A CNN-based keylogger using acceleration spectrograms

Cassiano S. N. C. Bueno[1], Bruno E. S. Oliveira[2], Josué Labaki[2]

[1]*Brazilian Synchrotron Light Laboratory - LNLS*
10,000 Giuseppe Máximo Scolfaro St, 13083-100, Campinas SP Brazil
cassiano.bueno@lnls.br
[2]*School of Mechanical Engineering, University of Campinas*
200 Mendeleyev St, 13083-970, Campinas SP Brazil
b165215@dac.unicamp.br, labaki@unicamp.br

Abstract. This work presents a model of supervised Convolutional Neural Network to identify keys pressed in a computer keyboard based solely on their acceleration signal. In this problem, vibration propagated through a desk from keystrokes were captured by an accelerometer. Individual keys have different acceleration signatures, and the network was used to tell keys apart by looking at their particular acceleration spectrograms. The proposed network showed an accuracy of up to 92% in the identification of individual keys.

Keywords: Deep learning, Signal processing, Convolutional neural networks, Keylogger, Keyboard attacks

1 Introduction

The use of machine and deep learning paradigms is gaining momentum in a variety of fields of engineering. Exceptional results from the application of these paradigms in areas such as signal processing, especially in problems with poor signal-to-noise ratios, are helping to establish these paradigms as staples in engineering analysis [1, 2].

One of the most common tasks executed by machine and deep learning algorithms is the classification of images of temporal data. In this work, we explored the application of these algorithms to create a remote-sensing keylogger [3]. In this problem, vibration propagated through a desk from keys struck in a nearby keyboard were captured by an accelerometer. Individual keys have different acceleration signatures – although remarkably noisy – and the idea was to use a deep learning scheme to tell keystrokes apart by looking at their particular acceleration spectrograms.

This work was inspired by different authors in the literature, who have worked with signals captured by cellphone and smartwatch acceleration sensors to identify and classify keystrokes. Based on the observation that each key makes a different sound when pressed, Berger et al. [4] have successfully used signal processing and machine learning to identify single words of 7-13 characters from acoustic emanations from typing. Yet another approach to remote keyloggers takes advantage of perturbations caused by typing in the surroundings of the keyboard. Marquardt et al. [5] have used signals read by the accelerometer of a smartphone placed near a computer keyboard to infer the contents of the typing. The idea is that each keystroke sends out vibratory waves that reach the smartphone through the desk and can be measured by its built-in accelerometers. Their technique to identify words from vibrations involves two steps. In the first step, the word is broken down into keystroke pairs, which are classified according to the distance each key within the pair is from each other, and to the side of the keyboard in which they are. In the second step, the authors incorporated semantics into the guess. Their scheme resulted in recognition accuracy comparable to that of Berger et al. [4], despite the drastically reduced sampling rate of the source device. The work of Marquardt et al. [5] has an important limitation, in that it requires a dictionary with which to compare stroke pairs, and because it relies on semantics, it is incapable of cracking passwords. Passwords are single-word strings of letters and numbers with no semantic meaning, the possible combinations of which are too abundant to be listed in a dictionary. A password-cracking remote keylogger must be able to identify the word accurately and solely from information contained within the input signal itself.

1.1 Problem statement

This paper presents a model of supervised Convolutional Neural Network (CNN) to classify the images of acceleration spectrograms of individual keys pressed in a computer keyboard. Acceleration signals were collected from keys using a fully configurable signal acquisition set up, and fed to the network in terms of three dimensional tensors of pixels. In order to investigate the performance of the network in dealing with keys of different levels of similarity between their acceleration signatures, two groups of keys were considered, which differ in the distance between the keys in the group.

2 Experimental setup

The experimental setup consists of an ordinary keyboard resting on a plain desk, near which an accelerometer was attached (Fig. 1). Other computer peripherals and cables were removed from the desk, so that they would not interfere with the measurements. The accelerometer is connected to an analog-to-digital converter, connected in turn to a computer for data collection and processing.

The keyboard is a model KM636 wireless Dell keyboard, which measured 441x128x29mm and weighed 549g. The keyboard rested on the desk through four rubbery feet, located near each of its corners. The accelerometer was a Kistler 8762A5 triaxial accelerometer, which was chosen due to its low noise-to-signal ratio, low mass, and to the fact that it can collect acceleration signals in three directions. The converter was a National Instruments USB-4431 model, which was chosen due to its relatively low internal noise, good discretization of measurements, and easy communication with available software. The specifications of the accelerometer and converter are given in Table 1.

Figure 1. Experimental setup: keyboard and accelerometer

Unfiltered signals were acquired at a 10 kHz rate for 2 minutes. Due to the converter's acquisition rate of 102.4 kHz, aliasing problems are only concerning above that frequency, which can be disregarded for this mechanical problem. Within the 2 minutes of acquisition, the keys were repeatedly pressed, one key per measurement time frame. An elapsed time of 2 s between key presses was defined to allow the vibratory signal from the key to decay to near background measurement noise levels. This resulted in about 60 signals collected from each key within the acquisition time frame. Figure 2 shows an example of acceleration signal, measured from key [Z]. Regularly-spaced peaks at approximately 2 s intervals and subsequent attenuation are clearly seen in these results. These results also show that the acceleration of the desk is much more pronounced in the y-direction (normal to the surface of the desk), than in the x- and z-directions, as expected.

Figure 3a shows a spectrogram obtained from the three first impacts (10 s) of key [Z] within the 1 to 100 Hz interval, within 100 ms windowing and superposition of 90%. This shows that some frequencies are excited soon after impact, and decay until the next impact is observed.

In order to standardize these signals for neural network training, they were saved as 1 second-long vectors of data beginning at the impact. This 1 s window was chosen in order to guarantee that only one key press was

Table 1. Specifications of the Kistler triaxial 8762A5 accelerometer and National Instruments USB-4431 analog-digital converter

Accelerometer	
Amplitude	$\pm$5g
Sensitivity	960 mV/g
Frequency band	0.5 to 6000 Hz
Mass	23 g
Internal ressonance	30 kHz
Excitation current	$>$2 mA

Converter	
AD resolution	24 bits (1.2 μV)
Max rate	102.4 kHz
Min rate	1 kHz
Amplitude	$\pm$10 V
Noise at 20 kHz max	75 μV rms
Noise at 20 kHz typ	55 μV rms

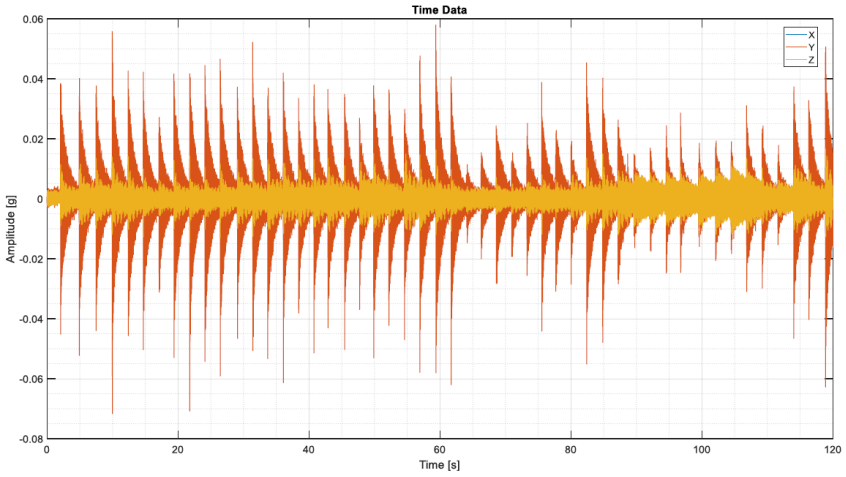

Figure 2. Example of acceleration signal from one full measurement of key [Z].

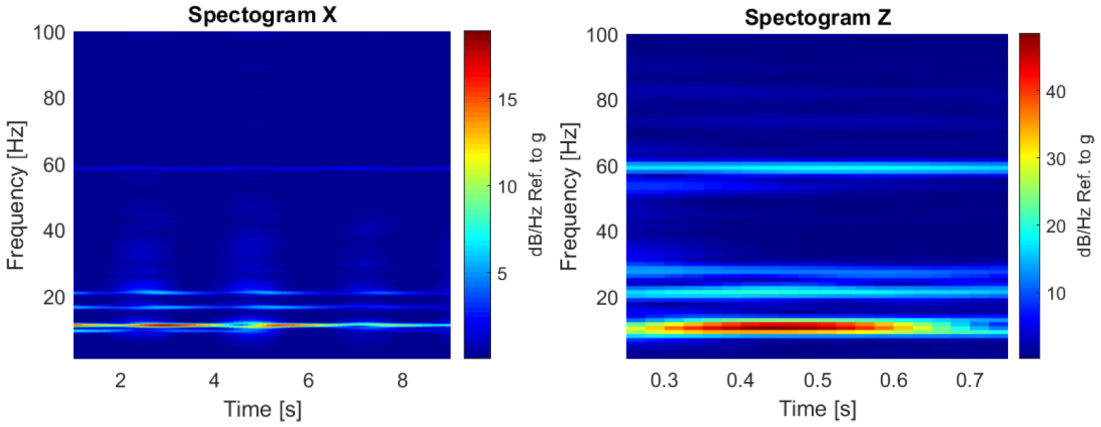

Figure 3. (a) Spectrogram of the 10 first seconds of key [Z] in the x-direction and (b) spectrogram of key [Z] in a single impact, in the z-direction.

present in each signal. Each impact resulted in an array of 60 lines by 10,000 columns per key, in which each line corresponds to a different measurement, and each column contains the discretized signal from impact to the 1 s mark. Figures 4 and 3b illustrate respectively the 1 second-long signal and spectrogram of 1 stroke of key [Z].

Two groups of four keys were considered in this experiment. The first group contained keys that were far from each other within the keyboard. This group contained the keys [1], [P], [:] and [Z], shown in blue in Fig. 5. The second group contained keys that were near each other. This group contained the keys [A], [S], [X] and [Z], shown in green in Fig. 5. The analysis was separated into these two groups because keys that are far from each other are expected to have more distinct acceleration signatures, therefore being more easily told apart by the neural network than keys that are close together in the keyboard. It is important to investigate the performance of

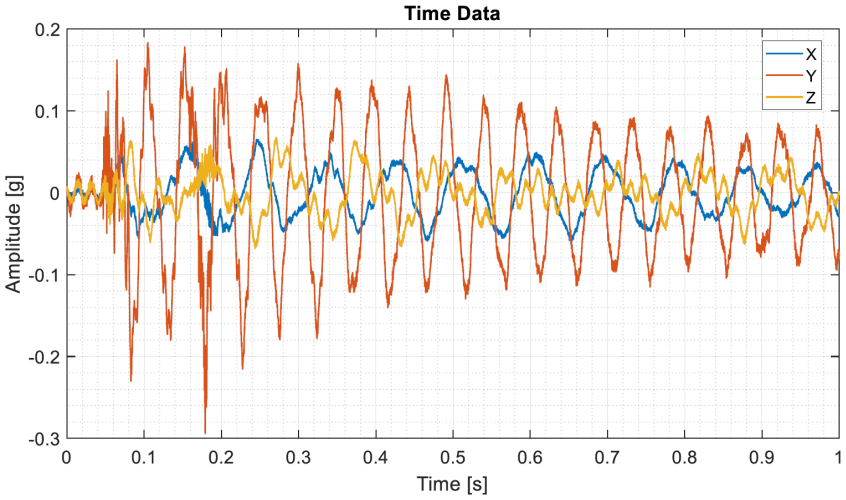

Figure 4. Example of acceleration signal from one press of key [Z].

Table 2. Accuracy of key identification in the two groups.

	Distant keys	Adjacent keys
Training	92%	81%
Validation	89%	74%

the network in dealing with each group.

Figure 5. Two groups of keys considered in this experiment. Blue keys are far from each other, and green keys are near each other. The key [Z] belongs to both groups.

3 Neural network setup

This work used a Convolutional Neural Network (CNN) of the *Keras* library, including image-specific fuctions *Conv2D* and *MaxPooling2D*, with two convolutional and one pooling layer, followed with three fully-connected additional layers. The intermediate layers were connected through ReLu neurons. The first layer had 15 5x5-pixel kernels, followed by a 2x2-pixel max pooling layer. The next layer had 30 5x5-pixel kernels, followed by a 2x2-pixel max pooling layer. The network ended with 4 neurons corresponding to the four keys to be identified in each group (blue or green). The CNN was fed with 50x50 pixel images, each of which comprising the spectrogram of individual key presses, resulting from the 1 s signal sample described in Section 2, windowed at 0.4 s and with 95% superposition. Tests involved 20 training epochs, with batch size 10. Training and validation were performed with 90% and 10% of the data set, respectively.

4 Results and discussion

Table 2 shows the accuracy in key identification obtained for each group of keys. The blue and green groups in Fig. 5 are referred to in Table 2 as ”distant” and ”adjacent” keys, respectively.

These results indicate that the current strategy is capable of identifying individual keys based on their acceleration signature with high accuracy, regardless of their position in the keyboard. As expected, the network performs better in the identification of keys that are far from each other, and this results from the acceleration signature of distant keys being more distinct from each other.

5 Conclusions

This work investigated the effectiveness of a CNN in identifying keys pressed in a computer keyboard from their acceleration signature. Spectrograms of the acceleration signals of selected keys were fed to the network in terms of three-dimensional tensors of pixels. The experiment considered a group in which the keys were far from each other, resulting in more distinct acceleration signatures from each key, and a group in which the keys were near each other, in which the similarity between the signals of each key could make it difficult for the network to tell them apart. The results showed that the network was capable of identifying the typed key with accuracy above 80% for the adjacent keys group, and above 90% for the distant keys group.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] D. Xie, L. Zhang, and L. Bai. Deep learning in visual computing and signal processing. *Applied Computational Intelligence and Soft Computing*, vol. 2017, 2017.

[2] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, and T. Sainath. Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, n. 2, pp. 206–219, 2019.

[3] R. Creutzburg. The strange world of keyloggers-an overview, part i. *Electronic Imaging*, vol. 2017, n. 6, pp. 139–148, 2017.

[4] Y. Berger, A. Wool, and A. Yeredor. Dictionary attacks using keyboard acoustic emanations. In *Proceedings of the 13th ACM conference on Computer and communications security*, pp. 245–254. ACM, 2006.

[5] P. Marquardt, A. Verma, H. Carter, and P. Traynor. (sp) iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers. In *Proceedings of the 18th ACM conference on Computer and communications security*, pp. 551–562. ACM, 2011.