Report 2: Rhythmic Keylogger for Authentication

Matriculation ID: 190021896

November 18, 2019

Word Count: 2242

CS4203: Computer Security

University of St. Andrews

Department of Computer Science

1 Introduction

A keylogger is a software or hardware which has a primary function of capturing a users' keystrokes (what keys are being pressed), the stream of characters is then saved to a log file. This can be on both traditional desktops, laptops and even virtual keyboards, such as smartphones. Keyloggers are used legally in some instances, such as with financial corporations logging their employees keystrokes to ensure that users are in line with secure practices and regulation. In most jurisdictions, employees have to be made aware that this software is installed on their desktops. However, the use of keyloggers is illegal when it is used with a malicious intent, to steal personal data, such as passwords and financial information, where users are not aware that the their software and/or hardware is infected with a keylogger. One common malicious use of keyloggers are to steal passwords, perhaps rendering the usual authentication of pairing usernames and password ineffective.

There are different types of keyloggers, but one analyses the the rhythm and manner in which a user types, and employs the fact that each key will sound slightly different based on its position, to reconstruct what the user has typed. Furthermore, researchers have found that different keys give off different frequencies of radio waves when keys are pressed. The biometric community has claimed that a password coupled with the normal rhythmic typing of the user is stronger than the password on its own. This paper will explore this claim and conduct a statistical analysis on data recorded by a rhythmic keylogger to test whether this truly improves the authentication of passwords.

The rest of this paper is organised as follows. Section 2 states the problem. Section 3 discusses relevant background and related work. Section 4 outlines the methodology. Section 5 defines the experiment. Section 6 discusses and evaluates the results of the experiment. Finally, Section 6 concludes.

2 Problem Statement

Users develop individual technique of typing, from which unique patters can be generated and profiled. This individual biometric characteristic is called keystroke dynamics. By analysing users' keystroke patterns, these can be used to authenticate users, if attackers are able to find a valid username and password combination.

3 Relevant Background

Although there has been a development in more secure forms of authentication, such as facial or fingerprint biometrics, these are normally only employed for authenticating users on devices, but not for accounts online. Password strings are more useable, but less secure, and this security-usability trade off is what keeps them popular, despite the fact that they're prone to social engineering attacks or spyware [1]. Long and complex passwords are better, but they are not user-friendly as they are hard to memorise. Although keystroke dynamics can be used by attackers to figure out what is being typed, it is hard for them to mimic this exact rhythm when typing. Therefore, Alsultan and Warwick [1] suggest that keystroke dynamics should be used to authenticate users, as well as their passwords.

Research has been conducted to generate keystroke dynamic algorithms that are successful in authenticating users. It is important to distinguish between the two forms of keystroke dynamics; free-text and fixed-text. Free-text refers to when the user simply types and continuous dynamic analysis refers to the user's authentication is based solely on matching their rhythmic profile continuously, not their text. Whereas, fixed-text refers to authentication of a certain string of characters, such as usernames or passwords. The biometric community has claimed that a fixed-text coupled with rhythmic matching is more secure than a password.

The analysis of digraphs and trigraphs have been used to distinguish or identify rhythm profiles. Sim and Janakiraman [2] conducted a study to evaluate whether digraphs are good for free-text keystroke dynamics. They collected keystroke data from 22 users over a period of two weeks, with different typing abilities; some trained typists, and others regular users of the keyboard. In total, their sample size was 9.5 million key events. A shortcut key allowed users to turn off the keylogger when typing sensitive information such as passwords or pin numbers, so their analysis did not record the digraph analysis of passwords. However, they found that digraphs are not discriminative with free-text, however, when word-specific digraphs are used, these can discriminate between rhythmic patterns. They suggested that combinations of held time (between a key press and key release of the same key) and inter-key times (time between two consecutive keystrokes) would yield better results.

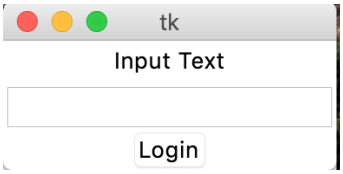
Snuzur [3] investigated whether it was possible to authenticate users based on fee-text keystrokes. The dataset they employed had a sample size of

around two and a half million digraphs, less than the above study [3], however, they achieved a 78 percent accuracy in identifying users using free-text. They used a keylogger called KEasyLogger, which has no malicious intentions and was designed to respect the privacy of the user, with only volunteered users. Their methodology involved building an algorithm that attempting to correctly cluster group sessions of the same user. For example, by comparing one user's keystrokes with the rest in that group over sessions to find the same rhythmic profile, and then checking against user ids to determine the success rate of the clustering algorithm.

4 Methodology

For this experiment, I designed a keylogger in python which utilised the package keyboard. This keylogger implemented a simple TkInter GUI interface to capture user input. This program monitors what is being being typed, and reports each string followed by the inter-key times between each keypress in nanoseconds, the entire string and the time it took to write the entire string in nanoseconds. Inputted text is then recorded and written to a log file. In order to do this, the package logger was used. As the provided test data was employed, there was no need to hash passwords for security.

The sample size was generated with data from three students. Each participant wrote out each of the twenty test cases (Table 1A) three times each, which resulted in 180 data tuples. Each participant conducted the trial three times in order to control and test for typing rhythm consistency. All statistical analysis was conducted on the statistical modelling package stata to analyse my data.



5 The Experiment

There is a claim in the biometric community that a password coupled with normal rhythmic typing of the owner is stronger than the password on its own. This paper will study this claim based on two hypotheses that analyse the rhythms and profiles of keystrokes using sample data:

H1: A simple rhythm can be determined more easily than a more patterned or dictionary based one.

A simple rhythm is defined as 1 or 2 distinct rhythms such as ta-ta-ta-ta or ta-tum-ta-tum where emphasis is given on either the ta or the tum keystroke, whereas a more patterned rhythm is one without consistent timings or more relational ones. In order to test this hypothesis, I grouped all the inter-key timings for individual by simple and complex, or more patterned text. This resulted in over 200 inter-key timings for simple text, and over 300 inter-key timings for complex words per user, resulting in a total over 1500 data points. I classified words as simple that had repeated rhythms such as alalal or AAAAA (a list of whether the username-password pair is simple or complex can be found in Table 1A).

Since H1 assumes there is a significant difference between rhythm patterns and lengths between complex and simple text, I used a t-test to determine whether there is a significant difference between these rhythms. The ttest was set out with the following hypotheses:

H0: U_1 - U_2 != 0

H1: U_1 - $U_2 = 0$

H2: There is a length L over which the timings of the password and ID are irrelevant. That is, L is more predicative of detection than rhythm when L $\geq$ N characters.

This hypotheses assumes that there is a certain threshold (length of username or password) above which the rhythmic timings are irrelevant. This can be represented by the following null and alternative hypotheses:

H0: There is no length that causes a change in rhythm.

H1: There is a length L which causes a change in rhythm.

In order to test this hypothesis, an OLS (Ordinary Least Squares) regression is conducted on all inter-key timings of each user against length. Since inter-key time is the main discriminator of rhythm profiling in this study, it is the main control variable. The regression is as follows:

$$y_i(\text{Length})_i = x_1(\text{User 1 Rhythmic Profile}) + x_2(\text{User 2 Rhythmic Profile}) + x_3(\text{User 3 Rhythmic Profile})\gamma + u_i \tag{1}$$

6 Results

6.1 Hypothesis 1

The p-values for difference in means for users 1, 2 and 3 respectively are 0.3549, 0.8192, and 0.2009 (Table 2,3,4). These are all greater ¿ 0.10, which means that the difference in means is not statistically significant at 10 percent. Therefore, the null hypothesis is rejected, indicating that there are no inter-key rhythmic differences between simple and complex rhythms. Unfortunately, this means that the data does not allow the study of whether a system can differentiate between simple and complex rhythms. False Acceptance Rate (FAR), which is an indicator of how often a biometric system will incorrectly accept an unauthorised user as authorised [2]. This result also indicates that if there is indeed no differentiating factors between simple and complex text, a system would not be able to validate a user on its rhythmic profile. This result may be attributed to the small sample size generated in this study, which means there is not enough data for the analysis to exploit and analyse. A future recommendation would be to increase the sample size of data points. Although my program was not able to calculate held key times, a compound analysis of held key times and inter-key times may provide richer data that may be able to differentiate between simple and complex rhythms.

Table 2: User 1 ttest

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
user1s~e	108	1053.015	152.0536	1580.188	751.5868	1354.444
user1c~x	108	909.5227	39.69285	412.5002	830.8363	988.2092
diff	108	143.4927	154.4231	1604.812	-162.6331	449.6184

mean(diff) = mean(**user1simple - user1complex**) t = 0.9292
Ho: mean(diff) = 0 degrees of freedom = 107

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
Pr(T < t) = 0.8226 Pr(|T| > |t|) = 0.3549 Pr(T > t) = 0.1774

Table 3: User 2 ttest

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
user2s~e	153	899.1341	33.05394	408.8546	833.8297	964.4386
user2c~x	153	910.0422	35.61265	440.5042	839.6825	980.4019
diff	153	-10.90804	47.64371	589.3201	-105.0374	83.22135

mean(diff) = mean(**user2simple - user2complex**) t = -0.2290
Ho: mean(diff) = 0 degrees of freedom = 152

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
Pr(T < t) = 0.4096 Pr(|T| > |t|) = 0.8192 Pr(T > t) = 0.5904

Table 4: User 3 ttest

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
user3s~e	150	953.6743	29.75027	364.3649	894.8874	1012.461
user3c~x	150	894.8644	36.08133	441.9043	823.5672	966.1616
diff	150	58.80992	45.77534	560.6311	-31.64275	149.2626

mean(diff) = mean(**user3simple - user3complex**) t = 1.2848
Ho: mean(diff) = 0 degrees of freedom = 149

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
Pr(T < t) = 0.8996 Pr(|T| > |t|) = 0.2009 Pr(T > t) = 0.1004

6.2 Hypothesis 2

Table 5 shows an OLS regression of length of string against each of the users' rhythms. This analysis shows that the constant length is 7.2. Furthermore, user one's rhythmic profile is negatively correlated to length. The relationship between length and user one and user two's rhythmic speeds are not significant. However, since user's rhythms are based on inter-key times, an increase of length by 1 decreases the inter-key speeds by 0.0001004 nanoseconds. Therefore, user three's rhythmic speed is negatively correlated to length at ten percent, and supports the alternative hypothesis. However, for users one and two, the results are inconclusive as we cannot fail to reject the null hypothesis that there is no length that causes a change in rhythm as the results are not significant. Although this regression shows the relationship between length and user's rhythmic profile, one of the disadvantages of

this model is that it does not have other data to control for other factors, such as keyboard type and whether that is the keyboard they are used to using and how experienced a user is with a keyboard. These are some controls which this study did not have the data for, but would provide a more rigorous analysis.

Table 5: OLS Analysis

Source	SS	df	MS	Number of obs	=	24
				F(3, 20)	=	0.52
Model	7.59887092	3	2.53295697	Prob > F	=	0.6755
Residual	98.0261291	20	4.90130645	R-squared	=	0.0719
				Adj R-squared	=	-0.0673
Total	105.625	23	4.5923913	Root MSE	=	2.2139

length	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
user1rhythm	-.0004701	.0006081	-0.77	0.449	-.0017386	.0007983
user2rhythm	.0005865	.0005973	0.98	0.338	-.0006595	.0018325
user3rhythm	-.0001004	.001074	-0.09	0.926	-.0023407	.0021399
_cons	7.218892	2.345842	3.08	0.006	2.325551	12.11223

7 Concluding Remarks

The first null hypothesis was rejected, that simple rhythms can be differentiated from complex ones. Furthermore, the results of estimating the relationship and length and users' rhythmic profiles are inconclusive as the results were not significant. Further studies may consider adding more controls into their regression to control for external factors that might affect this relationship. Furthermore, if held times of keys can be calculated, these can provide a better insight into how different users' rhythms differentiate, by controlling for an interaction term between held key times and inter-key times as suggested by Sim and Janakiraman [2]. This study was not able to conclude that a rhythmic keylogger can be used for authentication with fixed-text rhythm analysis.

8 Bibliography

[1] Alsultan, A., Warwick, K. (2013). Keystroke dynamics authentication: a survey of free-text methods. International Journal of Computer Science Issues (IJCSI), 10(4), 1.

[2] Sim, T., Janakiraman, R. (2007, June). Are digraphs good for free-text keystroke dynamics?. In 2007 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-6). IEEE

[3] Sznur, S. (2015). Advances in Keystroke Dynamics Techniques to Group Users Sessions. International Journal of Information Security Science, 4(2), 26-38

9 Appendix

Table 1: Test Data

Username	**Password**	Type of Rhythm
AAAAAA	123456	Simple
AAAAAA	alalal	Simple
AaAaAaAa	alalal	Simple
AAAAAA	c0mput3r sc13nc3	Simple / Complex
QWERTY	a2b3c4d5	Complex / Simple
QWERTY	mnbvcx	Complex
Azbycxdw	wdxcybza	Complex
Azbycxdw	lkjhgf12	Complex
zxcvbnm	qwdfvbhjio	Complex
zxcvbnm	pokjnbgfre	Complex

The rest of the code and data is included in the compressed file that has been submitted.