Keystroke forensics: Are You Typing on A Desktop Or on A Laptop?*

Ioannis Tsimperidis
Department of Elec. and Comp. Engineering
Democritus University of Thrace
67100 Xanthi, Greece
itsimper@ee.duth.gr

Vasilios Katos
Department of Elec. and Comp. Engineering
Democritus University of Thrace
67100 Xanthi, Greece
vkatos@ee.duth.gr

ABSTRACT

In this paper we investigate the potential of leveraging keystroke analysis - primarily used in user authentication - to user profiling and identification for forensic investigations. As such, the keystroke forensics approach proposed in this paper will support user profiling through integration with the offender profiling domain. Early findings show that it was possible to identify with significant probability the conditions and means a user is performing typing operations.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Abuse and crime involving computers; H.1.2 [**User/Machine Systems**]: Human Factors

General Terms

Measurement, Security, Human Factors

Keywords

keystroke dynamics, keystroke analysis, typing latency, user authentication.

1. INTRODUCTION

Keystroke dynamics is a relatively mature field of research, with the primary application domain relating to user authentication. However there exists a limited body of knowledge pertaining to recognition of emotional state and gender identification.

The biometric authentication approach involves the registration of a user where his biometric characteristics are stored in some form in a database. During the authentication phase the user is challenged to present similar characteristics (or behaviour in the case of dynamic biometrics) in order to successfully pass the authentication phase.

In this study we argue that components of user authentication based on keystroke dynamics can be utilised in an offender profiling context thus serving as a forensic tool. We examine whether it is feasible to perform a classification of users. More particularly we explore the possibilities of creating a system that could recognize if the user who enters a text is male or female, right-handed or left-handed, experienced or inexperienced computer operator, uses desktop or laptop, and so forth. The starting point of the proof is to explore whether it is possible to identify if a user has typed a text on a desktop or laptop.

The rest of the paper is structured as follows. In Section 2 we present the related work on keystroke dynamics. In Section 3 we develop our methodology, present the main goal, indicating the assumptions made and describe the developed software artefact required to conduct our primary data collection and analysis. Section 4 presents the model created and the results obtained. Lastly, Section 5 reports on the conclusions and further work.

2. RELATED WORK

The study of user typing behavior for identification purposes dates back to the 1970 [9]. The parameters used for analyzing keystroke dynamics can be divided into temporal and non-temporal. As temporal parameters are those that depend upon time (usually measured in ms). The most popular temporal parameter is keystroke duration, i.e. the time from pressing to releasing a button, and the digram latency, i.e. the time from releasing of a button to pressing next button [8]. Other variables relating to time can be found in other research initiatives, such as the time associated with the trigrams and tetragrams [11]. Non-temporal parameters relate to parameters whose values are not measured in time, such as typing speed (words per minute), the frequency of errors, error correction mode, which key is used when there are two or more options ("Shift", "Ctrl", "Alt", "Enter", etc.) [7]. Other non-temporal parameters revolve around typing and involve a multitude of factors such as the time of day when the user types, preferred applications and so forth.

The primary application domain of keystroke analysis in the literature is authentication control and a variety of different approaches and methods can be found, such as OneR, Random Tree, Decision Table, K-star, and Decision Tree [11, 3]. Naturally each method would have its advantages and drawbacks. For the user authentication stage the sample comparison was performed using different approaches,

*(Produces the permission block, and copyright information). For use with SIG-ALTERNATE.CLS. Supported by ACM.

BCI 2013, Thessaloniki, Greece

such as Near Neighbor [10], MLP, Euclidean Distance [2], Distance Manhattan. In all approaches a common modus operandi was the theoretical or empirical adjustment of the underlying coefficients in order to improve the results (that is reduce Type I & II errors) and remove potential outliers [2].

However, a portion of the work on keystroke dynamics has moved away from authentication, dealing with issues such as typing in different languages and the information that can be extracted from the user's textual artifacts [6], or identifying the gender of the author of a text [4, 1] based primarily on content (words, punctuation, emoticons, etc) and composition (number uppercase characters, number of digits, number of blank lines, length of the proposals, and so forth) of the text, while other work looked for relations between the psychological state of the user and the way in which she types [5].

As a general conclusion, it was established that modeling of typing is associated with many parameters and the data obtained from a typing sample can be significant. As such, most approaches focus on identifying the significant parameters, or alternatively, to discard data and parameters that are not significant to the modeling typing behavior.

3. THE FRAMEWORK

3.1 Preliminaries

In this paper we have adopted two terms to formally describe and empirically investigate the user and her behavior, namely "user" and "user classification". More specifically by the term "user" we refer to the person who is typing and is characterized by a multitude of variables (components), such as the gender, age, native language, the dominant hand, the experience in using computers, the medium in which typing, the level of education, and many others. In a sense, a user in our context is a snapshot of a physical person in a coven context. That is, the same person who is typing on a desktop, and after short time on a laptop, is two different users. Also, the same person who two years ago took its first steps with a computer, but today, after intensive and regular involvement, is an experienced typist. Under our scheme this is two different users. Essentially we assume that every observation exercise leads to recording a different, yet not necessarily unique and different user.

"User classification" is defined as the procedure by which, using the data from the typing behavior of a user, the user is attached to a series of characteristics that are present with some probability.

User classification through keystroke dynamics may be used as a forensic tool. For instance, if an e-mail account password has been stolen, it would be possible to disclose the fraud, since it is almost unlikely the legal and the illegal user have the same characteristics, and moreover to profile the offender. In that way, information of forensic interest can be disclosed, such as whether the offender is a man or a woman, right-handed or left-handed, over or under aged, which is his educational level, which is his nationality, and so forth.

3.2 Approach

One of the components mentioned earlier is the medium on which the typing is done. Nowadays due to the pervasiveness, flexibility and mobility of computers the desktop PC is not necessarily the preferred and standard medium and a user can type on a laptop, on a mobile phone, or on a tablet. In this study we investigated whether it is possible to identify the origin of a text, between a desktop and a laptop.

A further requirement was to create a method that was language independent. This was in order to allow the proposed method to integrate with other approaches that are language specific, analyse the userâĂŹs mood. Such requirement is in accordance to widely accepted forensic investigation approaches of correlating circumstantial evidence in a systematic way. This led to using temporal parameters but also maintains the benefit of avoiding more computationally expensive algorithms that deal with contextual and content interpretation of the user supplied data. As such the two parameters the method is built upon are the keystroke duration as well as the digram latency, with a slight differentiation from other approaches [8, 2, 6]. We define as digram latency the time between keydown of a button until the keydown of the next button, in order to avoid obtaining negative values.

The data collection apparatus was an application written in Visual Basic. The application upon recording the user with a suitable unique identifier it included a fixed text of 850 characters in Greek containing letters, digits, and other symbols a user may generate from typing to a keyboard. Upon completion of typing, a comma separated txt file with is created named after the subjectâĂŹs username with each line containing the character pressed, the keydown and keyup time in ms. The times recorded are measured as the time elapsed from the execution of the application. The records have a format as shown in the following excerpt:

"T",14961,15039
"O",15132,15257
"Y",15351,15429

From the data collected we export keystroke duration for all keys pressed, as well as every digram latency. It should be obvious that the above data contain information for high order n-grams, that is for $n > 2$.

The empirical data was collected from 17 volunteers during the period from 11.10.2012 to 21.11.2012. These volunteers were asked to type both on a laptop and a desktop and their typing profile was obtained. The selection of volunteers is anything but easy process, because apart from their desire to help, we should ensure the closest possible match of their characteristics with those of the general population. The volunteers, their usernames and the characteristics in respect of the general population representation attempt, are shown in Table 1.

The number of participants who were male was almost equal to number of female. The proportion of left-handed volunteers is about 11%, which closely reflects to the proportion of the whole population. The educational level of the participants corresponds to the ratio of the level of education of a population with a Greek nationality. Familiarity to a particular device is defined as the device that someone consumes more than 75% of the total time of interacting with a computing device. Volunteers who answered "Both" have a more balanced contact with both devices, whilst the participants answering "None" have limited exposure and interaction time with computers. The volunteers were given instructions not to use the mouse and to use the application in a setting where they are alone, in order to avoid loss of

Table 1: User profiles

User	Gender	Handed-ness	Level of education	Familiar with
anemos	Male	Left	High School	Desktop
basilis	Male	Right	High School	Both
chry19	Male	Right	University	Desktop
chrysa	Female	Right	University	Laptop
dnths	Male	Right	T.E.I.*	Desktop
elli	Female	Right	High School	None
giannis	Male	Right	T.E.I.	Desktop
giorgos	Male	Right	High School	Desktop
gwgw	Female	Right	High School	Both
kokopilas	Female	Left	High School	Laptop
lefteris	Male	Right	High School	Laptop
lelloo	Female	Right	T.E.I.	Laptop
mammy	Female	Right	T.E.I.	Laptop
mmmm	Female	Right	High School	Desktop
nik	Male	Right	University	Both
snuv	Female	Right	University	Desktop
teo	Male	Right	University	Laptop

*Technological Educational Institute

concentration. It should be noted that it was not confirmed on all runs whether the users complied with the instructions given.

The raw data (a total of 34 files about 1000 records each of them) were sanitized in order to remove inconsistencies (most of the inconsistencies were due to the users moving the mouse or continuous pressing keys) and the keystroke duration times and digram latency were exported. Initially, the data were macroscopically analyzed, observing the total duration of typing text, the number of errors made (entries with "Delete" and "Backspace") and the statistical descriptors of single characters and digram times (averages and standard deviations). The results provided some general indications, for example, that for most users it took longer to type the same text in the laptop than in the desktop, regardless of the device familiarity of the user, or as well as they performed more errors on a laptop, or the durations of the characters, digrams and trigrams were smaller on a desktop than on a laptop.

The obtained data were merged to form two datasets, namely the desktop and the laptop dataset. Such consolidation was performed as the main aim of the proposed research is the distinction of typing between a laptop or a desktop over all users, instead of a specific one. As such, the sample had to be carefully selected in order to reflect the whole population. The merging process resulted to files containing approximately 16,500 records, effectively reducing the statistical error.

From each dataset the data entries were clustered on a digram basis. The mean value of each digram was calculated and the digram instances exhibiting times greater than 3 times the mean were removed as they were most likely due to user pauses, which could artificially distort the results. From the remaining sanitized data the digrams that were eventually selected where those that appeared at least three times per user (corresponding to approximately 50 appearances in the consolidated data) [7, 11]. Finally, the means and standard deviations of these data were finally calculated.

The empirical findings revealed that there is no distinction of some digrams being typed on a desktop or a laptop. In Figure 1 the latency distribution probability of the digram 'E-I' is shown.

On the contrary, some digrams appeared to have significant statistical differences between a laptop and a desktop, such as the digram 'Y-(space)' as depicted in Figure 2.

Given that the sample size is relatively large âĂŞ over 300 appearances of each digram - the likelihood of statistical bias is small.

The reasons behind a user typing in a different way between a desktop and a laptop have been fairly studied in the literature [5]. Although it is challenging to express the influence of the particularities of desktop and laptop keyboards in a quantitative way, it appears that the main differences exacerbating the different behaviors are, among others, the lack of the right numerical pad on most laptops, the greater depth of keys on a desktop keyboard, which makes them more distinct, and the place is used by each of them.

The significant statistical differences of certain digrams when typed on a desktop or a laptop keyboard provide opportunities of evaluating the origin of a typed document. More specifically, the digram latency may directly map to a probability of a document being typed on a desktop or a laptop. For example, the "Y-(Space)" digram measured with a latency of 350ms would mean that it may have a 0.12 probability of being typed on a desktop and 0.085 being typed on a laptop. This means that it is 40% more probable that it was typed on a desktop. If such circumstantial evidence is extended across all available digrams, we can construct an overall indicator showing the origin of a document.

Based on the above observation of nontrivial significance, an application was designed and developed. The application accepts the txt files containing the keystroke latency data as generated by the logging application and the statistical descriptors are constructed for each available digram. Then the typing medium (desktop/laptop) probabilities are generated according to the method outlined above and using a scoring system that is used to aggregate and combine the evidence. In order to improve the accuracy of the proposed system different coefficients were investigated in order to assign different weights to the digrams depending on the distinguishing capabilities of each digram. For example, a digram showing large latency differences would contribute with a higher weight in the scoring system (say a factor of 2 or 3). Also, a digram showing minimal differences in time variables could be ignored or given to it a factor of 0.5. The use of a scoring system was proposed in [8]. Upon calculation of all relevant digrams and the final score, the application submits a guess "desktop" or "laptop".

4. RESULTS

The system was evaluated by adopting the following approach. The evaluation procedure consisted of three stages. During the first stage the model was created based on the keystroke logs provided by the users. In the second stage data from a new and different group of volunteers who were asked to type the same fixed text on desktop and laptop was obtained. During the third stage volunteers from both groups were asked to type freely text of a certain yet fixed size.

The results from the first stage are summarized in Table 2.

The overall success rate - approaching 80% - shows some

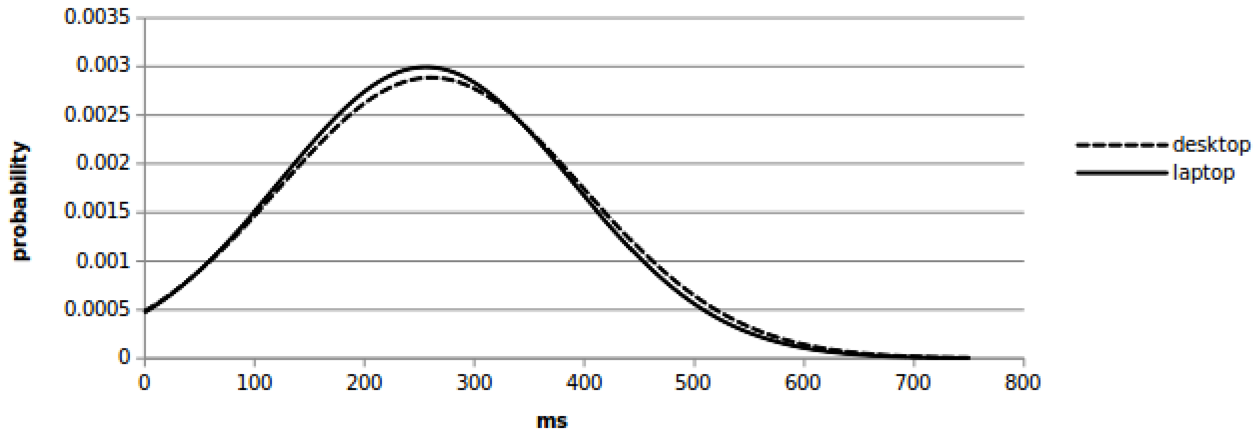

Figure 1: Digram latency probability for 'E-I'

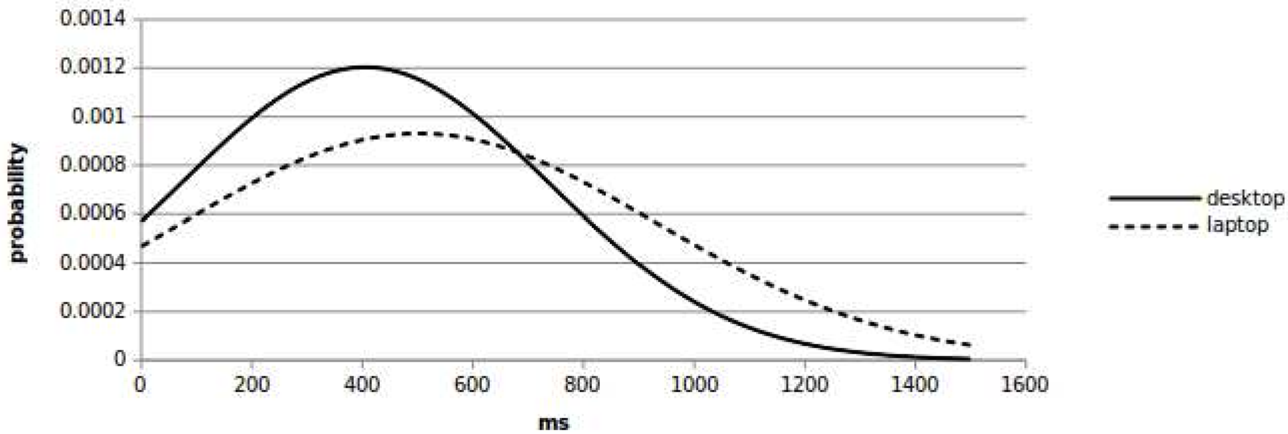

Figure 2: Digram latency probability for 'Y-(space)'

Table 2: First stage results

First group of volunteers (files that created the model)								
Desktop			Laptop			Total		
Successes	Failures	Success rate	Successes	Failures	Success rate	Successes	Failures	Success rate
17	0	100%	10	7	58.82%	27	7	79.41%

asymmetry regarding the prediction successes in desktop and laptop. With a closer inspection, it can be seen that from the 34 total tests, 24 pointed to a desktop, while 10 to a laptop. In other words when the system guessed "laptop", it was in fact completely successful, but when it guessed "desktop", it failed 29 percent of the time. A possible explanation of such asymmetry could be that not all users took advantage of the desktopâĂŹs numerical pad, and the typing seems to be more unnoticeable in relation to that of the laptop.

The second stage involved a new group of volunteers who typed the same fixed text. The usernames, files created and whether the system predicted correctly the origin of each file are shown in Table 3.

As it seems, not all volunteers provided keystroke measurements for laptops. The results are shown in Table 4.

The findings are almost the same with the first stage. That is, high percentage of correct prediction (near 77%) and asymmetry in the successful predictions between desktop and laptop.

Finally, in the third stage volunteers from both groups, as well as new volunteers were asked to type a text of 700 to 1000 characters of their choice. The volunteers were also asked to include digits. The usernames of the participants of this third testing stage, the files created and the correct and incorrect predictions of the system for each file are summarized in Table 5.

During this stage, all users provided keystrokes on both a desktop and a laptop, by typing the same text on both devices. The results are summarized in Table 6.

Despite the smaller number of volunteers, the success rates were similar to those of stage 2 and the conclusions remain

Table 3: Stage 2 tests

Username	File from		Correct perdiction in	
	Desktop	Laptop	Desktop	Laptop
0022	✓	-	✓	-
101210	✓	✓	✓	X
achilleas	✓	✓	✓	✓
aggelos	✓	✓	✓	✓
alina	✓	✓	✓	X
hacker4	✓	✓	✓	X
iwanna	✓	✓	X	✓
kain	✓	✓	✓	✓
makis	✓	-	✓	-
nina dobrev	✓	✓	✓	✓
pavlerm	✓	✓	X	✓
stergios	✓	✓	✓	✓
thanos	✓	✓	✓	✓
xampos1987	✓	✓	✓	X

Table 4: Stage 2 results

Second group of volunteers (same fixed text)								
Desktop			Laptop			Total		
Successes	Failures	Success rate	Successes	Failures	Success rate	Successes	Failures	Success rate
12	2	85.71%	8	4	66.67%	20	6	76.92%

Table 5: Stage 3 tests

Username	File from		text size	Correct perdiction in	
	Desktop	Laptop		Desktop	Laptop
elli	✓	✓	468	✓	✓
gz	✓	✓	824	✓	X
kain	✓	✓	984	✓	X
lelloo	✓	✓	537	✓	✓
oxos	✓	✓	833	✓	✓
snuv	✓	✓	475	X	✓

Table 6: Stage 3 results

Third group of volunteers (selected text)								
Desktop			Laptop			Total		
Successes	Failures	Success rate	Successes	Failures	Success rate	Successes	Failures	Success rate
5	1	83.33%	4	2	66.67%	9	3	75.00%

the same.

5. CONCLUSIONS AND FUTURE WORK

Leveraging keystroke dynamics research to construct user profiles in the context of a digital investigation is a promising area of research and a domain with practical importance to electronic discovery. Assuming that a user may not necessarily have his or her keystroke profile registered on an authentication system âĂŞchallenging thus the main assumption of keystroke authentication âĂŞ we investigated whether it would be possible to depart from a uniformal probability of identifying a particular user characteristic and build a system that would help an investigator make an informative decision with some degree of certainty. In this paper we focused on the feasibility of identifying whether a user typed a certain text on a laptop or a desktop keyboard, but a complete solution would need to consist of independent tests corresponding to a variety of characteristics or properties. Besides user authentication, keystroke dynamics may be useful to detect the emotional state of the user, or to identify his gender, or to assess whether the user is typing in their native language or not. In this paper we investigated whether the keystroke dynamics is characteristic of a class of users and if it is possible to rank a user, solely on the way she types.

Due to the preliminary yet encouraging results, the model will be extended to consider other user characteristics or properties in order to form a concise and concrete solution. The complete approach which is part of our ongoing research involves the identification of correlation of the user properties through latent variables in order to establish the mutual information between them and the construction of a formal evidence handling framework based on known evidence fusion constructs such as the Dempster-Shafer theory of evidence.

A limitation of the current research was the use of a fixed text to create the reference model, which departs from the realistic behavior of the users. An improvement would be to use an agent that logs the user in real working environment and we conjecture that this would increase the success rates. Another parameter increasing the prediction accuracy is a larger user sample. Finally, an improvement that will significantly raise the reliability of the system would be to create feedback mechanism, enhancing in this way the database that generated the equations for the possibilities export, and the weights of each digram.

6. REFERENCES

[1] N. Cheng, R. Chandramouli, and K. Subbalakshmi. Author gender identification from text. *Digital Investigation*, 8(1):78–88, 2011.

[2] M. Curtin, C. Tappert, M. Villani, G. Ngo, J. Simone, H. S. Fort, and S. Cha. Keystroke biometric recognition on long-text input: A feasibility study. In *Proc. Int. Workshop Sci Comp/Comp Stat (IWSCCS 2006), Hong Kong*. Citeseer, 2006.

[3] M. N. Danish Jamil and A. Khan. Keystroke pattern recognition preventing online fraud. *International Journal of Engineering, Science and Technology*, pages 1953–1958, March 2011.

[4] O. Y. de Vel, M. W. Corney, A. M. Anderson, and G. M. Mohay. Language and gender author cohort

analysis of e-mail for computer forensics. In *Digital Forensics Research Conference*. DFRWS, 2002.

[5] C. Epp. *Identifying emotional states through keystroke dynamics*. PhD thesis, University of Saskatchewan, 2010.

[6] D. Gunetti, C. Picardi, and G. Ruffo. Keystroke analysis of different languages: A case study. In *Advances in Intelligent Data Analysis VI*, pages 133–144. Springer, 2005.

[7] K. Hempstalk. You are what you type? In *New Zealand Computer Science Research Student Conference*, pages 24–31, April 2008.

[8] D. Song, P. Venable, and A. Perrig. ”user recognition by keystroke latency pattern analysisâĂİ, http://paris.cs.berkeley.edu / perrig/projects/keystroke, 1997.

[9] R. Spillane. Keyboard apparatus for personal identification. *IBM Technical Disclosure Bulletin*, 17(3346):3346, April 1975.

[10] D. H. H. Sungzoon Cho, Chinguen Han and H.-I. Kim. Web based keystroke identity verification using neural network. *Journal of Organizational Computing and Electronic Commerce*, 10(4):295–307, 2000.

[11] Y. Zhao. Learning user keystroke patterns for authentication. *World Academy of Science, Engineering and Technology*, 14:65–70, 2006.