exits constrained to real exits. [41]

Mobility data acquired from Probe Request packets has been used in [43] to conclude social relations between the users. The authors are studying a large number of mobility features for predicting the social tie strengths of the device owners present in the location at a given moment in time, and output an aggregate score of social connectedness for that location. Following features have been used: Overlap Only (a pair of devices visiting the same location, number and length of the visit is considered), Individual Only (individual visits of a device), Overlap and Individual (individual devices overall visiting patterns), Overlap and Location (how busy was the location when overlapping took place), Individual and Location (how busy was the location when individual visits took place).

Some form of Dynamic localization based on WiFi traces has been performed in many other papers as well. In [44] a novel framework named Jyotish is presented, which constructs a predictive model by exploiting the regularity of people movement found in the joint WiFi and Bluetooth trace. The prediction model is able to answer three questions: (1) where the person will stay, (2) how long the person will stay at the location, and (3) who will the person meet. Jyotish is based on Naive Bayesian classifier joining the WiFi and bluetooth traces which is later fed into the predictive model. The authors have tested the model on 50 participants, which were monitored for 20 to 50 days, ultimately managing to predict the movement of 80% of the participants with the correct position in 70% of the cases, and 60% of the participants with more

than 80% correct prediction.

WiFi traces have been used to perform vehicular tracking in [45]. By using only off the shelf equipment, the authors have built a system capable of concluding which path the vehicle has taken through the city. During a 9 month period, multiple sensing devices had been deployed through the city, and concluded that by using monitors spaced over 400 meters apart, the mean error is under 70 meters, as compared to GPS ground truth. The authors conclude that the practical deployment of such a system with the means of analyzing surface street traffic flow could be beneficial to commuters and traffic planners. Traffic analysis using WiFi as the basis has been covered in a number of papers and more detailed traffic analysis overview can be found in [46]. In a similar fashion, in [47] pedestrian tracking based on WiFi and bluetooth sensors has been performed. The authors bring an important point about dynamic localization: most of the traditional pedestrian monitoring technologies focus on counting pedestrians passing through a fixed location in the network, but it is not possible to anonymously track the movement of individuals or groups as they move outside each particular sensor's range. Results point out that it is possible to accurately estimate the number of pedestrians, pedestrian flows and average wait times based on a 2 month long collection of WiFi data on 6 public transportation terminals. Group movements of festival attendees have been tracked by deploying a network of RaspberryPI devices in [48].

However, the work proposed in this section so far focuses on mobility of devices, not the actual users. As the main identifier for data classification is the MAC address stored in the WiFi packets, it is possible to track the device and its location, predicts its movement or improve the estimated by including other datasets. What the MAC address can not give you is the name of the person holding the tracked device. We use the term MAC address deanonymization as the process of concluding the actual person behind the WiFi enabled device. Having gained such an information greatly increases the severity of location privacy attacks based on WiFi networks, as we are no longer tracking a device (mainly a smartphone) but we are tracking an actual individual. Further more, the tracking can usually be done by using cheap off the shelf equipment - meaning it is accessible to everyone. Question remains - how do we conclude which WiFi MAC address belongs to whom in the vast number of traces monitored on the WiFi channel?

Social networks are used as a side-channel in an attempt to deanonymize mobility traces in [49]. Social links are used as the basis of an algorithm which concludes that if two devices have been connected to the same Access Point within 600 seconds - they are related. Similarly - if two users are friends on facebook - they are, of course, related. A contact graph is used to identify meetings between anonymized users in the set of WiFi traces, and it is then structurally correlated with the social network graph thereby identifying anonymized users. The effectiveness of this approach is tested in [49] by using 3 separate mobility traces, each

mapped to its own related social network dataset (Facebook, university network and conference registration social network) and concluded that 80% of the users have been precisely identified, and 8% have been identified incorrectly.

Another approach at finding out the name of a person behind a MAC address has been done in [50]. Although the main purpose of the work was to explore semantics of SSID names, similar to [51], the authors have found that 1800 out of 120 000 SSIDs contained the pattern “s” which is in most times prefixed by a persons name (e.g. “Travolta’s Home”). In [52] WiFi assisted geolocation has been exploited where, by simulating a presence of a known WiFi access point, the authors have been able to make the victims device to geolocate the device in a wrong location. Social networks would use the wrong location to tag victim’s posts therefore letting the attacker know the name of the person behind the MAC address.

MAC address deanonymization is also one of the key focal points of our research, and in Chapter 4 an overview of our current endeavors is given.

3.3. Previous whereabouts and analytics

By the term *Previous whereabouts and analytics* we categorize the work based on using Probe request packets gathered off WiFi enabled devices using Active Service Discovery as explained in Section 2.4. Recall, a WiFi device using Active Service Discovery will periodically transmit Probe Request packets, each containing an Service Set Identifier (SSID) of an Access Point it previously connected to. As Probe Request packets are sent “in clear” (they are not encrypted), an attacker can easily monitor the WiFi channel, gather the Probe Request packets and group them by the origin WiFi card’s MAC address. Doing so over a period of time, the attacker has gained access to user’s Preferred Network List - the list of his previously used Access Points. In Figure 3.11 the WiFi connection settings screens are given for an Android device listing that devices PNL. One could argue that finding out victims PNL is not a big advantage for the attacker in terms of victims location privacy, however if we take a closer look at the SSIDs given in Figure 3.11, we can see that the SSIDs themselves can be quite revealing. Previously used Access Points “TZ Podstrana free internet” and “Welcome to Podstrana” obviously point out that this user has visited the city of Podstrana, Croatia sometime in the past, as well as the Vienna airport (“WirelessViennaAirport”). Using a web search tool such as Google, additional information can be obtained about other SSIDs for this user: “Solaris” is a hotel resort near Sibenik, Croatia, “MARCHE” is a food chain along the Croatian highway, “Medovdolac” is a small city, and “SRCE-GUEST” could refer that the user has visited SRCE - Central computing center of Croatia’s academic institutions.

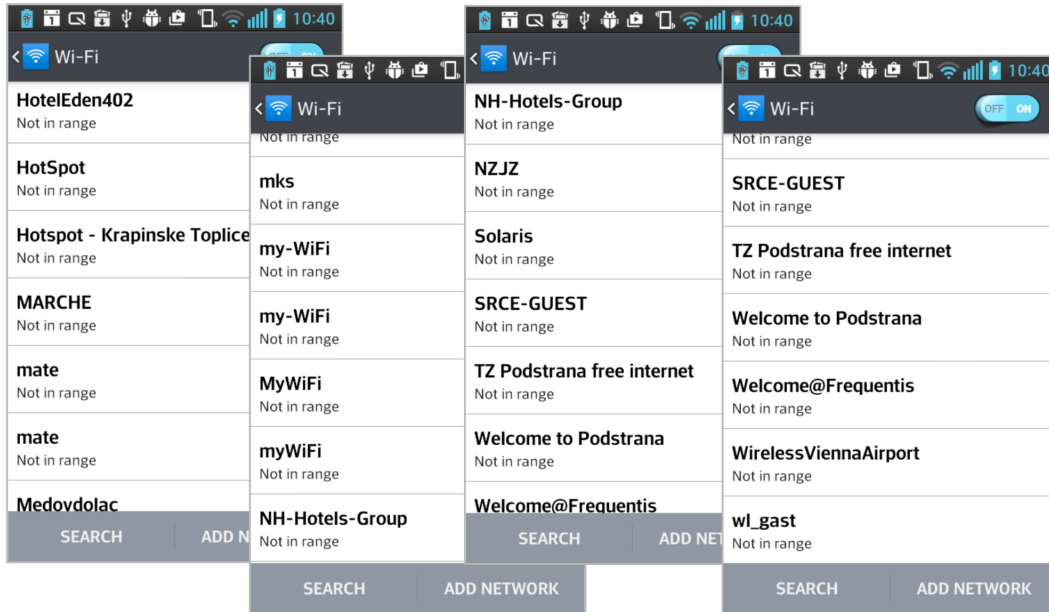


Figure 3.11. WiFi connection screen-shoots on an Android smartphone device listing that devices Preferred Network List

The origin of participants in large events has been concluded based on Probe Request packets in [20]. The authors have used publicly available Probe Request datasets containing more than 11M records collected citywide, national (two political meetings) and international religion-related relevance. The authors have exploited the semantic information from the SSIDs in order to discover with high accuracy the provenance of the crowds in each event. The quality of the match between the official voting results at the political meetings and the authors work is highlighted. Extracting additional information based on SSIDs has been performed in [51] where the authors have managed to extract points of interest names from 49% of SSIDs. The dataset used was gathered in university campus, residential suburb as multi location gathering using a mobile WiFi monitoring device.

Some devices will use MAC address randomization in order to prevent the tracking based on WiFi as we previously overview in Section 3.2. The authors of [53] are using the SSIDs from Probe Request packets, among other approaches, to fingerprint the device through its PNL in order to make the MAC address randomization technique obsolete.

SSID data is mapped to their actual locations in [54] where the authors have focused on many issues concerning such a mapping. For example, multiple SSIDs being present at the same location or it being a wrong match whatsoever (e.g. everyone can have the SSID “TimHorton’s”, however if it actually matches to the real location “Tim Horton’s Caffee” is uncertain). Additionally, the authors point out the lack of temporal information (when the targeted user has connected to the WiFi Access Point from his PNL) as well as the lack of the number of visits, as there is no proof of a relation between the Probe Request packets transmission and the number

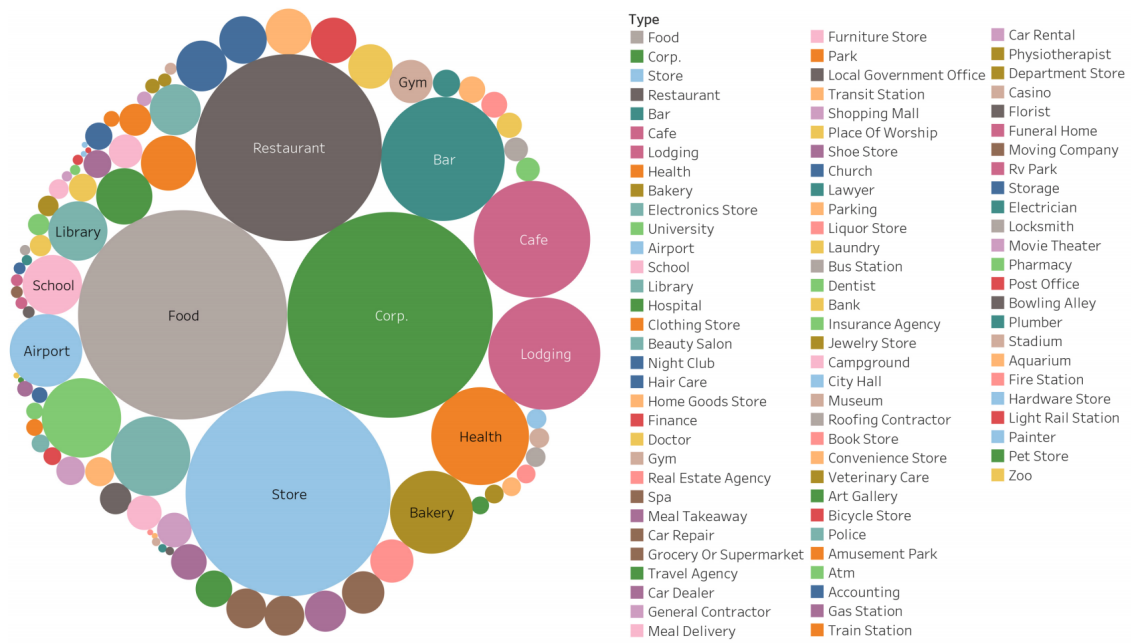


Figure 3.12. The entire location types extracted from the captured SSID. The bubble size corresponds to the frequency of the location type in dataset [54]

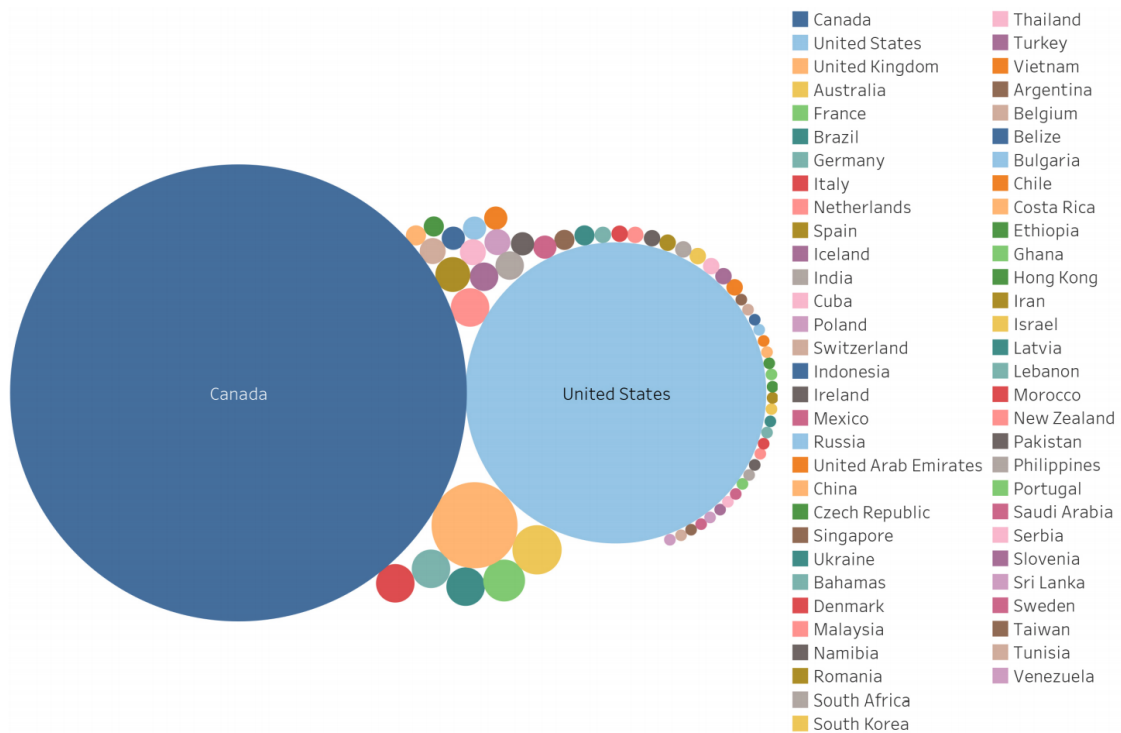


Figure 3.13. The countries of the captured SSIDs. The size of the bubbles corresponds to the frequency of a country (number of SSIDs matched to a location within that country) [54]

of times a particular AP has been used. The authors have proceeded to use several methods to match the SSID to their appropriate location. The previously mentioned WiGLE service [5] has been found to be unreliable, mainly because WiGLE does not remove abolished locations, and the fact it does not provide any semantic information about the mapped location. Next Google Place API has been used, [55] which is capable of providing more semantic information about a location compared to WiGLE, however it requires the SSID name to be similar to the location name. The authors have queried 3340 SSIDs using Google Places API, mainly with the focus of categorizing each SSID and gathering as much information about the location as possible. The results can be found in Figure 3.12 and Figure 3.13

4. Current and future work

In this chapter, we will briefly overview the current work our research group is focusing on. Based on the categories presented in Chapter 3, the research we are conducting would fall mostly in the Dynamic localization technology, mainly the MAC address deanonymization (finding out the person behind a MAC address) as well as disclosing users previous whereabouts.

Recall from Section 2.4, devices using Active Service Discovery will transmit their previously used Access Point Service Set Identifiers (SSID) in Probe Request packets in order to speed up the connection process to WiFi networks which the devices have connected to in the past. Nowadays, however, more and more devices no longer transmit their Preferred Network List (PNL) in clear by using Passive Service Discovery, or Active Service Discovery with Broadcast, thus mitigating passive attacks. Hidden PNLs are still vulnerable against active attacks whereby an attacker mounts a fake SSID hotspot set to one likely contained within targeted PNL. If the targeted device has this SSID in the corresponding PNL, it will automatically initiate a connection with the fake hotspot thus disclosing this information to the attacker. By iterating through different SSIDs (from a predefined dictionary) the attacker can eventually reveal a big part of the hidden PNL as depicted in Figure 4.1. Considering user mobility, executing active attacks usually has to be done within a short opportunity window, while targeting non-trivial SSIDs from user's PNL. In [56] we propose a simple mathematical model for analyzing active SSID dictionary attacks, allowing us to optimize the effectiveness of the attack under the above constraints (limited window of opportunity and targeting nontrivial SSIDs). Additionally, we showcase an example method for building an effective SSID dictionary using top-N recommender algorithm and validate our model through simulations and extensive real-life tests. The comparison of our model and achieved results is given in Figure 4.2.

In [21] we present the results of a Probe Request monitoring endeavor on a large scale music festival held in Croatia. A chronological overview of sensitive location data we collected in years 2014, 2015, 2017 and 2018 is provided. We conclude that using passive WiFi monitoring scans produces different results through years, with a significant increase in the usage of a more secure Broadcast Probe Request packets and MAC address randomizations by the smartphone operating systems. In 2014 the share of Probe Request packets containing

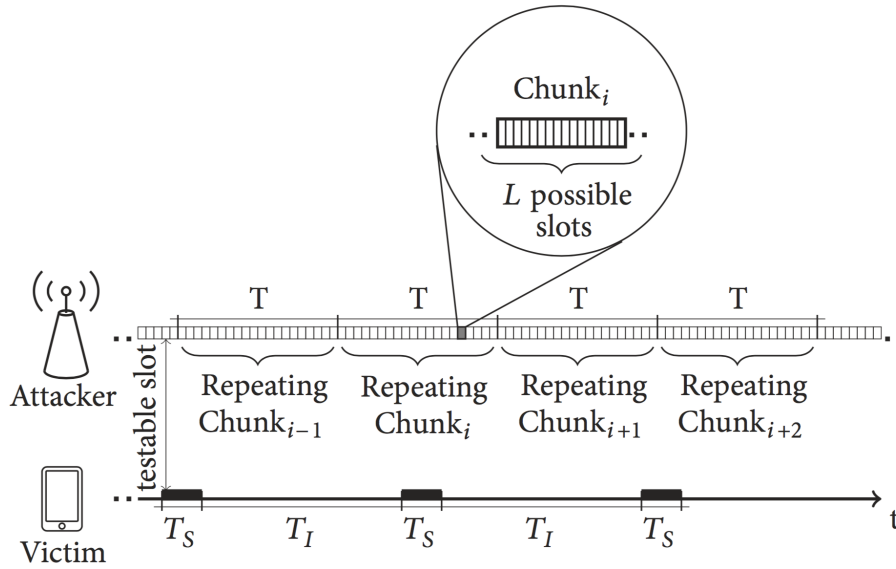


Figure 4.1. Attacker and Victim timelines when performing SSID Oracle attack. Each WiFi network scanning interval of the device is being matched to a chunk of fake SSIDs placed by the Attacker [56]

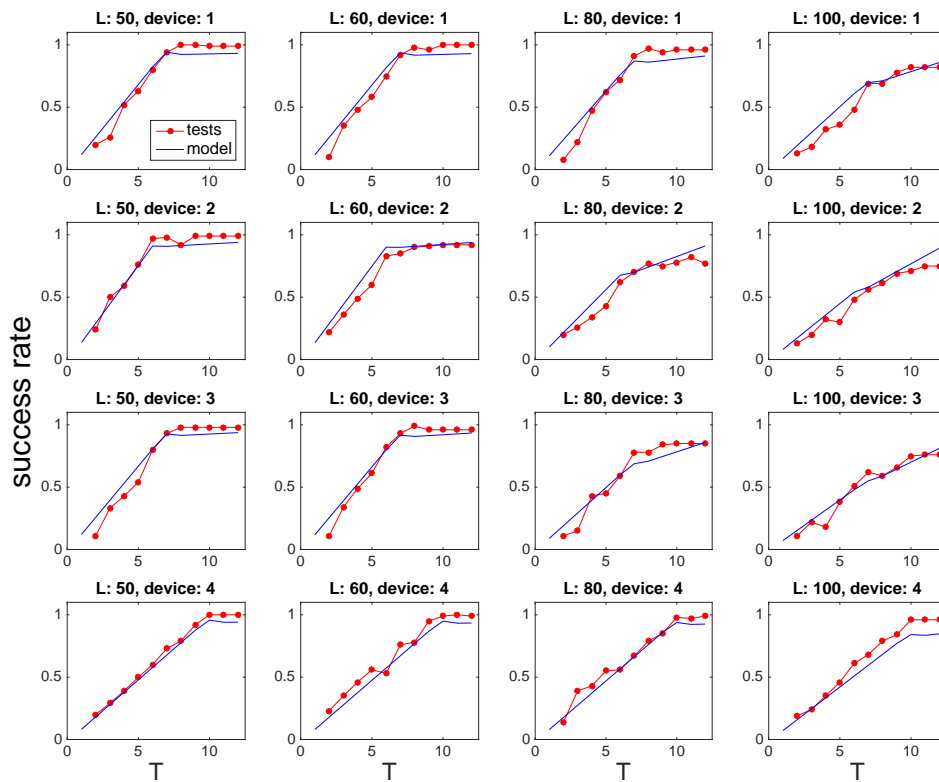


Figure 4.2. Model comparison with test results for different devices and chunk sizes L [56]

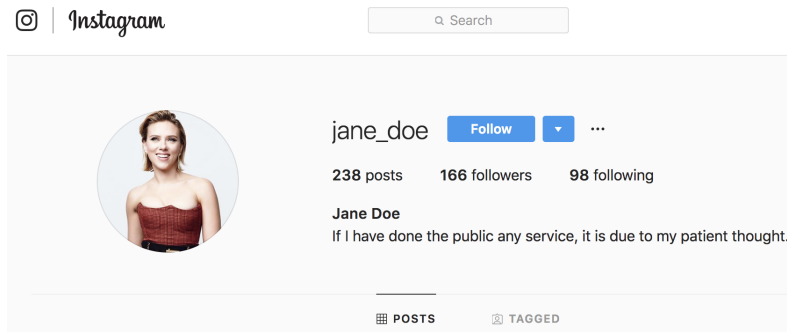


Figure 4.3. Jane Doe Instagram account

Table 4.1. Jane Doe matching profile

Instagram data	
Name	Jane Doe
City	Split, Croatia
Occupation	Pilates instructor / tourist guide
Total posts	238
Total locations	64
Device data	
Mac Address	3E:42:A2:XX:XX:XX
Device Manufacturer	Samsung Electro Mechanics co., LTD.
PNL size	57
Algorithm results	
Matched SSIDs	City Center one (<i>shopping center</i>) ISMILE (<i>dental center</i>) iSmile CAT
Matched Locations	City Center one Split (2 occurrences) ISmile Cat Caffe Zagreb
Score	78.7

the previously used SSIDs (using Active Service Discovery) was 46.7%, whereas in 2018 it was only 12.9%. This indicates a drop of more than 3 times in 4 years, however although the drop is certain, the exact drop percentage can not be firmly concluded since there is no data on the increase or decrease of devices using Passive scan.

Finding the person behind the monitored MAC address has been an interesting and challenging topic for our research group. In a still unpublished work, we introduce a novel SSID - location tag matching function, followed by an algorithm used for intersecting large PNL datasets with localization tags on Instagram social network. The algorithm enables us to match the users MAC address and PNL with his full name, photos and activities. We find that deanonymization of a MAC address provides serious implications for potential long term tracking. We tested our work in real life conditions on a large scale music festival. To approach the ground truth we conducted hand check tests performed by 10 testers who concluded that more than 50% of the proposed matches were correct. In Figure 4.3 and Table 4.1 we give an example of such a matching.

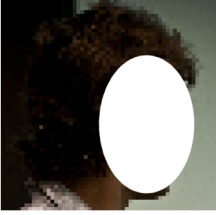
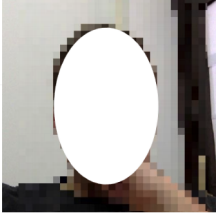
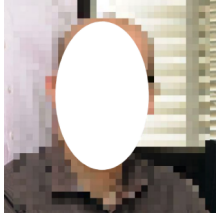
User face	Name	Designation	Username	MAC address
	John Doe	Assistant professor	████████@unist.hr	Apple: ██████████
	Jane Doe	Student	████████@fesb.hr	Samsung: ██████████
	John Smith	Full professor	████████@fesb.hr	Samsung: ██████████

Figure 4.4. Example set of deanonymization based on University contact book and leakage in eduroam network. Real name, face, designation and email can be linked to users device

Deanonymization has been covered in another unpublished work by targeting a known flaw in eduroam educational WiFi networks where a leakage of usernames is present. Security weakness of WPA2-Enterprise networks is well known and number of publications recently were trying to emphasize the problem of the security in these systems [57, 58, 59, 60]. Through the flaw in eduroam network, we managed to collect 1650 identities where more than 87% had wrongly configured their devices leaking usernames or usernames and eduroam passwords. Further more, we looked into the eduroam configuration tools of more than 1000 world-wide institutions and found that 67% of the profiles were sending usernames in clear. Since each WiFi packet is associated with a MAC address, the leaked usernames are automatically matched, and deanonymization based on the username can be easily performed (username is often associated with real name). In Figure 4.4 an example deanonymization based on this concept is given.

Future position of WiFi technology is very bright and opens up a lot of different angles for scientist to explore. Cellular and WiFi network inter-working is being standardized and proposes different approaches such as small cell offloading, which could present an interesting connection between cellular location privacy and WiFi location privacy. Also 5GHz WiFi standard will become more popular in the future, having the 2.4GHz band congested, and current localization performances can be researched. Another research foundation could be on different approaches on deanonymization of WiFi traces, and the connection to real users with the goal of providing easier tracking in the future. At the moment our goal is to further expand on the MAC address deanonymization as well as improve the success rate of the existing algorithm.

5. Conclusion

Data security and user privacy is one of the most discussed IT topics in 21st century with location privacy being in the very top. Lately there is a huge rise of mobile devices and different IoT projects such as wearable devices, sensor networks and similar, with almost all of them having some sort of a wireless connection, mostly WiFi. The expected growth is 20 billions by 2020 [61] with expectations that every user will have multiple carry-on device, and will live in an environment covered all over with different wireless sensing devices. WiFi is one of the most popular wireless technologies, mostly because of the big smartphone expansion, and is expected to stay. New WiFi standards are presented and LTE networks is expected to offload to WiFi [62] with will only substantiate the strong position WiFi is holding in IoT.

Having WiFi so widespread among users has set the foundation and interest in the IT community about WiFi location privacy. There has been a lot of scientific papers published in reputable journals and conferences covering different security implications for users location privacy within the WiFi network. Also a lot of entrepreneurship endeavors were focusing on using WiFi for user positioning and tracking in order to provide insightful data to be used for different marketing purposes.

This paper provides an insight into different techniques used to passively and actively disclose users sensitive information gathered from his WiFi enabled device. Focus was given to disclosing users previous whereabouts, current localization and dynamic tracking options using the WiFi MAC address. There is a wide variety of different real-world situations where Location privacy based research can contribute. Police can use such a technology to track suspects by either using some form of real time attack to disclose current location, or by using some historical location data in order to prove somebodies previous whereabouts. In the entrepreneurship world, you could find out if your client is seeing your competition. You could track you significant other in order to see if s/he is cheating on you. Journalists / paparazzi could use such technology to find out celebrities favorite restaurant / museum, or to be notified when they arrive at a particular location. As we are monitoring and storing more and more data, such data can then be used in different big-data researches, like user categorization, marketing purposes, behavior analysis and much more. All of this is done without any real knowledge by the user about being tracked. Also it can be simultaneously applied to large groups of people

(shopping malls, events, popular tourist destinations etc.). Nevertheless all of the implications are a violation of somebody's privacy. Some of them (suspect tracking, search and rescue etc.) could be justified on a moral grounds, but others are mostly privacy violation.

BIBLIOGRAPHY

- [1] SubPos - open source WiFi positioning system, <http://www.subpos.org>, accessed: 2017-05-21.
- [2] J. Yang, A. Varshavsky, H. Liu, Y. Chen and M. Gruteser, Accuracy characterization of cell tower localization, *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 223–226, ACM, 2010.
- [3] B. Dezfouli, V. Esmaelzadeh, J. Sheth and M. Radi, A Review of Software-Defined WLANs: Architectures and Central Control Mechanisms, *IEEE Communications Surveys & Tutorials*, PP, 1–1, 09 2018.
- [4] Google Wifi and your privacy, <https://support.google.com/wifi/answer/6246642?hl=en>, accessed: 2017-05-21.
- [5] Wigle - access point locations database, <https://wisle.net>, accessed: 2017-05-21.
- [6] M. Cunche, M. A. Kaafar and R. Boreli, I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests, *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium*, 2012.
- [7] A. E. Redondi and M. Cesana, Building up knowledge through passive WiFi probes, *Computer Communications*, 117, 1–12, 2018.
- [8] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta and J. Stefa, Signals from the Crowd: Uncovering Social Relationships through Smartphone Probes, *IMC '13 Proceedings of the 2013 conference on Internet measurement conference*, 2013.
- [9] M. D. Aime, G. Calandriello and A. Lioy, Dependability in wireless networks: Can we rely on WiFi?, *IEEE Security & Privacy*, 5, 1, 23–29, 2007.
- [10] Wi-Fi wiki, <https://en.wikipedia.org/wiki/Wi-Fi>, accessed: 2019-08-05.
- [11] IEE 802.11, https://en.wikipedia.org/wiki/IEEE_802.11, accessed: 2019-08-05.
- [12] Z. Hays, G. Richter, S. Berger, C. Baylis and R. J. Marks, Alleviating airport WiFi congestion: An comparison of 2.4 GHz and 5 GHz WiFi usage and capabilities, *Texas Symposium on Wireless and Microwave Circuits and Systems*, 1–4, IEEE, 2014.
- [13] Infrastructure mode, <https://www.sciencedirect.com/topics/computer-science/infrastructure-mode>, accessed: 2019-08-05.

- [14] V. Shah, R. Patel and R. Nayak, Short Range Inter-satellite Link for Data Transfer and Ranging using IEEE802.11n, *International Journal of Computer Applications*, 164, 23–25, 04 2017.
- [15] Karma tool, <https://www.offensive-security.com/kali-linux/kali-linux-evil-wireless-access-point/>, accessed: 2017-02-19.
- [16] M. Vanhoef, C. Matte, M. Cunche, L. S. Cardoso and F. Piessens, Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms, *ACM AsiaCCS*, 2016.
- [17] N. Sidiropoulos, M. Mioduszewski, P. Oljasz and E. Schaap, Open Wifi SSID Broadcast vulnerability, *SSN Project Assessment*, 24, 2012.
- [18] W. Wang, R. Joshi, A. Kulkarni, W. K. Leong and B. Leong, Feasibility Study of Mobile Phone WiFi Detection in Aerial Search and Rescue Operations, *APSys '13 Proceedings of the 4th Asia-Pacific Workshop on Systems*, 2013.
- [19] X. Hu, L. Song, D. V. Bruggen and A. Striegel, Is There WiFi Yet? How Aggressive WiFi Probe Requests Deteriorate Energy and Throughput, *IMC '15 Proceedings of the 2015 ACM Conference on Internet Measurement Conference*.
- [20] A. D. Luzio, A. Mei and J. Stefa, Mind Your Probes: De-Anonymization of Large Crowds Through Smartphone WiFi Probe Requests, *IEEE INFOCOM 2016, At San Francisco, CA, USA*.
- [21] A. Dagelić, T. Perković and M. Čagalj, Location Privacy and Changes in WiFi Probe Request Based Connection Protocols Usage Through Years, *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, 1–5, IEEE, 2019.
- [22] J. Freudiger, How talkative is your mobile device?: an experimental study of Wi-Fi probe requests, *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, 8, ACM, 2015.
- [23] G. Lui, T. Gallagher, B. Li, A. G. Dempster and C. Rizo, Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization, *2011 International Conference on Localization and GNSS (ICL-GNSS)*, 53–57, IEEE, 2011.
- [24] WireShark network protocol analyser, <https://www.wireshark.org/>, accessed: 2019-08-26.
- [25] A. Bose and C. H. Foh, A practical path loss model for indoor WiFi positioning enhancement, *2007 6th International Conference on Information, Communications & Signal Processing*, 1–5, IEEE, 2007.
- [26] M. Hata, Empirical formula for propagation loss in land mobile radio services, *IEEE Transactions on Vehicular Technology*, 29, 3, 317–325, 1980.
- [27] W. Zhang, X. Hua, K. Yu, W. Qiu, S. Zhang and X. He, A novel WiFi indoor positioning strategy based on weighted squared Euclidean distance and local principal gradient direction, *Sensor Review*, 39, 1, 99–106, 2019.

- [28] H.-Y. Hsieh, S. W. Prakosa and J.-S. Leu, Towards the Implementation of Recurrent Neural Network Schemes for WiFi Fingerprint-Based Indoor Positioning, *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 1–5, IEEE, 2018.
- [29] Y. Li, S. Williams, B. Moran and A. Kealy, Quantized RSS Based Wi-Fi Indoor Localization with Room Level Accuracy, *Proceedings of the IGSS Conference, Sydney, Australia*, 7–9, 2018.
- [30] S. Yousefi, H. Narui, S. Dayal, S. Ermon and S. Valaee, A survey on behavior recognition using wifi channel state information, *IEEE Communications Magazine*, 55, 10, 98–104, 2017.
- [31] I. Bisio, A. Sciarrone, L. Bedogni and L. Bononi, WiFi Meets Barometer: Smartphone-Based 3D Indoor Positioning Method, *2018 IEEE International Conference on Communications (ICC)*, 1–6, IEEE, 2018.
- [32] W. Wang, R. Joshi, A. Kulkarni, W. K. Leong and B. Leong, Feasibility study of mobile phone WiFi detection in aerial search and rescue operations, *Proceedings of the 4th Asia-Pacific Workshop on Systems*, 7, ACM, 2013.
- [33] H. Chen, Y. Zhang, W. Li and P. Zhang, Non-Cooperative Wi-Fi Localization via Monitoring Probe Request Frames, *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 1–5, IEEE, 2016.
- [34] F. Evennou and F. Marx, Advanced integration of WiFi and inertial navigation systems for indoor mobile positioning, *Eurasip journal on applied signal processing*, 2006, 164–164, 2006.
- [35] Z.-A. Deng, G. Wang, D. Qin, Z. Na, Y. Cui and J. Chen, Continuous indoor positioning fusing WiFi, smartphone sensors and landmarks, *Sensors*, 16, 9, 1427, 2016.
- [36] R. Ma, Q. Guo, C. Hu and J. Xue, An improved WiFi indoor positioning algorithm by weighted fusion, *Sensors*, 15, 9, 21824–21843, 2015.
- [37] Y. Du, D. Yang and C. Xiu, A novel method for constructing a WiFi positioning system with efficient manpower, *Sensors*, 15, 4, 8358–8381, 2015.
- [38] D. Bhatt, S. R. Babu and H. S. Chudgar, A novel approach towards utilizing Dempster Shafer fusion theory to enhance WiFi positioning system accuracy, *Pervasive and Mobile Computing*, 37, 115–123, 2017.
- [39] L. Sun, S. Chen, Z. Zheng and L. Xu, Mobile device passive localization based on iee 802.11 probe request frames, *Mobile Information Systems*, 2017, 2017.
- [40] C. Chilipirea, A.-C. Petre, C. Dobre and M. van Steen, Presumably simple: monitoring crowds using WiFi, *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, 1, 220–225, IEEE, 2016.
- [41] A. J. Ruiz-Ruiz, H. Blunck, T. S. Prentow, A. Stisen and M. B. Kjærgaard, Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning, *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 130–138, IEEE, 2014.

- [42] J. Andi3n, J. M. Navarro, G. L3pez, M. 1lvarez-Campana and J. C. Due1nas, Smart Behavioral Analytics over a Low-Cost IoT Wi-Fi Tracking Real Deployment, *Wireless Communications and Mobile Computing*, 2018, 2018.
- [43] T. Brugman, M. Baratchi, G. Heijenk and M. van Steen, Inferring the social-connectedness of locations from mobility data, *International Conference on Social Informatics*, 443–457, Springer, 2017.
- [44] L. Vu, Q. Do and K. Nahrstedt, Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace, *2011 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 54–62, IEEE, 2011.
- [45] A. Musa and J. Eriksson, Tracking unmodified smartphones using wi-fi monitors, *Proceedings of the 10th ACM conference on embedded network sensor systems*, 281–294, ACM, 2012.
- [46] Z. Liu, Y. Wang, S. Liu, Z. Liu and G. Li, Overview of Studies on the Wi-Fi Probe Data Analysis for Transport Problems, *CICTP 2019*, 5961–5972, 2019.
- [47] A. Kurkcu and K. Ozbay, Estimating pedestrian densities, wait times, and flows with wi-fi and bluetooth sensors, *Transportation Research Record*, 2644, 1, 72–82, 2017.
- [48] B. Bonn3, A. Barzan, P. Quax and W. Lamotte, WiFiPi: Involuntary tracking of visitors at mass events, *2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 1–6, June 2013.
- [49] Deanonymizing mobility traces: Using social network as a side-channel, author=Srivatsa, Mudhakar and Hicks, Mike, *Proceedings of the 2012 ACM conference on Computer and communications security*, 628–637, ACM, 2012.
- [50] S. Seneviratne, F. Jiang, M. Cunche and A. Seneviratne, SSIDs in the wild: Extracting semantic information from WiFi SSIDs, *2015 IEEE 40th Conference on Local Computer Networks (LCN)*, 494–497, IEEE, 2015.
- [51] M. Chernyshev, C. Valli and P. Hannay, On 802.11 access point locatability and named entity recognition in service set identifiers, *IEEE Transactions on Information Forensics and Security*, 11, 3, 584–593, 2015.
- [52] C. Matte and M. Cunche, Beam me up, Scotty: identifying the individual behind a MAC address using Wi-Fi geolocation spoofing, *1er Colloque sur la Confiance Num3rique en Auvergne*, 2014.
- [53] M. Vanhoef, C. Matte, M. Cunche, L. S. Cardoso and F. Piessens, Why MAC address randomization is not enough: An analysis of Wi-Fi network discovery mechanisms, *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, 413–424, ACM, 2016.
- [54] M. H. Sarshar, *Analyzing large scale Wi-Fi data using supervised and unsupervised learning techniques*, Ph.D. thesis, 2017.
- [55] Google Places API, <https://cloud.google.com/maps-platform/places/>, accessed: 2019-09-02.

- [56] A. Dagelić, T. Perković, B. Vujatović and M. Čagalj, SSID Oracle Attack on Undisclosed Wi-Fi Preferred Network Lists, *Wireless Communications and Mobile Computing*, 2018, 2018.
- [57] S. Brenza, A. Pawlowski and C. Pöpper, A Practical Investigation of Identity Theft Vulnerabilities in Eduroam, *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '15, 2015.
- [58] V. Ramachandran, Cracking WPA/WPA2 Personal and Enterprise for Fun and Profit, *Hacktivity 2012*, 2012.
- [59] A. Bartoli, E. Medvet and F. Onesti, Evil twins and WPA2 Enterprise: A coming security disaster?, *Computers & Security*, 74, 1–11, 2018.
- [60] A. Bartoli, E. Medvet, A. D. Lorenzo and F. Tarlao, (In)Secure Configuration Practices of WPA2 Enterprise Supplicants, *Proceedings of the 13th International Conference on Availability, Reliability and Security*, ARES 2018, 37:1–37:6, 2018.
- [61] Gartner, Inc, Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015, <http://www.gartner.com/newsroom/id/3165317>, accessed: 2018-08-04.
- [62] A. Pyattaev, K. Johnsson, S. Andreev and Y. Koucheryavy, 3GPP LTE traffic offloading onto WiFi direct, *2013 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 135–140, IEEE, 2013.

Labels:

AP	Access Point
CDF	Cumulative Distribution Function
IEEE	Institute of Electrical and Electronics Engineers
ISM	Industrial Scientific and Medical
IoT	Internet of Things
LOS	Line Of Sight
LSMT	Long Short Term Memory (deep learning)
MAC	Media Access Control
PHY	Physical Layer
PNL	Preferred Network List
RNN	Recurrent Neural Network (deep learning)
RSSI	Received Signal Strength Indicator
SSID	Service Set Identifier
UTP	Unshielded Twisted Pair (Cable)

Location Privacy and WiFi Networks

Abstract:

As the number of WiFi enabled devices is rising, so is the severity of location privacy vulnerabilities. Researchers have acknowledged this fact and a lot of work regarding the security of WiFi networks is focused on location privacy. In this paper we overview the location privacy within WiFi networks and categorize them in three categories: Current localization, Dynamic localization and Previous whereabouts and analytics. Current localization focuses on WiFi positioning solutions by using the signal strength and various other parameters in order to perform indoor positioning where error rates have proven to be in the 1 meter range. Dynamic localization focuses on tracking of a person mainly based on his/hers WiFi cards MAC address, where researchers setup a network of WiFi sensors and track or even assume somebodies location. Papers in previous whereabouts and analytics category use a flaw in the WiFi connection protocol to disclose users previously used WiFi access points - Preferred Network List. That data is then cross referenced with different databases, such as Google Places API, in order to conclude where the person has traveled and which locations s/he has visited. Furthermore, we overview the work performed by our research group and comment on the planned future work.

Keywords:

WiFi, location privacy, Probe Request, deanonymization, tracking, indoor positioning