Identification of Pressed Keys by Acoustic Transfer Function

Gerson de Souza Faria
Departamento de Engenharia de Sistemas Eletrônicos
Escola Politécnica
Universidade de São Paulo, Brazil
Email: gerson.faria@usp.br

Hae Yong Kim
Departamento de Engenharia de Sistemas Eletrônicos
Escola Politécnica
Universidade de São Paulo, Brazil
Email: hae@lps.usp.br

***Abstract*—The possibility of uncovering the sequence of pressed keys in a PIN-pad, ATM or other device with mechanical keyboard that processes sensitive information is a serious security threat. In this paper, we show that the Acoustic Transfer Function (ATF) of the device carries information that discloses the pressed key. To demonstrate this hypothesis, we glued two microphones in empty spaces of two different PIN-pads, collected the acoustic emission generated by the keystrokes and estimated the transfer functions of the system through the two acquired audio signals. A simple classification scheme using the ATFs yielded 99.6% of recognition accuracy in one device and 46% in the other. The results show that there are devices that can be easily attacked using ATF and those that are more difficult to attack using this technique. We analyzed again the data where we obtained 99.6% of accuracy, this time using the traditional frequency spectrum-based technique, obtaining only 52% of recognition rate. The results also show that the certification process has failed in identifying this vulnerability.**

Keywords*—*information security, side-channel attack, acoustic emission, transfer function, smart card skimming, EMV, PCI, Common Criteria, ATM, PIN-pad.

I. Introduction

Mechanical keypads are widely used for entering sensitive data. Passwords are typed in mechanical keypads in ATMs (Automatic Teller Machines) or PIN-pads (devices used in card transactions to input the cardholder's Personal Identification Number, also known as PIN entry devices or PEDs). In some countries, electors use electronic voting machines with mechanical keypads to choose the candidate. Thus, the possibility that someone finds out the sequence of pressed keys, without the user's knowledge or consent, is a serious security threat.

In card operations, the theft of card information in an otherwise legitimate transaction, known as "skimming", was responsible for 87% of attacks against ATMs in 2013, as reported in [1]. Card issuers estimate that U.S. merchants will acquire 575 million new chip-enabled payment cards and terminals by the end of 2015[1]. All of these new terminals are required to be certified by laboratories following PCI-PTS-POI[2] and/or Common Criteria[3] evaluation standards.

In this work, we use Acoustic Transfer Function (ATF) to identify the pressed keys of PIN-pads. Let us call our new attack ATF attack. In a PIN-pad certified by both PCI and Common Criteria, we could identify the pressed keys with $99.6 \pm 0.8\%$ of accuracy, where one person pressed the keys for the training and another person pressed the keys for the test, emulating a realistic attack scenario. Analyzing the same data with the traditional frequency spectrum-based technique, we obtained only 52% of recognition rate. However, in another PIN-pad (also certified by PCI and Common Criteria), we obtained only 48% of accuracy using ATF technique. The results show that there are devices that can be easily attacked using ATF and those that are more difficult to attack using this technique. We make some considerations for manufacturers to construct ATF attack-resistant devices.

Usually, modern ATM keypads are encrypted. They are sealed modules that encrypt the PIN soon after the entry. So, non-encrypted PIN numbers are not meant to be accessible from outside either by physically tapping onto wires or remotely sensing electromagnetic radiation. Any tampering of the keypad causes it to permanently disable itself. Similarly, PIN-pads are protected modules that permanently disable themselves if tampered. However, Drimer *et al.* [2] show cases where this mechanism fails. The possibility of identifying the sequence of pressed keys through acoustic emissions analyzed using ATFs is a serious security failure of secure keypads because they are designed to resist against any attempt of eavesdropping. The devices will continue functioning normally while confidential data are stolen. In the literature, there are some papers that identify the pressed key by sound, because each key usually emits a characteristic sound when pressed. Asonov and Agrawal [3] achieved 79% of key recognition success rate when identifying one out of 30 keys in a PC keyboard. Berger *et al.* [4] use keyboard acoustic emanations and a dictionary to recognize correctly 73% of the English words typed in a PC keyboard, without any training. Zhuang

[1] http://newsroom.mastercard.com/press-releases/nearly-half-u-s-merchant-terminals-accept-chip-2015/#prclt-sSxZ1V62

[2] PCI-PTS-POI stands for PIN Transaction Security - Point of Interaction, a set of requirements specific for PIN entry devices, proposed by the Payment Card Industry - PCI. Device compliance can be consulted at https://www.pcisecuritystandards.org/approved_companies_providers/approved_pin_transaction_security.php

[3] The Common Criteria for Information Technology Security Evaluation is an international standard for computer security certification. http://www.commoncriteriaportal.org

978-1-4799-8697-2/15/$31.00

et al. [5] take as input 10-minute sound recording of a user typing English text using a keyboard and recovers up to 96% of typed characters. Halevi [6] uses keyboard acoustic emanations for eavesdropping over random passwords, without using dictionary, achieving 40% to 64% recognition rate per character.

Similarly to acoustic emission, each key seems to emit a characteristic mechanical vibration when pressed. However, this idea has been much less explored in the literature. Marquardt and Verma [7] use this idea to recognize keystrokes of a computer keyboard. They use the accelerometer of a smartphone placed near the computer's keyboard to capture the vibrations. They do not actually identify the pressed key. Instead, they classify keystrokes in "left" or "right" and pairs of keystrokes in "near" and "far". They achieved classification rates from 65% to 91% making those binary decisions.

Faria and Kim [8] show that it is possible to identify the pressed key by gluing accelerometers in the device, acquiring mechanical vibration signals generated by keystrokes and analyzing them. They called it "vibration attack". They achieved 98.4% of accuracy in identifying keystrokes in ATM keypad, 76.7% in PIN-pad resting on a hard surface and 82.1% in PIN-pad hold in hand.

The approach used in this work is of a different nature: even if it were possible to have all the keys emit exactly the same sound and vibration, it would be still possible to identify the pressed key by estimating the acoustic transfer function of the system. The acoustic transfer function depends on device's empty space, the audio source location (the pressed key) and the location of microphones.

Researchers in acoustics engineering [9, vol. 2, p. 1381] use transfer function to estimate spatial frequency responses of rooms, reverberation times and other characteristics. In our case, the "room" is basically the SAM (Secure Access Module) card access compartment located just below the keypad, where the sound waves interact.

In theory, since the transfer function is a characteristic of the system and not of the excitation source, the ATF attack will be user-independent – no matter who presses the key, it will be correctly identified.

The rest of the paper is organized as follows. We present the basic theory on acoustic model in Section II. We present our preliminary experiment to test our hypothesis in (Section III). We tested ATF attack in two commercial PIN-pads designed to be secure and present the results in Section IV and Section V. We make some considerations on the results and discuss the certification processes in Section VI. We present our conclusions in Section VII. Appendix presents the implementation details: audio segmentation, feature extraction and classification methods.

II. Acoustic Model

A. Theory

Let us denote the signal of a microphone measuring air pressure at time t and position $\mathbf{r}$ as $y(t, \mathbf{r})$. Given some excitation $x(\tau, \boldsymbol{\rho})$ (also depending on time and position) and considering the system as linear, time invariant and spatially homogeneous, the response signal y can be described through a generalization of convolution [9, vol. 1, p. 14]:

$$y(t, \mathbf{r}) = \iint h(t - \tau, \mathbf{r} - \boldsymbol{\rho}) x(\tau, \boldsymbol{\rho}) \, \mathrm{d}\boldsymbol{\rho} \mathrm{d}\tau. \tag{1}$$

where $h(\cdot)$ is the impulse response of the entire system. The impulse response describes the linear transmission properties of any system able to transport or transform energy in a certain frequency range ([9, vol. 1, p. 65]). If the system is nonlinear, the equation describing the response signal given an excitation is not of the form of Eq. 1.

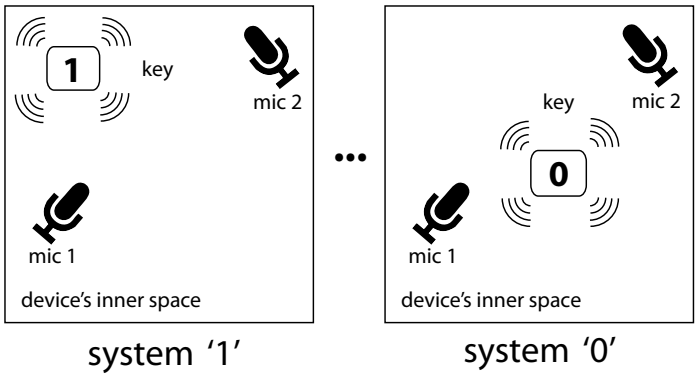

Fig. 1. The ensemble formed by the device's inner space, the position of the key and of the two microphones compose a system whose transfer function must be estimated. The keys are classified based on the transfer functions.

Since the impulse response is dependent on the spatial position of both the source and the measure locations, if we maintain the latter (microphones) fixed in the space, the impulse response can be used as a signature of the position of the audio source, providing that it can be estimated with a reasonable accuracy. There are several methods for estimating the impulse response of an acoustic system. All of them use as excitation sources deterministic signals or noise, in order to cover some desired range of the frequency spectrum. In our case, the excitation sources are the audio provided by distinct keys being pressed by distinct persons. Since the transfer function is a characteristic of the system and not of the excitation source, the attack will be user independent, i.e., no matter who presses the key, it will be, in theory, successfully classified.

Fig. 1 depicts the attack model proposed. The positions of keys define each system, since other factors are fixed (inner space of the device and the positions of microphones put inside it). Therefore, we have one system for each key, described by the transfer function of the system. Each transfer function is estimated from the signals captured simultaneously by the microphones, one taken as system's "input" and other as system's "output".

B. Transfer Function Estimation

For a given spatial configuration, Eq. 1 can be expressed through convolution notation:

$$y(t) = h(t) * x(t). \tag{2}$$

This relation has an equivalent representation in the frequency domain, as the product of the Fourier transforms of system's impulse response and input signal:

$$Y(f) = H(f) \cdot X(f). \quad (3)$$

The term $H(f)$ is called the transfer function of the system. Discrete estimates of this function will be used as the features that will characterize our system. However, as we use only discrete positions (the positions of keys), we can rewrite the transfer functions expression as:

$$H_i(f) = H(f, \mathbf{r_i}). \quad (4)$$

As stated before, there are several methods for its estimation. Here we will use the classic nonparametric Welch method [10], that can be computed with the Matlab procedure `tfestimate`. Resulting vectors H_i are used as features, through the method described in the Appendix. This method is not mandatory and other transfer function estimation methods can be used instead.

III. Box Model Experiment

If different audio source locations generate distinct transfer functions, we can use these functions as "location signatures". So, in order to observe the behavior of distinct source positions in the transfer function, we made an experiment with a small acrylic box with three holes (Fig. 2), using an earphone as audio source but changing its location among the three holes. We stress that this experiment (and all the subsequent ones) does not use the true transfer functions between the earphone and the two microphones, because in an ATF attack the true sound emitted by the key is unknown. We estimate the transfer function from the signals captured simultaneously by the two microphones, one taken as system's "input" and other as system's "output".

The audio signal is captured by two condenser microphones placed inside the box. Two distinct excitation signals were used[4]: (i) ≈30s of white noise and (ii) ≈15s of sinusoidal sweep increasing from 20Hz to 20KHz. T he dimensions of the box are approximately 40mm×65mm×85mm. The microphones used for recording have a pre-amplifier operating with gain of 50×. The output of the pre-amplifiers are connected in the "line-in" input of a PC and recorded at 96KS/s with the program "Audacity".

Fig. 3 depicts the absolute values (magnitudes) of the transfer functions for the three holes. The transfer functions are remarkably distinct between the different holes, and at the same time very similar using white noise or sweep tone, since the transfer function is a characteristic of the system and not of the excitation signal.

IV. Ingenico iPP320 PIN-pad Experiment

Fig. 4 shows the first PIN-pad we tested, an Ingenico iPP320, and the assembly of the experiment, where two condenser microphones were glued inside the SAM (Secure Access Module) card access compartment. We stress that a

[4]We used an iPhone application named "The Signal Generator" to generate the tones: https://itunes.apple.com/us/app/signal-generator-audio-test/id543661843?mt=8

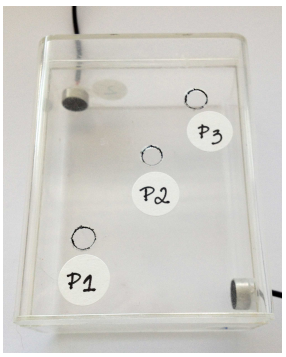

Fig. 2. An acrylic box was used to estimate three transfer functions, one for each position P1, P2, and P3. An earphone was used as excitation source playing tones through the holes, one hole at a time, keeping the others closed.

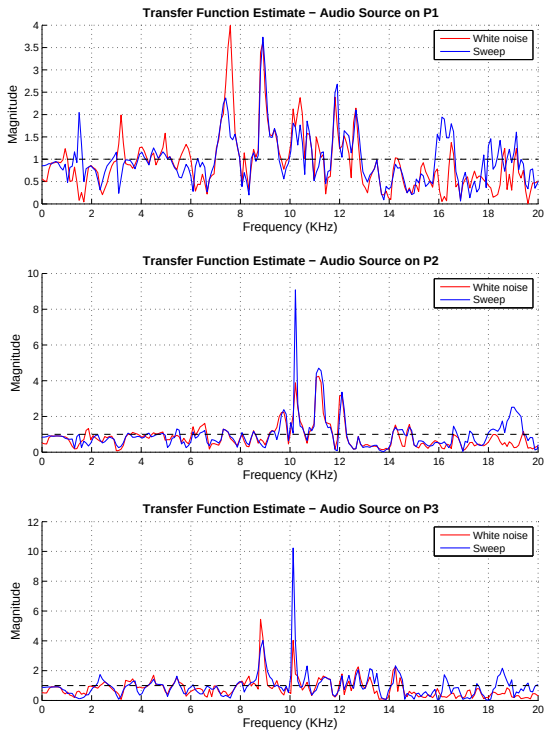

Fig. 3. The estimated transfer functions for the three holes of the box model. They are remarkably distinct between the different holes, but very similar using white noise or sweep tone as excitation source.

Fig. 4. (Left) Ingenico iPP320 PIN-pad used in the experiment. (Right) The bottom view showing the SAM compartment with the implanted microphones.

PIN-pad is a tamper proof device and it permanently disables itself if opened. However, the space in the SAM card compartment is designed to be freely accessible.

Ingenico iPP320 is a widely used model in retail stores. This device is Common Criteria Certified[5], PCI-PTS compliant, under 2.X and 3.X versions.

The choice of this model/brand was not guided by any vulnerability we could spot. We want to make it clear that most of PIN-pads with a SAM card compartment are potential targets of ATF attack and Ingenico iPP320 is not a special case. The SAM card compartment increases the vulnerability to ATF attack mainly because:

1) it provides room for implanting "hardware trojans" or "bugs", hidden within the compartment. ATF attack explores this space;
2) the compartment is normally located just below the keypad, the ideal place to capture the audio emanated from the keystrokes;
3) the SAM slots can eventually provide electrical power for the "bugs".

So, a hacker can put the microphones inside this compartment without disabling the device. A hacker can also easily put a wireless device to transmit the captured information. Thus, the attack can be executed in a undetectable way, without batteries and wires.

A. Data Acquisition for Training and Testing

To obtain the data sets for training and testing the system, one (training) person pressed 70 times each one of the "0" to "9" keys and both audio signals were recorded as a single stereo ".wav" file, with sampling rate of 96KS/s. Another (testing) person pressed the same keys 50 times. All pressings were done with PIN-pad's power off.

B. Features and Model Evaluation

Fig. 5 depicts the obtained features. Each color represents the "averaged" transfer function magnitude for the pressings of each user. They are very similar for the two users and distinct enough among keys, making them good features for the key identification.

Table I shows the confusion matrix. The recognition rate is extremely high (99.6±0.8%), where 0.8 is the standard deviation of the 5 cross validations. See Appendix for the methods used for features extraction and classification.

C. Comparison with the Frequency Approach

In order to compare the ATF approach with the traditional frequency spectrum classification[6], we used as features the estimated periodograms with normalization of total energy. We followed the same procedures used for the ATF feature dimensionality reduction and classification (See Appendix). This approach uses only one microphone. The average recognition rate using periodogram was (52±34%), a poor result.

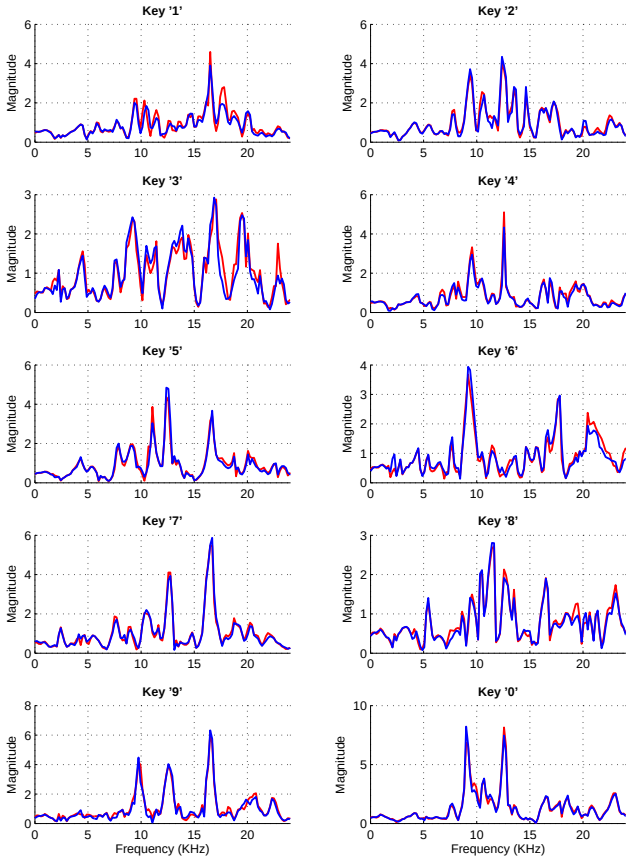

Fig. 5. The magnitudes of the transfer functions for each key obtained in Ingenico iPP320 experiment. Each color correspond to one user. The graphs show the averages for each key.

TABLE I. CONFUSION MATRIX OF THE INGENICO IPP-320 TRANSFER FUNCTION EXPERIMENT

Key	1	2	3	4	5	6	7	8	9	0	Acc.(%)
1	**244**	1									99.6
2		**245**			1	3				1	98.0
3			**230**								100.0
4				**250**							100.0
5					**205**						100.0
6						**235**					100.0
7							**240**				100.0
8								**245**			100.0
9		1			1	2	1		**245**		98.0
0										**250**	100.0

V. GERTEC PPC910 PIN-PAD EXPERIMENT

Fig. 6 shows the other tested device, a Gertec PPC910 and the assembly of the experiment, where the microphones were glued in the back of the bottom cover of the SAM card access compartment. This device is PCI-PTS compliant, under 2.X and EMV 2000 Levels 1 and 2 compliant[7].

[5]https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/Reporte08/0859a_pdf.pdf?__blob=publicationFile

[6]We did not reproduce faithfully Agrawal and Asonov's [3] experiment. We just used frequency spectrum as features, the approach these authors have introduced.

[7]http://www.gertec.com/produto.aspx/produtosdetalhe/57/PPC_910

Fig. 6. (Left) Gertec PPC910 PIN-pad used in the experiment. (Right) The bottom cover with the implanted microphones.

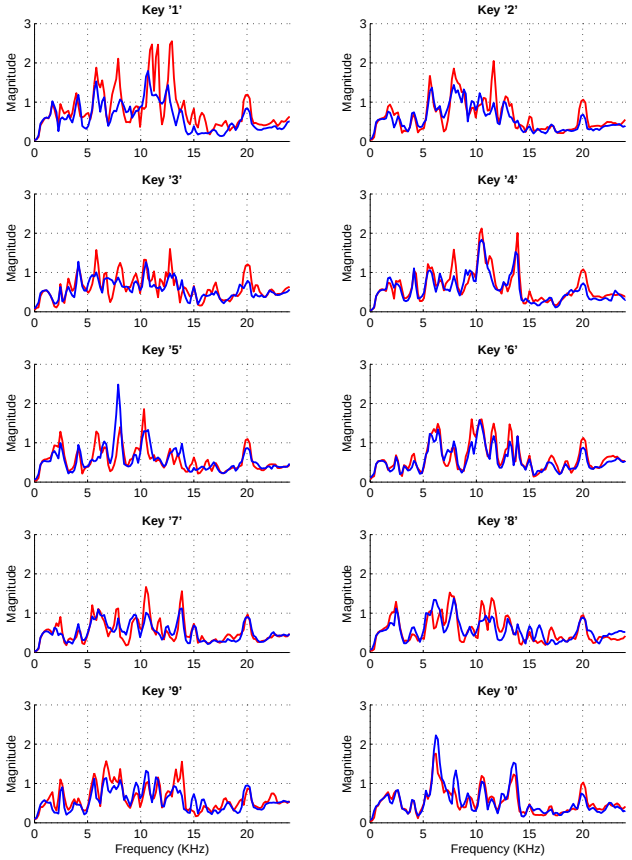

Fig. 7. The magnitudes of the transfer functions obtained in Gertec PPC910 experiment. Each color correspond to one user. The graph shows averages for each key. The transfer functions obtained from the keystrokes of the two users have poor agreement.

The signal acquisition process was the same as in the previous experiment, but not all the generated keystrokes were detected by our segmentation method because of the low acoustic energy emitted. Fig. 7 depicts the obtained features. As in the previous experiment, each color represents the "averaged" transfer function magnitude for the pressings of one user. Table II shows the confusion matrix. The average recognition rate is very poor (46±22%). The real classification rate would be even lower if we consider that not all the generated keystrokes were detected by the segmentation method.

TABLE II. CONFUSION MATRIX OF THE GERTEC PPC910 TRANSFER FUNCTION EXPERIMENT

Key	1	2	3	4	5	6	7	8	9	0	Acc.(%)
1	**20**	28	10	32	9	2	7		21	1	15.4
2	3	**119**		13	11	7	6	7	26	13	58.0
3		3	**41**	11	8	1	1	4	12	9	45.6
4	1	10		**128**	9	1	3		6	7	77.6
5	3	8		23	**56**	1	19	15	50	5	31.1
6	1	14	2	11	5	**103**	9	3	7	10	62.4
7		17		20	16	3	**51**	1	5	17	39.2
8	1	14		6	12	1	9	**17**	3	37	17.0
9		33	3	15	3	10	3	48	**64**	1	35.6
0		1			4	3		1	16	**70**	73.7

VI. CONSIDERATIONS

A. *Qualitative observations*

We demonstrated experimentally that Ingenico iPP320 is much more vulnerable to ATF attack than Gertec PPC910. Even so, some keys have considerable recognition rates in Gertec's, deserving further analysis. Some qualitative observations regarding the operation of both equipments follow:

- Ingenico's keys leak an audible and inevitable "click". If, by one hand, this click is useful as acoustical and mechanical feedback for the operator, by the other hand it also furnishes acoustic information. Gertec's can be pressed in a way that no click is heard, so reducing the energy of the audio signal. We think this is the main cause of the low average classification rates for Gertec's.
- Ingenico's keys have a much more regular operation than Gertec's. The force used to press a key in Gertec's is greater and the key movement is very unstable.
- Gertecs's empty space are much more "corrugated" (see Fig. 6 right) and structurally complex than Ingenico's. This property can introduce nonlinearities affecting negatively the ATF linear model.

Another fact that may influence the success of the attack is how much dependent are the keys on a common support structure. If the keys are more closely dependent on a common structure (e.g. a loosely assembled printed circuit board where the contacts lie), it will be more difficult to distinguish systems for each key, because that structure may "blur" the origin of the sound source.

B. *PCI (Payment Card Industry)*

PCI requires that an attack such as described in this work should be possible only with very high cost of 26 for identification and 13 for exploitation [11, p. 16]. Nevertheless, the ATF attack costs only 12.5 for identification and 3 for exploitation (Table III). The method used to calculate the costs can be found in [11, p. 142].

C. *Common Criteria*

Many documents used in the certification process warn against acoustic emission attacks. Ingenico's Security Target

TABLE III. CALCULATION OF THE COST OF PIN-PAD ATF ATTACK

Factor	Identification	Exploitation
Attack time	Beyond 160 hours = 5.5	$\leq$1 hour = 0
Expertise	Expert = 4	Layman = 0
Knowledge of the PIN entry device	Public = 0	Public = 0
Access to the PIN entry device	Mechanical Sample = 1	Mechanical Sample = 1
Equipment required for the attack	Standard = 1	Standard = 1
Specific parts required	Standard = 1	Standard = 1
Total cost	**12.5**	**3**

for iPP-320 states clearly which assets are being certified and PIN is one of them [12, sec. 4.1]. The same document [12, sec. 8.1.1.1] states that the device shall not emit sound, electromagnetic emissions, power consumption etc.

From the citations above, we can conclude that attacks based on acoustic emissions are explicitly covered by both PCI and Common Criteria requirements. And we can also conclude that the devices we attacked were evaluated and certified throughout these requirements. However, the certification process failed to identify the vulnerability explored in the ATF attack.

VII. CONCLUSION

In this paper, we have demonstrated that the sound emitted by the keystrokes of a PIN-pad can be used to identify the sequence of pressed keys by analyzing the audio captured at two different points inside the device's SAM compartment. We have shown that the acoustic properties of the empty space, characterized by its acoustic transfer function (ATF), can be used as "location signatures". These signatures are used in our "ATF attack" to identify the pressed keys. In this preliminar research, a simple classification scheme using these signatures yielded 99.6% of key recognition success rate in one device and 46% in another one. This shows that there are some devices completely vulnerable to ATF attack, while others are less vulnerable. We made some considerations regarding what makes a device vulnerable to ATF attack. Our findings indicate that the certification process was not able to detect that sensitive information leaks from the device.

VIII. ACKNOWLEDGEMENTS

We would like to thank Ross A. Anderson for guidance on certification processes, specially Common Criteria, and also for information on responsible disclosure.

APPENDIX

We made the segmentation by (i) finding the N highest peaks of $\sqrt{l^2(t) + r^2(t)}$, where $l(t)$ and $r(t)$ are respectively the audio signals of left and right channels; (ii) taking 4096 sample points, 80% after the peak, 20% before it; (iii) maintaining the segments with correlation coefficient between any pairs better or equal to 85%, in order to avoid spurious signals. The audio segmentation is not the main focus of this work and other techniques can be used as well.

In the procedure `tfestimate`[8], one can set the size of the FFT calculation in the power spectrum estimation. We set this value to 512 points for the PIN-pad experiments, generating a transfer function vector of 257 points. From these, the first 120 were chosen as features, representing a frequency range from 0 to 22.5KHz. In order to reduce the dimensionality, we used the classic PCA (Principal Component Analysis) approach and selected the 30 eigenvectors with the largest eigenvalues, corresponding to $\approx$97% of the sum of all eigenvalues. More elaborate criteria for this procedure can be found in [13].

We used a standard artificial neural network for the classification. We used the configuration with input layer of size 30, one hidden layer of size 50 and the output with size 10, encoding the 10 keys.

In order to avoid overfitting, the data of one person was used exclusively for training the network and the data of another person was used for the testing, emulating a realistic attack scenario where the victim's pressing behavior is unknown. We repeated training/testing process 5 times, each time using a different training set and tested using all available testing data.

[8] http://www.mathworks.com/help/signal/ref/tfestimate.html

REFERENCES

[1] Verizon 2014 Data Breach Investigations Report. [Online]. Available: http://www.verizonenterprise.com/DBIR/2014/

[2] S. Drimer, S. J. Murdoch and R. Anderson, "Thinking inside the box: system-level failures of tamper proofing," in *Proc. IEEE Symp. on Security and Privacy*, pp. 281-295, 2008.

[3] D. Asonov and R. Agrawal, "Keyboard acoustic emanations," in *Proc. IEEE Symp. Security and Privacy*, pp. 3, 2004.

[4] Y. Berger, A. Wool, and A. Yeredor, "Dictionary attacks using keyboard acoustic emanations," in *Proc. 13th ACM Conf. Computer and Communications Security*, pp. 245–254, 2006.

[5] L. Zhuang, F. Zhou, and J. D. Tygar, "Keyboard acoustic emanations revisited," *ACM T. Information and System Security*, vol. 13, no. 1, pp. 3:1–3:26, Oct. 2009.

[6] T. Halevi and N. Saxena, "A Closer Look at Keyboard Acoustic Emanations: Random Passwords, Typing Styles and Decoding Techniques." [Online]. Available: http://eprint.iacr.org/2010/605.pdf, 2010.

[7] P. Marquardt, A. Verma, H. Carter, and P. Traynor, "(sp)iPhone: decoding vibrations from nearby keyboards using mobile phone accelerometers," in *Proc. 18th ACM Conf. on Computer and Communications Security*, 2011.

[8] G. S. Faria, H. Y. Kim, "Identification of Pressed Keys From Mechanical Vibrations," *IEEE T. Information Forensics and Security*, vol. 8, no. 7, pp. 1221-1229, July 2013.

[9] D. Havelock, S. Kuwano, M. Vorländer *Handbook of Signal Processing in Acoustics*, Springer, 2008.

[10] S. M Kay *Modern Spectral Estimation - Theory & Application*, Prentice Hall, chap. 4, 1988.

[11] Payment Card Industry - Security Standards Council LLC, *PIN Transaction Security (PTS) Point of Interaction (POI) Modular Derived Test Requirements 4.0*, Jun 2013.

[12] Ingenico Group, "IPP3xx Security Target Lite - ICO-OPE-00719-V1-EN." [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/Reporte08/0859b_pdf.pdf?__blob=publicationFile

[13] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE T. Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 387-392, Apr. 1985.