TP 2: Wireless Privacy

1 Introduction

The open medium used by wireless communications makes them vulnerable to eavesdropping and collection of all kind of sensitive information. Despite security mechanisms, some data are not encrypted. For instance in Wi-Fi, the headers as well as all management frames are not encrypted. The objective of this practical work is to observe potentially sensitive information available in clear in Wi-Fi communications and in second time to study fingerprinting and tracking techniques.

2 Private information in Wi-Fi communications

2.1 MAC address

The header of Wi-Fi frames contains the MAC address of the source, which is always available in clear since the header is not encrypted. This unique identifier can be used for tracking the device location and activity.

Question: 1 *Using tshark, extract the MAC address of the wireless stations (excluding the Access Points).*

Question: 2 *Select a subset of those MAC addresses and resolve their OUI (*`http://www.coffer.com/mac_find/`*).*

Question: 3 *What is the OUI ? What information can be inferred from the OUI ?*

Question: 4 *Count the number of distinct MAC address in your dataset. How does this information relates to the number of people in the area ?*

2.2 SSIDs

Wi-Fi probe requests are used for service discovery: discovering surrounding Wi-Fi APs. A probe requests include a SSID field that can be left empty. When it is not empty, it contains the SSID of the network searched by the device. As all management frames, probe requests are never encrypted, leaving their content exposed to eavesdropping. Wi-Fi devices thus reveal the SSIDs of the network they have been connected to in the past, potentially exposing information about the owner.

Question: 5 *Using tshark or wireshark, display SSIDs found in probe requests.*

Question: 6 *Select 3 SSIDs and infer information about them. Information includes: the location of the network (GPS coordinates), the nature of location (home, restaurant, hotel) ...*

Note: Wigle `https://wigle.net/` is a crowdsourced service hosting a database of Wi-Fi networks. The attributes of this database include the SSID, the BSSID, the GPS coordinates, ... You can use this service to retrieve a location from a SSID.

2.3 Other: DNS, mDNS

When a device is associated to an AP, its traffic can also reveal sensitive information. Several protocols are can reveal information about the owner and its activity. This is for instance the case with DNS and service discovery protocol like mDNS (a.k.a. bonjour).

Question: 7 *Using tshark or wireshark, display the content of DNS and mDNS queries. What information can be found in those packets ?*

3 Fingerprinting Wi-Fi devices

A solution to prevent tracking based on the link layer identifier is to use a random MAC address. This is for instance the case with some Wi-Fi devices that uses a pseudonym MAC address in their probe requests and this pseudonym is changed periodically. In this cases, the MAC address cannot be used for tracking and the tracker needs to use more advanced techniques such as fingerprinting. An efficient fingerprinting techniques based on Information Elements has been developed by one of Privatic's PhD student [4]. Information elements are included in Wi-Fi frames and are used to communicate extended information on the device and its capabilities. Based on a real world dataset you will evaluate the fingerprinting potential of information elements by computing and plotting the anonymity sets with `kmap`.

3.1 The dataset

The Sapienza dataset [1] is a large dataset of probe requests that have been collected in Rome in 2013. It contains more than 11M probes sent by 160K devices. It is publicly available on the website Crawdad (`http://www.crawdad.`

org/). A detailled description of this dataset is available in the corresponding research paper [2]. A subset of this dataset will be used: the Vatican1 dataset that has been collected in the city of Vatican.

Question: 8 *Has this dataset been anonymized ? If yes how has it been anonymized.*

Question: 9

Is the source MAC address field a reliable information in this dataset ?

3.2 Information elements

We will focus only on probe requests and on a subset of Information elements. The considered Information Elements are the following (`wlan.sa` being the source MAC address will be used as a unique identifier but will obviously not be used for fingerprinting):

```
wlan.sa
wlan_mgt.tag.number
wlan_mgt.supported_rates
wlan_mgt.extended_supported_rates
wlan_mgt.ht.capabilities
wlan_mgt.ht.ampduparam
wlan_mgt.ht.mcsset.rxbitmask
wlan_mgt.htex.capabilities
wlan_mgt.txbf
wlan_mgt.asel
wlan_mgt.extcap
wlan_mgt.interworking.access_network_type
```

Question: 10 *What is the purpose of the information element* `wlan_mgt.supported_rates` *? respectively* `interworking.access_network_type` *and*

3.3 Preparing the data

Before generating the anonymity sets, we need to extract the information from the datasets and to preprocess it so that it matches the format expected by `kmap`.

Question: 11 *Using* `tshark`*, parse the pcap files and extract the information elements. A sketch of the corresponding script can be found in the file* `extract_IE.sh`*.*

The generated files contains one entry (line) per frame. In order to compute anonymity sets, only one entry per device is required.

Question: 12 *Using* `sort`*, process the previous file in order to have only one line per device, ie per MAC address. Then use the script* `convert_to_pickle.py` *to convert this dataset in the pickle format.*

3.4 Generating the anonymity sets

Question: 13 *Using* `kmap`*, generate and plot the anonymity sets corresponding to all those attributes. A sketch of the script can be found in the file* `make_plots_IE.py`

Question: 14 *Are the selected IE enough to generate a globally unique fingerprint ? Why ?*

Question: 15 *Are they enough to create a locally unique fingerprint, i.e. are they enough to uniquely identify someone within a small crowd (100-1000 individuals) ?*

4 MAC address re-identification from UUID

One of the information element found in probe requests contains a UUID (universally unique identifier). According to the RFC 122 [3] the UUID is typically derived from one of the devices' MAC address using a specific algorithm. Even if the real MAC address of a device is not available in the frames (randomized address, anonymized dataset), it may be possible to reverse the UUID generation process to go back to the original MAC address. This attack have been performed in [4] and lead to the re-identification of the MAC address in around 70% of the cases. The objective of this exercise is to reproduce the re-identification attack on the Sapienza dataset.

4.1 Preparing the data

Question: 16 *Using tshark, extract the source MAC address and the UUID for all the frame containing a UUID (the field corresponding to the UUID is* `wps.uuid_e`*). Your output file should contain one pair* $< MACaddr, UUID >$ *per line. A sketch of the script can be found in the file* `extract_uuid.sh`*.*

Question: 17 *Using* `sort`*, process the previous file in order to have only one line per device, ie per MAC address.*

4.2 Re-identifying MAC addresses

Question: 18 *Write a script that takes as input a file containing UUID and attempt a re-identification attack on each of them. A sketch can be found in* `reverse_uuid.py`*.*

Note: The re-identification attack can use the information found in the MAC address field.

Question: 19 *Would this attack be feasible without any information on the real MAC address ? How much times would it take ?*

Question: 20 *Design an efficient method to re-identify the MAC address from a UUID assuming that you have no prior information on the MAC address.*

References

[1] Marco V. Barbera, Alessandro Epasto, Alessandro Mei, Sokol Kosta, Vasile C. Perta, and Julinda Stefa. CRAWDAD dataset sapienza/probe-requests (v. 2013-09-10). Downloaded from http://crawdad.org/sapienza/probe-requests/20130910, September 2013.

[2] Marco V. Barbera, Alessandro Epasto, Alessandro Mei, Vasile C. Perta, and Julinda Stefa. Signals from the crowd: Uncovering social relationships through smartphone probes. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, pages 265–276, New York, NY, USA, 2013. ACM.

[3] P. Leach, M. Mealling, and R. Salz. A universally unique identifier (UUID) URN namespace. RFC 4122 (Proposed Standard), July 2005.

[4] Mathy Vanhoef, Célestin Matte, Mathieu Cunche, Leonardo Cardoso, and Frank Piessens. Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In *ACM AsiaCCS*, Xi'an, China, May 2016.