Keyboard Acoustic Emanations Attack

An Empirical Study

SRAVANTHI PONNAM

KTH Information and
Communication Technology

Master of Science Thesis
Stockholm, Sweden 2013

TRITA-ICT-EX-2013:267

Keyboard Acoustic Emanations Attack – An Empirical Study

Sravanthi Ponnam

Department of Computer and Systems Sciences

Degree project 30 HE credits
Degree subject (Computer and Systems Sciences)
Degree project at the master level
Spring term 2011
Supervisor: Iskra Popova
Reviewer: Louise Yngstrom
Swedish title: Akustik strålning från tangentbord – en empirisk studie

Akustisk strålning från tangentbord – en empirisk studie

Sravanthi Ponam

Abstract

The sounds produced from the keystrokes when a user types on the keyboard are called keyboard acoustic emanations. These sounds can be recorded with a microphone and stored as a file on the computer. Different techniques can be used to retrieve each keystroke. In this way sensitive information, such as passwords used to unlock the system or enter various protected cyber spaces can be collected and misused. This study investigates the seriousness of the keyboard acoustic emanations attack and possible threats from this type of eavesdropping. The aim of the research is to show this type of attack can be performed using simple equipment and easy to use signal processing techniques and to suggest protective measures against the threat from the attack. We use empirical methodology and perform experiments under different scenarios. Unlike the previous research, the experiments are performed in a moderately noisy environment. Our attack includes two phases, training and recognition phase. The structure of the attack is created considering views of previous research and having in mind the aim of the study. Six scenarios are created based on how the characteristics of the waveforms are presented and what types of techniques are used at the recognition phase. A separate procedure for identifying which scenario produces the highest recognition rate is designed. The results show that the waveform of the acoustic signal in presence of noise has similar shape as in silent environment and that an attacker can easily perform our experiment with keyboard acoustic emanations attack. We achieved 60% recognition rate that can be considered as satisfactory. The experiment is compared with similar ones from the previous research. Easy computation, analysis and simplicity are the advantages of our approach. At the end of the thesis we suggest preventive measures for mitigating the attack.

Keywords

Keyboard acoustic emanations, Side channel attacks, Key logging

Table of Contents

List of Figures

List of Tables

1 Introduction

This chapter presents the importance of the sounds the keyboards produce for retrieving passwords typed to enter a protected cyber area, unlock a computer or get cash from ATM (Automatic Teller Machine). The threat for these passwords being revealed by an attacker equipped with a simple microphone and a laptop is defined as acoustic keyboard emanations attack. The first section of the chapter clarifies this type of attack. It is followed by the description of the problem addressed with this research, the aim, objectives, and research questions. At the end of the chapter the limitations and the target audience are portrayed.

1.1 Background

Computer security is considered to cover a wide area of Computer Science related to confidentiality, integrity and availability of information in digital format. Gross and Rosson (Gross & Rosson, 2007) use "End-User Security Management (EUSM)" to refer to the information security from end-users' perspective. Their research showed that one of the aspects users are usually concerned about is keeping their passwords private. This implies that users are aware of the importance of their passwords. If this confidential information is exposed to an attacker, he/she can get full access to the system which usually results in financial loss and other damages.

Nowadays the most common way of accessing computer systems is still through entering the combination of a user name associated with the user identity, and a password. Combination of both is used to authenticate the user. According to Bishop (Bishop, 2003) the simplest attack against password-based systems is to guess the password by repeated trial and error. This is known as a dictionary attack. The speed at which the password is cracked depends on the length and the elements in the password. Besides this, there are other ways to get this confidential information, especially for systems connected to the Internet. For example, vulnerabilities of the system can be used to install malicious code that records user's keystrokes. Key logger is the name of the malware that monitors users' keystrokes or other activities and is undetectable by most antivirus software (Sagiroglu & Canbek, 2009). It can record the password on the user's machine and send it to the attacker via the network. In the early days of Internet, before encryption was introduced for keeping the confidentiality of the information conveyed via the network, passwords were transmitted as plain text. Hence attackers equipped with software for packet capturing could discover the sensitive information by simple wiretapping (Gait, 1978).

Another, rather new and not that well known way of getting unauthorized access to a user's password, is by recording the sounds from the keyboard when a user is typing. Both, the desktop and laptop computer keyboards make unique sounds when a particular key is pressed and released. Touch tone telephones and ATM keypads also produce sounds unnoticeable to the human ear. The user usually does not care about these sounds, and is only trying to hide his/her credentials from being overlooked by other people. An attacker who is near the user can record the keystroke sounds using an ordinary microphone and a laptop. The recorded sounds can be used to discover the characters typed. This type of attack is known as keyboard acoustic emanations attack. Emanations are the wave signals

Define what is Keyboard acoustic emanation attack, differences etc.

corresponding to the sounds generated when the keys are pressed and released. The advantage for the attacker is that any security measure, such as encrypted passwords, is not a hindrance for the attacker. Users are often not aware of this type of attack and do not use any tools to defend against it. If the attacker discovers the address and the password to a secure site on the Internet, she/he does not even need to have access to the user's computer again. The secure site can be accessed from any computer.

This attack is more dangerous compared to other password attacks. This is primarily because it is inexpensive and anyone can perform it. The only additional equipment needed besides an ordinary computer is a better quality microphone. The software necessary to retrieve the passwords from the recordings is often available through the Internet at no or very little cost.

The same type of attack is applicable to discover authentication credentials of a user in front of ATM or the code necessary to enter the house. Nowadays, codes are widely used as authentication mechanism. An attacker needs only a microphone to record the sounds of the keys being typed in order to be able to discover the code.

1.2 Description of the research problem

All electronic devices emit unintentional electronic waves, usually referred to as emanations, during their normal operation. Both, acoustic and optical emanations can at some point be used for all sorts of attacks. Khun's research (Khun, 2003) shows that by using optical emanations from both, cathode-ray tube and liquid-crystal monitors it is possible to eavesdrop on contemporary computer displays. In his report he also demonstrates how to use the information emitted via the video signal for recovering the plain text on the screen.

Acoustic emanations from various devices could also be utilized to reproduce all kinds of information. For example, Zhuang, Zhou and Tygar show that the text printed on the matrix printer could be retrieved using acoustic emanations from the printer and there is a possibility to find CPU operations using acoustic emanations, (Zhuang, et al., 2005).

The keyboard acoustic emanations attack is a relatively new type of attack. Users are not aware about the possibility of the sounds emitted by the keyboard as the user types being intercepted and recorded in order to retrieve sensitive data. Even if the security of the system is rather high and involves encryption of passwords, this type of attack can discover the credentials being typed by the user and use them for malicious purpose. Therefore this type of attack is of a serious security concern. It is very difficult to detect it, identify the attacker and take legal actions.

To the best of our knowledge, the first paper on the possibility of this type of attack, appeared in 2004 (Asonov & Agrawal, 2004), followed by another one in 2005 (Zhuang, et al., 2005). Both of them demonstrated that the keyboard acoustic emanations can be easily recorded using an ordinary microphone and a laptop. The main problem these researches had was to distinguish the different acoustic signatures different keys produce, especially when passwords are not ordinary English words.

Two issues were not addressed with the experiments presented in these two papers. The first is the physical environment of the real scenario. The experiments were conducted in a laboratory, which is considered to be a controlled environment. In a real scenario the environment where the victim and the attacker are physically located is some public space where other sounds, considered as noise, are present. The second issue is the ability of the attacker to identify the target keystrokes and the

complexity of the recognition software needed. In both experiments rather complicated methods for recognising the keys typed were used.

1.3 Research, aim and objectives

The aim with this research is to present empirically how easy it is to breach the security of a computer user using keyboard acoustic emanations and to suggest precautions users can take to protect themselves from this attack. The objectives are the following.

- Perform an experiment in which the keyboard acoustic emanations will be recorded in the presence of some environmental noise (real scenario) using a very simple equipment;
- Implement different techniques to characterize the acoustic emanations and easy to use signal processing techniques to extract the keystrokes typed from the signal recorded;
- Analyze the accuracy in recognizing the typed keystrokes under different scenarios and identify the scenario with the best performance
- Compare the experiment conducted within this research with those conducted by other researchers;
- Raise awareness of the users on how serious and harmful this attack is.

1.4 Research questions

The research focus is on performing an experiment with the keyboard acoustic emanations attack. The findings in this experiment are supposed to answer the following research question.

How difficult is it to perform a keyboard acoustic emanations attack?

Several sub-questions can be derived from this research questions.

1. What kind of equipment is necessary to record keyboard acoustic emanations?
2. Can the typed keystrokes be extracted accurately under various circumstances such as different typing styles and presence of noise?
3. How to use simple signal processing techniques to remove the noise and extract the keystrokes?
4. What is the accuracy in recovering the correct key strokes when compared to more complicated techniques for their extraction?
5. What should the users do to mitigate the threat from the keyboard acoustic emanations attack?

1.5 Limitations

Due to the limited time, not all the keys on the keyboard were taken in consideration. Our experiments are performed using only seven keys on the keyboards. The keys selected are *A, C, R, K, I, M* and *P*. Hence, we cannot confirm that results obtained for these keys will be the same for all other keys. Due to the use of seven keys only, a limited number of words created with the some of the seven letters could be used for the tests performed with the keyboard emanations attack where the user (the victim) types the whole password.

Our area of research on keyboard acoustic emanations uses simple statistical methods for extracting the characters typed from the recorded signals. However, in some real time scenarios different than the one used here, these statistical methods may produce different results. The environmental conditions and typing styles can differ very much depending on the situation and the user. Therefore we cannot claim that the results obtained with this empirical research are completely replicable.

The sounds produced depend on the type of keyboard. The experiments were conducted on Dell and HP keyboard, therefore the results obtained in the same environment and using the same statistical methods can differ if keyboards from other manufacturers are used.

We have not performed experiments with ATM although the literature review performed shows that acoustic emanation attacks are possible and can be done in a similar way to the experiment we performed.

The approach used in this research assumes that most of the users have English words for their passwords. Therefore for achieving better accuracy in retrieving the sensitive information, language tools such as online dictionary is proposed.

1.6 Target audience

The target audience for this research comprises all security professionals, those working on research in the area of IT security, as well as employees in the IT security departments. Having in mind that these attacks can appear inside organizations' IT departments, security experts can find this research useful especially for their own education and when offering security training courses to the users. The results presented here should be of particular interest to those who are investigating side channel attacks. Besides them, any computer user interested in protecting sensitive information can make use of the findings in the thesis through becoming informed about how dangerous keyboard acoustic emanations attacks are and how to protect against these attacks.

1.7 Outline

The thesis is structured into 6 chapters. The first provides short background of the area, describes the research problem, the goal of our research, and the research questions. The limitations and constraints, as well as the target audience are presented in the last sections of chapter 1.

The second chapter is called "Extended background". We discuss the concepts related to keyboard acoustic emanations and some of the techniques used when performing these types of attacks. The literature review section presents the articles where different experimental procedures of keyboard acoustic emanations attacks are described.

In part three, we explain the methodology we followed, and present the reasons behind the preferences chosen in our research. This chapter contains a section where the implementation of the methodology we followed is described.

Chapter four presents the results achieved with this research. It presents the approach to the keyboard acoustic emanations attack we took, the novelties introduced, as well as all the graphs and tables obtained through in this study.

In chapter five we critically discuss our work and compare the results achieved with those of the previous research. The experimental errors, the summary of all the findings of our experiments and reflection on our study is presented.

The final part consists of concluding remarks and directions for the future work in this area.

2 Extended Background

The intention of this chapter is to introduce the terminology and short explanation of techniques used in the keyboard acoustics emanations attacks, and to present the current research in this area. It consists of three sections. The first one explains keyboard acoustic emanations and some of the related issues. The second section gives an overview of the Fast Fourier Transform (FFT), a procedure used to analyze recorded sounds, and presents the distance metric technique used in the process of recognizing the keystrokes from the recorded acoustic waves. The last section provides literature review on the research related to the keyboard acoustic emanations attacks.

2.1 Keyboard acoustic emanations

When a person is typing the pin number on the automatic teller machine (ATM) or on the keyboard, the ATM keypad or the computer keyboard produces sounds. Each key produces different sound when pressed. The user rarely pays attention to these sounds and feels as if no one knows what was typed. The human ear can notice same sound while typing. However, this is rarely paid attention to despite the fact that the acoustic sounds coming from the keyboard/keypad can be used by attackers. They can capture them and have possibility to extract pins/passwords to be used later for malicious purposes. In order to capture the sounds the attacker needs a microphone, computer, and software to process the gathered information.

Figure 1 depicts the scenario of an attack on an ATM keypad. Here, the user is typing ATM pin number while the other two people are standing in the queue to withdraw money from the ATM. The person who is standing next to the victim has microphone in the hand. Using the microphone the keypad sounds are recorded. In this the way attacker can extract the PIN number being typed by the victim. Once the attacker has the PIN number, he/she can easily reproduce a magnetic strip smart card with this pin and use the pin to redraw cash from the ATM. (Audri & Lanford, 2008)

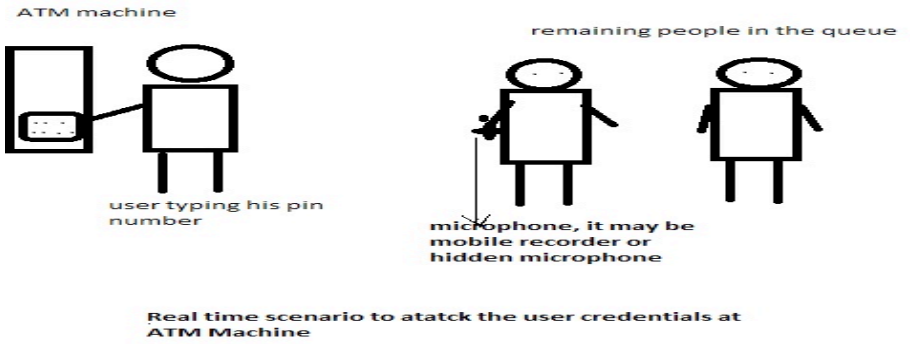


Figure 1: Attack scenario on ATM keypads

The sounds produced by the keyboard whenever a typist presses the keystrokes on the keyboard or the keypads are called keyboard acoustic emanations. These keyboards/keypads are the common input devices that have been the main hardware equipment used to enter the authentication credentials for a

long time. Although the number of touch screen devices in a form of smartphones and pads is increasing, the keyboards are still the prevalent users' input devices.

2.1.1 Keystroke logging

Keystroke logging, also referred as key logging is a technique to retrieve sensitive information via recording the key strokes. However, this technique requires having a software program or a hardware device which monitors the key strokes being typed by the user on the computer's keyboard. These hardware and software tools capture the data transferred from the keyboard. Rouse defines the technique as "It is the action of tracking the key struck on a keyboard" (Rouse, 2010).

Key loggers are available in five categories: hardware, software, wireless, Bluetooth and acoustic key loggers. These five key logging techniques store captured data in the log files. Their way of capturing data and the implementation differ as per the tool which is used in key logging. Software and hardware key loggers store log files related to key logging information on victim's machine or the machine the attacker wants to compromise. Wireless and acoustic key logging techniques store these files in a device which can be useful for data collection (Rouse, 2010).

2.1.2 Acoustic key loggers

This is different type of key logging technique when compared to hardware and software key loggers. One advantage of this approach is that there is no need of any access to the victim's computer. The sounds emitted from the keystroke can be recorded remotely. Only a small microphone and a laptop to record the signals are needed. The extraction of the keystrokes from the captured data can be done by using appropriate software. Acoustic key loggers work with emanations produced form the keyboard. Hence, they are referred to as acoustic emanations. Acoustic emanations can be observed not only in keyboards, but also from printers, CPU, and other computer hardware (Olzak, 2008)

2.1.3 Keystroke dynamics

Keystroke dynamics is the dynamics or the typing pattern of the user. The dynamics include speed variation while moving between particular keys, and duration of the key being pressed. Recently it has become one of the biometrics techniques to recognize the user.

Biometrics is a discipline concerned with uniquely identifying humans based on their physical and behavioral attributes. When acceptable level of reliability in identifying users is achieved the identity access management and access control to various cyber areas is performed using biometrics instead of ordinary passwords.

Biometrics can be physiological and behavioral. Keystroke dynamics comes under behavioral biometrics which relates to the behavior of the person (Ross, et al., 2007). Behavioral biometrics of a person can be identified based on how he/she performs some activity (Olzak, 2006). For this purpose a template of behavior is created for each user. In case of keystroke dynamics the template is based on the unique typing pattern of the user. In order to form the template the duration and latency are estimated (Crawford, Aug. 2010). They are defined in the following way.

1. Latency: How long a key has been pressed?

2. Duration: How much time it takes to move from one key to the other key (Anonymous, 2010).

In order to work as a biometric feature, keystroke dynamics needs to calculate the average latency between the keystrokes and the duration of each keystroke. This is done by having the user typing the

authentication credentials several times. Based on these details one reference template is created and stored in the database for future verification. This template is used to compare the latency and the duration with the one calculated when the current information for identity verification of the user is entered. The main advantage of the key stroke dynamics used for biometrics identification is that it does not utilize additional hardware equipment like other biometric techniques. It uses only the keyboard on which the user types his/ her authentication credentials. So it reduces the overall cost of the system, thus making the keystroke dynamics one of the cheapest biometric techniques. In addition keystroke dynamics is a very user-friendly technique because it does not require additional effort from the user (Olzak, 2006). In some areas organizations are already using keystroke dynamics as an authentication mechanism. (Olzak, 2008).

2.1.4 Keyboard Acoustic Emanations Attacks

Keyboard acoustic attacks are performed by recording the typing sounds of keys on the keyboard or keypad. Attacker uses these sounds to obtain the text being typed by the user. The attacks can be performed not only on a regular keyboard connected to a desktop computer. It was shown that it is possible to conduct the same attack on variety of keyboards, such as wireless, USB, PS/2, and laptop keyboards (Frankland, 2011). This type of attacks may jeopardize the use of keystrokes dynamics as biometrics technique to become a vital option in many computer based organizations because of the vulnerabilities of remaining authentication techniques. Keystroke dynamics can also be retrieved with this type of attack, making this biometrics technique vulnerable.

Attackers are always challenging the security professionals by attacking secure systems using different techniques to breach any type of security. The keyboard acoustic emanations attacks are considered to belong to the so called side channel attacks. They are mainly based on side channel information which can be easily retrieved from the system in nonstandard way. This information can be obtained even when encrypted.

Side channels attacks observe the information leaks from the computing devices. These types of attacks rely mainly on physical implementation of the cryptosystem rather than algorithm or theoretical weaknesses. Their main focus is on electromagnetic emanations, acoustic emanations, power consumption, and diffuse lights from the CRT displays. They are very dangerous threats because the attackers can capture information without the knowledge of internal operation of the system. Besides information obtained from acoustic emanations, other known side channel attacks are power monitoring, timing analysis and electromagnetic as presented in *Syhw's Posterous* (Syhw, 2011). Different attack techniques, such as power and communication links, microwave radiation etc. that are also considered as side channel attacks follow different approach as presented in *Side channel attacks* (Quisquater & RiZK, 2002).

One example of power monitoring is the power analysis attack on ultrasonic waves which were recorded near the system while performing RSA encryption. They were used to find patterns in power consumption. These patterns were compared to the CPU operations and helped in breaking the RSA key (Quisquater, 2002 October).

The side channel attacks are not characteristics of the modern times. They have been known much earlier. During the World War I the side channel attacks proved their capacity by tempest attacks on telephone lines. Recently, many side channel attacks have been found and are published regularly (Shamir, 2011). However, the standard crypto applications cannot protect users from the side channel

attacks. Every time countermeasures to protect the system from the specific attack are developed, a new type of attack arises. It is very difficult to predict upcoming attacks.

Obtaining sensitive information through side channels is a big security breach since attacker can find the confidential information without user intervention or accessing the system. The keyboard acoustic emanations attack is inexpensive and non-invasive. Attackers can attack user activity by placing microphone somewhere in room without any physical contact to the system. It is considered to be very portable attack because the attacker needs only microphone to capture data. Because of these characteristics this attack has been of an important concern with respect to security and privacy.

2.2 Procedures in extracting keystrokes

Capturing the acoustic emanations from the keyboard/keypad and storing them in a file is the first part of the keyboard acoustic emanations attack. The other, more challenging part is the extraction of the typed keystrokes.

The literature reviewed shows that the accuracy in recognizing the keystrokes is higher when two phases are used for the attack. The first one needs to be performed in advance and is usually called training phase. In this phase the features or characteristics for the sounds emanated from the keys on the keyboard are extracted. They are needed for achieving higher accuracy in the second phase when it is necessary to recognize the keystrokes from the sounds recorded. The second phase is called recognition phase and it is focused on finding a way to recognise the typed keys. In this phase the same features as in the training phase are extracted for the recorded sounds. Then the keystroke classifier is used to retrieve the key based on the features extracted and the information from the training phase.

This section contains two subsections explaining techniques used in the research presented in this thesis. During the two phases, training and recognition phase, it is necessary to analyze the sound waves recorded with the microphone and extract the features or characteristics for each sound. This can be performed by using Fast Fourier Transform (FFT) or other techniques. Literature on the keyboard acoustic emanations attacks has shown that FFT features, cepstrum or using the raw audio signal as a set of features can be used. This research uses FFT and raw audio signals. Hence, the subsection 2.2.1 gives a short explanation about FFT.

?? what is this?

Researchers used different procedures for the keystroke classifiers, some more and some less complicated. Our research uses a rather simple procedure known as Distance Metric Approach together with language tools. It is our assumption that most passwords are taken from a dictionary, hence the language tool is introduced to improve the accuracy. The Distance Metric Approach (DMA) is selected because of its simplicity. It calculates distances between the features of the sounds for the known and unknown keys and is described in the subsection 2.2.2.

2.2.1 Fast Fourier Transform (FFT)

The microphone converts the sound waves in electric waveforms stored in a file. The electric waveforms obtained from the acoustic emanations cannot be analyzed immediately. These waveforms are complex and consist of many sinusoidal components. Therefore, it is necessary to obtain the frequency spectrum for these waves. The frequency spectrum displays the frequencies and relative intensities of the sine waves that constitute the complex wave. For example, in case the time-varying signal consists of a single frequency (tone), the time waveform has a shape of a sinusoid as shown on

a) b)

ure 2 Time-varying signal of a single
e (a) and its frequency spectrum (b)

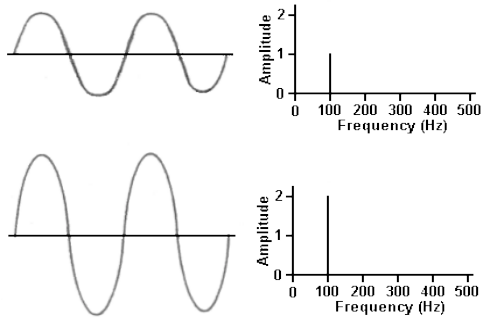


Figure 2a and the frequency spectrum consists of a single vertical line placed at the appropriate frequency. The height of the line represents the amplitude of the wave as shown on Figure 2b. In case of complex waves the frequency spectrum is obtained using Fourier Transform (FT). It is a widely used mathematical transformation that represents complex waves as an infinite sum of their sinusoidal components. The constant in front of each component, called Fourier coefficient, determines the height of the component on the frequency spectrum. In order to remove the components with smaller amplitudes, which are not of interest, it is sometimes necessary to filter the signal.

Discrete Fourier Transform (DFT) combines FT and filtering and makes possible to implement the algorithm for computing FT coefficients. A procedure that is very efficient in performing DFT and obtaining the frequency spectrum is called Fast Fourier Transform (FFT). FFT decomposes the time signal of a waveform in a sequence of values for different frequencies and performs it in an efficient way. The order of the number of operations necessary for DFT to compute the values for N frequencies is O(N^2) arithmetical operation, whereas FFT computes the same in O(N log N) operations. Compared to DFT, FFT has substantially higher speeds. This is very important in applications with a huge amount of data (Anon., n.d.). FFT is usually a part of many software packages.

2.2.2 Distance Metric Approach

In the first phase of the procedure for extracting the keystrokes, FFT is performed and the characteristics or the features for all the keys on a keyboard are determined. In the second phase, it is necessary to perform the same steps for the recorded acoustic emanations. The goal is to organize and analyse both collections in order to be able to define and compute meaningful similarity between the shapes of the datasets. Then the features extracted can be compared using the distance metric approach.

Distance metric approach is used for similarity estimation hence pattern recognition and machine learning are the fields where it is implemented the most. The goal is to get high recognition rate on unknown samples from the similar ones. For example, face recognition, speech recognition, and visual object recognition methods are concerned with the distance metrics approach. During the process of recognizing a pattern it is common the outcome to be used for further training. In context of the distance metric approach this is referred to as metric learning. The objective of metric learning is to learn an optimal mapping in the original learning space. The algorithms for distance metric learning can be categorized as supervised distance metric learning, and unsupervised distance metric learning based on the way we use our side channel information. Unsupervised metrics takes input data set, and get an embedding of that data in the training set. Supervised metric learning learns distance metrics which keep distance points to the same classes close. At the same time it separates data points from different classes (Jin, 2008). The supervised learning gives better metrics for classification compared to unsupervised learning particularly when information is not fully structured (Jin, 2008).

For the research performed in this thesis Euclidean distance metric and Manhattan distance metric were used for the key recognition and for supervised learning. These metrics are also used in biometrics when it is necessary to compare extracted features with features already available that are

slightly different every time they are used. The comparison of the extracted features is based on the distance measured between the extracted features from the trained samples and those from the pattern observed. At the same time the comparison is used for metric training.

A distance metric function *d* for set *X* can be defined as the following.

$$d: X*X \rightarrow R \quad (1)$$

X is characterized by some number of elements, $X = \{x_1, x_2, \ldots, x_n\}$. In case of the sound waves the values would correspond to the features characterizing the sound wave. *R* is the set of real numbers and the relation shows that the operation defined with *d* usually produces a real number. *d* is the distance function which defines the distance between the elements in the set X. Distance functions usually have the following properties.

- Non-negativity property, meaning that the distance is always a positive number,
- Property of symmetry, meaning that the distance between two elements in *X* is the same regardless of the direction in which it is measured,
- Identity meaning that the function has a value equal to 0 only when two elements are equal.

There are many distance functions. In this section we define only the Euclidean and Manhattans distance functions because they are used for the research presented in the thesis.

Let's assume we have two data sets, P and Q. Then we can define Euclidean distance metric and Manhattan distance metric as follows.

Euclidean distance metric

The Euclidean distance metric between data sets *P* and *Q*, *where* $P = \{p_1, p_2, p_3 \ldots p_n\}$ and $Q = \{q_1, q_2, q_3 \ldots q_n\}$ is

$$E = d(P, Q) = d(Q, P) = \sqrt{(\Sigma (p_i - q_i)^2)}. \quad (2)$$

Manhattan Distance Metric

Manhattan distance metric provides a very fast way to find the distance between two points. It is defined in the following way.

For two data sets *P* and *Q*, *where* $P = \{p_1, p_2, p_3 \ldots p_n\}$ and $Q = \{q_1, q_2, q_3 \ldots q_n\}$, the Manhattan distance metric is

$$M = \Sigma |p_i - q_i| \text{ for } i = 1, \ldots n \quad (3)$$

(Anon., 2010).

Figure 3 illustrates graphically the differences between the Euclidean and Manhattan distance in a two-dimensional space.

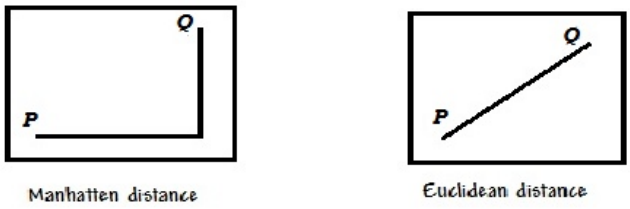


Figure 3 Euclidean and Manhattan distance in a two-dimensional space

??? There are 2 ways to measure the distance: using intra person distance and inter person distance. Inter person distance must be large, and intra person distance must be small in order to get the expected value. (Facundo, 2011)

2.3 Literature review

The research in the area of keyboard acoustic emanations is rather new and there are not too many publications. To the best of our knowledge, the first article was published in 2004. Through the experiments performed, for the first time it was shown that the sounds generated when different keys are pressed on a computer keyboard slightly differ from each other. The extensive search of literature showed that very few publications produced later used different methodology to understand the keyboard acoustic emanations attacks than the one presented in the first published article. In this literature review section, we include only those making analysis of the keyboard acoustic emanations attack through experiments and using different methodology than in the first article. This is because this research is focused on performing experiments to find out how difficult it is to perform this kind of attack. Therefore, the idea was to combine methodologies which are known in such a way that the simplest one is selected for each phase of the attack.

The subsections below give an overview of the articles which we reviewed to understand the keyboard acoustic emanations attacks and methodologies used in each phase of the attack.

2.3.1 The seminal paper by Dmitri Asonov and Rakesh Agrawal

In 2004, the first research paper on keyboard acoustic emanations was published by Dmitri Asonov and Rakesh Agrawal from IBM Almaden Research Center (Asonov & Agrawal, 2004). The authors showed that each key on a computer keyboard makes a different sound when struck and explained that the clicks sound differently because the keys are positioned at different position on the keyboard plate. This discovery was the bases for the idea of inspecting the possibility in identifying which key has been pressed based on the audio recordings of the acoustic emanations. Their experiment in performing a keyboard acoustic emanations attack employed a neural network to classify clicks since this had already been successfully used for speaker identification. They used expensive computations to extract the FFT coefficients from the keystrokes sounds and used them to identify relevant features for each key. The neural network was trained with pairs {key, feature}.100 clicks were recorded for each key and features extracted accordingly. Then they were tested before the final step of comparing the output produced by the neural network and the actual key pressed was performed.

Besides the complicated procedures they implemented, the recognition rate they were able to get was only 80%. The main drawback in this research is the fact that they performed their experiments on a single type of keyboard, and with only one person typing the keys. Performance would have been

affected if they would have tried to recognize the typed keys on a different keyboard and with another typing style.

Besides showing that different keys on the computer keyboard produce different sounds, the big contribution of this research is the proof that the same type of attack they performed with the computer keyboard can be done on any push button devices like ATM pads, or telephones.

2.3.2 Research performed by Li Zhunag, Feng Zhou, and J. D. Tygar

In 2005, Li Zhunag, Feng Zhou, and J. D. Tygar from University of Berkeley accomplished research on keyboard acoustic emanations in order to solve limitations of the previous research (Zhuang, et al., 2005). Their concern with the previous research was the use of a labelled training sample in the recognition phase. Obtaining this type of sample might not be an easy task to do for the attacker since it would require very long surveillance of the victim. Instead of using a labelled training sample (the pairs {key, feature} for all the keys on the keyboard), Li Zhunag, Feng Zhou, and J. D. Tygar showed that the attack could be performed using as an input the recordings obtained for 10 minutes of a user typing English text on the keyboard. They considered that obtaining this type of sample is easier and therefore makes the keyboard acoustic emanations attack more dangerous. Their assumption was that the typical text typed by the user is not random. According to the authors "when one types English text, the limited number of English words limits the possible temporal combinations of keys, and English grammar limits the word combinations" (Zhuang, et al., 2005). Based on the sound produced by each click, the keystrokes are clustered in classes. In the recognition phase the most-likely mapping between these classes and the typed characters is established using language constraints. Using a word as the unit of extraction instead of a single character, spelling and grammar rules can be used to further increase the recognition rate.

The researchers performed their own experiment using the unlabelled training data. Instead of using FFT coefficients as features they used Cepstrum[1], a technique developed by researchers in voice recognition. They showed that the recognition rate using Cepstrum features was higher than that with FFT. In addition, they showed the neural networks perform worse than linear classification on their data sets. They performed unsupervised clustering on the Cepstrum features, splitting the keystrokes into K classes where K was slightly bigger than the number of keys on the keyboard. This was done using clustering algorithms which are slightly imprecise. Each key can belong to a certain class based on a certain probability. A key can belong to one or two classes. To recover text from these classes, the researchers used HMM (Hidden Markov Model)[2] language model. With HMM the correlation between the keys typed in a sequence can be determined based on the sequences of letters that are next to each other in the English language. To achieve better accuracy spelling and grammar checking is performed and feedback from the recognition phase is fed in.

In the recognition phase researchers used trained keystroke classifier to recognize new sound recordings. The recognition rate achieved for English text was 96% and 90% for 5-character random passwords consisting of letters only.

The main contribution with this research is the use of unsupervised training phase (recordings of typed English text analyzed with HMM are needed), the proof that Cepstrum features could be used instead

---

[1] *Mel-Frequency Cepstral Coefficients (MFCCs) were used. These are calculated from the audio wave spectrum in a different way than FFT and widely used in many speech recognition applications.*

[2] *HMM: Is finite set of states each associated with probability distribution. Each state generates observation as outcome which is hidden to the outside observer.*

of FFT coefficients and the higher accuracy achieved. The main disadvantage of this approach is that it is language specific, hence requires different algorithms, dictionaries and grammar rules for different languages. In addition the passwords typed that could be recognized were limited to letters from the English alphabet.

2.3.3 Dictionary attacks conducted by Yigael Berger, Avishai Wool, Arie Yeredor

Yigael Berger, Avishai Wool and ArieYeredor (Berger, et al., 2006) built their attack on the results from the previous research. They simplified the attack through avoiding the use of any training samples and used combination of signal processing, efficient data structures, and algorithms in order to re-construct the single words from the recordings made during the attack. Analysis was performed by taking audio signal recordings of keystrokes and dictionary of words and assuming that the typed word would be present in the dictionary. The researchers utilized a dictionary-based attack to predict individual words from the recorded audio samples. They exploit the fact that the sounds produced by keys correlate to their physical positioning and that the distance between keys in feature space is proportional to the physical distance between them on the keyboard. For example Q, W, E produce sounds alike that are different from far separated keys like M, N. They used this similarity to create constraints for each word analyzed. The distance between every combination of key pairs is categorized as either the same key, adjacent keys, near keys (at most 2 keys apart) or distant keys. Then these constraints are applied to dictionary algorithms to discover which of them are the most probable. Different feature types were tested to find out the levels of precision and recall. Precision means the fraction of constraints that are true for the real world relative to the number of constraints. Recall measures the fraction of true constraints relative to all possible constraints. They tested the performance using different features obtained from FFT, cepstrum and raw audio signals which gave the best results. They found that one of the signal processing techniques “cross correlation primitive” is more effective method compared to FFT and cepstrum which were used by previous researchers.

The attack consists of 5 stages, signal processing and feature extraction, keystroke processing, constraint formulation, constraint evaluation, and outcome prioritization. The demonstrated success rate was 90% of finding the typed word in the top 50 candidate words identified with the dictionary, and 73% over all the words they tested. They also showed that the factors affecting the success were the length of the word and the number of repeated characters in the word. The attack they have shown is also known as “Acoustic based password hacker”.

The main advantage of this research is that the attacker does not need any previous training sets, thus making the attack easier to perform. The disadvantage of this approach is the fact that it is applicable only for passwords taken from the dictionary, it is limited to the English language only and is not applicable in case of strong passwords.

2.3.4 Triangulation attacks by Hui ShunPak

In order to evaluate the seriousness of this type of attack Hui ShunPak, a student from the Chinese University of Hong Kong performed extensive research on acoustics. He analyzed the previous research on keyboard acoustic emanations attacks and suggested a new approach “Time Difference Approach” also referred as “Triangulation attack”.

In this study two microphones were used for detecting the sounds from the keyboard. By measuring the time differences in arrival times of the keystrokes at the two microphones placed near the keyboard the attacker can easily reconstruct the typed key. The research is based on time difference approach.

The attacker can attack the system by placing two microphones in the proximity, but on the different side of the keyboard. When the user types the key it emits the sound that travels through the air and reaches the microphone. The two microphones are reached with different timings due to the different distance of the key from the two microphones. By finding the time difference of the received keystrokes by two microphones, knowing the distance and the sped of the sound, the typed key can be located.

The main advantage of this attack is very easy computations involved making the attack to be very cheap. Hence, it is more dangerous compared to other kinds of attacks since it could be easily implemented and anyone can use it. The disadvantage of the attack is the need of some knowledge in sound physics and having to use two microphones. However, it may work fine if the attacker places these two microphones in a correct place. Since this research is based on time difference approach, it may not be applicable in all types of scenarios (Pak, 2006).

2.3.5 Evaluation of strong passwords and typing styles by Tzipora Halevi and Nitesh Saxena

The research presented in previous sections lack the capability of recognizing strong passwords with big accuracy. One of the main concerns nowadays is to avoid vulnerabilities in password maintenance. Therefore users are instructed to use very strong passwords because weak passwords are susceptible to dictionary attacks. Strong passwords consists of high bit entropy, and apply random selection of characters. In order to retrieve the strong passwords with keyboard acoustic emanations attack Tzipora Halevi and Nitesh Saxena form Polytechnic Institute of New York University proposed a new approach (Halevi & Saxena, 2006).

In their research their main focus is on typing styles. They consider the typing style to be neglected in the previous research and make it crucial in their research. They define two different typing styles which are considered in the research, one is referred to as “hunt and Peck typing”, the other as “touch typing”. These typing styles severely affect the eavesdropping algorithms.

They used different signal processing techniques and machine learning for the keystroke detection. They preferred “Dynamic Time Wrapping” technique which is mainly used for speech recognition. Based on the time frequency classification they introduced key detection method which achieved 83% recognition rate.

They argued that there are some limitations in previous acoustic eavesdropping techniques under security- sensitive and realistic scenarios that means different typing styles and also on strong passwords. However, their work is limited to 6 character strong passwords only.

Research scenario

Deduction

Induction

3 Method

Predictions

Test of prediction

3.1 Research approach and methodology

Observation

The role of the research is to follow scientific methodology when performing some particular study and create useful knowledge based on empirical evidence and conclusions made from the results obtained.

Figure 4: Srages in thr methodology for our research

One of the objectives of this research is to perform an experiment in which the keyboard acoustic emanations attack will be performed. This includes identifying all necessary equipment for this type of attack, creating scenarios for the experiment, conducting necessary experiments, gathering data, making observation and analysis and deriving conclusions. Hence, the experiment as a research strategy is the best choice having in mind the objectives of the research.

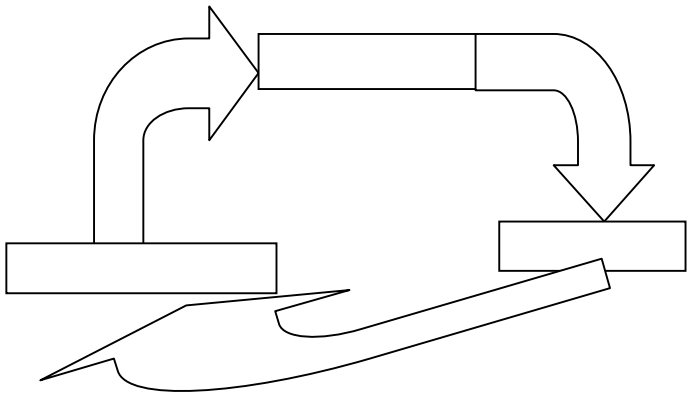
The work presented in this thesis is an empirical study meaning that actual experiments are taking place together with objective observations and experimentations. Empirical methodology involves observation and analysis of the previous experimental results, creating the scenario for the experiment and applying particular methods for gathering and data analysis that will lead to answering the research questions. Using empirical research methodology one can gain knowledge by observing the research area and also form experiences which researchers call “evidence”. In this particular study the evidence are the sound waves recorded, the feature extracted and values obtained through various computations.

The way of doing empirical research varies based on the field of research and the research questions posed. If the research question cannot be answered by performing experiments, some other research strategies might be considered.

We follow empirical methodology cycle to perform our research and collect empirical data from our observations. The main corpus of research performed on keyboard acoustic emanations attack consists of publications that present experiments performed using different techniques for the two phases of the attack. This is an additional reason for feeling more comfortable in the use of empirical methodology.

Our research has three stages as presented on Figure 4. The stages are the following.

1. Induction: It is the process of collecting ideas and information from the literature to form our own research scenario.



ature ew, edure ction	ntal setup, testing	gathering data	scenario selection	conclusion

2. Deduction: It is the act of forming our own research process, by collected data as per our convenience.
3. Observation: It is the act of testing the research scenario (Trochim, 2006).

The research begins with the research question and the knowledge gathered from the literature review. Induction process is used for observations of the research already performed and for consolidation of our own ideas on how to formulate the research scenario to reach our goal. Once the research scenario is defined we perform the experiments to test the prediction. Observation is necessary to find out whether there is an answer to the research questions of this study. In case there is not or the answer is not good enough, the whole procedure is repeated until a satisfactory answer is obtained. By using deductive reasoning we narrow down our research as per our convenience using statistical methods. Finally we test our own experimental results and evaluate those results (Trochim, 2006).

We define six phases in our empirical study, as presented on Figure 5.

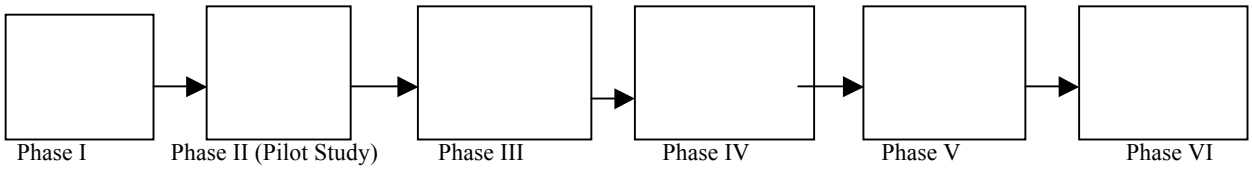


Figure 5: Research phases

Phase 1 consists of literature review in which we analyse the previously performed research and studies related to the keyboard acoustic emanations attack. This phase assists researchers in building their theoretical knowledge and getting an overview of the options for selecting procedures to be used in the experiment. The first ideas about the scenario of the experiment are created in this phase.

In phase 2 the pilot study, we test the system with only two keys of a single keyboard to analyze the features of the keys in normal circumstances. The output of this phase is used to modify some settings in the equipment. In this phase the result obtained by Asonov and Agrawal about each key on the keyboard producing a unique sound is confirmed.

During phase 3 tests with multiple keys and words are conducted. We form six different scenarios based on the key features selected and the type of the distance metric procedure used. By running several experiments, data for all six scenarios are gathered.

Phase 4 is dedicated to the analysis of the data gathered during the experiments performed. This is necessary to understand under which circumstances the best performance of the system can be expected. The system is considered to have the best performance when the EER (Expected Error Rate)[3] value is minimal.

Phase V is necessary to derive conclusions and disseminate it to the target audience.

3.2 Methods for data collection and analysis

Two types of data are gathered in this study. The first type are the procedures used for performing the keyboard acoustic emanations attack used by other researchers, and the second type is the data collected during the experiments made within our own research.

[3] *EER describes the performance of the system, which in turns describes most preferable feature set in order to get the expected result*

Information on how the attacks were performed by other researchers is available in the articles presented in section 2.3 of this thesis. The attacks describe two different approaches to the attack. One approach is to include the training phase in the attack and collect training data to be used for recognizing the keys typed from the sounds recorded. Several ways for gathering training data are suggested in the literature, some of them requiring more effort than the others. The other approach does not require training data and uses other information for performing the attack. Both of these approaches extract features from the sound waves recorded when the user types on the keyboard. The features characterize the sound obtained when specific keys on the keyboard are typed. Some procedures for feature selection are more complicated than the others, some lead to better accuracy than the others. Several of them are described in the documents studied and are eligible for including them in our research.

Another procedure to be selected is the one that finds the similarity between the sound waves recorded and the training data or other available information to help recognize the keys pressed. Some of the attacks described in the literature based the recognition on the assumption that passwords are English words and besides other procedures use different language tools do improve the accuracy.

The data gathered throughout the experiment are the audio signals generated by the key clicks during the experiments performed within this research, captured by the microphone and stored in a file. These data are in a form of multiple .wav files which are later used for various types of transformations.

Both, the qualitative and quantitative methods are used for data analysis.

Qualitative analysis is required for dissection of different procedures for the keyboard acoustic emanations attack presented in the literature in order to select those most appropriate for this research. The criteria for selection are extracted from the objectives presented in section 1.3. The procedures are to be simple, have the necessary software and hardware easily accessible and user friendly, computationally light (do not require large computational resources) and at the same time produce good performance or accuracy comparable to that achieved by other researchers. The option to combine different and select other procedures for the different phases of the attack is open providing that the aforementioned criteria are satisfied.

Quantitative analysis is needed for the data collected during the experiments performed. Numerical transformations are needed when features are extracted from the recorded sound waves and when the distance metric approach is used to estimate the similarities between two waves. Both, quantitative and qualitative methods are used when comparing the performance of the different experimental scenarios within our research alone and with the performance achieved by other researchers.

The method for presentation of results of the research is also important and is discussed in this section. One can perform research and present the results as a pure text making them not very clear to the audience. Pictures and tables are more user friendly way of presenting results. For our research, we use pictorial research methodology. This means that we present the pictures obtained throughout the experiments. The nature of our study makes possible do that because of the experimental diagrams obtained using signal processing. (Levy, 2006).

3.3 Implementation of Methodology

The implementation of the methodology is presented in five sections, each corresponding to the phases presented in Figure 5.

3.3.1 Phase I: Literature review, procedure selection

The literature review was already presented in section 2.3 of this thesis. All the research articles are carefully studied and analysed in order to gather ideas of how to perform keyboard acoustic emanations attack using simple equipment and easy to use procedures, and at the same time to achieve accuracy comparable to those achieved by other researchers. It served as assistance in framing our own approach to this attack having in mind the research aim and the objectives.

The first objective required the attack to be performed in real scenario and the second objective required implementation of simple equipment and easy to use signal processing techniques (see section 1.3 in this thesis). Hence, it is required to select environment where some environmental noise exists, and relatively cheap hardware and software. The aim is to show that even attacker with not so high level of expertise in signal processing can perform the attack. On the other side, these simple procedures can produce much lower accuracy in identifying the keystrokes compared to the accuracy obtained in the attacks described in the literature. To achieve comparable accuracy we combine procedures described in different articles for the different phases in the attack, and suggest some procedures of our own. In this way our own approach necessary to reach the research goal is developed. Here are the decisions we made by analyzing the research articles and documents and performing procedure selection.

1. Real scenario: The real scenario environment for the experiment is achieved by performing the tests at the Cyber System Security laboratory, Department for Computer and system science, KTH and Stockholm University. The room is located on the seventh floor in the Forum building in Kista, Sweden and most of the time teaching assistants and students work there having occasionally discussions and producing low level sounds. Therefore, during the experiments the noise was not coming only from the keyboard itself. A moderate noise originated from the environment, too. To get confidence in proceeding with the experiments, the decision was made to perform a test phase where the keyboard acoustic emanations are recorded in presence of moderate noise and the extracted waveform is compared with those available in the literature Agrawal (Asonov & Agrawal, 2004).

1. Training phase: Based on the research performed by Asonov and Agrawal (Asonov & Agrawal, 2004) and Zhuan, Zhou and Tyger (Zhuang, et al., 2005) we decide to include the training phase in our approach. In our research we have separated the keystrokes first from the single large wave signal, and later we have extracted features of each key to train the system. These features are the unique characteristics of each key on the keyboard. Using these features we have developed templates as part of training phase. Due to the limited time we have, the templates were developed for seven keys only and not for all keys on the keyboard.

2. Feature extraction: Asonov and Agrawal, (Asonov & Agrawal, 2004) as well as Berger, Wool and Yerdor (Berger, et al., 2006) use FFT for feature extraction in the sound waves. Zhuan, Zhou and Tyger use cepstrum (Zhuang, et al., 2005), and in the approach taken by Halevi and Saxena the time waveforms are used (Halevi & Saxena, 2006).

Selecting what features to be used is a crucial phase since the choice made may produce lower accuracy in the recognition phase. The decision what features to use was based on the knowledge obtained from the literature on the keyboard acoustic emanations attack, as well as the previous knowledge the author had about the characteristics of sounds. Although the literature presents cepstrum as a feature with prospect for good accuracy, it was considered as too complicated. FFT was used in many of the attacks performed earlier, hence it was considered as a good choice. Using only a single feature would not provide opportunities to compare how attacks with different features perform.

Sometimes it is useful to normalize the output of FFT so that a unit sinusoid in the time domain corresponds to unit amplitude in the frequency domain. Hence, we consider using normalized FFT as the second type of feature.

Several researchers have used the amplitudes from the time-varying waveforms of the sounds as features when conducting their attack. Therefore, this feature can also be considered.

At last, the decision is made. Three different features are to be used to be able to see which one produces the best results. The features selected are ordinary FFT values, normalized FFT values and the amplitudes of the time waveforms. The idea behind having three features is to find out if different features yield better accuracy when retrieving the typed keystrokes.

3. Recognition phase: Asanov and Agrawal implemented neural nets4 for training the system and for the phase of recognition (Asonov & Agrawal, 2004). This procedure is complicated and computationally intensive hence it does not comply with the criteria to be simple and computationally light. Instead we implement distance metric approach described in section 2.2.2. We select distance metrics approach to compare trained samples with the recorded ones. This is because it is a very light computational procedure and very easy to use. As described in section 2.2.2, there is a possibility to use two different distances with distance metric approach, Euclidian and Manhattan distance. Our experiment uses scenarios with both of these distances to find out which one produces better results. The experiments in the recognition phase are for a single keystroke and simple words constructed of the seven letters. This is because only these seven keys were involved in the training phase. In cases when it is necessary to perform recognition of English words the accuracy is improved by using integrated dictionaries added to make spelling corrections. This decision was inspired by the fact that almost all the previous research except the seminal work by Asanov and Agrawal included some form of language tools.

4. Measuring performance: Due to the inclusion of three types of features that characterize the sound of each key and two metrics in the recognition phase, the performance of the six scenarios which are created are measured. The detailed description of the scenarios is presented in the next subsection. The features of the extracted keystroke are compared with the features of all seven keys selected in each of the six scenarios based on the distance metric approach described in section 2.2.1. Both, the Manhattan and Euclidian distance metric are used.

The Manhattan metric takes the absolute difference between the vectors and calculates the sum of these values. We can calculate Manhattan distance between two vectors *x* and *y*, using the following formula.

$$D = sum(abs(x-y))$$

i.e. $D = sum(abs(test\ input-template))$

The Euclidean distance metric takes two vectors of same length, and calculates the square of all the differences of their elements. Then the sum of those values is calculated. Let's assume a vector *A* of length *M*D* and another vector *B* of length *N*D*, then pairwise distance between two sets of observations is the distance metric. Euclidean distance between two sets of observations in signal processing using Matlab can be calculated as

$$D=pdist\ (A,B)$$

[4] *Neural Network is an adaptive system, its structure based on external or internal information that flows through the network during the learning phase*

i.e. *D=pdist(test input, template)*

Output vector of the Euclidean distance metric is of length *M*N* i.e. for each pair of test input and template there is one distance metric value. Based on the distance metric values the processing of the gathered data is performed.

Several test inputs are measured and distance metrics between each test input and templates for all seven keys are calculated. This is necessary to estimate the performance for each scenario.

For example, let's assume that the keystroke recorded during the attack is the key *A*. Then the comparison with all the 7 templates is necessary. It is natural to expect that the distance metric in this case for the key *A* should be lower when compared to the other keys. This means that if the template and the recorded keystroke are from different keys then the distance metric is generally larger than in case the template for the same key. (Asonov & Agrawal, 2004). This means that,

Distance metric between *A* ⟶ *A* is smaller,

Distance metric between *A* ⟶ *M* is larger.

Following the logic explained on these arguments the keystroke can be easily recognized. However, the experiments performed show that this is not always the case. In some of the experiments the distance metric has larger value for the keys different that the key corresponding to the keystroke recorded.

4. Novelties in this research: This research proposes a novel structure of the acoustic emanations attack performed with a real scenario. It is based on the analysis of previous research and complying with the aim of this study to make the attack as simple as possible. In order to make the accuracy in the recognition phase, six scenarios are used and their performance compared. A novel procedure for the comparison is introduced. In addition the different attributes of the attack performed are compared with the attacks performed by other researchers.

3.3.2 Phase II: Experimental setup, testing

The equipment for the test bed was rather cheap and simple. Below is the list of the hardware and software used together with the specifications for each item.

Microphone: Logitech Desktop Microphone, which has sensitivity -67dbV/Pascal+/-4db, and frequency response 100 -16Khz.2.5m shielded cord.

Keyboard: Dell Desktop keyboard, with 104 keys, external unit connected to the Philips desktop computer with USB, and HP desktop keyboard (only in Pilot phase).

Computer system: Philips desktop computer.

Matlab software: Matlab R2011a, which is the most common tool for signal processing as a fourth generation programming language. It gives numerical environment. When it comes to user-friendly software and understanding the signal features Matlab is the preferable tool.

WinFF software: WinFF version 1.4.0 is software with a command line GUI for audio conversion. We have chosen this software because of its multi functionality. It can convert multiple files at a time into multiple formats.

The drawing on Figure 6 illustrates the experimental setup for the keyboard acoustic emanations attack.

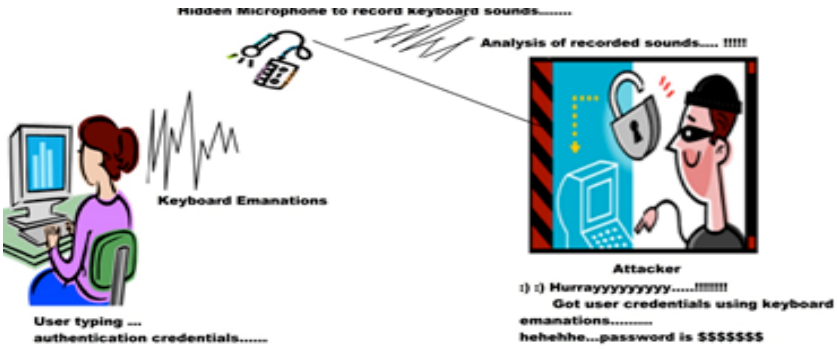


Figure 6: Schematic presentation of the experimenta setup

The user (victim) is working on a desktop computer. The user is typing her authentication credentials in order to login the system. The attacker who has intention to retrieve her credentials has a hidden microphone near her desk. The keystroke sounds which are being typed by the user are wave signals which can be captured by the microphone and processed to find out what exactly she types. Here the hidden microphone records those keystrokes signals. After the attacker process these signals using signal processing techniques authentication details necessary for the actual attack to be performed are retrieved.

The actual experimental setup for the experiments is based on the description of how the attack is performed. The difference between the actual attack and the experimental setting is that instead of retrieving the full credentials that usually consists of multiple keystrokes, a single keystroke was retrieved. This is because of the limited time dedicated to this research in which it was not possible to perform the training phase of the attack for all keys on the keyboard. Instead only seven keys were included.

The equipment setup complied with the second objective in the research. As described above the equipment used is very simple.

The experiments are carried out following the steps given below.

1. Record keystroke sounds (key clicks) of several keys when user is typing, using a microphone and store them in a file on the hard disk of the computer. The file format is .wma.
2. Pre-process the sound recordings using WinFF. The software is converting the .wma file into .wav type of file. This file type can be used as an input in the Matlab software.
3. Remove the noise from the wave signal. The wave signal is a combination of original keystrokes and some background noise. It is important to remove the noise in the signal, and also the silence between the keystrokes.
4. Segment the total wave signal into individual clicks/segmentation using Matlab. Audio segmentation is the artificial process that can identify the boundaries between audio signals. Matlab reads the wave file using “wavread” function. The instruction used for this is *[t,fs]=wavread(file),* where *fs* is the sampling frequency of the signal and *t* is the original data. The sampling frequency refers to the number of samples per second gathered from a signal. Usually sampling frequency value is 44100 Hz.
5. Extract main features of the individual key clicks.
6. Train the system /create templates based on the extracted feature sets.

7. Recognition/Compute distance metric values between different templates with multiple test inputs.
8. Find out the performance of the system based on the computed values, using genuine and imposter scores.
9. Output the results as per a given scenario
    a. Assume the typed word based on the distance metric values.
    b. Evaluate the probability of each letter, and compute probability of whole word.
    c. Spell and Grammar checking using online dictionaries.

Besides the experimental setup this phase includes testing of the system. The first test consists of recording sounds from eleven keystrokes in order to check the application for storing the raw sound waves, extracting the noise and segmentation of the key clicks. This is necessary to obtain the waveform for a single key.

Figure 7, depicts the time domain signal of 11 key clicks together with some environmental noise.

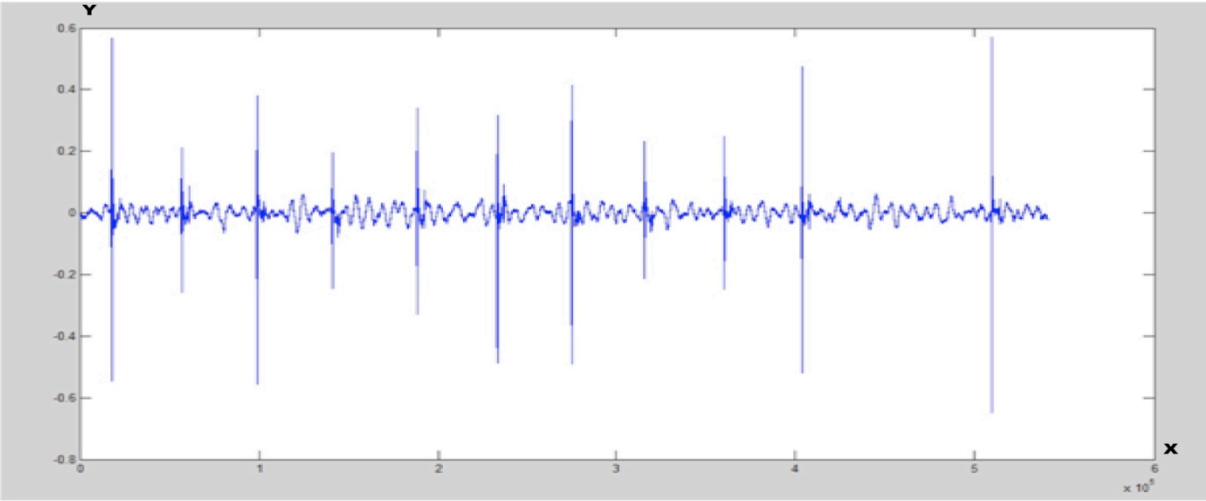


Figure 7: Recorded signal with 11 clicks together with noise

The wave from the recording obtained during the test and shown on Figure 7 does not clearly present the typical waveform of acoustic emanations as described in the literature. Because of the environmental noise, as well as the noise from the surrounding keys, the wave does not contain the higher peak that appear when the key is pressed and the lower peak corresponding to the release of the key. The noise can be a problem for correct extraction of the features of the waves representing particular keystrokes. The noise could be a reason for the wave corresponding to the same key to contain different frequency values. Therefore, it is not convenient to perform any analysis without the noise component in the signal to be diminished from the recorded signal. The signal obtained after removing the noise is presented on Figure 8. This signal appearance is slightly better since the original wave with the hit peaks and release peaks can be recognised even before the waves of individual keys are extracted.

Figure 9: Keys A and P on the keyboard

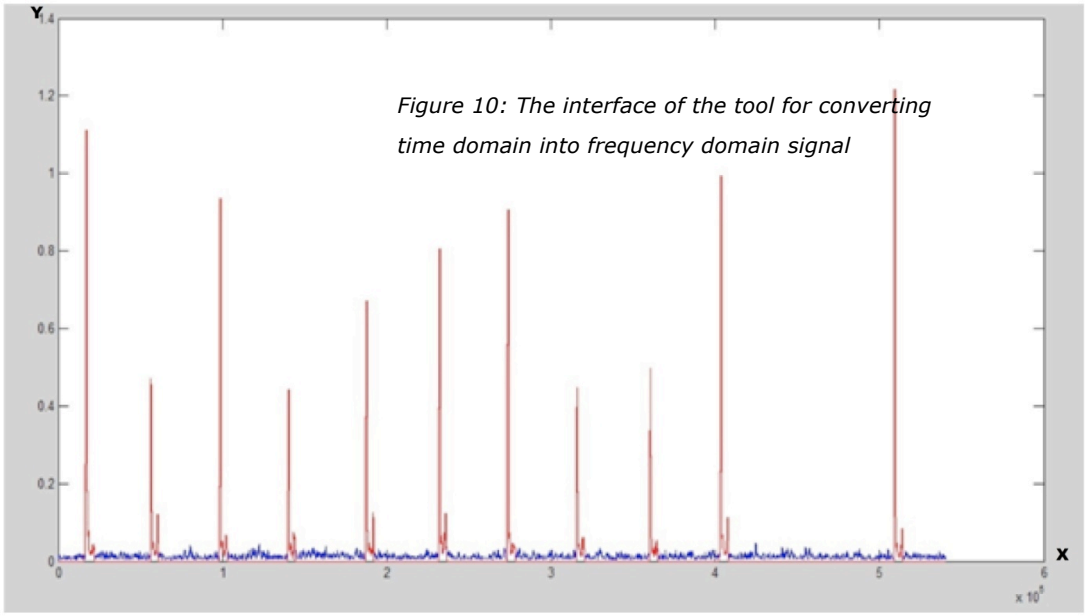


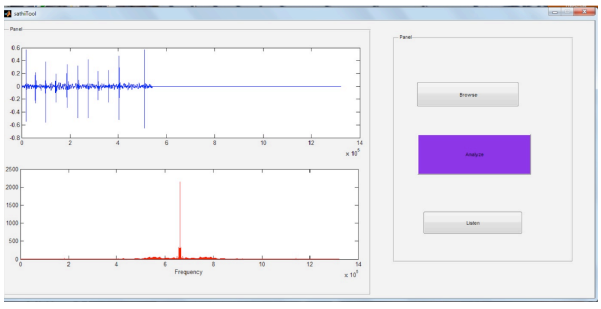
Figure 10: The interface of the tool for converting time domain into frequency domain signal

Figure 8: The wave after removing a large portion of the noise

The extraction of the signals for the individual keys produced the expected waveforms with the higher peak corresponding to the sound when key is struck and the lower peak corresponding to the sound when the key is released. The wave obtained is presented in the chapter showing the results from this research.

The second test is performed to find out whether the wave signals produced by the keys comply with those described in the literature. The other is to check the differences in the features extracted from different keys. The first two keys on the keyboard that were tested are *A* and *P*. As shown on Figure 9 these two are very far from each other on the keyboard.

The test phase includes observation of the differences of the features extracted from two of the sound waves in the time and frequency domain. For the easy analysis a Graphical User Interface Tool that converts the time domain signal into frequency domain signal is used (Figure 10). With a single click on a button the time signal is converted into FFT signal and plot of both waveforms can be observed.



On the time domain plot the variation of the amplitude of the signal is shown. In the frequency plot the frequencies containing the most of the energy of the signal are visible. These presentations are important in order to identify the differences in the features for the two keys using different representations of the signal. This is even more

important for the keys that are located closer on the keyboard plate since from the previous research it is known that the differences in the features are largest for the keys that are far away from each other.

The code snippet to *callback* pushbutton function which converts the time domain signal into frequency component signal is given below.

```
xt=get(handles.pushbutton1,'UserData');
fs=get(handles.pushbutton1,'UserData');
xt=xt(:,1);
XF=fft(xt);
```

3.3.3 Phase III: Defining scenarios, gathering data

In this phase the scenarios for the experiments were defined and gathering of necessary data was done.

Due to the limited time not all the keys on the keyboard are included in the experiments. The experiments are performed only for the keys *A, C, R, K, I, M* and *P*. The selection of the keys is made in such a way that three of them are located on one side of the keyboard and the other four on the other.

As explained in section 3.3.1 for our approach to the keyboard emanations attack we defined three types of features, FFT, normalized FFT and the amplitudes for the time-varying waves. At the same time the distance metric approach is implemented using two differently defined metrics, the Euclidian and Manhattan metric. Hence, we use all three feature sets and both types of metrics in order to be able to make comparison among them. First we create templates for the selected keys, *A, C, R, K, I, M* and *P,* based on 3 feature sets, so we have in total 3 templates for each key. Besides, we use 2 distance metric techniques in order to compare their performance. Hence the total number of scenarios for which experiments are performed is six or two for each feature. They can be identified as the following.

1. Feature set of FFT and Manhattan distance metric
2. Feature set of Normalized FFT and Manhattan distance metric
3. Feature set of Amplitude and Manhattan distance
4. Feature set of FFT and Euclidean distance metric
5. Feature set of Normalized FFT and Euclidean distance metric
6. Feature set of Amplitude and Euclidean distance metric

For each of the six scenarios the recognition phase needs to be performed. Instead of performing the attack with having the user type words, the experiments are performed by the user typing a single key. In this process the keystroke being typed by the user is taken as a test input, assuming the key typed is one for which templates are available. For example if the keystroke is from the key *M*, the three features extracted from the sound recorded are compared to the three features for all seven keys for which the templates exist using Euclidian and Manhattan metric. One distance metric value for each template and each metric is obtained and used to test the accuracy.

The setup presented in Figure 11 is used for gathering data necessary to create the templates for the keys.

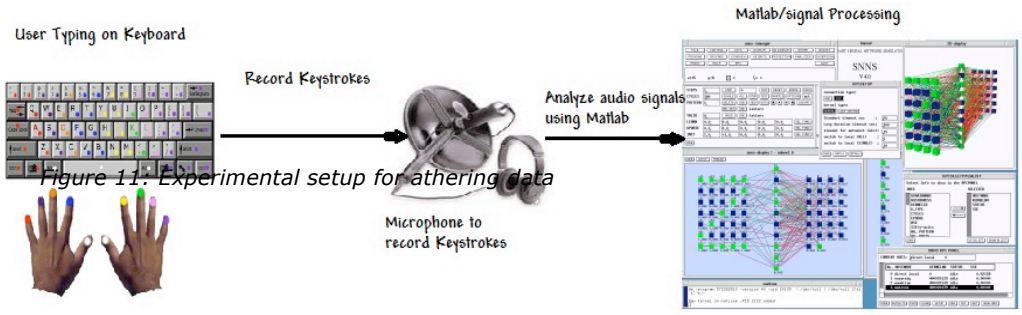


Figure 11: Experimental setup for athering data

The setup shows that it is first necessary to go through the training phase. We first train the system with the three feature set (FFT, normalized FFT and amplitude) of the key sound for the seven selected keys to create a template of features set for each key. The training is conducted through gathering data for each key using the setup presented. Particular keys are pressed a number of times to generate the sounds. In our experiments 50 clicks were performed for each key. The acoustic waves are intercepted by the microphone and stored in a file on the computer. The WinFF and Matlab software are used respectively, to transform the files in appropriate format, and to extract the features for each pressed key.

As mentioned previously limited amount of data is gathered since only the templates for the seven keys, *A, C, R, K, I, M* and *P* are created. Here, only the data for the keys *A, P, C* and *R* are presented.

The graphs on Figure 12 show the FFT of 4 clicks of keys A and P. The left graph shows plots of key A, and the right one of P. If these plots are observed carefully four curves can be noticed for each key. Among the four, two curves are similar for key A. All four curves for key P are similar. This means that each individual key has almost the same FFT values. This was shown by tests on 50 to 100 clicks of a key for further analysis.

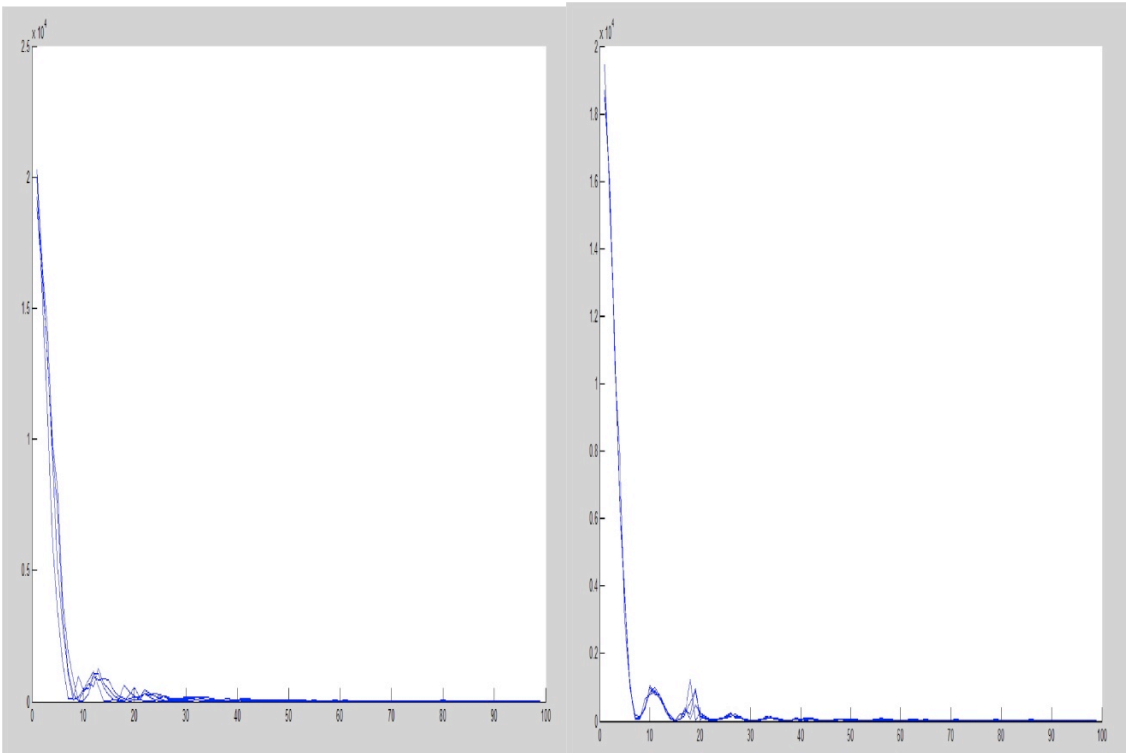


Figure 12: FFT features of a key signal for key A and key P

The graph on Figure 13 shows the amplitude plots of the same keys, *A* and *P*. In the same way as with FFT, tests with 4 clicks of key *A* and key *P* are performed. The same observations made for FFT can be repeated here. Two clicks of key *A* are showing almost the same amplitude values, and four clicks of *P* show the same amplitudes.

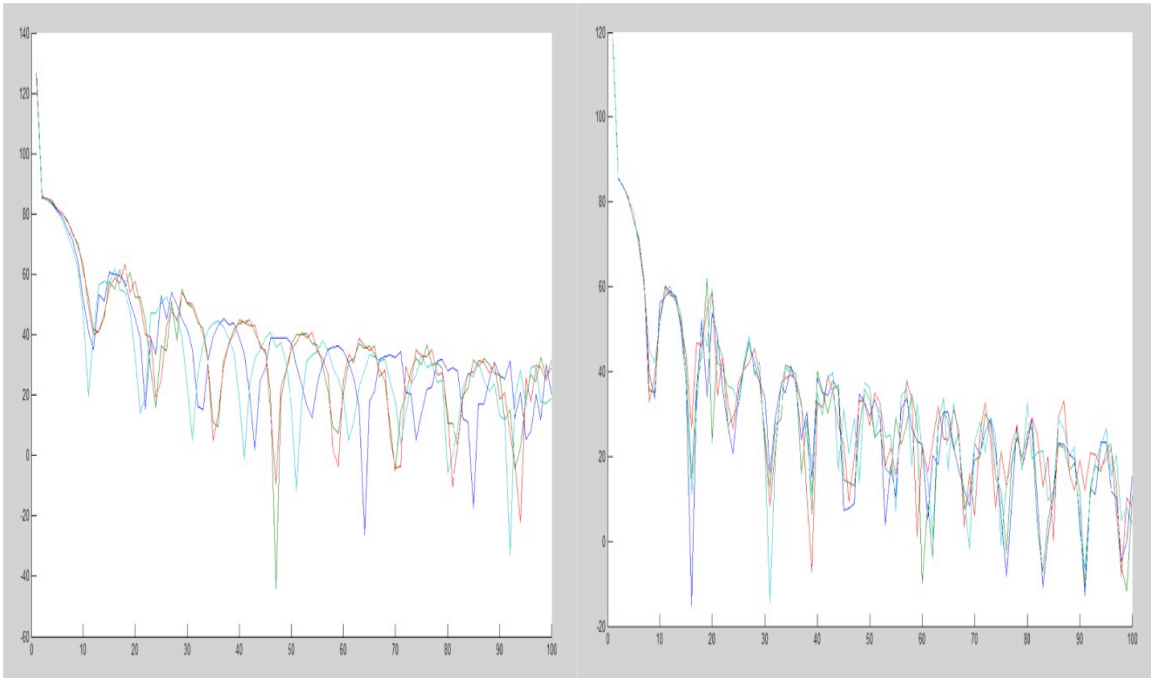


Figure 13: Amplitudes of four clicks for keys A and P

On Figure 14 the normalized FFT for the same keys are presented. Similar observations can be made for these plots as for those on Figure 12 and 13.

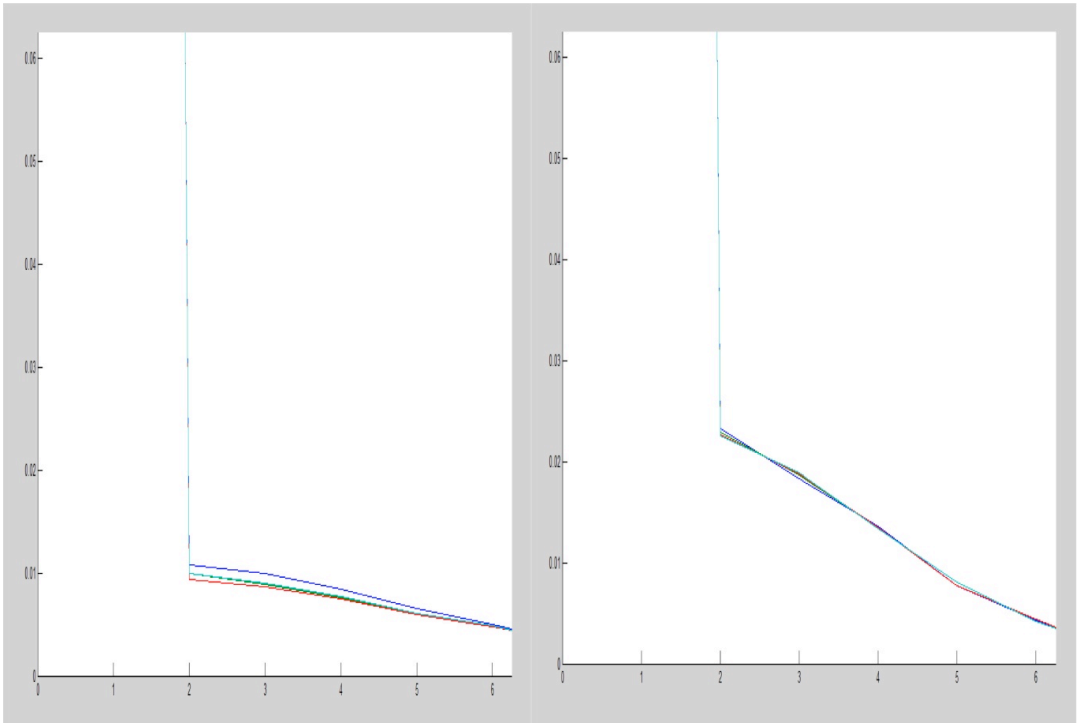


Figure 14: Normalized FFT components for keys A and P

In the same way the FFT, normalized FFT and amplitudes of the time-varying waveform for the keys *C* and *R* are obtained. Figure 15 shows the FFT values of keys *C* and *R*. The graph for *C* shows that the four clicks have similar FFT values. However, on the graph for the key R significant differences can be observed.

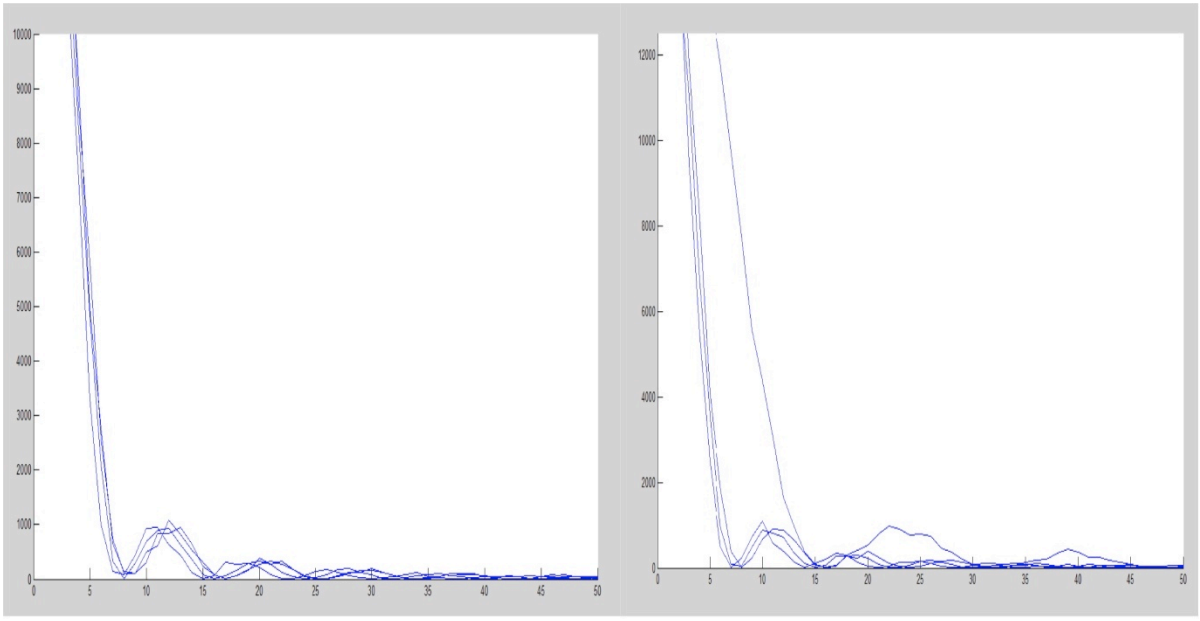


Figure 15: FFT components of 4 clicks for keys C and R

In the same way as in Figure 15, the plot for key C on Figure 16 shows more or less similar values for the amplitudes of the time-varying wave, where as the one for key R shows one curve that is different from remaining the three remaining curves.

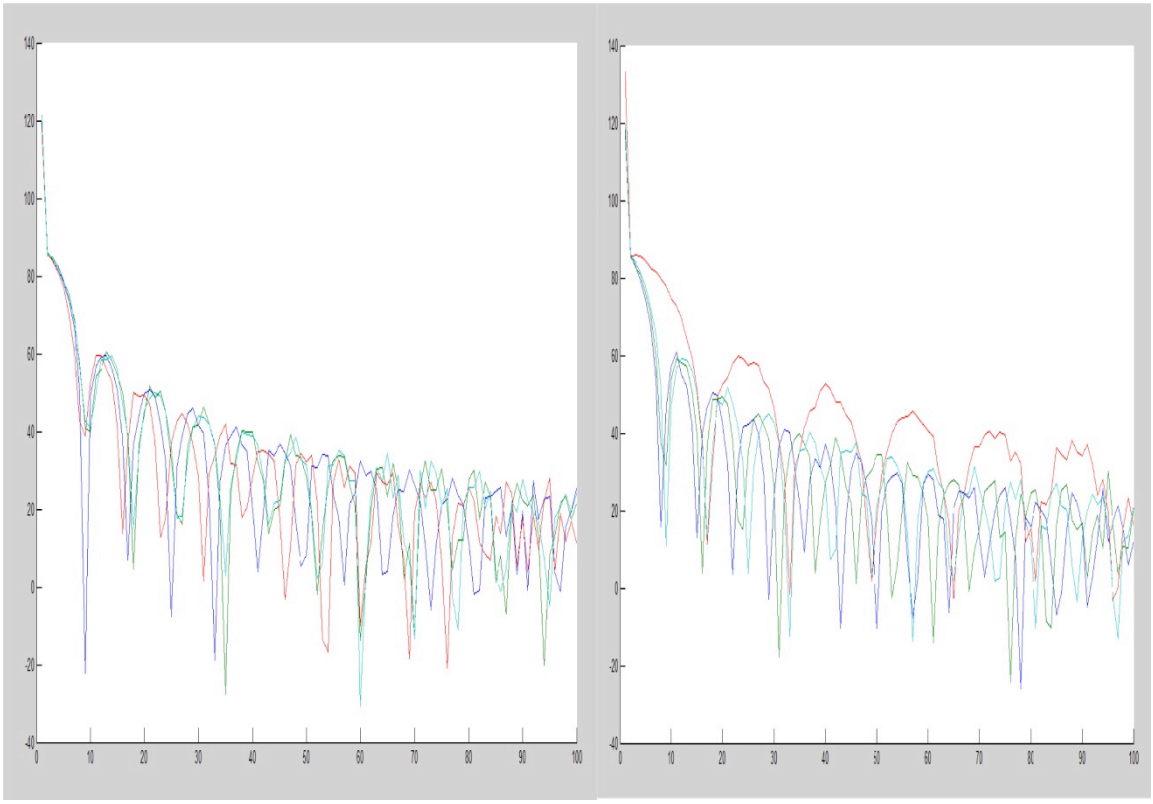


Figure 16: Amplitude components of 4 clicks of keys C and R

Figure 17 shows the plotting of normalized FFT values for four clicks of key *C* and *R*. Key *C* has more or less same normalized FFT values for four clicks, but key *R* has variation in one click.

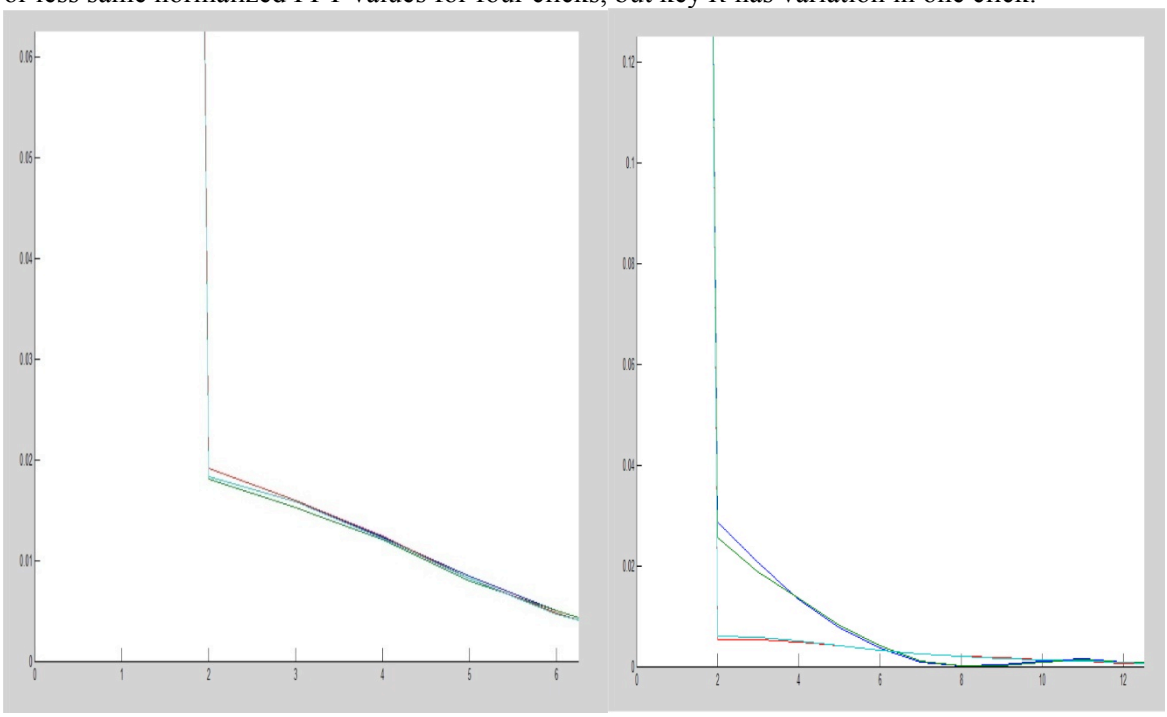


Figure 17: Normalized FFT components of 4 clicks for keys R and C

The data gathered show that based on the experimental analysis we can say that each key on the keyboard gives unique features most of the time when user is typing. Sometimes we may get variations in the features. The differences are usually due to the various typing styles or pressure on the key or may be the environmental noise.

We derive FFT values of each key typed 50 times by considering 100 sample periods. Fourier theory explains that each complex periodic waveform can be decomposed into set of sinusoids with varying phases, amplitudes and frequencies in the Fourier analysis process. The result of this Fourier analysis would be set of frequencies, phases, and amplitudes for each sinusoid. In our analysis, it gives 1*100 vectors of FFT values for a single click. The following code snippet first calculates the FFT of the signal for 100 sample periods contained in the original wave signal and then the normalized FFT.

```
% Calculates FFT
times=(0:length(tick)-1)/fs;
user = tick;
audiofft = fft (user);
audiofft = audiofft.*conj (audiofft);
audiof = audiofft (1:100);
%Calculates normalized FFT
audiofn = audiof/sqrt(sum (abs (audiof).^2));
```

In a similar way amplitude and normalized FFT are derived and stored for each key. Table 1 shows the characteristics of the files Matlab software creates as templates for each feature set of a particular key.

Name of the file	Size	Bytes	Class
Audiof	1x100	800	Double array
Audiofn	1x100	800	Double array
Amplitude	1x100	800	Double array

Table 1: Files for the feature set of individual key

Each key has the same set of features, FFT, normalized FFT, and amplitude for the acoustic signal for the key. In the final step of feature extraction we exported these entire feature set values into one Excel sheet with the name *Features.xls*.

This Excel sheet is acting as a database. We can use this database for further processing to train and classify the system. At this point the feature extraction is over. The output of the feature extraction is "Vector of feature set".

Once the features are extracted it is necessary to gather data showing the similarity between the features for the keystroke recorded and the templates available.

3.3.4 Phase IV: Data analysis, performance selection

The data analysis in our experiments refers to comparing the sound waves of the keystroke recorded during the attack (test input) with the templates derived for the seven keys. In order to find out which of the six scenarios created gives the best performance for the keystroke to be recognized, 50 clicks are used for each key in the training phase. 7 templates, one for each of the seven keys are made and 7*50 =350 tests are performed to create the templates. Comparisons are made using the distance metric approach as explained in section 3.3.2.

The tables presented below show the distance metric values for the six scenarios defined in 3.3.3. Each table contains the test input that is the keystroke recorded during the attack, and the distance metric for all seven keys for which the templates are created. The expectations are the test input key to have the lowest distant metric when the template of the same key is used. However, the experiments show that this is not always the case. Therefore, comments below each table are provided.

Scenario 1: Feature set of FFT, and Manhattan distance metric

Two keys, *P* and *C*, are used as test inputs. Tables 2 and 3 present the values obtained for the distance metrics.

Number	Test input	Template	Distance Metric Value
1	P	R	2.8434e+05
2	P	A	1.3323e+06
3	P	C	2.5686e+05
4	P	I	2.1465e+05
5	P	K	1.4049e+05
6	P	M	6.5539e+05
7	P	P	1.6256e+05

Table 2: Comparison of test input P with 7 templates (Scenario 1)

Table 2 shows that the distance metric value for K is lower than for P that is opposite of the expectations.

Number	Test input	Template	Distance metric value
1	C	R	2.6914e+05
2	C	A	1.3171e+06
3	C	C	2.4156e+05
4	C	I	3.3889e+04
5	C	K	1.2526e+05
6	C	M	6.4012e+05
7	C	P	1.4733e+05

Table 3: Comparison of test input C with 7 templates (Scenario 1)

Table 3 shows that the distance metric value for I is lower than for C that is opposite of the expectations.

Scenario 2: Feature set of Normalized FFT and Manhattan distance metric

Three keys, C, P and *H*, are used as test inputs. Tables 4, 5 and 6 present the values obtained for the distance metrics.

Number	Test input	Template	Distance Metric value
1	C	R	0.0161
2	C	A	0.0236
3	C	C	0.0108
4	C	I	0.0238
5	C	K	0.0236
6	C	M	0.0123
7	C	P	0.0273

Table 4: Comparison for test input C with 7 templates (Scenario 2)

Table 4 shows that the lowest distance metric value for C is with the same key which complies with the expectations.

Number	Test input	Template	Distance Metric Value
1	P	A	0.0534
2	P	C	0.0247
3	P	I	0.0193
4	P	K	0.0193
5	P	M	0.0442
6	P	P	0.0158
7	P	R	0.0357

Table 5: Comparison for test input P with 7 templates (Scenario 2)

Table 5 shows that the lowest distance metric value for P is for the same key which complies with the expectations.

Number	Test input	Template	Distance Metric
1	K	R	0.0225
2	K	A	0.0403
3	K	C	0.0090
4	K	I	0.0098
5	K	K	0.0088
6	K	M	0.0309
7	K	P	0.0085

Table 6: Comparison for test input K with 7 templates (Scenario 2)

Table 6 shows that the lowest distance metric value for K is with the same key which complies with the expectations.

Scenario 3: Feature set of Amplitude and Manhattan distance metric

Four keys, *P, C, K* and *I*, are used as test inputs. Tables 7, 8, 9 and 10 present the values obtained for the distance metrics.

Number	Test Input	Template	Distance Metric value
1	P	A	911.5163
2	P	C	849.3066
3	P	I	772.8619
4	P	K	765.1158
5	P	M	851.8761
6	P	P	725.82
7	P	R	774.6665

Table 7: Comparison for test input P with 7 templates (Scenario 3)

Table 7 shows that the lowest distance metric value for *P* is with the same key which complies with the expectations. Note, that several other keys have distance metrics very close to this value.

Number	Test input	Template	Distance metric value
1	C	R	758.6771
2	C	A	784.5345
3	C	C	600.6520
4	C	I	803.2974
5	C	K	721.7895
6	C	P	660.7806
7	C	R	758.6771

Table 8: Comparison for test input C with 7 templates (Scenario 3)

Table 8 shows that the lowest distance metric value for *C* is with the same key which complies with the expectations.

Number	Test input	Template	Distance Metric value
1	K	R	784.8390
2	K	A	833.9128
3	K	C	763.0578
4	K	I	789.6526
5	K	K	737.5174
6	K	M	729.2329
7	K	P	753.9850

Table 9: Comparison for test input K with 7 templates (Scenario 3)

Table 9 unexpectedly shows lower distance metric value with template M. There is slight difference between lowest value and second lowest value.

Number	Test input	Template	Distance Metric value
1	I	M	925.1257
2	I	A	966.9736
3	I	C	908.9795
4	I	I	827.1056
5	I	K	855.4632
6	I	P	914.4068
7	I	R	847.7117

Table 10: Comparison for test input I with 7 templates (Scenario 3)

Table 10 shows that the lowest distance metric value for *I* is with the same key which complies with the expectations.

Scenario 4: Feature set of FFT and Euclidean distance metric

Two keys, *C* and *P*, are used as test inputs. Tables 11 and 12 present the values obtained for the distance metrics.

Number	Test input	Template	Distance metric
1	C	A	1.4088e+06
2	C	C	3.5638e+05
3	C	I	9.8675e+04
4	C	K	2.4387e+05
5	C	M	7.4605e+05
6	C	P	2.6519e+05
7	C	R	3.8357e+05

Table 11: Comparison for test input C with 7 templates (Scenario 4)

Table 11 shows that the distance metric value for *I* and *P* is lower than that for *C* that is opposite of the expectations.

Number	Test input	Template	Distance Metric
1	P	R	1.7712e+05
2	P	A	8.4820e+05
3	P	C	2.0431e+05
4	P	I	4.6202e+05
5	P	K	3.1681e+05
6	P	M	1.8542e+05
7	P	P	1.6549e+05

Table 12: Comparison for test input P with 7 templates (Scenario 4)

Table 12 shows that the lowest distance metric value for *P* is with the same key which complies with the expectations.

Scenario 5: Feature set of Normalized FFT and Euclidean distance metric
Three keys, *R, P* and *C*, are used as test inputs. Tables 13, 14 and 15 present the values obtained for the distance metrics.

Number	Test input	Template	Distance metric
1	R	A	0.0233
2	R	C	0.0098
3	R	I	0.0070
4	R	K	0.0070
5	R	M	0.0190
6	R	P	0.0058
7	R	R	0.0152

Table 13: Comparison for test input R with 7 templates (Scenario 5)

Table 13 shows that values for *I, C, P* and *K* are lower than that for *R.* It is opposite of the expectations.

Number	Test input	Template	Distance metric
1	P	R	0.0151
2	P	A	0.0233
3	P	C	0.0096
4	P	I	0.0070
5	P	K	0.0070
6	P	M	0.0189
7	P	P	0.0057

Table 14: Comparison for test input P with 7 templates (Scenario 5)

Table 14 shows that the lowest distance metric value for *P* is with the same key which complies with the expectations.

Number	Test input	Template	Distance metric
1	C	R	0.0086
2	C	A	0.0168
3	C	C	0.0029
4	C	I	0.0019
5	C	K	0.0016
6	C	M	0.0123
7	C	P	0.0017

Table 15: Comparison between test input C with 7 templates (Scenario 5)

Table 15 shows that the distance metric value for *I, P* and *K* is lower than that for *C* that is opposite of the expectations.

Scenario 6: Feature set of Amplitude and Euclidean distance metric

Three keys, *C, P* and *I*, are used as test inputs. Tables 16, 17 and 18 present the values obtained for the distance metrics.

Number	Test input	Template	Distance metric
1	C	R	137.0586
2	C	A	131.6961
3	C	C	131.9620
4	C	K	142.1621
5	C	M	141.9618
6	C	P	138.4848
7	C	I	144.4036

Table 16: Comparison for test input C with 7 templates (Scenario 6)

Table 16 shows that the distance metric value for *A* is lower than that for *C* that is opposite of the expectations. However, a very slight difference is observed.

Number	Test input	Template	Distance metric
1	P	I	120.0736
2	P	A	85.8151
3	P	C	104.9217
4	P	K	106.5904
5	P	M	97.6959
6	P	P	84.4479
7	P	R	103.0073

Table 17: Comparison with test input P with 7 templates (Scenario 6)

Table 17 shows that the lowest distance metric value for *P* is with the same key which complies with the expectations.

Number	Test input	Template	Distance metric
1	I	R	122.2455
2	I	A	138.7008
3	I	C	122.5747
4	I	I	112.5416
5	I	K	116.0057
6	I	M	124.1698
7	I	P	112.5639

Table 18: Comparison for test input I with 7 templates (Scenario 6)

Table 18 shows that the distance metric value for I is with the same key which complies with the expectations. However, the distance metric value for *P* is very close to that for *I*.

From the data analysis performed above, it can be observed that in some cases the comparisons give satisfactory values as per our expectations. However, among it was very difficult to select the scenario that gives really good performance among the six scenarios. Similar results are obtained for all six scenarios. Even though some values might be irrelevant because of the different typing style, the selection of the scenario with the best performance cannot be made based on the tables presented. Therefore, a novel procedure for selecting the scenario with the best performance based on the expected error rate is suggested. This procedure as well as the data obtained is presented as a part of the next chapter.

3.3.5 Phase V: Presenting results

The main results from this study are presented in the next chapter.

3.3.6 Phase VI: Discussion, conclusion

Discussion of the results is given in Chapter 5. The concluding remarks as well as the directions for the future research are shown in Chapter 6.

3.4 Ethical considerations

All the experiments within this study are performed by the author of this thesis. Hence, there are no ethical issues to be considered with respect to the experiments. However, it is necessary to address the ethical issues regarding the information contained in this report.

Publishing information on any kind of security attack in which sensitive information typed on the keyboard can be retrieved can be in some circumstances considered as unethical. This is because the description on how the attack can be performed may be misused. The intention of this research is not to help attackers in performing the keyboard acoustic emanations attacks. On the contrary, the knowledge gathered with this research is supposed to raise the awareness about these types of attacks among the users and the security community and increase the knowledge on possible measures to be taken to avoid these attacks. The solely interest of the author is to show the threat of these types of attacks and to suggest protective measures for the users.

4 Results

This chapter presents the main results from this research. In the first section the waveform for the keyboard acoustic emanation obtained in a real scenario is presented. The second part shows the structure of the keyboard acoustic emanations attack performed. This is followed by the novel procedure introduced to measure performance for the six scenarios.

4.1 Waveform obtained from the real scenario

Previous research (Asonov & Agrawal, 2004), showed that the acoustic signal emitted when pressing a key on the keyboard has two distinct peaks corresponding to pushing the key and releasing the key. This finding was for the laboratory environment without considering the noise. Our experiments were performed in environment with moderate noise. The extracted waveform from the noisy environment has the same shape as the one obtained in laboratory environment. The time waveform and its corresponding frequency domain plot obtained for an example key during our experiments are shown in Figure 18.

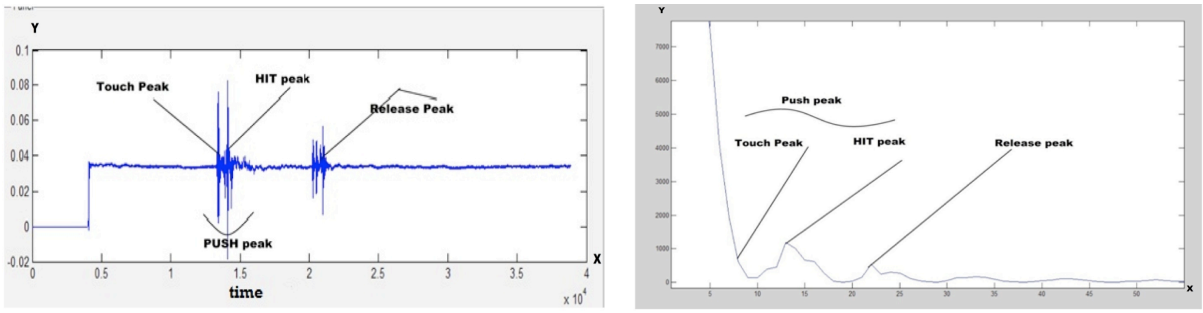


Figure 18: Push and release peaks of a key signal in time and frequency domain

The time domain graph on the left side presents the higher peak of the signal when the key is pressed, and the lower peak when the key is released. The frequency domain graph on the right side displays the strength of the signal. It shows the power level at each frequency.

4.2 Structure of the attack

A novel approach for conducting the acoustic emanations attack was created. Some of the procedures used by previous researchers as described in section 3.3.3 are used, while others are selected having in mind the aim with this study. The structure of the attack using this novel approach is presented on the figure 19. It presents the training phase (in the upper part of the figure), and the recognition part (in the lower part of the figure).

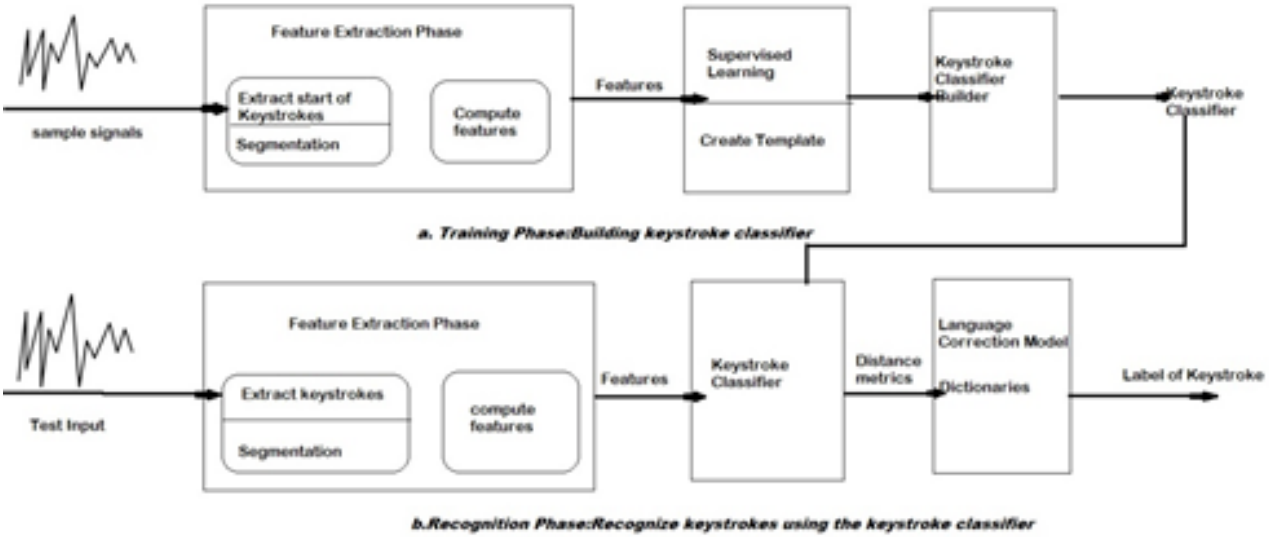


Figure 19: Structure of the attack performed for this research

The recognition phase of our attack includes the Language Correction Module. It is necessary for improving the accuracy in case English words are used as passwords. For example, assume the user has typed the word "PAPER". The analysis on these 5 unknown time-varying acoustic signals will most likely retrieve the keystroke and the typed word. Instead the words such as "LAPER" which is not in the dictionary might be obtained. It is very likely the first letter retrieved to be "L" instead of "P". In some cases the third letter is likely to be "T" instead of "P". Based on this analysis it can be concluded that the typed word is "LATER" or "PAPER".

Sometimes we can get inappropriate probabilities for the letters hence the best way is to use online dictionaries. By integrating online dictionaries into the system, we can find the mismatch of letters in the word. The system reads word by word from the dictionary by comparing known most likely letters. So we would get the existed word in the dictionary which has been typed by the user.

In our experiment we use 7 key templates which created in training phase. We have formed the words with the combination of these 7 keys. Tests on "PACK" and "PICK" and also "MAC" and "MAP" were made to test the role of the dictionary in retrieving the correct word.

4.3 Performance measuring procedure

Tables 2 to 18 presented in section 3.3.4 show similar results for almost all six scenarios. Therefore, none of them could be claimed to be the best of all. This required development of a new procedure that will be able to correctly identify which scenario will produce the highest accuracy in identifying the typed keystrokes. The text below presents this procedure.

The first step is to divide the total distance scores into two sets, one of genuine scores, and other one of imposter score. The definitions of these terms are the following.

Genuine score: The distance metric value for the test input and the template for the same key. This can be thought of to belong to the genuine scores. For example, genuine score of *A* would be:

A=distance (template of A with test input of A)

Imposter scores: This is the distance metric value for the test input and the templates for keys different than the test input. For example, imposter score of A would be:

A=distance (template of A with test input of M)

The second step is to group all genuine scores into one vector and all imposter scores into another big vector. The experiments are done with 7 keys, therefore the genuine vector is of length 350 (50 clicks are performed for each key), and the imposter vector is of length 2100 (6*7*50).

After grouping the distance metrics into clusters, we sorted both of these two vectors using below mathematical equation.

Gen=sort (genuine vector)

Imp=sort (imposter vector)

These sorted vectors are then merged into one big vector, which is of length 2450.That merged vector is the combination of both genuine vector of length 350 and imposter vector of length 2100.

Merge vector SC=(Genuine vector of length 350+ Imposter vector of length 2100)

SC=(sc1, sc2, ……sc2450)

In the next phase we calculate threshold values for all the scores in the merge vector in order to find the performance of the system over different threshold values.

For the performance evaluation of the system, we need to set the threshold values to find the error rates. The distance scores which are above the threshold are genuine scores, and below the threshold are imposter scores.

We have total of 2451 threshold values for total of 2450 distance scores.

Threshold T1=Sc1-1(This is the smaller value than remaining scores)

Ti=((Sci+Sci+1)/2)

T2451=(Sc2450+1)(larger value among all scores)

From the above equations we can say that no Threshold value will be equal to genuine or imposter scores. Using these threshold values we have calculated FMR (False Match Rate), FNMR (False Non Match Rate) values. The definitions of these rates are given below

False Match Rate: False match rate can be described as probability of the system to incorrectly match the input pattern to another template in the database. This error rate calculates the percentage of incorrect inputs which are correctly accepted. FMR occurs when classification algorithm classifies an imposter comparison value as genuine. FMR of a particular value I can be calculated as a number of imposter comparisons with score higher than I divided by the total number of imposter comparisons.

False Non Match Rate: False non match rate is the probability that the system fails to detect a match between the input pattern and a matching template in the database. FNMR occurs when genuine value is classified as imposter value. FNMR can be calculated as number of genuine comparisons with score lower than I divided by the total number of genuine comparisons.

Figure 20 illustrates how we can visualize the evaluation of the performance of the system.

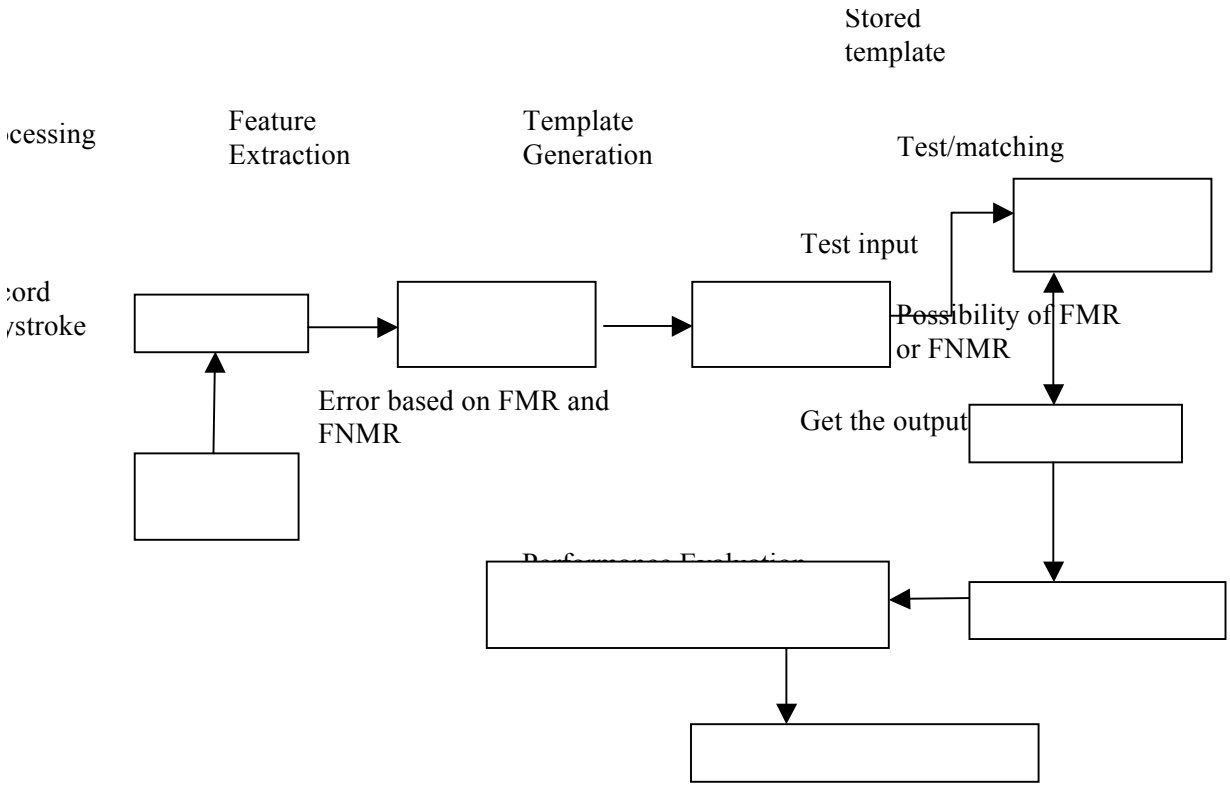


Figure 20; Procedure for evaluating the performance of the system

In our procedure, we introduce FNM and FM related to FNMR and FMR respectively. FNM presents false non match distances and FM false match distances. Then for FNMR and FMR we obtain the following.

FNMR=number of FNM (false non match distances) distances above threshold/360

FMR=number of FM (false match distances) below threshold/2100

For each *i=1 to 2451* , we calculate

$$FNMR_i = \#(GEN_j > T_i) / 350$$

$$FMR_i = \#(IMP_j < T_i) / 1200$$

The result is the following.

$$FNMR = (FNMR1, FNMR2, ..., FNMR2451)$$

$$FMR = (FMR1, FMR2, ..., FMR2451)$$

Finally we calculate the performance for each scenario by calculating the Equal Error Rate(EER) using FMR and FNMR.

EER is calculated as $EER = (FMR_i+FNMR_i+FMR_i+1+FNMR_i+1)/4$

If $FNMR_j = FMR_j$ then

$EER = FNMR_j = FMR_j$

The process of calculating EER for the six scenarios involves sequential and middle search, as described below.

Start = 1 Stop = 2451

Middle = floor ((Start+Stop)/2)

If FNMRMiddle > FMRMiddle then Start=Middle

If FNMRMiddle < FMRMiddle then Stop=Middle

Repeat until Stop=Start+1

Finally the value of EER for each scenario was obtained and it can be used as a measure of the performance of the scenario. This means that based on the value for EER the most preferable feature set to be extracted and the best distance metric approach selected.

The values for EER for the different feature sets and different distance metrics are given in Table 19. Unfortunately the values for all six scenarios are lower than 50 %. The lowest EER value among all of them is 0.4523 or 45.23%. It is obtained for scenario number 3 where the feature set consists of the amplitudes in the time-varying signal and Manhattan distance metric is used in the distance metric approach.

Template	Distance metric	EER value
Amplitude	Manhatten	0.4523
Normalized FFT	Manhatten	0.4970
FFT	Manhatten	0.4799
Amplitude	Euclidean	0.4656
Normalized FFT	Euclidean	0.4868
FFT	Euclidean	0.4808

Table 19: Performance Evaluation values of the six scenarios

Additionally, ERR for a case where a combination of all three feature sets and the Manhattan metric are used. The value of this ERR is 0.5902 that is higher than 0.4523. Hence, the combination of the features does not guarantee better performance.

4.4 Comparative analysis

This section compares the attack performed for the purpose of our study with the experiments of other researchers. The following attributes of the attacks described by various researchers presented in section 2.3 are identified: cost, difficulty and accuracy. The cost is related to the equipment used to perform the attack. According to this attribute, the attack can be considered as cheap or expensive. The difficulty corresponds to the technical expertise required by the attacker, especially with respect to the required knowledge in the area of signal processing. The gradation we use is from very difficult to very easy. The accuracy refers to the precision with which the typed keystrokes are recovered from the recorded signal. It varies from very high to low.

Table 20 presents the comparative analysis. Based on the above mentioned attributes, the last column of the table answers the following question: Is it possible to perform the attack in reality?

The year and researchers performing the attack	Cost	Difficulty	Accuracy	Possibility of real attack?
2004, Dmitri Asonov and Rakesh Agrawal	Cheap	Difficult	High	No
2005, Li Zhunag, Feng zhou, J.D.Tygar	Cheap	Very difficult	Very High	Yes
2006,Yigael Berger and Avishai Wool	Cheap	Difficult	High (not applicable for strong passwords)	Yes

2006, Hui Shunpak	Cheap	Moderately easy	Low (not applicable for all scenarios)	Yes
2010, Nitesh Saxena and Tzipora Halevi	Cheap	Difficult	Average(only for 6 character word)	Yes
2011, Sravanthi Ponam	Cheap	Easy	Moderately high	Yes

Table 20: Comparison of the attack performed within this study with other research

4.5 Protecting users

Although it is very difficult to defend against the keyboard acoustic emanations attack, some defending methods still exist.
Suitable defending methods found within this research are the following.

- Increasing the environmental noise levels while typing on the keyboard. Our experimental results show that when noise is higher than some moderate level it affects the recognition rate. Hence when there is more noise within the signal the attacker will experience difficulties in recognizing the recorded keystroke.
- Using virtual keyboard which are newly evolved in the market is another way of protection. This virtual keyboard does not produce any sound. Nowadays because of technological advancements people are more and more accepting the touch screen technology used for the pads and smart phones. This means that everyone can work anywhere without the need of chair, table and desktop. If a person is in the crowd and using touch screens, it is impossible to attack the user using acoustic emanations.

Another method suggested by Snyder (Snyder, 2011) is by modifying the signals to fool the hacker. The signal can be modified by changing the perceived release peak and push peak times. Then the attacker program would fail to understand which key has been typed.

5 Discussion

5.1 Critical overview

Our empirical study of possible ways in using side channel information to attack a system is inspired by the area of keystroke dynamics and previous research on keyboard acoustic emanations attacks. The specific of these types of attacks is that the attacker does not need to interact with the system of the user. We studied and learned the techniques used in keyboard acoustics and other biometric techniques, such as speech recognition. The seriousness of the attack is estimated based on the analysis of the experiments presented in the literature and our own experiments.

Unlike the previous research that used completely silent laboratories, our experiments are performed in a room with moderate noise. Noise plays important role in this type of acoustic experimental analysis. It can originate from computer systems of other people, air conditioning or other sources. The recordings made during the experiments are performed with few people working on their systems hence we consider our attack to take place in a nearly real environment. We show that the noise can be removed and our extracted waveforms are not too much different than those of sounds emanating from the user's keyboard in completely silent environment.

Another objective of our research is to show that simple equipment and signal processing methods can be used for the keyboard acoustic emanations attack. Artificial intelligence techniques which are used by the previous research are avoided due to the fact that they are considered as complicated and difficult to use. Much simpler signal processing techniques are suggested and several scenarios created. Statistical methods are used to analyse the scenarios and the performance. By doing so many experiments, we have shown that keyboard acoustic emanations are dangerous security threat.

We could not achieve 100% recognition rate. Our rate was 50 to 60% recognition that is not very low having in mind the use of simple statistical methods, normal microphone, and university environment. We believe that the main reasons for not getting higher recognition rate are the following.

- Environmental noise,
- Air sound,
- Noise within the keystroke, because of surrounding keys,
- Typing pattern,
- Pressure applied on the key,
- Manufacturing changes in the keyboard models,
- Use of touch screens and Virtual keyboards.

5.2Weak points in our research

Previous research as well as our observations shows that the sound emitted from the keystroke depends on the manufacturing model of the keyboard. Different manufacturing models have different construction method. So if we test on Dell keyboards and HP keyboards, both will give different results. We cannot compare the template of Dell keyboard with test input of keyboard HP because we will get irrelevant values. Similarly, laptop keyboards and ATM key pads each has its own construction pattern and sound emanations. The structure of our attack required the training phase which needs to be performed on the same keyboard/keypad on which the attack is to be performed.

Variation in the area where the user hits the key is a concern in the previous research. The keys on the different position of the plate produce different sounds. However, even the same key if presses in the different area may produce slightly different waveform. Users do not hit the keys always in the same area. Hence, this can sometimes affect the accuracy of the recognition phase.

In our study a limited number of keys were used .Any of the thirty keys on the keyboard could be a part of the sensitive information that needs to be recorded for our experiments. Our approach can find the exact word, only if it is an English word. We can say that, it is difficult to retrieve other language words using this approach. Strongly typed password with randomized keys like punctuations and numbers cannot be retrieved.

Typing pattern is a biometric feature that can affect the keyboard dynamics. If a person uses the same typing rhythm every time then the probability of getting the same sound every time is high. However, in case the user changes the body position or the typing is performed with one hand due to the other one being injured the typing pattern changes. These changes have impact on the total performance of the attack.

6 Concluding remarks

6.1 Conclusions

We have successfully performed the keyboard acoustic emanations attack which uncovers the confidential information being typed by the user from audio recordings of keyboard clicks. The performance and efficiency of our attack mainly depends on how well we can develop the attacking tool, which is the classifier in our case by extracting real features of the individual keys.

We successfully achieved 50 to 60% recognition rate using our own methodologies and statistical methods. We believe that this attack can be useful as a keyboard acoustic password cracker with some limitations. Since our approach only works for English dictionary words it is difficult to crack strong and random passwords. These days every system is forcing the user to choose random passwords for strong authentication, in that case we can´t achieve the results with this system.

In order to improve the efficacy of our research, we need to extract some more relevant features of the keys, and need to develop the tool using digits, shift keys, and punctuations. Then we may crack the strong passwords as well.

The hurdles attacker should face, in order to perform the attack are getting the high quality clear audio recording of key clicks for the analysis, and training the recognizer with appropriate key model. This attack is very spectacular, since ordinary person also can perform attack, because he can get all experimental tools from the internet or shops.

We believe that, through this project we have developed the awareness for the public about keyboard security. If people know about the possible attacking methods, they could protect themselves from the hackers.

6.2 Future work

Nowadays, attackers are finding ways to compromise secured system. Keyboard acoustic emanations attacks may easily compromise the users' security. These types of attacks could be performed not only on computer keyboards , different kinds of keypads are also prone to the acoustic attacks.

One way of extending our research is to perform experiments with creating templates for all letter keys on the keyboard and use keyboards from several manufacturers. Including an ATM or any other type of keypad in the investigation will show these attacks are possible on different kind of input devices.

In order to improve this system, more advanced features like cepstrum features and MFCC (Mel Frequency Spectral coefficients) could be added.

References

Anon., 2010. *iOS, improved Outcomes software.* [Online]
Available at:
http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/Clustering_Parameters/Manhattan_Distance_Metric.htm
[Accessed March 2011].
Anon., n.d. s.l.:s.n.
Anon., n.d. *FFT Tutorial.* [Art] (University of Rhode Island Department of Electrical and Computer Engineering).
Anonymous, 2010. *Biometric-Solutions.com.* [Online]
Available at: http://biometric-solutions.com/
[Accessed February 2011].
Asonov, D. & Agrawal, R., 2004. *Keyboard Acoustic Emanations.* s.l., IEEE, pp. 3-11.
Audri & Lanford, J., 2008. *ATM Theft: 8 Tips to Protect Yourself From the 5 Most Common ATM Scams,* NC: Internet ScamBusters.
Berger, Y., Wool, A. & Yeredor, A., 2006. *Dictionary Attacks Using Keyboard Acoustic Emanations,* ISRAEL: Tel Aviv University.
Bishop, M., 2003. *Computer Security.* Boston MA 02116: Pearson Education, Inc..
Briol, R. L., 1991. *Emanation: How to Keep Your Data Confidential.* s.l., s.n.
Crawford, H., Aug. 2010. *Privacy Security and Trust.* Glasgow, UK , Conference Publications , pp. 205-212.
Facundo, M., 2011. Distances and the Metric approach to object mapping. *Foundations of computtaional Mathematics.*
Frankland, R., 2011. *Side Channels, Compromising Emanations and Surveillance,* England: Royal Holloway University of London.
Gait, J., 1978. Easy entry: the password encryption problem.. *SIGOPS Oper. Syst. Rev. 12, 3,* July.pp. 54-60.
Gross, B. J. & Rosson, B. M., 2007. *Looking for teouble: Understanding User-Security Management.* Cambridge, Massachusetts, ACM, New York, NY, USA.
Halevi, T. & Saxena, N., 2006. *A Closer Look at Keyboard Acoustic Emanations:,* s.l.: iace.org.
Jin, S. W. a. R., 2008. *An information geometry approach for distance metric learning,* Michigan: Dept. of Computer Science and Engineering Michigan State University.
Khun, G. M., 2003. *Compromising Emanations: Eavesdropping Risks of Computer Displays,* Cambridge, UK: University of Cambridge, Computer Laboratory.
Levy, D., 2006. *Qualitative Methodology & Grounded Research,* s.l.: University of Auckland.
Mary, S., 1998. *Strategies for Empirical Research in Writing.* s.l.:The University of Memphis.
Olzak, T., 2006. Keystroke Dynamics: Low Impact Biometric Verification. 10.
Olzak, T., 2008. Keystroke Logging (Key Logging). *Adventures in Security,* p. 13.
pak, H. s., 2006. *keyboard acoustic Triangular Atatck,* s.l.: s.n.
Pak, H. S., 2006. *keyboard acoustic Triangular Atatck,* s.l.: s.n.
Quisquater, J.-J., 2002 October. *State of the Art Regarding Side channel Attacks,* Japan: s.n.
Quisquater, J.-J. & RiZK, M., 2002. Side channel attacks. *state of the art- Scientific Report,* p. 50.
Ross, A., Flynn, P. & Jain, A., 2007. *HandBook of Biometrics.* s.l.:Springer.
Rouse, M., 2010. *keylogger (keystroke logger, key logger, or system monitor),* s.l.: searchmidmarketsecurity.
Sagiroglu, S. & Canbek, G., 2009. Keyloggers. *Technology and Society Magazine*, No. 3 Vol. 28, pp. 10-17.
Shamir, A., 2011. *A Top View of side channel attacks.* Paris, Computer Security.

Snyder, R. W., 2011. An Active Defense From the Attack of a Keyboard-Emanations Classifier Algorithm.. *Robert Wesley Snyder's Posts*.
Syhw, 2011. Two Amusing Side Channel Attacks- Syhw'w Posterous. *independent and identically-distributed random experiments,* pp. 10-12.
Trochim, W. M., 2006. Research Methods Knowledge Base. *Deduction & Induction*, 20 10, p. 150.
Tzipora, H. & Nitesh, S., 2006. *A Closer Look at Keyboard Acoustic Emanations:,* s.l.: iace.org.
Zhuang, L., Zhou, F. & Tygar, J. D., 2005. *Keyboard acoustic emanations revisited.* s.l., ACM, pp. 373-382.

partment of Computer and Systems Sciences
ɔckholm University
rum 100
-164 40 Kista
one: 08 – 16 20 00
w.dsv.su.se

TRITA-ICT-EX-2013:267

www.kth.se