Wirtschafts-
wissenschaftliche
Fakultät

The Full Title of your Thesis

Your Name

Bachelor Thesis – 0, 2019.
Cyber Security
Department of Information Systems
University Münster, Germany

Principal Supervisor: Prof. Dr.-Ing. Thomas Hupperich

Abstract

The usage of smartphones has become quite popular in the last decade. Every new smartphone model contains new hardware features, thus making people more attracted to benefit from these features. But the more features a smartphone contains, the more potential privacy and security issues arise from these features. Many of these features utilize the various built-in smartphone sensors like the accelerometer. Accelerometer data lead to the ability for installed applications to track the user's actual condition and activity. Previous laboratory studies have proved that hardware imperfections during the accelerometer manufacturing process, provide the possibility to recognize smartphones by utilizing signal feature extraction. In our approach, we demonstrate whether this method is applicable to recognize device models just as well as unique devices.

Contents

1 Introduction

In this chapter, we give an overview on our motivation and the related work done so far on fingerprinting and recognizing accelerometers, followed by our contribution to this field of studies.

1.1 Botivation

The ability to recognize devices such as tablets and desktops has been discussed quite often in the last decade. Methods such as supercookies and tracking static IP addresses, and tools like panopticlick [**?**] have been utilized to recognize this kind of devices. With the growing popularity of mobile devices, advertising companies have also developed an interest in this category of devices. The challenge of recognizing mobile devices remains in the great similarity between the hardware and software specification of device models. Additionally, traditional tracking options have been restricted to prevent mobile devices being fingerprinted. For instance, in case of tracking cookies, the cookie law [**?**] which regulates storing and retrieving information on a computer, smartphone, or tablet of visitors to a website in EU countries was passed on May 26 2011. User-defined settings in case of privacy-concerning features, e.g., GPS sensors prevent applications from tracking the current device location. Another example is Apple's improved policy since the presentation of iOS 7. Apple's iOS application developers are prohibited from using unique device features such as the **UDID!** (**UDID!**), **IMEI!** (**IMEI!**), and **MAC!** (**MAC!**) address of the Wi-Fi interface as recognition mechanism. These effective preventive measures have caused adversaries to look for fingerprinting methods that do not require any sort of permission. Getting deeper into the modern features of mobile devices, it turns out that the accelerometer does not need any of the mentioned permissions, proved by various gaming and fitness applications that have never required permission to access the accelerometer readings so far. Gaming applications need accelerometer readings to track device movements and perform predefined actions, according to recognized patterns from the accelerometer readings. Fitness tracking applications use accelerometer readings to distinguish between user's physical activities and extract the resulting heart rate [**?**]. Previous studies have illustrated that calibration imprecisions and hardware imperfections during the manufacturing process lead to unique hardware-tolerances. We refer to these studies in Section **??**. By combining

these unique imprecisions and lack of necessity of obtaining permission to read accelerometer readings, an adversary can calculate unique accelerometer fingerprints of a mobile device.

1.2 Related Work

The power of extracting time and frequency domain features in accelerometers has been shown by W. Dargie and M. Denko [**?**]. Their study includes random placement of accelerometers on moving humans and cars, and investigating the behavior of accelerometers during similar movements. They conclude that the extracted frequency domain features remain generally more robust than time-domain frequency features. In our study, we applied accelerometer readings that were gathered from both resting and moving devices. The features that were analyzed in this paper differ from the features that we used. Yet, several time domain features such as mean, standard deviation, and highest and lowest value, as well as frequency domain features such as spectral centroid are common in both works.
S. Dey et al. [**?**] created fingerprints for mobile devices by extracting time and frequency domain features from the accelerometer readings of vibrating sensors and mobile devices. Their study illustrates the existence of such unique fingerprints by conducting a series of training and test set scenarios on 107 different stand-alone chips, smartphones, and tablets under laboratory conditions. This paper is the main basis of this thesis. In this thesis, the duration of the time window of the accelerometer readings amounts 10 seconds while the AccelPrint paper used accelerometer readings with a duration of 2 seconds. A shorter time window leads to a more stationary signal which also provides more accurate feature extraction results and also more fingerprints for the training and test phase. Other than the applied bagging tree classifier in AccelPrint, we also tested random forests, extra-trees, and gradient boosting to have a better overall overview on device and model recognition. Additionally, we could compare the behavior of different ensemble methods.
Bojinov et al. presented accelerometer-based fingerprinting by applying JavaScript code to read out the accelerometer readings [**?**]. They attempted to recognize smartphones by calculating two bias parameters from the z axis while the device was facing up or down on a resting surface. Their best and worst recognition rates were 8.3 and 15.1%.
Application of the machine learning library *scikit-learn* on recognizing human activities has been shown by H. He [**?**]. This study used the accelerometer and gyroscope readings of 6 different human activities. The number of features that were extracted in this study were 561, which is much more than our 17 extracted features. In a PCA (principal component analysis) dimension reduction, He reduced the number of features to train the applied classifiers to 50 and 20. In the later number of features which is closer to the number of features used in our study, the achieved results in the random forests classifier are similar to our result (cf. Chapter **??**). The other

applied classification methods were different from our methods. We preferred to apply ensemble methods as classification methods and also test clustering methods for device and model recognition purposes.

1.3 Contribution

Our contribution can be summarized as follows:

1. We implemented the signal feature extraction process utilizing the *NumPy* and *SciPy* libraries and, where necessary, developed our own implementation of the formulas. This process was especially essential regarding the spectral feature extraction.
2. We show that besides recognition of unique mobile devices, it is also possible to recognize the device model. For this purpose, we conducted two similar tests where the same minimum number of fingerprints was applied to have a precise comparison between the results.
3. The comparison of 4 ensemble classification methods of the *scikit-learn* library to recognize unique mobile devices and device models is also part of this thesis. We compare the results of the classification methods and discuss their efficiency on data sets that are either free of noise or contain fingerprints created from faulty accelerometer readings.
4. To the best of our knowledge, this is the first work where clustering is attempted to recognize mobile devices or the model of a device through their accelerometer fingerprints. We conducted tests with two different clustering methods of *scikit-learn* to find out if clustering is also applicable to group accelerometer fingerprints correctly.

1.4 Organization of this Thesis

The rest of this thesis is organized as follows. Chapter **??** discusses the accelerometer briefly and introduces the time and frequency domain features and libraries that were used in this thesis. In Chapter **??**, we provide the implementation details and explain the tested machine learning methods. Chapter **??** covers the data analysis of the data set and the results of applying machine learning on the accelerometer fingerprints. Chapter **??** presents the conclusion and possible future work.

2 Background

3 Approach

4 Implementation

5 Evaluation

6 Conclusion

A Acronyms

IMEI international mobile station equipment identity

MAC media access control

UDID unique device identifier

List of Figures

List of Tables

List of Algorithms

List of Listings

Declaration

I hereby declare that, to the best of my knowledge and belief, this Bachelorthesis titled "The Full Title of your Thesis" is my own work. I confirm that each significant contribution to and quotation in this thesis that originates from the work or works of others is indicated by proper use of citation and references.

Date	Signature

Consent Form

for the use of plagiarism detection software to check my thesis

Full Name: Your Name
Student Number: 000000
Course of Study: Information Systems
Title of Thesis: The Full Title of your Thesis

What is plagiarism? Plagiarism is defined as submitting someone else's work or ideas as your own without a complete indication of the source. It is hereby irrelevant whether the work of others is copied word by word without acknowledgment of the source, text structures (e.g. line of argumentation or outline) are borrowed or texts are translated from a foreign language.

Use of plagiarism detection software The examination office uses plagiarism software to check each submitted bachelor and master thesis for plagiarism. For that purpose the thesis is electronically forwarded to a software service provider where the software checks for potential matches between the submitted work and work from other sources. For future comparisons with other theses, your thesis will be permanently stored in a database. Only the School of Business and Economics of the University of Münster is allowed to access your stored thesis. The student agrees that his or her thesis may be stored and reproduced only for the purpose of plagiarism assessment. The first examiner of the thesis will be advised on the outcome of the plagiarism assessment.

Sanctions Each case of plagiarism constitutes an attempt to deceive in terms of the examination regulations and will lead to the thesis being graded as "failed". This will be communicated to the examination office where your case will be documented. In the event of a serious case of deception the examinee can be generally excluded from any further examination. This can lead to the exmatriculation of the student. Even after completion of the examination procedure and graduation from university, plagiarism can result in a withdrawal of the awarded academic degree.

I confirm that I have read and understood the information in this document. I agree to the outlined procedure for plagiarism assessment and potential sanctioning.

Date	Signature