Improving Security Of Keystroke Dynamics By Increasing The Distance Between Keys

Purvashi Baynath Electrical and Electronics Engineering University of Mauritius Reduit, Mauritius p.baynath@gmail.com K.M.Sunjiv Soyjaudah Electrical and Electronics Engineering University of Mauritius Reduit, Mauritius ssoyjaudah@uom.ac.mu Maleika Heenaye-Mamode Khan Computer Science & Engineering Dept. University of Mauritius Mauritius m.mamodekhan@uom.ac.mu

Abstract- Keystroke dynamics is gaining popularity and researchers are striving to improve existing techniques or to explore aspects that have not been given much attention. In this paper, a new means of authentication for keystroke dynamics has been provided, by using a password with different distances between the keys. The classifier used in this paper is neural network. The mean square error has been used to compute the performance of the classifier. After the analysis and evaluations of the results, it was deduced that distance of keys on a keyboard affect the reliability of the password. The mean square error of the most space digraph was in the range of 15.5×10-3 to 107.6 ×10-3 and the least distant digraph has a mean square error range of 8.5×10-8 to 3.7×10-9. In this way, it is observed that the smaller the distance between the keys of the password used, the easier is the keystroke pattern compromise compared to larger distance between keys. Hence, it can be concluded by the larger is the distance between the keys, the more the security increases.

Keywords: Authentication; Keystroke dynamics; Neural Network; Classifier.

I. INTRODUCTION

Keystroke dynamics is a behavioural biometric authentication system [1]. It analyses the manner a user types at a station by using a keyboard. Keystroke dynamics is one of the efficient and inexpensive techniques that can authenticate computer users. This is typically characterized as flight time and dwell time. The advantage of keystroke dynamics is that it is readily deployable at the user end without the requirement of any additional hardware, as only a keyboard is required. Keystroke dynamics is a cheap and reliable biometric mechanism that has been proven accurate in distinguishing individuals [2][3][4].

Keystroke dynamics came in the early 1980's with the primary work done by Gaines and Lisowski [6]. Gaines and Lisowski [6] work was on statistical significance test of flight time for 87 lowercase letters using the T-test on digraph features. Even though the study attained remarkable performance rates, 0% False Accept Rate, the password used was long. Since then, various works continued using flight time as the principal feature extracted. The novel classification methods adopted a tendency of short phrases or set of words as password [7][8][9]. Obaidat and Sadoun, in 1997 [10], combined the duration of keystrokes known as the hold time in addition to flight time. Using a range of machine learning algorithms, the work attained the most remarkable performance results to date (0% FAR, and 0% FRR) with a comparatively minimalistic input condition of only a username. Despite the impressive results, the study included only 15 users in a controlled lab environment. In 2001, Monrose et al. [11] proposed the use of keystroke dynamics as a password hardening scheme with a password of eight characters. Although the hardening effect was demonstrated, the performance level was not as high as published in previous work. In 2006, Barlow and Cukic [12] provide a reliability of Credential Hardening through keystroke dynamics by incorporating the shift-key pattern to the password used.

Digraph is the feature used to be able to analyze the keystroke pattern. This feature has been explored in many researches and different novel techniques have been discovered through digraph. Monrose and Rubin [13] have extracted keystroke features of digraph and tri-graph using the mean and variance. The result yield an identification of 92% for a dataset containing 63 users by the application of Euclidean distance metric with Bayesian-like classifiers. Bergadano et al. [14] and Gunetti and Picardi [15] proposed the extract keystroke features by using the relative order of duration times for different n-graphs. They demonstrated that the authentication performance using free text is improved by the combination of new relative feature with features using absolute timing.

In 2013, Mondal et al. [16] introduce a new complexity metric based of the position of the keys on the keyboard, however the focus of the researchers were based on the complexity metrics taking various component like the length of the password, the distance between the keys, bigram frequency and consecutive letters with each hand. Since there were various variables in this study, it becomes difficult to demonstrate how the distances between keys affect the reliability of the password. In 2014, Senathipathi et al. [17], uses Dwell time, Flight time, Digraph, Bigraph and Virtual Key to make a comparative analysis of Particle Swarm Optimization and Genetic algorithm has been shown with respect to keystroke dynamics.

This paper extends the previous work carried on keystroke dynamics by Mondal et al. [16]. Inspired from the previous work conducted using flight time and digraph, in the study, flight time of keys are being used to show that the greater the



distance between the keys of a password, the higher is the reliability of the password. In this study a QWERTY keyboard is used for the data capture and strong and non-obvious password is used which contains the standard requirement of password [18].

The paper is organized as follows. Section II provides a literature on keystroke dynamics, and the method and metrics used in keystroke dynamics. The methodology of the method used in the design of the system and details on data collection and feature extraction have been detailed in section III. Section IV shows the results of the simulation and Section V provides ground for discussion and future work while Section VI gives an insight into how the technique could be improved further.

II. KEYSTROKE DYNAMICS

Keystroke dynamics is considered as a strong behavioural biometric based authentication system [1]. It is a process of analysing the way a user types at a station by observing the keyboard in order to recognise the users based on habitual typing rhythm patterns.

Keystroke dynamics systems can run in two different modes namely the Identification mode or Verification mode [1]. Identification is the process of finding out an individual's identity by investigating a biometric pattern calculated from the individual's biometric features. A larger amount of keystroke dynamics information is collected, and the user of the computer is recognized based on formerly collected information of keystroke dynamics profiles of all users. For each user, a biometric template is calculated in the training stage. A pattern that is going to be recognized is matched against every known template, yielding either a score or a distance describing the likeness between the pattern and the template. The system allocates the pattern to the person with the most alike biometric template. To prevent impostor patterns from being correctly recognized, the resemblance has to exceed a certain level commonly known as the threshold level. If this level is not reached, the pattern is rejected [1].

Digraph latency is the metric that is most commonly used in keystroke analysis and it typically measures the delay between the key-up and the subsequent key-down events, which are produced during normal typing (e.g. pressing letter A-B). Salthouse [19] adopted the digraph analysis in which the text was divided into easily remembered chunks. The conclusion was that for the first keystroke in a word, the typing speed is generally slower than that of subsequent keystrokes in the word. This word-initiation effect has been documented clearly by Salthouse [20], where the latency of the first keystroke in a word is found to be approximately 20% longer than the latency of the following keystrokes.

The effectiveness of digraph and n-graphs for free text keystroke dynamics were investigated by Sim and Janakiraman [21], and concluded that n-graphs are discriminative only when they are word-specific. Syed et al. [22] shows that the digraph and n-graph features do depend on the word context that they are computed in. The digraph of users contains distinguishing information for user authentication, while being independent of typing proficiency.

Roth et al. [23] explored keystroke dynamics in an interesting way by applying keystroke acoustics for user identification. A virtual vocabulary based on keystroke sound was built and then the digraph latency features were extracted using the learned virtual keyboard. An EER of 11% on a dataset of 50 subjects were obtained. Epp et al. [24] also conducted analysis on emotional states using keystroke dynamics features such as digraphs.

The most important advantage of Keyboard dynamics is that it requires no special hardware, since only a standard computer keyboard is needed. Additionally, the monitoring and capture of the keystroke pattern can be run in background.

A. Methods and Metrics for Keystroke Dynamics

From previous studies, a list of data acquisition techniques and typing metrics has been identified which can be used in the keystroke analysis [1][3][5]. The section below summarises the methods and metrics that has been used:

1) Static at login: During Static keystroke analysis, the typing pattern is authenticated based on a known keyword, phrase or some other pre-set text. During the system enrolment, the typing pattern taken is compared against a previously recorded typing patterns stored. Static methods for user verification is more robust compared to simple passwords, but do not provide continuous security, that is, they cannot detect an exchange of user after the initial verification.

2) *Digraph latency:* Digraph latency is the metric that measures the delay between the key-up event and the following key-down events, which are produced during normal typing (An example is pressing alphabet l and n).

B. Measurement Used

The biometric template used to recognize an individual, in keystroke dynamics, is grounded on the typing pattern, the regularity and the speed of typing on a keyboard. Dwell time and flight time are the measurements used for keystroke dynamics [16].

- Dwell time is the time period that a key is pressed
- Flight time is the time period in between releasing a key and pressing the next key

While typing a sequence of letterings, the time which the subject needs to find the right key (flight time) and the time which he holds down a key (dwell time) is specific to that subject, and it can be calculated in such a way that it is independent of overall typing speed. The rhythm with which some sequences of characters are typed is person dependent. For example someone used to typing in English will be quicker at typing certain character sequences such as 'the' than a person with French roots [16].



III. METHODOLOGY

The insight of the steps followed in the study has been provided Fig. 1. The Keystroke pattern consisted of flight time of the collected data. Each digraph used, has passed through the steps in Fig. 1.

A. Password and Data Collection

In this research work, data were collected from 100 subjects, each typing 20 repetitions of a password for a total of 2000 password-typing samples. The data was collected at different intervals so that ageing does not occur. The flight time was extracted from the raw data. The user was asked to use only one hand (their strong hand) during the data collection activities.

The first step in the evaluation was to collect a sample of keystroke-timing data. In next section, it has been explained how the password for data collection has been chosen, designed a data-collection apparatus, and extracted a set of password-timing features.



Fig. 1. Flowchart of the design of the system

B. Choosing a Password

For this research a static password is chosen as the basic idea of the statistical approach. A reference set of typing characteristics of a user is compared with a test set of typing characteristics of the same user or a test set of a hacker. A strong and non-obvious password is chosen, which contains the standard requirement of password. A strong password has a minimum of with eight characters and the characters constitute of an upper case, lower case, numerals, and special characters. The password also contains at least four unique characters and each character has not been repeated more than four times consecutively.

To make the password better match the strong passwords characteristics, a 10-character password containing numbers, letters, and punctuation, and has been used. The result of this procedure is the following password: .tie5Roalnb. For the special character and numerical value, the numerical pad of the keyboard was used. Fig. 2, shows the position of the key used on the keyboard used. For the password chosen, as you can see from Fig. 2, the distance between keys are much far way when using the numerical pad and there is a variation of distance between the other keys.

C. Data-Collection apparatus

A laptop with an external QWERTY keyboard has been set up to collect data. A Windows application has been developed which prompts a user to type the password. fig.2 show the interface to capture the data. The environment is set as consistent as possible for all subjects. The application displays a text-entry field on the screen and the password to be type is also displayed on the screen. The user must type the 10 characters of the password accurately, in sequence, and then press Enter. If any errors in the sequence are detected, the user is prompted to type the password again. Whenever the user presses or releases a key, the application records the occurrence (i.e., key down or key up), the name of the key involved, and what time the event occurred. The flight time was calculated from the data captured. The data flow of the typing process is shown in figure 3.

🖳 Data Capture	
Usemame	
Password	
Capture	Clear

Fig. 2.Interface to capture data





Fig. 3. Flowchart of data capture of keystroke timing

IV. SIMULATION AND RESULTS

The neural network toolbox of MATLAB has been used to perform the simulation of the features. The number of neurons, training set, and testing sets were initially chosen at random until a good and consistent result was obtained. The abovementioned parameters were eventually set to be fixed.

The Levenberg Marquart algorithm was used as the training algorithm and the results were computed in terms of mean square error (MSE). The mean square error shows the error scored while the neural network was trained to learn the flight time for different digraph. Four digraph of the password has been tested, after z-score normalisation applied to the pattern.



Fig. 4. Keyboard Pattern with the key used highlighted

In this study, the digraph of .-t, e-5, l-n and n-b has been studied since there is a variation between the distances of the keys. Fig. 2 shows the picture of the keyboard used. The alphabet and number and special character used for the password are highlighted. Hence the positioning of the keys can be seen. From the Fig. 2, it can be concluded that the distance between the digraph .-t is the largest followed by digraph e-5 then l-n and at last n-b.

Fig. 3 shows, the mean square error graph against the number of trainings for all the digraph chosen during the study. The blue continuous line (-), representing the .-t digraph shows the highest mean square error of range 15.5×10-3 to 107.6×10-3. The blue green dots line (.) shows the digraph of e-5 with a mean square range of $1.1 \times 10-3$ to $25.1 \times 10-3$. The red star line (\bigstar) shows digraph of 1-n with a lower mean square error (Range $7.4 \times 10-4$ to $5.0 \times 10-6$) compared to the digraph .-t. The green plus sign line represented by (+) which is hardly visible represents the lowest mean square error of digraph n-b (Range $8.5 \times 10-8$ to $3.7 \times 10-9$). To have a better view of the difference of the mean square error of the l-n and n-b digraph, since they are overlapping, the graph of Fig. 4 was plotted. The mean square error of digraph n-b is considerably less than the digraph .-t. These graphs were obtained after the 50 simulation for each digraph.



Fig 5. Mean Square error (MSE) against number of trainings of different digraph





if keystroke dynamics is a reliable means of authentication, with the evolution of technology, hackers have become so intelligent that they tend to mimic user's tying pace. In this research, it has been shown that the greater the distance of the keys, the more difficult it becomes for hacker to replicate the pattern. User were encouraged to use the new conform way to design password, proposed in this paper, which makes it more robust. Password must compromise of one alphabet, numbers, special character to increase its robustness and it is advised to use more distant keys so that it is more difficult for imposters to hack the keystroke pattern. It is advisable to change password every one month.

NNGT Int. J. on Information Security, Vol.5, February, 2016

REFERENCES

- D. Shanmugapriya and G. A. Padmavathi, "Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges," International Journal of Computer Science and Information Security, Vol. 5, No. 1, 2009.
- [2] N. Bartlow, "Username and Password Verification through Keystroke Dynamics," M. S. thesis, College of Engineering and Mineral Resources at West Virginia University, 2005.
- [3] L. O'Gorman, "Comparing Passwords, Tokens, and Biometrics for User Authentication," Proceedings of IEEE, vol. 91, no. 12, pp. 2019-2040, December 2003.
- [4] F. Monrose and A. Rubin, "Keystroke Dynamics as a biometric for authentication," Future Generation Computer Systems, DOI: 10.1016/S0167-739X(99)00059-X, vol. 16 no. 4, pp. 351-359, February 2000.
- [5] R. M. Lourde and D. Khosla, "Fingerprint Identification in Biometric Security Systems," International Journal of Computer and Electrical Engineering, Vol. 2, No. 5, October 2010.
- [6] R. Gaines, W. Lisowksi, W. Press, and S. Shapiro, "Authentication by keystroke timing: Some preliminary results," Rand Report R-256-NSF, The Rand Corporation, Santa Monica, CA, 1980.
 R. Javes and G. Gunta, "Identity Authentication Based on Kaystroke

R. Joyce and G. Gupta , "Identity Authentication Based on Keystroke Latencies," Communications of the ACM, DOI: 10.1145/75577.75582, vol. 33 no. 2, pp. 168-176, February 1990.

- [7] M. S. Obaidat and D. T. Macchairolo, "An online neural network system for computer access security," IEEE Transactions on Industrial electronics, DOI: 10.1109/41.222645, Vol. 40, no. 2, pp.235–242, April 1993.
- [8] J. R. Young and R. W. Hammon, "Method and apparatus for verifying an individual identity," Patent 4,805,222, U.S. Patent and Trademark Office, Washington, D.C., 1989.
- [9] M. S Obaidat and B. Sadoun, "Verification of computer users using Keystroke dynamics," IEEE Transactions on Systems, Man and Cybernetics, Part B, DOI: 10.1145/75577.75582, Vol. 27, no. 2, pp. 261-269, April 1997.
- [10] F. Monrose, M. Reiter and S.Wetzel, "Password Hardening Based on Keystroke Dynamics," International journal of Information security, pp. 1-15, 2001.
- [11] N. Bartlow and , B.Cukic, "Evaluating the Reliability of Credential Hardening through Keystroke dynamics," 17th international Symposium on Software Reliability Engineering, pp. 117-126, 2006.
- [12] F. Monrose, and A. Rubin, "Authentication via keystroke dynamic," Proceedings of the 4th ACM Conference on Computer and Communications Security, Zurich, Switzerland, April 1997, pp. 48-56.
- [13] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics" in ACM Transactions on Information and System Security, 5(4):367–397, 2002.
- [14] D. Gunetti and C. Picardi, "Keystroke analysis of free text," ACM Transactions on Information and System Security, 8(3):312–347, 2005.
- [15] S. Mondal, P. Bours, and S.Z. Idrus "Complexity of a Password for Keystroke Dynamics: Preliminary Study," in 6th International Conference on Security of Information and Networks (SIN'13), 2013, pp.301-305.

Fig. 6. Mean Square error (MSE) against number of trainings of ln and nb digraph

From the graph, it clearly shows that greater is the distance of the keys, it becomes more difficult for any system to learn the pattern. This conclusion has been drawn from the graph since the mean square error is more, for more distant keys and the vice versa. Hence to avoid hackers to intrude any system, it is advisable to use a password with a long distance between the lengths.

V. DISCUSSION AND FUTURE WORK

The results obtained were encouraging. From Mondal, et al.[15] work, a new metric for determining complexity has been found which were about the distance between the keys on the keyboard. Mondal et al. [15] have combined several metrics along with the distance metric, to measure complexity of password and observe the performance of the combined metric. This study was focused only on the distance between keys, even if a strong password was used. The results which were obtained show evidence that the greater the distance between the keys on the keyboard for the keystroke, the reliability of the password increases.

The caveats was that the data was collected without taking errors taking into consideration (i.e. the user was requested to type the password again if any letter was typed by mistake without taking the values of what will happen in real life). By the addition of the error correction, the accuracy of the keystroke digraph would be affected. Choosing password with distant keys in real life requires the user to analyse his own typing habit as in case the user uses the two hands to type, he tends to approach some keys faster compared to other keys. The door is open for researches to continue to analyse the distance between keys of password using dynamic password and taking the errors into consideration.

VI. CONCLUSION

Keystroke dynamics authentication excels compared to other biometric system in terms of supervision requirement, location independence, decentralisation and replicability. Even



- [16] K. Senathipathi, "An Analysis of Particle Swarm Optimization and Genetic Algorithm with Respect to Keystroke Dynamics," in Green Computing Communication and Electrical Engineering (ICGCCEE), Coimbatore, 2014, pp. 1–11.
- [17] C. Castelluccia, M. Durmuth and D. Perito, "Adaptive Password-Strength Meters from Markov Models," In Proceedings of Network and Distributed Systems Security Symposium (NDSS), The internet society, 2012.
- [18] T. A. Salthouse, "Perceptual, cognitive, and motoric aspects of transcription typing," in Psychological bulletin, 99(3):303, 1986. W. E. Cooper. Cognitive aspects of skilled typewriting. Springer, 1983.
- [19] T. A. Salthouse, "Effects of age and skill in typing," Journal of Experimental Psychology: General, 113(3):345, 1984
- [20] T. Sim and R. Janakiraman, "Are digraphs good for free-text keystroke dynamics?," In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–6, 2007.
- [21] Z. Syed, S. Banerjee, and B. Cukic, "Leveraging variations in event sequences in keystroke-dynamics authentication systems," In 15th International Symposium on High-Assurance Systems Engineering (HASE), pages 9–16, 2014.
- [22] J. Roth, X. Liu, A. Ross, and D. Metaxas, "Investigating the discriminative power of keystroke sound," in IEEE Transactions on Information Forensics and Security, 2014, pp.333-345
- [23] C. Epp, M. Lippold, and R.L. Mandryk, "Identifying emotional states using keystroke dynamics," In SIGCHI Conference on Human Factors in Computing Systems, pages 715–724, 2011.
- [24] Keystroke Dynamics., Biometric-Solutions.com," [online] 2013,http://www.biometricsolutions.com/solutions/index.php?story=keystroke_dynamics (Accessed: 5 November 2014)
- [25] A. K. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition," IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, vol. 14, no. 1, pp. 4-20, January 2004.
- [26] H. Mohabeer, K.M.S. Soyjaudah and N. Pavaday, "Enhancing The Performance Of Neural Network Classifiers Using Selected Biometric Features," SENSORCOMM 2011: The Fifth International Conference on Sensor Technologies and Applications, 2011.
- [27] E. Yu and S. Cho, "Keystroke dynamics identity verification- its problems and practical solutions," Computers & Security, DOI: 10.1016/j.cose.2004.02.004, Vol. 23, No. 5, pp 428-440, July 2004.
- [28] R. Fussell, "Authentication: The Development of Biometric Access Control," The ISSA journal, July 2005.
- [29] A. Meszaros, Z. Banko, and L. Czuni, "Strengthening Passwords by Keystroke Dynamics," in Proceedings of the 4th IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp.574-577, September 2007.
- [30] K. Nandakumar, Y. Chen, A. K. Jain and S. C. Dass, "Quality-based score level fusion in multibiometric systems," Proc. of ICPR, August 2006, vol. 4, pp. 473–476.
- [31] A. K. Jain, "Biometrics: Proving Ground for Image and Pattern Recognition," IEEE Image and Graphics Fourth International Conf. (IGIG07), DIO: 10.1109/ICIG.2007.195, August 2007, pp. 3-3.
- [32] W. Shen and T. Tan, "Automated biometrics-based personal identification," Proceedings of the National Academy of Sciences (PNAS99), DOI: 10.1073/pnas.96.20.11065, Vol. 96, pp. 11065-11066, September 1999.
- [33] L. Fernando, F. L. Podiol and J. S. Dunn, "Biometric Authentication Technology: From the Movies to YourDesktop," NIST, 100 Bureau Drive, Stop 1070, Gaithersburg, MD 20899-1070 [US Department of Commerce, 1401 Constitution Avenue, NW, washingtom, DC 20230], 2005.
- [34] A. Guven and I. Sogukpinar, "Understanding users' keystroke patterns for computer access security," Computers and Security, Elsevier, Vol. 22, pp. 695-706, 2003.
- [35] R. A. Wasniowski, "Using Data Fusion for Biometric Verification," World Academy of Science, Engineering and Technology International

Journal of Social, Management, Economics and Business Engineering Vol.1 No.5, 2007.

- [36] K. S Killourhy and R. A Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in Proceedings of IEEE/IFIP International Conference on Dependable Systems and Networks (DSN '09), July 2009, pp. 125-134.
- [37] S. Bleha, C. Slivinsky and B. Hussien, "Computer-Access Security Systems Using Keystroke Dynamics," IEEE Transactions On Pattern Analysis And Machine Intelligence, DOI: 10.1109/34.62613, Vol. 12, No. 12, pp. 1217-1222, 1990.
- [38] S. Cho, C. Han, D. H. Han and H. Kim, "Web-Based Keystroke Dynamics Identity Verification Using Neural Network," Journal of organizational computing and electronic commerce, DOI: 10.1207/S15327744JOCE1004_07, Vol. 10, No. 4, 295-307, December 2000.
- [39] J. A. Robinson, V. M. Liang, J. A. M. Chambers, and C. L. MacKenzie, "Computer User verification Using Login String Keystroke Dynamics," IEEE Transactions on systems, Man, and cybernetics, Part A: Systems and Humans, DOI: 10.1109/3468.661150, Vol. 28, No. 2, pp. 236-241, March 1998.
- [40] J. Owens, and J. Matthews, "A Study of Passwords and Methods Used in Brute-Force SSH-Attack," Technical Report, Department of Computer Science, Clarkson University, 2008.
- [41] X .De Carné de Carnavalet and M. Mannan, "From Very Weak to Very Strong: Analyzing Password-Strength Meters," Network and Distributed System Security (NDSS'14), San Diego, CA, US, 2014.
- [42] F. Bergadano, D. Gunetti and C. Picardi, "User authentication through keystroke dynamics," ACM Transactions on Information and System Security (TISSEC), DOI: 10.1145/581271.581272, vol. 5, no. 4, pp. 367-397, November 2002.
- [43] R. Gaines, W. Lisowski, S. Press, and N. Shapiro, "Authentication by keystroke timing: some preliminary results," Technical Report Rand Rep. R-2560-NSF, RAND Corporation, 1980.
- [44] F. Monrose and A.D. Rubin, "Keystroke dynamics as a biometric for authentication," Future Generation Computing Systems, 16(4):351–359, 2000.
- [45] D. Gunetti and C. Picardi, "Keystroke analysis of free text," in ACM Transactions on Information and System Security, 8(3):312–347, 2005.

