Computer science department, Technion
Project in computer security (236349)
Spring 2018-2019

Project proposal: Acoustic Keylogger

Advisors: Danny Tylman, danny.tylman@citi.com, 050-9386265
Amir Schwartz, amir.schwartz@citi.com, 054-3055607

Background

Recording keystrokes made by a user is known as “keylogging” and is a well-known method of cyber-attack. Keyloggers pose a significant threat, since they are a common tool in the hands of cyber criminals trying to steal user credentials – which are then used to leak and manipulate sensitive data. A keylogger may be active in specific short time-intervals or on a continuous basis, monitoring and recording keystrokes made by the user and delivering them to the attacker.

In most cases, keyloggers are implemented by malware installed on the target computer. In some cases this is done unintentionally and unknowingly by the user himself while browsing the Internet or opening malicious files, in other cases it can be part of a larger and more sophisticated cyber campaign.

The purpose of this project is to explore the viability and effectiveness of a new type of keylogger. Instead of using a malware installed on the target computer itself to record keystrokes, an acoustic keylogger can be deployed on a separate dedicated device and monitor keystrokes on another keyboard based on the acoustic sound fingerprinting. Such a keylogger can be installed on a small “system-on-chip” device or a smartphone, thus eliminating the need to infiltrate the target computer itself and expanding the attack surface.

Scope

The project will consist of building a prototype of a working acoustic keylogger. This keylogger will analyze sounds of keystrokes and reconstruct the corresponding text that was typed. The keylogger should use a machine learning model trained using sound samples of keystrokes, and may improve its model while more and more keylogged data is received.

Some of the main research questions that should be addressed during the project:

- Which machine learning algorithm is best suitable for this task?
- What is the quality of the reconstructed text compared to the original keystrokes?
- Is the model generic or dependent on the type of keyboard?
- Is the model generic or dependent on the specific user?
- Are there any factors that affect the quality of the keylogger: physical distance of the recording from the keyboard, quality of the sound recording, etc?
- Can the keylogger learn during its run, without a specific configuration for each keyboard and user?

Delivery

The goal of the project is to deliver a working prototype of an acoustic keylogger and provide statistics on the quality of its results – i.e., reconstructing text based on analyzing the sounds of the corresponding keystrokes.

The minimal requirement is to deliver the keylogger as a standalone Linux or Windows-based application that can analyze a file containing the recording of keystrokes and reconstruct the corresponding text accordingly in offline mode. The application can be written in any computer language preferred by the students.

Delivering the keylogger as a real-time application that can monitor keystrokes as they are being typed and reconstruct the text on-the-fly, either on a system-on-chip device such as Raspberry Pi or as a smartphone application, would be a significant enhancement to the project.

High-level breakdown to phases/tasks:

1. Review the existing status of research in this field and previous attempts to build acoustic keyloggers.
2. Build an initial machine learning model and train it using sound samples.
3. Test the model to identify impacting factors and tune it accordingly. If needed, examine different types of machine learning algorithms to select the best one.
4. Design a standalone Windows or Linux-based application for implementing the model, analyzing sound files of keystrokes, and reconstructing the corresponding text in offline mode. Alternatively – design a real-time system-on-chip or smartphone application.
5. Build the keylogger application according to stage #4.
6. Run multiple tests on the keylogger to measure the quality of the results, provide statistics, and identify impacting factors.
7. Improve over as best as you can.

Prerequisites

- Students should have sufficient background in machine learning.
- Basic knowledge in audio files processing is preferred.

Remarks

- We have no specific preferences for the programming language you use.