Poster: Fingerprinting Smartphones Through Speaker

Anupam Das
PhD Student
University of Illinois at Urbana-Champaign
Email: das17@illinois.edu

Nikita Borisov
Associate Professor
University of Illinois at Urbana-Champaign
Email: nikita@illinois.edu

***Abstract*—The widespread use of smart devices gives rise to privacy concerns. Fingerprinting smart devices can jeopardize user privacy by enabling remote identification without user awareness. We propose a novel fingerprinting approach that uses the speakers of smart phones to uniquely identify an individual device. During fabrication, subtle imperfections arise in device speakers which induce anomalies in produced sounds. We exploit this observation to fingerprint smart devices through playback and recording of audio samples. We use audio-metric tools to analyze and explore different acoustic features and analyze their ability to successfully fingerprint smart devices. Our experiments show that not only is it possible to fingerprint devices manufactured by different vendors but also devices that have the same vendor and model; we were able to accurately distinguish over 94% of all recorded audio clips from 15 different units of the same model.**

I. Introduction

Mobile devices, including smartphones, PDAs, and tablets, are quickly becoming widespread in modern society. In 2012 a total of 1.94 billion mobile devices were shipped, of which 75% were smart and highly-featured phones [1], [2]. Canalys predicted that the mobile device market will reach 2.6 billion units by 2016, with smartphones and tablets continuing to dominate shipments [1]. The rapid uptake of intelligent mobile devices is not surprising due to the numerous advantages they provide consumers, from entertainment and social applications to business and advanced computing capabilities. However, smartphones, with all their interactive, location-centric, and connectivity-based features impose threatening concerns on user privacy and information security.

In this work we propose a novel technique for fingerprinting the *hardware* of smartphones. The observation is that even if the software on mobile devices is strengthened, hardware-level idiosyncrasies in speaker can be used to fingerprint physical devices. During manufacturing, imperfections are introduced in the analog circuitry of speakers, and as such, two speakers are never alike. Through an observational study, we find that these imperfections are substantial enough, and prevalent enough, that we can reliably distinguish between devices by passively observing audio, and conducting a simple spectral analysis on the recorded audio. Our approach can substantially simplify the ability for an adversary to track and identify people in public locations, for example, an adversary can use the short ringtones produces by mobile device speakers to reliably track users in public environments.

Our approach centers around the development of a novel fingerprinting mechanism, which aims to "pull out" imperfections in device circuitry. Our mechanism has two parts: a method to extract auditory fingerprints and a method to efficiently search for matching fingerprints from a database. To generate fingerprints of speakers we record audio clips played from smartphones on an external device (i.e., laptop/PC). To match fingerprints we use two different classifiers. We also test our fingerprinting approach for different genre of audio clips. Moreover, we study various audio features that can be used to accurately fingerprint smartphones.

II. Methodology

The key insight behind our work is that imperfections in smart device hardware like speaker induce unique signatures on transmitted audio, and these unique signatures, if identified, can be used to fingerprint the device. Our approach consists of two main components. The first task is acquiring a set of audio samples for analysis in the first place. To do this, we have a *listener* module, responsible for receiving and recording device audio. We implement the listener module as a stand alone application recording audio signals (e.g., the adversary has a microphone in a public setting to pick up device ringtones). The next step is to effectively identify device signatures from the received audio stream. To do this, we have an *analyzer* module, which leverages signal processing techniques to localize spectral anomalies, and constructs a 'fingerprint' of the auditory characteristics of the device. For fingerprinting speakers we record audio clips played from smartphones onto a laptop and we then extract acoustic features from the recorded audio excerpts as shown in Figure 1. We experiment with smartphones produced by both different and same manufacturer.

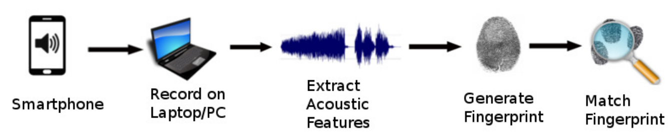

Fig. 1: Fingerprinting speakers embedded in smart devices.

III. Evaluation

A. Experimental Setup

Our experimental environment consisted of a 266 square foot (14'x19') office room with ambient background noise produced by hallway footsteps, air conditioning, desktop computers, and florescent lighting. To emulate an attacker, we placed an ACER Aspire 5745 laptop in the room and used the laptop's built-in microphone to collect audio samples.[1]

Devices and Tools: We test our device fingerprinting approach on devices from five different manufacturers namely – Apple (iPhone5), Google (Nexus 4G), Samsung (Galaxy Note 2), Motorola (Droid A855) and Sony Ericsson (W518). We also

[1] An attacker with a higher-quality microphone may attain better accuracy

investigate three different genres of audio excerpts as listed in Table I. Duration of the audio clips varies from 3 to 10 seconds. The sampling frequency of all audio excerpts is 44.1kHz. All audio clips are stored in WAV format using 16-bit pulse-code-modulation (PCM) technique. For analysis we leverage the following audio tools and analytic modules: *MIRtollbox* [3], *Netlab* [4] and *Audacity* [5].

TABLE I: Types of audio excerpts

Type	Description
Instrumental	Musical instruments playing together, e.g., ringtone
Human speech	Small segments of human speech
Song	Combination of human voice & instrumental sound

Algorithms and Evaluation Metrics: We use two alternate classification algorithms: *k-nearest neighbors* (associates an incoming data point with the device corresponding to the nearest "learned" data points), and *Gaussian mixture models* (computes a probability distribution for each device, and determines the maximally-likely association). We use standard multi-class classification metrics like *precision*, *recall*, and *F1-score* [6] in our evaluation.

Acoustic Features: We extract *acoustic features* from an audio stream, and use these features to construct a *fingerprint* of the device. To gain an understanding of how a broad range of acoustic features are affected by device imperfections, we investigate the following five acoustics features: root-mean-square (RMS) value, spectral entropy, spectral spread, mel-frequency cepstral coefficient (MFCC) and chromagram. All of these features have been well studied and documented by researchers [7]. We adopt a well known machine learning strategy known as *sequential forward selection* (SFS) [8] to determine the dominating subset of acoustic features.

B. Fingerprinting Devices From Different Vendors

We first look at fingerprinting smartphones manufactured by five different vendors. We found fingerprinting smartphones manufactured by different vendors relatively easier compared to fingerprinting devices manufactured by the same vendor. The main reason behind this is that the sensitivity of the speaker volume of different smartphones are quite different, thus making it easy to discriminate them. Simple acoustic features like RMS value and spectral entropy are good enough to obtain good clusters of data points. Figure 2 shows a plot of spectral entropy vs. RMS value for 50 samples of an audio excerpt (10 samples per handset). We also test our fingerprinting approach using three different types of audio excerpt as listed in Table I. Each audio sample is recorded 10 times (50% used for training and 50% used for testing). Table II summarizes our findings (values are reported as percentages). From Table II we see that we can successfully identify (with 100% precision) which audio clip originated from which smartphone.

TABLE II: Fingerprinting different smartphones using speaker output

Audio Type	k-NN			GMM		
	SFS Features: RMS, Spectral entropy			SFS Features: RMS, Spectral entropy, MFCCs		
	$AvgPr$	$AvgRe$	$AvgF1$	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	100	100	100	100	100	100
Human speech	100	100	100	100	100	100
Song	100	100	100	100	100	100

C. Fingerprinting Devices of The Same Model

Next, we now look at fingerprinting smartphones manufactured by the same vendor. For these set of experiments we use 15 Motorola Droid A855 handsets. Table III highlights our findings. We again test our fingerprinting approach against three different forms of audio excerpt. We use sequential feature selection technique [8] to obtain the dominating subset of acoustic features. From Table III, we see that we can achieve an F1-score of over 94% in identifying which audio clip originated from which handset. Thus fingerprinting smartphones through speaker seems to be a viable option.

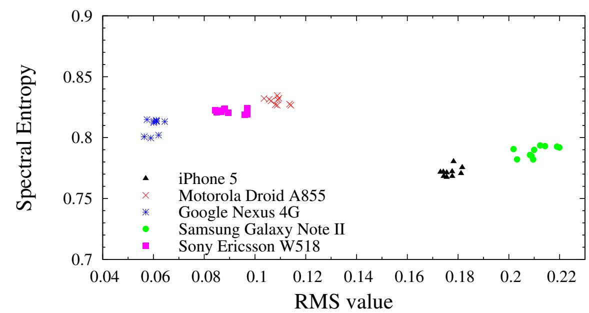

Fig. 2: Plotting audio samples taken from five different handsets using acoustic features — signal RMS value and spectral entropy.

TABLE III: Fingerprinting similar smartphones using speaker output

Audio Type	k-NN			GMM		
	SFS Features: MFCCs, Spectral spread			SFS Features: MFCCs, Chromagram		
	$AvgPr$	$AvgRe$	$AvgF1$	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	96.7	96	96.3	98.3	98	98.1
Human speech	98.9	98.7	98.8	98.9	98.7	98.8
Song	93.7	92	92.8	95.6	93.3	94.4

IV. Limitations and Future Work

Our approach has a few limitations that we plan to address in future. Firstly, we only explored five acoustic features. A rich set of acoustic features exist which we plan to investigate thoroughly in future. Secondly, we did not investigate the sensitivity of our fingerprinting approach against different environmental factors like—distance between audio source and recorder, and impact of different ambient background noise. Lastly, we plan to test our approach on a larger number of smart devices.

References

[1] Mobile device market to reach 2.6 billion units by 2016. http://www.canalys.com/newsroom/mobile-device-market-reach-26-billion-units-2016.

[2] Global mobile statistics 2013. http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a.

[3] MIRtoolbox. https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox.

[4] Netlab: Algorithms for Pattern Recognition. http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/book/.

[5] Audacity is free, open source, cross-platform software for recording and editing sounds. http://audacity.sourceforge.net/.

[6] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009.

[7] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A Review of Audio Fingerprinting," *J. VLSI Signal Process. Syst.*, vol. 41, no. 3, pp. 271–284, Nov 2005.

[8] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, no. 26, pp. 1157–1182, mar 2003.