PLATFORM FOR COLLECTING USER IDENTIFIERS ON IEEE 802.11 WIRELESS NETWORKS

Petr Ilgner
Doctoral Degree Programme (1), FEEC BUT
E-mail: xilgne00@stud.feec.vutbr.cz

Supervised by: Dan Komosný
E-mail: komosny@feec.vutbr.cz

Abstract: At the present time the widely used Wi-Fi networks provide comfort by zero configuration settings. However it leads to serious privacy issues. The article is focused on user information which can be extracted from service communication. The article presents several ways how to collect these information and introduces the platform for collecting user data. In the first part there is described the available network discovery mechanism followed by section about designed methods for obtaining user information. The last part describes the implemented platform in detail.

Keywords: wifi, wireless, privacy, beacon frame, probe request, user identification, browser fingerprint

1 INTRODUCTION AND RELATED WORK

Using a wireless data network, often referred to as Wi-Fi, to connect end devices such as mobile phones, tablets or laptops has become widespread among common users. Most of the new mobile devices are not equipped with a wired network card which can be used for connecting to a data network.

Some users use public wireless for connecting to the Internet almost anywhere, and it is common that they perform available wireless networks scanning as one of the first activities after entering a public space such as restaurant, coffee house, airport hall or library.

Users except from Wi-Fi networks zero configuration and do not want to solve technical details or specify connection parameters, it leads to the general public not know the ways how is the connection to wireless access points done. The method used for scanning available networks in IEEE 802.11 prefers comfort over privacy of users.

This article discusses the principles and methods used for wireless clients associating to the access points with emphasis on the process of searching for available networks. Several parts which can provide certain data about user resulting in the disclosure user identity are considered.

Several researches related to the use of service information in IEEE 802.11 wireless networks has been published. Their authors have identified privacy issues associated with Wi-Fi probe requests, such as leaking past access points identifiers and user mobility. Some researches also refer to the use of radio waves of a wireless Wi-Fi connection for disclosing information about nearby users, such as [6].

The way how a mobile device and its apps can inadvertently broadcast user personal information through wireless network is introduced in [7]. The paper firstly discusses wireless network security and side-channels which can be used to gain private information. In its main part, authors selected mobile apps and explored plus described the character and content of broadcast data.

A more detailed description of privacy hazards caused by preferred network list leaking is introduced in [4]. Authors work with almost 3000 entries data set of MDNS and SSID broadcasted by mobile devices as a mobile phones, computers and various multimedia devices. They find out that more than half of collected names includes the real name of user. They also performed an online survey and found that about 29 % users do not know their exact device name.

The aim of this work is to design and implement platform for logging data acquired from wireless traffic to obtain data set about captured network information based on discussed principals. The resulting platform should provide an environment to capture described data from various capturing points. The whole platform should consist of low-power capture points devices and server part which collects all captured data and performs the requested analysis. On the basis of the captured data it was observed how tested platform are working with preferred networks list during association to wireless networks.

2 WIRELESS NETWORKS DISCOVERY MECHANISM

In our work we focused on an infrastructure mode, in which require at one wireless access point (AP) and one or more associated wireless client (further marked as client). This topology is called as BSS (Basic Services Set), sometimes called as a cell. This cell is identified by BSSID (BSS Identifier) which is generated from access point physical link address. More BSS can be interconnected together creating an ESS (Extended Service Set) as result. Because in this set more various acccess points are typically present, the ESS bears also its logical name, which is called as ESS Identifier (ESSID), often marked just as SSID. Using this name provides to user the easy remember network identifier and allows logical network definition independently on physical network devices and its location. [2] One BSS can also use more ESSID.

To ensure simple connection process IEEE 802.11 wireless network provides the available networks list. For generating this list there is used a simple principle that in Target Beacon Transmission Time (TBTT), each BSS access point transmits a Beacon frame. It includes information about timestamp, beacon interval, capability info, ESSID identifier and some network specific data as supported speed rates profiles, QoS capabilities. [3]. It can also include certain specific vendor data as information about neighboring APs. The probing interval depends on device platform or administrator network settings. In location where the wireless network radio range highly is utilized, TBTT is reduced. On the other hand, the client requires a quick reconnecting to the network, typically when user wakes up the suspended laptop, he requires the fast reconnection. Also in roaming, where the client is moving across different APs, is required as little time as possible. The typical process of client association to BSS with beacon frame is shown on the figure 1.

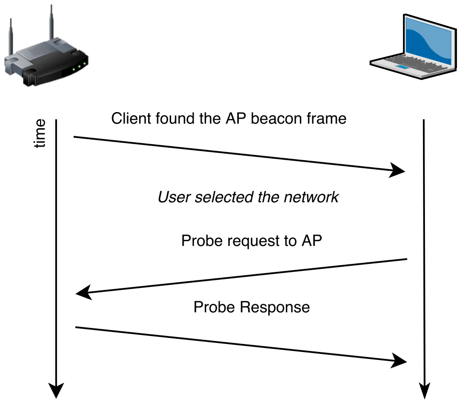

Figure 1: Client association to network diagram.

To ensure automatic and fast connecting to known networks network managers integrated in operating systems hold the list of preferred networks with all of necessary information for connecting. When client is not connected, network manager is periodically trying to connect these preferred networks. As a optimization most of platforms do not wait for beacon packet with corresponding ESSID but send the probe requests directly to all saved networks. It helps the devices to make quick connection. After receiving some valid probe response from any AP, the devices are associated to network.

Although modern wireless security methods provides an encryption of transmitted data with different encryption keys for all BSS clients, services frames which include probe request and response, beacon frames are not encrypted [2]. This data allow the user identification – summarized, from this service data, we are able to get for all clients in neighbor extract the unique identification profile, which includes the set of user preferred networks and user's wireless radio unique physical address [3].

3 OBTAINING USER GEOLOCATION AND OTHER DATA

Each saved network can be further examined. If the ESSID network name is enough unique we can try to find the saved network location. For extract this data we can use some SSID public database, for example the WiGLE service [5]. This service provides an API data access and allow using this data for research. This database contain about more than 420 millions saved network with precious coordinates entries aquired by more then 200,000 stumblers.

3.1 SOCIAL AND MEMBERSHIP DATA

When some user uses wireless connection on his device often it is able to compile his connection location history from the detected network geographical location. The implemented logging platform perform the location queries to database for all received preferred network. By analyzing the most frequently occurring location we can guess the user estimated location.

If user is using the free WiFi hotspot on different places while traveling we can find out which kind of business attends, eg. class of restaurant, cinema, stores. It can provide the indices to determinate user estimated age or interests. As network managers stores networks associated successfully in the past, we can infer that user which sends probe requests to employers only network is its employer. In a similar way if we captured probe request to academic network called 'eduroam' we can say that this user is academic related.

4 IMPLEMENTED CAPTURING PLATFORM

In the Introduction were listed the requirements for the logging system. The designed architecture consist of collecting device. For ensuring the low-power requirement and low price of collecting devices we used the ESP8622 based boards. It comes with integrated IEEE 802.11 b/g/n wireless chip. The board is also equipped with 4 MB flash memory which can be used for storing the captured data. The collecting program is written in C++. The board in sniffing mode can be easily powered by power bank what enable to place it to hard-to-reach places for example corridors or another public places. The reached battery live is about 2-3 days on 10.000 mAh USB power bank.

As a second type of capturing device was used the TP-LINK TL-WDR3600 wireless router with installed OpenWRT Linux distribution for embedded devices. The collecting program has been ported to this platform – the probe requests data are retrieved from `aircrack-ng` suite. Thanks to the presence of Ethernet interface, the device can be connected to local network. This device is equiped with two independent radios, which allows to scan on 5 GHz band at the same time.

As a server part is used the created Python application which is connected to PostgreSQL database. It provides TCP server for data storing and interfaces for displaying and export the collected data.

4.1 Capturing process

Due the European rule, the probe request can be received on overall 13 channels [[2]] on 2.4Ghz band. The capturing on all channels is ensured by frequency hopping technique. It is implemented as infinite loop in which is every one second changed the channel. It can result in potentially missed beacon frames. In the environment where users only pass through with their mobile devices should be better to divide the spectrum into two half and use dedicated devices to every half.

The collected data delivery is made by sending over another Wi-Fi network around. The capturing board stores the captured probes for every device in local stack with time stamp. When is the stack filled or after one hour, all saved data are send over TCP session to collecting server. During sending stack, the collecting process is interrupted. It typically takes about 6 second including the board wireless connection time. The architecture of capturing devices is shown on figure 2.

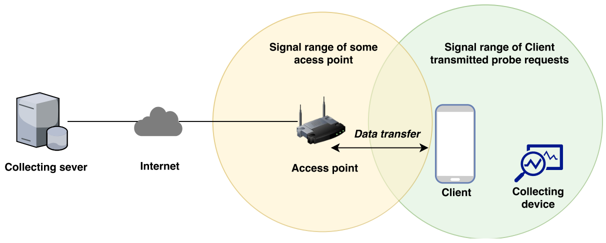
Signal range of some acess point
Signal range of Client transmitted probe requests
Collecting sever
Internet
Access point
Data transfer
Client
Collecting device

Figure 2: Collecting boards architecture.

5 TEST SCENARIO AND COLLECTED DATA

This section looks at performed platform deployment which aims to verify the platform design usability and provide some data about its performance. They were used two ESP8266 capture boards and one TP-LINK router. First board together with the router was placed into the computer lab. The test showed that devices are capable to grab probe requests from devices in the neighboring rooms. Second board was installed to the main corridor close to the faculty library and was powered by power bank. Every student who accesses the computer lab should go through the main corridor. This testing was performed for two working days. Overall collection statistics for capturing devices can be found in Table 1. 36 % of devices detected in computer lab was found and identified in the main corridor, which is lower then excepted. It can be caused by insufficient time of presence in the range of capture board. It is possible to observe that the router solution was more effective in the count of detected devices, which is caused by more powerful radio part at the price of much higher power requirement. On the basis of performed test, can be said that the platform is able to track the wireless devices completely passive.

Using the Wigle API [5], it was examined for how many of captured SSID can be determined a probable position. We considered as location-determinable these ones for which API returned acess points with identical location. Overall, 14 % of founded SSID was location-determinable. Although it is a relatively small number, for 67 % of detected clients was found more than two location-determinable SSIDs.

6 CONCLUSION

Users of wireless mobile devices are exposed to passive tracking without detection possibility. The information which can be grabbed are personal, it can include the employer or estimated place of residence. Using an short active attack can be on some platform acquired the user browser fingerprint which can be identified with fingerprint acquired on any web page.

Table 1: Number of probe requests captured and devices detected

	COMPUTER LAB		MAIN CORRIDOR
	TP-LINK router	1st capture board	2nd capture board
Totally captured requests	42 124	31 059	98 468
Unique devices detected	681	520	1 406

The practical output of work is implemented collecting platform. Using their data we can track the location of persons which are occurring near the collecting points. By longer term operation of this platform can provide data which can be used in various research or and analyze differences in the behavior of different operating systems as future work.

REFERENCES

[1] *Free WiFI map: List of public places with free WiFi* [online]. [cit. 2018-03-10]. Dostupné z: https://wifispc.com/

[2] *Telecommunications and information exchange between systems Local and metropolitan area networks: Wireless LAN Medium Access Control and Physical Layer Specifications*. New York, USA: IEEE, 2016. DOI: 10.1109/IEEESTD.2016.7786995. ISBN 978-1-5044-3645-8.

[3] CALHOUN, P., M. MONTEMURRO a D. STANLEY. *RFC416: Control and Provisioning of Wireless Access Points (CAPWAP) Protocol Binding for IEEE 802.11*. IETF, 2009. Dostupné také z: https://tools.ietf.org/html/rfc5416

[4] KONINGS, Bastian, Christoph BACHMAIER, Florian SCHAUB a Michael WEBER. Device Names in the Wild: Investigating Privacy Risks of Zero Configuration Networking. In: *2013 IEEE 14th International Conference on Mobile Data Management* [online]. IEEE, 2013, 2013, s. 51-56 [cit. 2018-03-15]. DOI: 10.1109/MDM.2013.65. ISBN 978-0-7695-4973-6. Dostupné z: http://ieeexplore.ieee.org/document/6569062/

[5] *WIGLE.net: All the networks. Found by Everyone.* [online]. [cit. 2018-03-03]. Available from: https://www.wigle.net

[6] ZHANG, Jin, Bo WEI, Wen HU a Salil S. KANHERE. WiFi-ID: Human Identification Using WiFi Signal. In: *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)* [online]. IEEE, 2016, 2016, s. 75-82 [cit. 2018-03-15]. DOI: 10.1109/DCOSS.2016.30. ISBN 978-1-5090-1460-6. Available from: http://ieeexplore.ieee.org/document/7536315/

[7] ATKINSON, John S., John E. MITCHELL, Miguel RIO a George MATICH. Your WiFi is leaking: What do your mobile apps gossip about you?. *Future Generation Computer Systems* [online]. 2018 [cit. 2018-03-15]. DOI: 10.1016/j.future.2016.05.030. ISSN 0167739X.

[8] MARTIN Jeremy, MAYBERRY, Travis and DONAHUE Collin. A Study of MAC Address Randomization in Mobile Devices and When it Fails. *Proceedings on Privacy Enhancing Technologies* [online]. De Gruyter Open, 2017, 365-383 [cit. 2018-03-15]. DOI: 10.1515/popets-2017-0054. ISSN 2299-0984.

[9] SCHRITTWIESER, S., P. RESCHL a M. LEITHNER. Fast and Reliable Browser Identification with JavaScript Engine Fingerprinting. *Web 2.0 Security&Privacy 2013* [online]. 2013 [cit. 2018-03-15]. Available from: https://www.sba-research.org/wp-content/uploads/publications/jsfingerprinting.pdf