Probe Requests

Connected Mobility Basics, Spring 2018

Assignment 3

Submitted by Johannes Seiler

Filed on Munic, 26.02.2018

Contents

1. Introduction

The main-goal of assignment 3 was to collect probe requests traces from different locations and scenarios. This should be done by using a provided probing kit consisting of a Raspberry Pi with four WiFi adapters and a software to capture the required packages. Finally, after analysing the obtained data sets from varied environments, the privacy and ethical implications that arise in such an experiment should be discussed.

1.1. Collecting Probe Requests

Although the overall process of getting familiar with the provided Raspberry setup and its use for collecting data was pretty much straight forward, the understanding of the concept behind probe requests proved to be the first high challenge for me. Therefore, my first task was to gain sufficient knowledge about probe requests so that I'm able to answer the question, why collecting them could be interesting for researchers, malicious people and other stakeholders.

1.1.1. Probe Request and Probe Response

In general, probe requests are small frames which are sent by mobile phones and other client devices while they are searching for WiFi networks. They consist of components like the signal strength (RSSI), a unique MAC address of the device and a list of previous SSIDs encountered. More precisely, these packets allow the clients to request a so-called broadcast beacon frame with information from all the access points within reach. As soon as an access point receives such a probe request frame, it will answer with a probe response. These include information required to begin with exchanging data between two stations, such as the network name (SSID) of the access point as well as authentication and encryption details.

Since every probe request includes the hardware address of the WLAN adapter of the client device, they can be easily leveraged to perform WLAN-tracking, for example by determining the movement patterns of the device owner. My thoughts concerning the resulting privacy issues are summarized in the end of this report.

1.1.2. Capture-Settings

Since the device and the corresponding software to capture probe requests was already provided, the only remaining task was to configure the capture settings. In order to adjust the available options, the web browser “Firefox” was used to remote-control the Raspberry Pi via the graphical user interface. The following settings were used for the capture-scenarios described in section 2.

Time Horizon: Time span a device is considered “present” after being detected. The setting has no effect on the capture, just on the real-time display in the section Device Details.

The Time Horizon was set to 2 minutes.

Channels: Selection of channels or bands to be scanned in the capture.

The choice for the WLAN channel fell on the non-overlapping channels 1, 6 and 11.

Channel Hopping: Method to be applied for switching between the channels during the scan process.

For the channel hopping method, the option “FIFO” (First In – First Out) was used.

Figure 1 shows the user interface of the Raspberry Pi with all mentioned capture-settings, which were used to perform this experiment.

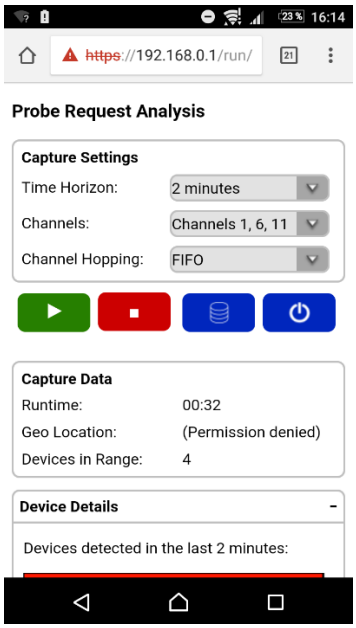

Figure 1: capture-settings for the experiment

2. Capture-Scenario

It is well known that probe requests are being collected in airports, malls, hotels and other public locations to track the moving of passersby. In fact, this method is used to gain profit without the passersby knowing about it, for example by selling the collected data to other companies or using it to perform market research. In a same manner, I wanted to figure out if it's possible to gain an advantage over other people, just by using the provided probing kit and some basic computer science knowledge. Furthermore, I wondered if it's possible to track one specific person, although the hardware-address of each probe request was anonymized by the probing kit. Hence, I started to set up my own tracking-scenarios.

2.1. Scenario 1 – train ride

My first idea was to use the probing kit during one of my train rides from Munich to my hometown Kempten, in order to figure out, how mobility influences the captured data. Unfortunately, whenever I wanted to start a session in the train (both from Munich to Kempten and Kempten to Munich), the process was aborted either immediately after starting it or after a random period of time without my knowledge. With a total amount of 34 collected probe requests, I couldn't perform any rational analysis. Hence, I started to focus on scenario 2, whereby the same termination-problem also appeared in this scenario.

2.2. Scenario 2 – student dorm

The "public location" in scenario 2 was the dorm I currently live in with 19 other students. As we have only one common washing machine, it is very difficult to find a free time slot to wash your clothes. Thus, my goal was to gain the mentioned "advantage" over my housemates by creating something like a time schedule, which reveals the best time span to wash my clothes. Moreover, I tried to figure out who is using the washing machine the most. Before I could start my experiment, I had to make some assumptions, which do not represent a realistic image of the reality, but were necessary to perform a suitable analysis with the Raspberry Pi.

First of all, I assumed that my housemates always carry their mobile phones with them, especially when they are using the washing machine (e.g. to set a timer when the washing process is done) and when they leave the house. Since the washing room is right next to the main-entrance and therefore in the capture radius of the WLAN antennas, I hoped, that I also

could find out, if certain persons leave the building. In addition, I assumed that the WLAN-function of the mobile phones is always activated such that the antennas can receive their corresponding probe requests.

3. Main challenges

One of the two main problems I encountered when I started the experiment, was the anonymization of the hardware addresses of each probe request, which was done by the capture software to ensure privacy regulations. This precaution severely restricted the tracing of specific individuals in my dorm, as the Raspberry Pi generates a new random seed to anonymize the MAC addresses stored in the database. In order to bypass this privacy measure, my idea was to execute long-term capture sessions so that I was able to trace the behaviour of one specific device by tracing its anonymized hardware address during multiple days.

However, as already mentioned in section 2.1, it turned out that in many cases the recording process was interrupted unintentionally, sometimes already after less than 30 minutes. I couldn't figure out if that was just a random crash-pattern or consequence of the unreliable hardware and operating system, as mentioned in the introduction-slides by Mr. Kessner. These crashes made the execution of my experiment much more difficult and the tracking of one specific device impossible. Nevertheless, I was able to collect some data sets for scenario 2. Their analysis is described in detail in section 3.

4. Analysing the captured Data

The following table lists the total amount of capture-sessions I started during the assignment. Each session is associated with the corresponding scenario (1, 2 or test-session) and marked accordingly if the capture-process terminated unintentionally.

No.	Start-Time	End-Time (time of last received probe request)	Channel	Scenario	Unintended Termination
1	2017-12-18 17:05:54	~ 2017-12-18 17:06:15	1-13	Test	
2	2018-01-22 16:31:05	~ 2018-01-22 17:41:57	1, 6, 11	2	
3	2018-01-25 12:19:38	~ 2018-01-25 12:20:37	1, 6, 11	2	X
4	2018-01-25 12:23:38	~2018-01-25 19:37:33	1, 6, 11	2	
5	2018-01-26 11:30:37	~ 2018-01-26 14:00:48	1, 6, 11	2	
6	2018-02-01 15:22:27	~ 2018-02-01 15:23:10	1, 6, 11	2	X
7	2018-02-04 18:45:17	~ 2018-02-04 18:45:29	1, 6, 11	1	X
8	2018-02-09 19:47:50	~ 2018-02-09 19:47:54	1, 6, 11	1	X
9	2018-02-09 19:48:09	.	1, 6, 11	1	X
10	2018-02-09 19:49:00	.	1, 6, 11	1	X
11	2018-02-09 19:53:43	.	1, 6, 11	1	X
12	2018-02-09 19:54:18	.	1, 6, 11	1	X
13	2018-02-09 19:54:27	~ 2018-02-09 19:54:42	1, 6, 11	1	X
14	2018-02-12 12:37:50	~ 2018-02-12 14:50:48	1, 6, 11	2	X
15	2018-02-12 15:03:59	~ 2018-02-12 17:11:22	1, 6, 11	2	
16	2018-02-22 16:14:09	~2018-02-22 16:27:11	1, 6, 11	2	X
17	2018-02-25 16:49:02	~ 2018-02-25 16:49:48	1, 6, 11	Test	

I started my experiment by conducting some basic analysis on the data sets. First, I discovered that channel 1, which uses a frequency of 2412 MHz, was used the most, as shown in figure 2.

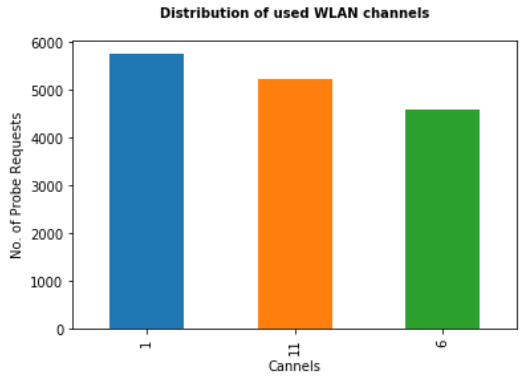

Figure 2: Usage of the WLAN channels stated in the capture-settings

Next, I wanted to know, how many probe requests I collected during the sessions with the longest duration. Therefore, I split the data into different sessions according to each session start-time. Unfortunately, only session 4, 5, 14 and 15 contained enough probe requests to perform a useful analysis on the data. The total amount of collected probe requests during each session is shown in the table below.

Session	No. of collected probe requests
4	4801
5	1622
14	1129
15	463

As every MAC-address contains a vendor prefix of the device manufacturer (OUI – Organizationally Unique Identifier), I tried to determine the most popular manufacturers/devices in my dorm. However, recent attempts of WLAN-tracing have led many suppliers into anonymizing the hardware-address of the probe requests. Hence, I wasn't able to fully determine the distribution of the devices, but I could observe that anonymized probe requests were clearly the most common. Since session 4 includes the most probe requests, it can be used as a representative of the overall distribution of manufacturers, as shown in figure 3.

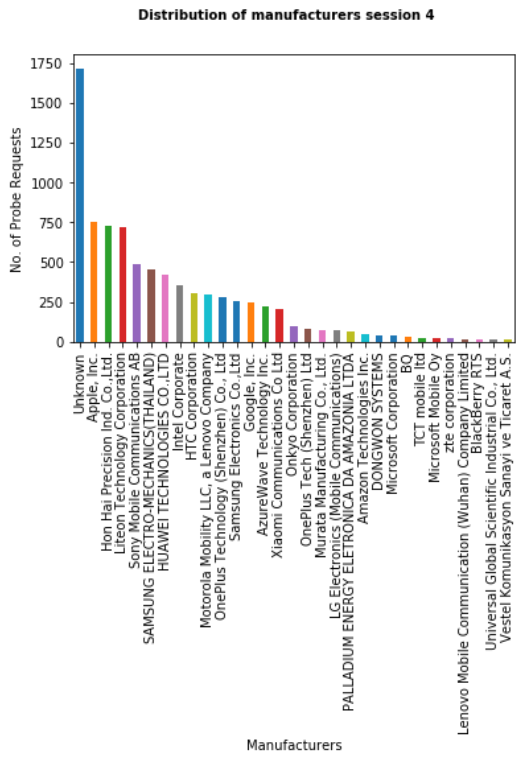

Figure 3: Distribution of manufacturers among the captured probe requests in session 4

Since Apple is one of the largest player in probe request anonymization, it can be conjectured that most of the WLAN-adapters of the devices which anonymize their MAC addresses – represented by the blue "Unknown" bar in the chart– are made by Apple. Yet, it seems like that the devices also send out the real hardware address every once in a while, as "Apple Inc." also represents the second largest bar. Besides that, three other vendors dominate the chart: "Liteon Technology Corporation", "Hon Hai Precision Ind. Co. ,Ltd" and "Sony Mobile Communications AB".

My next goal was to discover if there is a time span in which the washing machine is the least used. Thus, I calculated the number of probe requests received over time for each session individually. Since session 4 had the longest duration as well as the most entries, it revealed the most realistic pattern. As shown in figure 4, the average number of least received probe requests were observed during the hour between 3 p.m. and 4 p.m.

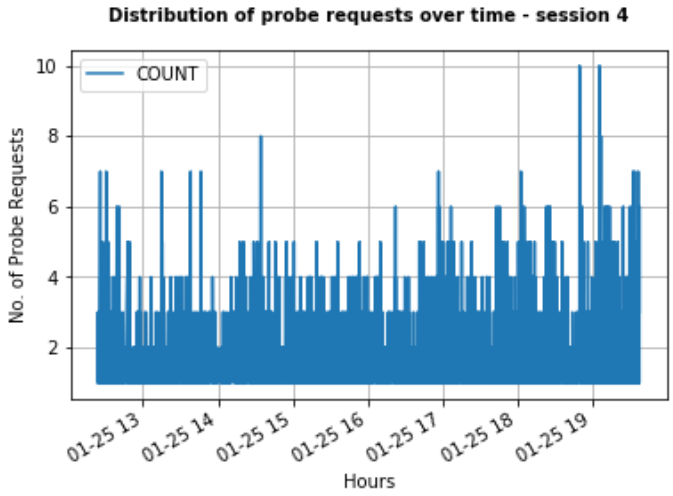

Figure 4: Distribution of probe requests received during session 4

I performed the same analysis for the remaining three sessions, as shown in the following three figures.

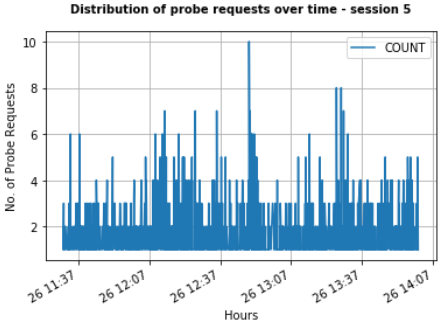

Figure 5: Distribution of probe requests received during session 5

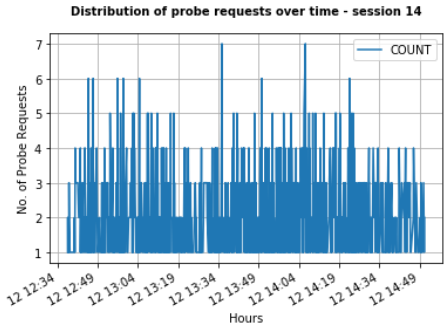

Figure 6: Distribution of probe requests received during session 14

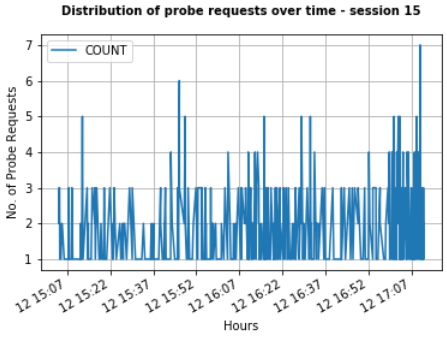

Figure 7: Distribution of probe requests received during session 15

Eventually, I was able to accomplish my overall goal by discovering the best time to use the washing machine. However, I couldn't figure out who is using the machine the most. The tracking of specific roommates was impossible due to the challenges described in section 3.

By looking at the graphs, I concluded that the best time to use the washing machine is between 1:30 p.m. and 4 p.m., as the standard washing process takes an average of 2 ½ hours. I assumed, that the main reason for the higher number of received probe requests in the morning and in the evening is the common schedule-based behaviour of the students in my dorm. In the morning they actively use their devices, e.g. to check the news or to the received messages. Due to lectures, most students are not in the house during lunch time. This means that their phones are also not in the house and their other devices, such as routers or laptops, might be also turned off in order to save electricity costs. In the evening, they return to the house and start using their devices again.

5. Conclusion – Privacy Essay

Although I wasn't able to gain a lot of information from the collected data itself, due to many different challenges such as the anonymization of the hardware-addresses or the truncation of the capturing-sessions, I could gather a lot of interesting information just by performing the experiment.

Since most humans own a mobile phone nowadays, tracking them has become relatively easy by tracking their devices instead. Experiments like the one described in this report show, that everyone with a little knowledge in basic analysis tools can perform WLAN-tracking just by investing a negligible amount of money (the provided probing kit sums up to approximately 80 euros). However, the intentions might differ depending on the person performing the experiment.

On the one hand, malicious people can actively exploit probe requests, for example by recording packages at different locations. Since WLAN clients such as mobile phones are usually used by one specific person, a malicious person is able to create a movement profile of one certain victim. This profile could then be used to predict future locations and lead to other crime like theft.

On the other hand, researchers try to follow ethical guidelines and try to minimize harm, that could emerge by performing these experiments. Their goal is to achieve advantages for the greater good, for example by using the results to improve technologies or by pointing out possible threats for the privacy of humans, which was also my intention in this assignment. Still, MAC-addresses are considered to be individual-related data and are therefore subject to data-privacy laws. In fact, a legally compliant processing of the recorded date is usually not possible without the consent of the persons concerned. The crucial question that arises from this and similar experiments is whether the result can justify the means of the research. Moreover, the border between malicious and benevolent intentions are often set subjectively by humans with different ethical beliefs and backgrounds. While the collection of probe requests to improve the shopping experience of customers (e.g. by analysing movement patterns to install WLAN-hotspots in the most suitable places) might be done with the best intentions of the mall operators, customers may consider this as a deliberate violation of their privacy.

The same problematic nature manifested itself during my experiment. Since I wanted to collect as much data as possible in direct proximity to the washing machine, I placed the probing kit in

the same room. As soon as I started the experiment, my housemates started to question the weird looking device next to the washing machine and asked for its meaning. Although I explained my intentions, the reason for collecting their probe requests and the goal of this assignment, they politely asked me to refrain from doing so. The reason that I did not tell them about my project beforehand was to see how they would react if they find out that somebody is collecting their data and obviously violating their privacy. I concluded that people do indeed appreciate their privacy and try to protect it in the best way they can, although it is almost impossible for a single person to know every available method that can be used to violate it. A good example is again my experiment, since I simply proceeded with capturing data in my room instead of the washing machine room which was only 10 metres away. This reveals another significant issue. The likelihood of catching somebody collecting probe requests is pretty low, as WLAN-tracking can be done passively.

To sum things up, I firmly believe that the user himself should always try to be fully aware of what could happen to his data when he's using a WiFi-capable device. To prevent malicious attacks on his privacy, for example through WLAN-tracking, it should also be his task to support manufacturers and researchers that aim to provide countermeasures like the anonymization of the hardware-addresses, for example by participating in researches and experiments. However, it is also up to the researchers to ensure transparency, clear information about the research for data subjects and the balance between gained user-information and user-privacy to gain the trust of the persons concerned.