Mixing and Matching Human Traits using Hand Typing

Manoj Devare
Associate Professor
Department of MCA
PK Technical Campus Pune

ABSTRACT

In this paper we have discussed few findings with respect to hand typing. The user behavior of simple typing for the lower case letters, and key stroke behavior during combination of lower and upper case letters is observed. During the entire cases user's keystroke latencies are measured and the use of combination of "shift key" and "caps lock" are observed. Here, the targeted (clear) text is in front of the user and key strokes were recorded. The results are illustrated for the traits finding which will be useful for verification of the user. The deliverable outcome of this work is a timer based program and a deployable library for running in background polling towards the keystrokes.

Keywords

Biometrics, Security, Hand Typing, Keyboard Typing.

1. INTRODUCTION

The human natural traits are possibly used during the biometric system implementations, and application areas. The retina scanning, fingerprint recognition, DNA testing, hand geometry, face recognition are few of the characteristics for user authentication. In few cases, the heart beat measurement and ear shapes are also useful for finding human identity based on their body structures and natural habits. However, most of these techniques rely on the image processing tools and techniques. Here, the keystroke dynamics is under the observation, for different habits of typing of the users. Although much of the directional research work and commercial products are available in the market, the new insights have been placed in this research work.

This is interesting to observe the two types of users, i.e. "trained or habitual" and "un-trained". However, the technological usage of Computer, Laptops or mobile devices is mostly for habitual users, having good knowledge of usage of these devices. On the other side, if the habitual user is trying to work on different device which may differ slightly, then typing rhythms may vary. For example, if someone using 'X' company Laptop cannot have same typing rhythms when same user using the 'Y' company laptop, due to different keyboard layout. Due to hesitation, the habitual user's key pressing latencies perhaps show statistical variations.

The typing patterns of the user who has taken training of the "typing institute", or completed the typing certifications course such as 30 WPM (Words Per Minutes), 40 WPM or 50 WPM , are habitual with typing and having the better speed than other "un-trained" users. The un-trained users, i.e. not certified, having the good habits of using the keyboard can have faster typing speed. The common conversion factor between WPM and CPM (Characters per Minutes) is five [8].

It is physically observed that, few person using their "initial fingers" and "middle figures" of two hands. Whereas, the trained users can use their all ten fingers of two hands during typing. The "training" specific observations will be considered for developing the authorization system. For authentication of the user with keystroke dynamics perhaps characterized differently for "trained", "untrained" users.

The page-up, page-down, home and end keys could be observed with the help of intentional spell mistakes checking and corrections. The use of arrow keys, tab key for navigation and data entry will be possibly observed. There are two shift, ctrl, and alter keys, because the user may use one of them in typical cases. The use of special characters while typing the constrained "passwords" will be useful for certain cases.

Whereas, the "right hand thumb" is mostly used by right hand masters for pressing the "spacebar" on the keyboard. Moreover, the pressure on the keyboard can be measured using similar kind of sensors. To detect, which finger is used by the person for typing, there is need of special sensing instruments to sense their keystrokes. Hence, the analysis is not addressing the solution of these kinds of problem, but habit of using the keys.

The frequently used keys in Windows specific operating systems like "ctrl + s" (short cut key for saving the document contents), "ctrl + c" for copying text or documents, "ctrl + x" to cut, and "ctrl + v" for pasting perhaps used by the user. Hence it could be used as the habits of using the keyboards. These things can be observed during the day to day use of the computer. The desktop users not having power back up or uninterrupted power supplies, having precautionary habits for pressing "ctrl + s", frequently. These are the few examples for usage of the keys. But there are certain keys that are not identified manually, but needs the learning algorithms to invent the keys which will be identifying the particular person. The static short cut key analysis algorithm can be cracked easily. Hackers can learn the keystroke habits and mitigate the developed algorithm.

The person having laptop can behave comfortably with his instrument, whereas the same person may behave differently when using some another keyboard attached to the personal computer. Several keyboards and several person's behavior on the different keyboards can be analyzed. The user of the laptop becomes comfortable after particular number of days. If the same user is working with another computer system, then he may hesitate to use the keyboard and perhaps do more mistakes. The observations during such a changing conditions will be interesting to note down.

From hardware point of views, the different ports are available to connect the keyboards e.g. the PS2, USB etc. which may affects the capturing of key codes. The operating system specific speed may affect the decision making, due to

continuous verification. The key stroke and their driver specific recognitions may provide different speeds. For examples different Linux distributions may provides different key typing speeds due to their open source drivers. The knowledge of device drivers can be useful for developing adaptive algorithms.

It will be interesting to measure the latencies during the numeric keys usage, while typing the known text and un-know text typing. In case of different keyboards the e.g. laptop keyboard, simple keyboard connected to the desktop PCs may given variations in the results.

The human mood like anxiety, fatigue, drowsiness, relaxes, freshness can be detected using the typing rhythms. The emotions and sentiments can be identified, with typing habits. Interestingly, such mood detection can inform to the user for his betterment of the health. Moreover, the algorithmic development needs to checked the day or night time, for making conclusions regarding the verification of the user. For example, the person can type with specific rhythm in the morning time, different rhythm in the noon, or evening.

If the system user arrived after short or long holiday, the typing rhythms perhaps different than that of the regular usage. The sustainable environment to accept such a right user in such a complicated situations need to be tackled. Provision to handle such a cases need intelligent machine learning algorithms.

2. BACKGROUND

Lívia C. F et. al., describes the inputs of the key down, key up ASCII codes captured while the user is typing a string. The key code, keystroke latencies, and key duration were analyzed. The results of the experiments evaluated with three types of user: the legitimate, the impostor and the observer impostor users. The False Rejection Rate (FRR) obtained 1.45% and a False Acceptance Rate (FAR) is of 1.89%. This approach can be used to improve the usual login-password authentication when the password is no more a secret [1].

Koichiro Niinuma et. al., discusses continuous user authentication methods from usability and security. The webcam records user's face and color of clothing. This method can authenticate users regardless of their posture in front of the workstation (laptop or PC). Color information of users' clothing as an enrollment template in addition to their face information. The system cannot pre-register the clothing color information because this information is not permanent. The system automatically registers this information every time the user logs in and then fuses it with the conventional, password identification system [2].

Fabian Monrose et. al., discusses non-static biometric technique that aims to identify users based on analyzing habitual rhythm patterns in the way they type using template matching and Bayesian likelihood models. The use of digraph-specific measures of variability instead of single low-pass filters. It has been suggested that the use of structured text instead of allowing user to type arbitrary text during the identification process, will give better results [3].

User identification by typing samples written in different languages like Italian and English is found [4]. The Imposter Pass Rate(IPR) and False Alarm Rates (FAR) have been calculated. [5] Typing dynamics of free text provide useful information for user identification and authentication even when a long time has passed since typing profiles of users were formed, and even when ascertaining users are writing in a language different from the one used to form their profiles. [5]

Romain Giot et. al., [6] shows multimodal biometric system combining keystroke dynamics and 2D face recognition. Different fusion methods like sum configured with genetic programming on the scores of three keystroke dynamics algorithms and two 2D face recognition. This multimodal biometric system improves the recognition rate in comparison with each individual method. On a database composed of 100 individuals, the best keystroke dynamics method obtains an EER of 8.77%, the best face recognition one has an EER of 6.38%, while the best proposed fusion system provides an EER of 2.22%.

The pressure-based biometric authentication system (PBAS) has been designed to employ force sensors to measure the exact amount of force a user exerts while typing. Signal processing is hen carried out to construct a waveform pattern for the password entered. In addition to the force, PBAS measures the actual timing traces, which are often referred to as "latency". Two approaches to construct user typing pattern have been implemented with PBAS. It also eliminates the security threat posed by breaching the system through online network as the access to the system is only possible through the pressure sensor reinforced keyboard "bio-keyboard".

A continuously monitoring of genuine user using the concept of Penalty-and-reward function was done using key typing behavior of a genuine user, and system can be locked if a different user is detected. The static and continuous evaluation of the performance of a biometric authentication system differs greatly. Static biometric systems are generally evaluated in terms of False Match Rate (FMR) and False Non-Match Rate (FNMR), and the overall performance is often only reported with a single value: Equal Error Rate (EER). A continuous biometric authentication system is for faster detection of an impostor. [9]

Typing rhythms detecting the insider-threats, accessing backdoors, using shared accounts, or masquerading as other users. The multitude-of-factors problem is solved to evaluate multiple classifiers with linear mixed-effects models (LMMs). The classifier error rate factor is used validate the models and demonstrated that they accurately predict error rates in subsequent evaluations. For different classifiers different error rates are found, which is dependent on the user of the system. [10]

Keystroke dynamics is a biometric mainly used for verification, but also identification is possible. Keystroke dynamics is a very cheap biometric verification method because there is no need for any additional hardware besides a normal keyboard. Existing words can be cracked by dictionary attacks. The short length passwords can be easily cracked. In such a cases the keystroke dynamics can be useful. Secure shell (SSH) based systems may face problem of password cracking. SSH is designed to provide a secure channel between two hosts. As the mechanism of sending IP packets immediately after the key is press, the keystroke timing information of the users typing is reveled at the other end. The timing differences detected by the eavesdropper can cause serious problem of security, even by knowing root password. The statistical study is done and Hidden Markov model and key sequence prediction algorithm developed in this work. The SSH system is monitored and collection of the timing information is done. The application to the general class protocols for encrypting interactive traffic is done. The

suggestions to develop the new protocol by considering the timing attacks is given in this work. [11] [16]

The validity of the password using a keystroke dynamics enclosing statistical analysis program is developed. The information gathered from the user trials of this program is used to confirm that Typing Dynamics Biometric Authentication is a valid method for identity verification [12].

Currently, the smart phone users' security limited to the Personal Identification Number (PIN), a secret knowledge based technique that has historically demonstrated to provide ineffective protection from misuse. The numerous pattern classifiers have been used with the trade-off with computation versus performance. The statistical classifiers are found most effective. [13]

The comparison of typing samples of free text used to verify personal identity. The technique tested with a wide set of experiments on more than two hundred individuals, obtaining a False Alarm Rate(FAR) of less than 5% and an Impostor Pass Rate of less than 0.005%. The samples have been collected in different working sessions. [14] As the use of the keystroke dynamics is absolutely suitable for different application areas of e-commerce like Amazon.com. It can direct or indirect control the access to company resources and verifying the billing of a customer. [15]

3. METODOLOGY

The simple thread based program has been written for the analysis of the above four observations. The data is collected in the simple flat-file. The background program has noted down all observations. The sampling is collected with the office faculties with the help of the developed software. The obscure results and samples can be predictive if taken in large quantities. The data entry screen has been created for typing mixed small and capital characters to recognize the use of SHIFTKEY (KEYCODE 16) and CAPSLOCK (KEYCODE 20).

The user interface created for typing mixed small and capital characters and numbers to recognize the use of SHIFTKEY (KEYCODE 16), CAPSLOCK (KEYCODE 20), and use of NUMLOCK (KEYCODE 144). It is being observed that, for the keyboard layout as shown in figure 1 is not suitable for using NUM-LOCK. Hence user is using the keys running from left to right , below the function keys. The analysis will be different if the user is using the keyboard attached with the personal computer, where the number keys are also available on the right hand side. The keyboard layout for analysis is shown in figure 1.

Table 1: Latency and key pattern observation in keyboard typing

User	Training Status	Initial Latency	Delay in (a-z) typing	Delay in Paragraph typing	Delay in (a-z) & (A-Z) Typing	Shift / Caps lock Ratio	Use of NUMLOCK
1	Habitual	1082	1832	25143	1484	10/6	Not used
2	Habitual	770	900	31658	2468	12/4	Not used
3	Habitual	1112	1513	23996	4584	13/0	Not used
4	Trained	2605	818	24659	2016	12/0	Not used
5	Trained	614	682	22901	1625	10/4	Not used
6	Habitual	920	1118	42587	4418	8/4	Not used
7	Habitual	1225	795	35180	2202	9/4	Not used
8	Habitual	1204	925	45236	3606	13/0	Not used
9	Habitual	820	2404	29960	3174	12/1	Not used

```
Start timer
Initialize Timer counter
Start recording
        Open File
If <key pressed> then
Write TimerCounter and KEYCODE
End if
        End recording
        Close File
Stop timer
```

Pseudo code for analysis of the keystrokes:

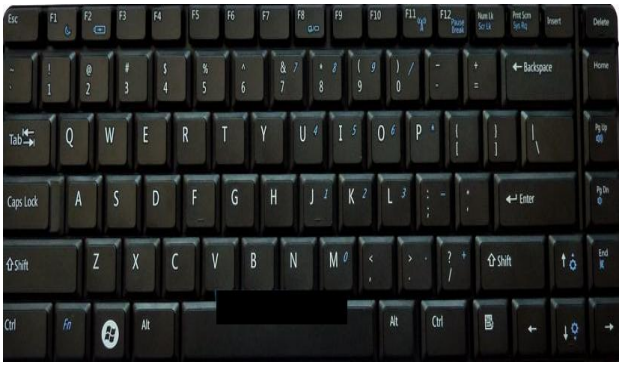

Fig 1: The keyboard used for measurement of latency.

4. RESULTS

As shown in Table 1, the initial latency is the time taken to start typing. This is measured on first screen only. Other delays are measured as, the time on last key press minus the time lapsed on first key press. For typing delay measurement, the training of the user matters, hence mentioned as "trained". The trained perhaps a person taken training from the typing institute or the person who did some practice using some training software available online, and installed on personal computer. Whereas, few persons are "habitual" of using computer but not trained. The "trained" person does not look towards the keyboards and "habitual" may look towards the keyboard. Unfortunately, there was no measurement technique, for observing users vision towards the keyboard. It perhaps could be done using some "webcam" capturing the movement of eyelashes. While entering the data in this software, if the user had done the mistakes, all is considered as the habit, and not ignored.

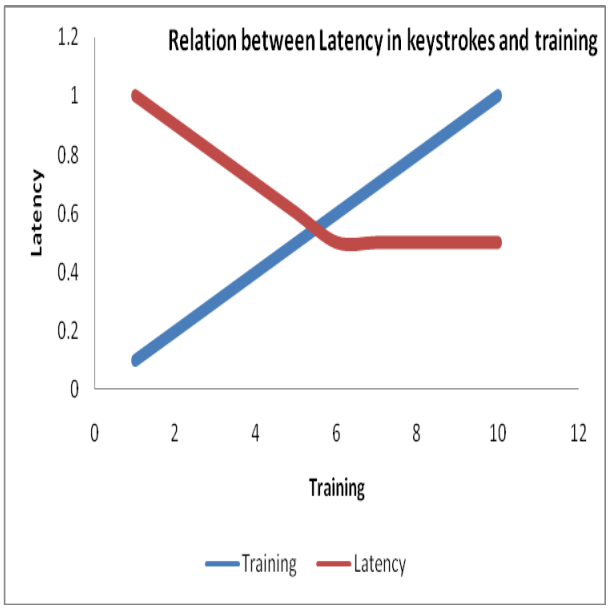

Fig 2: Relation between latency in keystrokes and training duration

The learning habits of the user are "adaptive in nature" hence it is suggested that the algorithms must be "adaptive" to learn the user's level of expertise while using the system.

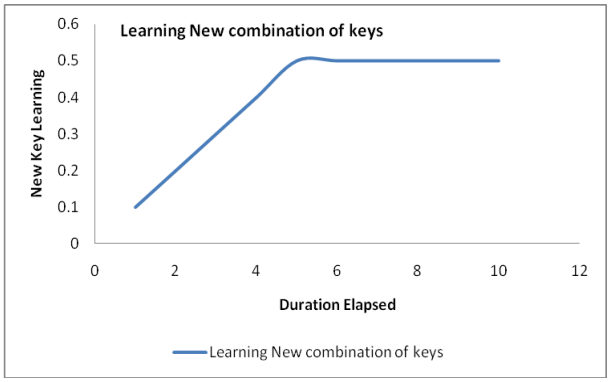

Fig 3: Learning new combinations of keys

The observation of the user, about what kind of keys he/she is pressing frequently, gives some good information for his continuous authentication, in case of sensitive and critical applications. For the purpose of the training of the System, it is necessary to ride certain kind of users, with different typing scenarios. In case of multi-user (like Linux administrators working) system this becomes very useful.

5. CONCLUSION

Succinctly, the learning habits of the user are "adaptive in nature" hence it is suggested that the algorithms must be "adaptive" to learn the user's level of expertise while using the system. The observation of the user, about what kind of keys he/she is pressing frequently, produces some suitable information for his continuous authentication, in case of sensitive and critical applications. The use of CAPSLOCK and SHIFTKEY ratio will be useful for developing the adaptive algorithm. Above work is more analysis oriented for the development of the keystroke based verification algorithms.

6. REFERENCES

[1] Lívia C. F. Araújo, Luiz H. R. Sucupira Jr., Miguel G. Lizárraga, Lee L. Ling, Andjoão B. T. Yabu-Uti (February 2005) "User Authentication Through Typing Biometrics Features" , IEEE Transactions On Signal Processing, Vol. 53, No. 2, , 851-855.

[2] Koichiro Niinuma, Anil K. Jain 2007, "Continuous User Authentication Using Temporal Information", Fujitsu Laboratories.

[3] Fabian Monrose, Aviel Rubin (2000) "Keystroke dynamics as a biometric for authentication", Future Generation Computer Systems, 351–359.

[4] Daniele Gunetti, Claudia Picardi, and Giancarlo Ruffo, 2005, "Keystroke Analysis of Different Languages: A Case Study", IDA 2005, LNCS 3646, 133–144.

[5] Daniele Gunetti, Claudia Picardi, and Giancarlo Ruffo , 2005, "Dealing with Different Languages and Old Profiles in Keystroke Analysis of Free Text", LNAI 3673, 347–358.

[6] Romain Giot, Baptiste Hemery, Christophe Rosenberger 2010, "Low Cost and Usable Multimodal Biometric System Based on Keystroke Dynamics and 2D Face Recognition", IEEE Computer Society, International Conference on Pattern Recognition 2010, 1128 -1131.

[7] Wasil Elsadig Eltahir, M. J. E. Salami, Ahmad Faris Ismail, and Weng Kin Lai 2008, "Design and Evaluation of a Pressure Based Typing Biometric Authentication System", Hindawi Publishing Corporation EURASIP Journal on Information Security Volume.

[8] http://en.wikipedia.org/wiki/Words_per_minute [accessed on date 10 February 2013].

[9] Patrick Bours, "Continuous keystroke dynamics: A different perspective towards biometric evaluation" , Information Security Technical Report, 2012, pp. 36-43, Elsevier Publications.

[10] Kevin S. Killourhy 2012, "A Scientific Understanding of Keystroke Dynamics" , Phd Thesis, School of Computer Science Computer Science Department Carnegie Mellon University.

[11] Jarmo Ilonen, "Keystroke dynamics", Lappeenranta University, Finland.

[12] Darren Clifford D'Souza 2002, "Typing Dynamics Biometric Authentication", Bachelor of Software Engineering Report, Department of Information Technology and Electrical Engineering, University of Queensland.

[13] A. Buchoux, N.L. Clarke 2008 , Deployment of Keystroke Analysis on a Smartphone, Edith Cowan University.

[14] Daniele Gunetti, Claudia Picardi August 2005, "Keystroke Analysis of Free Text ", ACM Transactions on Information and System Security, Vol. 8, No. 3, , Pages 312–347.

[15] Sam Hyland, April 7, 2004, "An Analysis of Keystroke Dynamics Use in User Authentication", Software engineering report.

[16] Dawn Xiaodong Song, David Wagner, Xuqing Tian, "Timing Analysis of Keystrokes and Timing Attacks on SSH", DARPA, NWSCSD, NSF Research Project report.