HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD
www.heig-vd.ch

Department: Information and communication technologies
Faculty: Information technology and communication systems
Orientation: Software engineering

Bachelor Thesis

RF fingerprinting on NFC devices

Intermediary report

Not confidential

Student: Luc Wachter

Project proposed by: Joël Conus
Kudelski Group SA
22-24, Route de Genève
1033 Cheseaux-sur-Lausanne

Teacher in charge: Alberto Dassatti

Academic year: 2019-2020

Yverdon-les-Bains, 12th June 2020

1 Scope statement

Scope statement for RF fingerprinting on NFC devices (Bachelor Thesis)

Project purpose: Develop a tool to identify NFC devices by analysing the RF spectrum of their transmitted signals

Duration: 450 hours (ends 31st July 2020)

Detailed description

RF fingerprinting is a technique that allows the identification of radio transmitters (such as IoT devices) by analysing the spectrum of their transmissions. This analysis can typically be performed using machine learning algorithms.

NFC technology is often used in access control and payment applications but many implementations are vulnerable to relay attacks with research and tools that facilitate such attacks being publicly available.

The goal of this project is to determine if RF fingerprinting could be used as an authentication technique against relay attacks.

The main steps of this project are the following:

- Build a simple lab setup with Software-Defined Radio (SDR) equipment to acquire signals between an NFC device and its reader
- Acquire RF spectrum data of various NFC devices
- Analyse the signals
- Classify the signals of the devices by using supervised machine learning classification techniques in order to differentiate trusted devices from attacker / relay devices
- Determine if this identification technique could be used as an authentication feature against relay attacks

As the receiving equipment (SDR) has an influence on the recorded signals, for this project we consider a single receiver to record the RF samples. Similarly, the lab setup should be built to provide an ideal low-noise & low-interference environment to simplify the analysis phase.

The expected deliverables are the following:

- A tool able to identify NFC devices by analysing the RF spectrum of their signals, at least in an ideal environment and with a small number of devices
- A detailed account of the steps taken and the setup used (as part of the report)
- An analysis of the results (as part of the report)

Collaboration with other researchers in this field is wished (EPFL, ElectroSense).

Table of contents

2 Introduction

2.1 Project description

Radio Frequency (RF) fingerprinting is a technique that allows the identification of radio transmitters by extracting small imperfections in their spectrum. These imperfections are caused by tiny manufacturing differences in the devices' analog components. Using Software-Defined Radio (SDR) equipment, we can analyse this spectrum in order to extract the aforementioned differences and identify a device.

Such techniques can be used on any type of radio transmission: Bluetooth, BLE, WiFi, LTE, etc. This project aims to use RF fingerprinting on NFC devices. Indeed, NFC is often used in access control and payment applications but many implementations are vulnerable to relay attacks. Spoofing the imperfections in an emitter's radio spectrum is close to impossible at the present time, since it is essentially a hardware signature. This is why a technique like the one described here would be a valuable additional security layer.

The goal of this project is to determine whether applying machine learning techniques to the problem of RF fingerprinting NFC devices could be used as an authentication technique, in order to prevent relay attacks. If a dataset of sufficient quality and variety can be produced, it could be another outcome of the project. Indeed, while some exist for 802.11 communications, no dataset seems to be available for raw recordings of NFC transactions.

2.2 Context

This project is conducted in the context of my bachelor thesis at HEIG-VD.

- Department: Information and communication technologies
- Faculty: Information technology and communication systems
- Orientation: Software engineering

It was proposed by Mr Joël Conus of Kudelski Group SA.

2.3 Document description

This intermediary report marks the middle of the project. Because of this, it is firmly anchored in the analysis and conception phases, which means much of what is presented is subject to change in the second half of the project.

Nevertheless, this document describes the research done while studying the state of the art. It then presents the acquisition setup and the results it brought, before showing the steps undertaken to validate the captured signals through decoding. Finally, it showcases the first conception ideas and decisions made for the learning model, in light of our study of the state of the art.

3 State of the art

3.1 Taxonomy

It is certainly useful to start with a review of the different ways to categorize the features and algorithms used by researchers in the field of Radio Frequency Machine Learning (RFML). The goal is to define our needs precisely and select the important things to consider.

3.1.1 Taxonomy for features

The features we select must allow us to identify a precise device among potentially very similar devices. We need what Delgado et al. [1] describe as a Physical Unclonable Function (PUF). PUFs are physical distortions that are unique to a specific system. They are another way of talking about fingerprints.

Xu et al. [2] propose three ways to categorize radio signal features:

- based on the specificity of the feature (from vendor specific to device specific),
- based on the layers (PHY, MAC, Network and higher),
- and based on the acquisition method (passive or active).

Features from the MAC and higher layers typically require in depth knowledge of the protocols in play. Not only that, but they also tend to be less specific than we would like (either vendor specific or depending on the type of device). This indicates we should probably focus on the physical (PHY) layer features, which rely on imperfections in the manufacturing process of the devices.

3.1.2 Taxonomy for fingerprinting algorithms

Riyaz et al. [3] provide a visual categorization of fingerprinting approaches, which we adapt in figure 1. We take a look at these approaches in the following paragraphs.

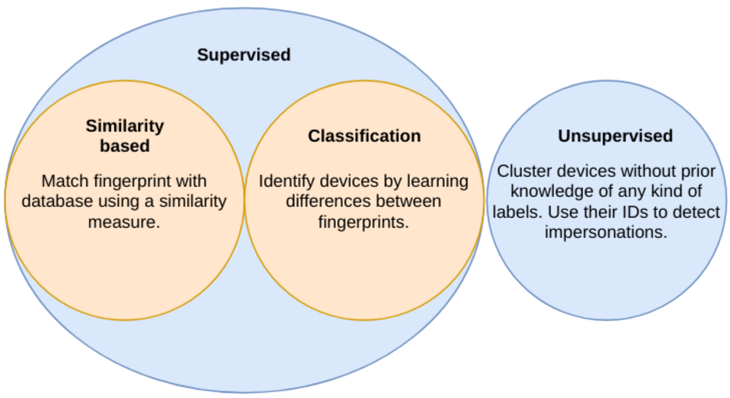

Figure 1: Fingerprinting algorithms taxonomy

Supervised approaches: Supervised approaches use features from labelled data to generate a function capable of separating the different classes. These approaches can be further categorized in similarity based and classification techniques.

Similarity based techniques are white-list algorithm that use a database of fingerprints and a similarity measure to determine whether a device is legitimate. Developing a technique like this usually requires prior knowledge of vendor specific device features [3].

Classification systems can also require deep knowledge of the features and protocols used, in the case of "traditional", manually tuned classifiers. Those are built to extract predetermined features and work similarly to other white-list algorithms afterwards.

In this age of deep learning though, research seems to be more interested in classification techniques that are able to extract the features they need by themselves. This can be done through deep Multi-Layer Perceptrons (MLP) [1, 4] or through more advanced techniques like Convolutional Neural Networks (CNN) [3, 5, 6, 7, 8]. The latter have proved very powerful in domains like computer vision, natural language processing and recommendation systems. This success is one of the reasons experimentations on CNNs are common in RFML research.

Unsupervised systems: Unsupervised approaches cannot by themselves discriminate a legitimate device from an illegitimate one. They don't have that information, since they work with unlabeled data. In order to detect attempts of impersonation, such a system has to keep a record of fingerprints and linked identifiers (MAC addresses, serial numbers...). It can then throw an alert and update a black-list when multiple fingerprints are linked to the same ID or when multiple IDs are linked to the same fingerprint [2, 9].

Xu et al. [2] specify that unsupervised approaches are appropriate when the fingerprints of legitimate devices are not available.

3.2 Acquisition

The vast majority of the research considered for this project uses USRP systems to record transmissions as raw I/Q signals. The number of devices can be anywhere from 5 to 500 (but most often less than 20). They all use data acquired from WiFi (802.11) or Zigbee (802.15.4) devices. [3, 5, 6, 7, 8, 9]

WiFi and Zigbee both use Quadrature Phase Shift Keying (QPSK) modulation schemes. They differ from NFC's modulation scheme: On-Off Keying (OOK) which is a form of Amplitude Shift Keying (ASK). (Except for NFC type B, which uses BPSK in target to initiator mode.)

Some, like Sankhe et al. [8], also describe how they add artificially induced impairments to simulated signals with MatLab.

3.3 Features

3.3.1 Features selection

Even if we don't plan to manually select and extract the features that will form the fingerprints of our devices, it is useful to learn about them. It will allow us to make sure they are present for the algorithm to extract and also allow us to design preprocessing methods that magnify the features.

Whether we end up with a system that is able to identify many devices uniquely, or one that only tries to separate a specific device from the others, we will need device specific accuracy. We don't want to make relay attacks impossible only if the attacker doesn't use a device from the same vendor as the victim's device. This is why in section 3.1.1, we concluded that the features we are most interested in are from the physical layer.

Because of their nature, these features should be appropriate no matter the protocol used. The following list is composed of features described by Riyaz et al. [3] and also used in other works.

- I/Q imbalance: The amplitude and the phase are not exactly the same on the in-phase and the quadrature signals, because of the imperfections in the quadrature mixers.
- Phase noise: When the baseband signal is up-converted to the carrier frequency, it is sensible to phase noise which creates rotational vibrations.
- Carrier frequency offset: The difference between the carrier frequency of the transmitter and the carrier frequency of the receiver.
- Harmonic distortions: The Digital-to-Analog Converters (DAC) cause harmonic distortions because of imperfections.

3.3.2 Preprocessing

It is clear that preprocessing the data appropriately can greatly increase the accuracy of a model and reduce its complexity.

The first question to ask is how should the data be partitioned, in order to be fed to the learning algorithm. This question will be discussed in section 3.4.3 since it is identical to choosing the input of our model.

Several works mention wavelet transforms (either discrete or continuous) as effective means to amplify the characteristic features of a wireless device [2, 5, 6]. They report increased accuracy and scalability, and reduced complexity. It is certainly worth it to explore this preprocessing method.

3.4 Machine learning applied to radio frequency

Difficulties (ref)

3.4.1 Supervised or unsupervised

This describes the work done by Nguyen et al. [9] on such a system. They use a Nonparametric Bayesian model to cluster devices based on their fingerprints. This technique allows them to discriminate between an unknown number of devices that the model never encountered before.

They show good results with four devices using two features strictly from the PHY layer.

In the context of this project, this would require us to be able to record the NFC serial number of a device at the same time as we extract the fingerprint.

3.4.2 Comparing supervised approaches

Several articles have compared the performance of different machine learning approaches.

SVM, DNN, CNN, MST A-LM

CNNs -¿ good shift invariance thanks to pooling layer?

3.4.3 Neural network architecture

Inputs (windows, sliding?...) partition length: [6] ...

4 Dataset creation

4.1 Radio setup

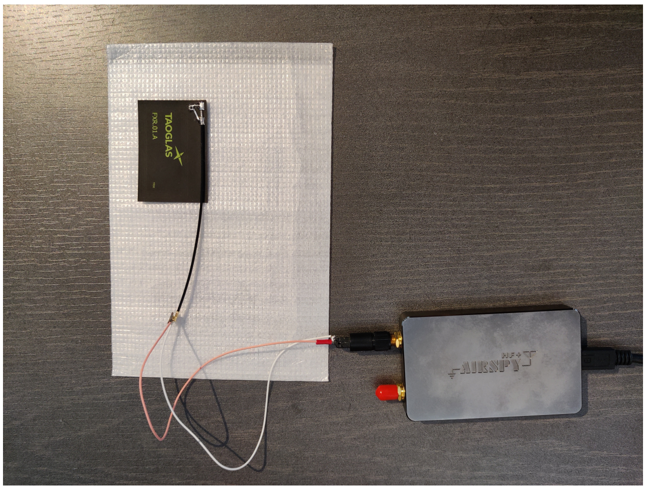

Figure 2: SDR and antenna setup

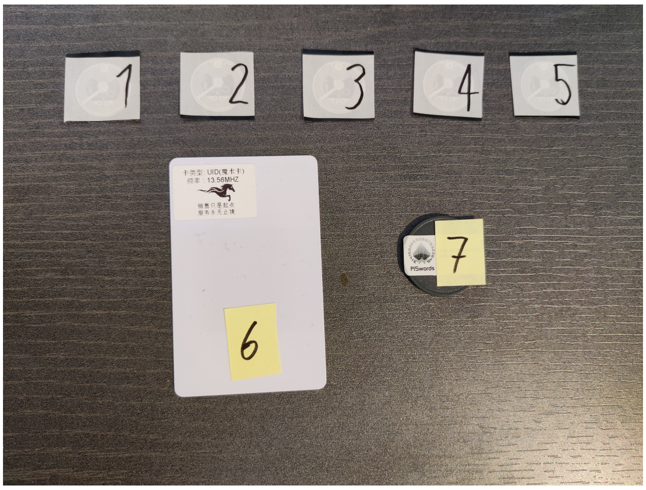

Figure 3: NFC tags 1-7

4.2 Inventory of devices

When possible, the content of the tags is harmonized to ensure the algorithm won't use the content as a feature to identify devices.

4.3 Dataset description

An early observation here is that interferences are not as troublesome as they are when working on WiFi fingerprinting. Because of the frequency used by NFC and the necessity for proximity, a lot of interference sources... *Furthermore, channel variation problems are tied to coding methods that use subcarriers (like OFDM) and don't affect NFC.*

Name	Serial number	Type	Chip	ATQA	SAK
tag1	04:5A:F5:2A:37:60:80	ISO 14443-3A (NFC-A)	NTAG213	0x0044	0x00
tag2	04:7A:F6:2A:37:60:80	ISO 14443-3A (NFC-A)	NTAG213	0x0044	0x00
tag3	04:7B:F6:2A:37:60:80	ISO 14443-3A (NFC-A)	NTAG213	0x0044	0x00
tag4	04:5B:F6:2A:37:60:80	ISO 14443-3A (NFC-A)	NTAG213	0x0044	0x00
tag5	04:99:F6:2A:37:60:80	ISO 14443-3A (NFC-A)	NTAG213	0x0044	0x00
tag6	08:72:8A:04	ISO 14443-3A (NFC-A)	Mifare Classic 1k	0x0004	0x08
tag7	6E:13:66:01	ISO 14443-3A (NFC-A)	Mifare Classic 1k	0x0004	0x08
tag8	CF:6C:B1:23	ISO 14443-4 (NFC-A)	Mifare Classic 4k	0x0002	0x38
tag9	01:27:04:98:3A:B6:5F:9B	JIS 6319-4 (FeliCa)	RC-S967	-	-

Table 1: Inventory of PICC devices

Name	Serial number	Type	Chip	ATQA	SAK

Table 2: Inventory of PCD devices

5 Model conception

5.1 Model architecture

Decision on the metaparameters...

5.2 ...

6 Conclusion

Good

Bibliography

1. DELGADO, Oscar; KECHTBAN, Louai; LUGAN, Sebastien, et al. Passive and active wireless device secure identification. *IEEE Access* [online]. 2020, pp. 1–1 [visited on 2020-05-08]. ISSN 2169-3536. Available from: https://ieeexplore.ieee.org/document/9082628/.
2. XU, Qiang; ZHENG, Rong; SAAD, Walid, et al. Device Fingerprinting in Wireless Networks: Challenges and Opportunities. *arXiv:1501.01367 [cs]* [online]. 2015 [visited on 2020-04-24]. Available from arXiv: 1501.01367.
3. RIYAZ, Shamnaz; SANKHE, Kunal; IOANNIDIS, Stratis, et al. Deep Learning Convolutional Neural Networks for Radio Identification. *IEEE Communications Magazine* [online]. 2018, vol. 56, no. 9, pp. 146–152 [visited on 2020-04-24]. ISSN 0163-6804 1558-1896. Available from: https://ieeexplore.ieee.org/document/8466371.
4. STANKOWICZ, James; ROBINSON, Josh; CARMACK, Joseph M., et al. Complex Neural Networks for Radio Frequency Fingerprinting. In: *2019 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)* [online]. Rochester, NY, USA: IEEE, 2019, pp. 1–5 [visited on 2020-02-24]. ISBN 978-1-72814-352-1. Available from: https://ieeexplore.ieee.org/document/8923089.
5. OYEDARE, Taiwo; PARK, Jung-Min Jerry. Estimating the Required Training Dataset Size for Transmitter Classification Using Deep Learning. In: *2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)* [online]. Newark, NJ, USA: IEEE, 2019, pp. 1–10 [visited on 2020-03-13]. ISBN 978-1-72812-376-9. Available from: https://ieeexplore.ieee.org/document/8935823.
6. YOUSSEF, Khalid; BOUCHARD, Louis; HAIGH, Karen, et al. Machine Learning Approach to RF Transmitter Identification. 2017. Available also from: https://www.researchgate.net/publication/320890695_Machine_Learning_Approach_to_RF_Transmitter_Identification.
7. MORIN, Cyrille; CARDOSO, Leonardo; HOYDIS, Jakob, et al. Transmitter Classification With Supervised Deep Learning. *arXiv:1905.07923 [cs, eess]* [online]. 2019 [visited on 2020-03-13]. Available from arXiv: 1905.07923.
8. SANKHE, Kunal; BELGIOVINE, Mauro; ZHOU, Fan, et al. No Radio Left Behind: Radio Fingerprinting Through Deep Learning of Physical-Layer Hardware Impairments. *IEEE Transactions on Cognitive Communications and Networking* [online]. 2019, pp. 1–1 [visited on 2020-02-24]. ISSN 2332-7731, 2372-2045. ISSN 2332-7731, 2372-2045. Available from: https://ieeexplore.ieee.org/document/8882379.
9. NGUYEN, Nam Tuan; ZHENG, Guanbo; HAN, Zhu, et al. Device fingerprinting to enhance wireless security using nonparametric Bayesian method. In: *2011 Proceedings IEEE INFOCOM* [online]. Shanghai, China: IEEE, 2011, pp. 1404–1412 [visited on 2020-06-12]. ISBN 978-1-4244-9919-9. Available from: http://ieeexplore.ieee.org/document/5934926/.

List of Figures

List of Tables