Investigating the Feasibility of Keyboard Acoustic Attacks

Gavin Neale
Department of Computer Science
University of Auckland
gnea005@ec.auckland.ac.nz

Abstract

A keyboard acoustic attack is a relatively new type of attack where an attacker intercepts the sounds emitted by a keyboard as the user types. An attacker can then identify what is being typed by recognising different keys by the slight differences in the keystroke sounds. Although this type of attack has been shown to be possible in a controlled environment, it is not yet clear how effective this attack would be in a real world environment. This paper discusses some different attack scenarios and possible future experiments that could show whether keyboard acoustic attacks are feasible.

1 Introduction

Nearly all computer users type sensitive or confidential information on a keyboard at some point. Often this is user-names and passwords but can potentially also be emails, URLs, documents, instant messaging conversations or any other personal or proprietary information typed in by the user. When ever there is some value in this information to those who do not have it, there will be a motive for an attacker to gain access to it.

Several methods have been devised to intercept this information as it travels from the keyboard to its final destination. A common attack is to install a keystroke logger on the victim's machine. Intercepting network traffic can also reveal this information if it is being transmitted to another location in clear text. Both of these are well known types of attack and as they are happening within a computer system they can be defended against or detected by a security system.

An acoustic attack is a type of attack based on the emanations of computer hardware into the surrounding environment. As a person types on a keyboard the keys emit similar sounds. However, because of the design of most keyboards, there are slight differences in these sounds which can be used to identify

individual keys. This type of attack is not detectable by a security system running on the computer.

Two recent papers have explored the use of keyboard acoustic emanations to discover what is being typed on the keyboard. Asonov and Agrawal, in [1], first showed that it was possible to identify keys from their sounds using a neural network as a recogniser. Zhuang et al, [2], then significantly improved the recognition rate using machine learning techniques and statistical knowledge of the English language.

Although these two papers have shown that it is possible to identify individual keys by their sound, the limited scope of attacks simulated in the experiments raises the question of whether or not a keyboard acoustic attack is a major threat. In order to determine the feasibility of a keyboard acoustic attack we should have some understanding of the type environments in which an attack could take place, the capabilities of the attacker and the ability of a recogniser to identify the target keystrokes.

Section 2 will give a background on the previous research conducted on keyboard acoustic emanations. An overview of the results will be given and an explanation as to why the experiments are not sufficient.

Section 3 will discuss possible real world environments and conditions that may reduce the effectiveness of this attack. This will include some requirements a keystroke recogniser may need, such as the ability to recognise the shift key, in order to successfully identify keystrokes in a real attack. The types of attacker will also be discussed and why they may consider using an acoustic attack.

Section 4 will then propose an extension of the research done by Zhuang et al. This will focus on expanding the scope of the experiment by testing in realistic environments, increasing the number of variables and extensions to the recogniser software.

Section 5 discusses the feasibility of this attack and section 6 will mention ways in which acoustic attacks could be detected or defended against.

2 Background

Previous works done on keyboard acoustic emanations have shown that the small differences in the sound of keystrokes are sufficient for a software recogniser to identify individual keys. This was first shown by Asonov and Agrawal who were able to achieve a recognition rate of around 80%. This

recognition method was based on recognising each keystroke independently, with no knowledge of previous keystrokes. The main problem with their recognition method was that they used labeled training samples to train the recogniser. Each letter on the target keyboard was recorded and given to the recogniser to learn, while knowing what the correct output should be. This means that an attacker would either need to know what the target was typing, or have access to the keyboard to record samples, prior to training the recogniser. Both of these constraints could render a real world acoustic attack unnecessary. Assuming that an attacker did have access to the keyboard to record a sample, it was shown that having a different typist and typing style from the training sample decreases the recognition rate. This would further decrease if both the typist and the keyboard were different from those used to record the training sample.

Zhuang et al have shown that having a labeled training sample is unnecessary and developed a recogniser that could identify up to 96% of typed text from just a 10 minute sound recording. This improved attack is based on the assumption that the text being typed is not random but a language that has constraints. Identifying which sounds correspond to which keys is similar to frequency analysis of a substitution cipher. This is complicated by the fact that keystroke recordings are analogue data and can be different each time the same key is pressed or similar for different keys. This improvement greatly increases the threat posed by a keyboard acoustic attack. However, there are limitations to the recogniser that could lessen the severity of a real attack. The recogniser only works with character keys and does not take special keys, such as shift, caps lock and backspace into account. There must also be a sufficient amount of language based text for the recogniser to learn the keys.

There was also not a large range of experiments done with different environmental conditions and only a limited number of variables were tested. A couple of tests were done in a noisy environment but the type of noise was not specified. Background noise from other keyboards, which could be quite common in office buildings, could possibly make identifying keystrokes a lot harder. Only four keyboards were tested, all of them desktop keyboards, three of which were made by the same manufacturer. This does not seem to be sufficient to say that all keyboards are vulnerable to this attack. The typist was the same, as was the document that was typed, in all experiments. While this is good for testing the keyboards, additional tests could have been done to determine how the typist and

the text typed affect recognition. The typing speed was not mentioned although a document of 2273 characters was recorded in around 12 minutes. This works out to just under 40 words per minute, which is average but still slow for an experienced typist.

3 Real Attacks

There are several reasons why an attacker may want to obtain information that has been typed into a keyboard. The attacker may be after specific information or may just be watching to see if anything interesting turns up. Other than an acoustic emanation attack there are several other methods of capturing keystrokes that an attacker could use.

If the attacker has physical access to the keyboard or computer they could install a hardware based key logger which would not be detectable by the system. A software based key logger could be installed if the attacker could obtain the necessary permissions to install it. A key logger would give the best results since every single keystroke is logged. There is a chance these could be detected and as there is a need to retrieve the captured keystrokes, perhaps traced back to the attacker.

A hidden camera would allow the attacker to watch the keys as they were typed. Some keys may be obscured by the typist but by knowledge of the language and of which keys were not typed, a relatively high recognition rate could be achieved. With this method there may be problems with finding a good place to hide the camera so that it has a line of sight to the keyboard and doing so without being seen.

The final method is simply watching the typist's fingers as they type or looking at the screen. While this method requires the least equipment, it is not suited for long term attacks but is useful when the typist is entering a password.

In order to use a keyboard acoustic attack the attacker must be able to get close enough to receive the sounds produced by the target keyboard. An advantage of this attack is that it is a passive attack because it does not interact with the target system and so can not be detected by the target system. It is also quite an unexpected type of attack so there may be little or no effort made to detect or defend against it. The acoustic attack also offers the attacker a lot of anonymity as no electronic evidence is left and as long as the recording process is not discovered the target typist may have no idea their security has been

compromised.

3.1 The Attacker

To carry out an acoustic attack, the attacker must record a long enough sample to be able to start the recognition process. Currently this must be at least 10 minutes long. The recording device would either need to be hidden near the keyboard or held by the attacker. Unless the attacker knows when the desired information will be typed in, the recording device would potentially need to be recording for hours or days. This would lead to more information being collected but increase the chance of exposure. For a long term recording it would not be practical for the attacker to wait nearby and record the keystrokes. This would require a hidden recorder or perhaps a microphone and radio transmitter. If a device like this was discovered it would almost certainly arouse suspicion. Recording from a distance would be possible with a parabolic microphone but not very discrete, especially in public places, unless the attacker could conceal themselves.

In the case where the attacker can be close to the target, a recording device that is hidden on the attacker or disguised as something else is unlikely to be noticed. Some cell phones and mp3 players come with built in recorders; these would be a good way to discreetly make a recording while the attacker sits nearby.

3.2 Environments

Due to the nature of this attack, an attacker is going to have to get physically close to the target, as opposed to network based attacks where the attacker can potentially be anywhere in the world. Environments for this attack can be divided up into two types, public and private. A public environment would be one where anyone can enter, such as an Internet café or shopping mall. A private environment such as an office building would normally only be accessible by someone with permission to enter, such as an employee. An exception to this is when the acoustic emanations can be picked up from a distance by say, an attacker with a parabolic microphone or over the phone. The environment determines what type and how much background noise there is and how close the attacker can get to the target.

In a computer lab, Internet café or kiosk, people are not assigned to a computer so it would be hard to target a particular person. Instead, an attack in these public environments would most likely be intended to harvest any valuable information that is typed in. Login details for email accounts would be quite common and

possibly Internet banking or credit card information. At public computers people do not normally type large documents, although they may send some emails, so the number of keystrokes may not be sufficient to identify the keys. In this case the attacker may need to train the recogniser on the target keyboard. Successful attacks in these environments would rely heavily on the recogniser's ability to identify keys based only on the keyboard and not the typist.

A private environment, most at risk from an inside attacker, would be in locations such as office buildings, private offices or reception areas. In these locations people are often assigned to a single computer, meaning a single typist on the keyboard. The target information could be proprietary documents or other employee's login details.

The final type of environment is a public location where the target keyboard is private but the sounds are emanated into the public location. Examples of these are people with laptops in a public place or an office where sounds can be picked up through a window. This gives the attacker the opportunity to attack a location they may not normally have physical access to.

These environments could also have a lot of different types of background noise such as talking, music and other keyboards. Music and talking vary a lot and it is hard to say how it will affect the recogniser. If the music is at a low volume, as it usually is in public places and offices, then it may not be much of a problem. Obviously other keyboard sounds could pose a problem to the recogniser; depending on how loud they were they could be mistaken for the target keyboard.

4 Proposed Extensions

There were some important factors in the experiments done by Zhuang et al that were not fully tested or discussed. They conducted one test in a noisy environment but they did not mention what type of noise, such as other keyboards or talking, it was. They also did not say how loud the noise was or what signal to noise ratio the recogniser could tolerate. The typing speed and how it affects the recognition process has not been mentioned. Further experimentation also needs to be done to test a wider range of keyboards, typists, documents, noise and distances.

The proposed extensions have been split up into four parts. The first two parts deal with further and more in-depth testing of the recognition software. The third part describes possible improvements that could be made to the recogniser to

increase the recognition rate. The final part proposes some mock attacks that could show how effective an acoustic attack might be.

4.1 Part 1: Target Variables

Keyboards can vary in manufacturer, design, model, size, materials, laptop or desktop, and age. A wide selection of keyboards should be chosen as this attack is based on the sounds coming from these keyboards. Keyboards from as many different manufacturers as possible should be chosen because the manufacturing process and materials, which could be the same in all models from that manufacturer, may be what determines the differences in sound. Age is also an important factor because it has not been shown how age and use of a keyboard affects the sounds it makes. The age could affect both the variance in sound between different keys and the variance in sound between different strokes of the same key.

Styles of typing can greatly differ between typists and the targets of this attack should not all be considered experienced typists. Speed, force and error rate are important factors in classifying a typist. The experience of typist is related to these, for instance touch typists tend not to hit the keys as hard as a two finger typist. Fast typists could prove hard to attack because the push peak of one key could occur before the release peak of the previous key, in effect overlapping the sounds from each key. A selection of typists should be chosen ranging from beginner to expert. This selection does not need to be large (perhaps five), just enough to show a good range of experience levels.

The document typed by each typist, on each keyboard should be the same. Several runs of the experiment can be conducted with different documents. Taking into account the current state of the recogniser the documents should use few numerical and special keys as these may not conform to the language model.

It would not be practical to test all combinations of these variables especially considering the number of possible keyboards. To reduce the number of keyboards to a manageable amount, a subset that represents the distribution of keyboards could be selected. To do this a typist can type the same document on all keyboards with a uniform force and speed. This is so that as much of the recognition is based on the keyboard as possible. The keyboards can now be ranked by recognition rate and a smaller set can be selected.

4.2 Part 2: Environmental Variables

The environmental variables are the amount of background noise and the

distance from the keyboard to the recorder. In previous experiments these have only been briefly tested. These tests are to show the effects of noise and distance on the recognition process and can be done independently of the target variable tests. The keyboard, typists and document that gave the best recognition should be used to best show the effects of noise and distance. A range of levels should be set for both noise and distance and then all combinations tested.

Distances should reflect where an attacker might be able to record from. For example a hidden microphone would probably be within a meter of the keyboard. An attacker who is sitting near the target with a hidden recorder could be within around five meters. Any other longer distances that may be possible with a parabolic microphone would be a good idea to test the limits of recording range. To speed things up all distances could be done at once for each level of background noise using several microphones and the recordings processed separately.

Noise can be of different types and volumes. A type of background noise that should definitely be tested is other keyboards. These should be tried at different volumes starting from zero and increasing in steps until the recogniser can no longer distinguish keystrokes.

4.3 Part 3: Recogniser Improvements

As discussed in [1], Song et al, [3], have shown that it is possible to use inter-keystroke timing information to determine likely combinations of keys. For example the time between the keystrokes “er” can be much smaller than the time between “qz”. Adding this functionality into the keystroke recogniser could further improve the key recognition rate. Using this method would require knowledge of the target keyboard's character mapping. This can generally be easily guessed.

Support for special keys, especially the shift and backspace keys, would make interpreting the output of the recogniser a lot easier. There are some difficulties with adding this functionality. The shift key is normally pressed and held while other keys are pressed. The recogniser would have to identify the shift key by its push peak, and then shift each subsequent key until the release peak is heard. A possible way to implement recognition for the backspace key is discussed in [2].

Natural language processing could be used when there is a partially recognised word with more than one possible correction to determine the most natural word for the sentence. This should improve the word recognition rate.

4.4 Part 4: Mock Attacks

Simulations of real attacks, while not conducted under experimental conditions, would be the best way of determining how feasible a real acoustic emanation attack would be. One attack could be for the attacker to sit near the target while pretending to listen to an mp3 player which is actually set to record. This method of recording could be possible in public environments like an Internet café or unrestricted computer lab.

A second attack could be to hide recorders near (or in) computers in an organisation that may be vulnerable to insider attacks and request the users to continue with their normal day to day typing. As well as determine the effectiveness of this attack, it would also help to raise awareness of it.

5 Discussion

5.1 Current Limitations

A limitation of keystroke recognisers which has not been mentioned before is that in order to identify a keystroke, that key must have been typed or is able to be inferred by the language model. If some keys are rarely used and are not part of any language, then when they are used the recogniser will have no idea what they are. An example is when a document is typed using the character keys but none of the numeric keys. Then when the same typist enters a numerical password the recogniser will not be able identify the keys. In the best case it will know that a new key has been pressed and not return a character key. Still, this can help reduce the search space by knowing which keys were not pressed.

Some documents may not conform to any language model, such as a spreadsheet or input into forms. In order for the recogniser to learn which sounds are from which key without labeled training samples it requires text that conforms to some language that offers enough constraints to help reduce the possible key combinations. If not enough language based text is given for training then it may be impossible to identify much of the keystrokes.

Another limitation is the accuracy of recognising random text. The tests done to recognise random text such as passwords show that 80% of 10 character passwords can be obtained in less than 75 attempts. This is an extremely good reduction of the number of possible passwords the attacker would have had to try before the attack. One fault with this is when the attacker has to try the possible passwords by interacting with the computer system, as opposed to an offline

attack where the attacker has a password hash or other information that can verify the password. An intrusion detection system should alert the administrator after too many incorrect logins and web based accounts often lock the user out after too many unsuccessful login attempts.

The special keys can not yet be recognised however this may not be too much of a problem in the case of shift and backspace since it may be quite possible for a human to understand what the text should have been. A human could help the recogniser identify shift and backspace keys by pointing some of them out manually until the features of the sound can be learnt. Most other special keys do not affect the other characters and so can be ignored or replaced with a space as mentioned in [2].

5.2 Feasibility

If further experimentation shows that it is possible to recover a good portion of text from the acoustic emanations of most makes and models of keyboard then an acoustic attack is really quite feasible. Even with a recognition rate of 50% it is normally possible for a human to reconstruct the rest of the information. With the lower recognition rates for random text such as passwords, an attacker still has far more information than they should have. For passwords that are not random data the number of guesses an attacker has to make could be even smaller.

The ease and success of an attack are highly dependent on the environment. The attacker would need to record enough language based text to be able to train the recogniser. This may not be a problem if the target is often typing long documents but on public computers the attacker may have to train the recogniser with their own typing. This could lead to a lower recognition rate when targeting another typist.

The training period might be able to be reduced if the attacker knows or can predict what some of the keystrokes are. This would be similar to a known plaintext attack and would help the recogniser quickly identify certain keys. The known plaintext would have to be synchronised with the correct keystrokes to prevent miss-identification which may not be obvious until well into the training period.

6 Defense

The best way to mitigate the problem of acoustic attacks is to stop the problem at its source: the keyboards. Quiet keyboards have been shown to be vulnerable

but as they are quieter, noise will be more of a problem for the attacker. Virtual keyboards that project the image of the keyboard onto a surface using a laser are also an alternative. Noise such as music could be added to the environment to make it harder to distinguish the differences in keystrokes. The volume of noise required to do this may not be suitable for the environment and may be disrupting to people.

To reduce the risk of a stolen password, authentication could be done via an alternative method such as one time passwords, biometric features or smart card. Ideally a combination of these could be used.

As mentioned in section 3, the victim of this attack may be unaware of any disclosure of their confidential information. Depending on the attacker's goal, this attack may be just one step in a larger attack. Information gained from the acoustic attack, especially passwords, may be used to further compromise the system. If the attacker makes an observable use of the stolen information there is a chance for an auditor, intrusion detection system or human to detect the misuse of this information. This would be best suited for detecting insider attacks.

7 Conclusion

The fact that somebody can identify any portion of a typist's keystrokes without detection is a serious security concern. The hurdles faced by an attacker are getting a clear sound recording for analysis and being able to capture enough language based text to train the recogniser with. Knowing what type of information is likely to be typed in different environments is an important factor in predicting the success of an attack.

Further experimentation should show just how much of a threat this is by showing how many keyboards and typing styles are vulnerable to this attack. There are also some ideas for improving the recogniser to increase both key and word recognition. With the public release of recognition software acoustic attacks could become quite popular as there is little cost and risk for the attacker.

Unless the stolen information is used when interacting with the compromised system it may not be possible to ever detect whether an emanation based attack has occurred. Computer users should be aware that this type of attack is possible and not rely only on the computer system to ensure confidentiality.

References

[1] Asonov, Dmitri, and Agrawal, Rakesh. 2004. Keyboard Acoustic Emanations, *Proc. of IEEE Symposium on Security and Privacy*: 3-11

[2] Zhuang, Li, Zhou, Feng and Tygar, J.D. 2005. Keyboard Acoustic Emanations Revisited, *12th ACM Conference on Computer and Communications Security* (CCS'05): 373-382

[3] Song, D.X. Wagner, D. Tian, X. 2001. Timing Analysis of Keystrokes and Timing Attacks on SSH, *Proceedings of the 10th USENIX Security Symposium*

[4] Lampson, Butler, 2004. Computer Security in the Real World, *Computer* 37 (6): 37-46

[5] Wikipedia. October 2006. Frequency Analysis. URL: http://en.wikipedia.org/wiki/Frequency_analysis