Keystroke Dynamics authentication using data filtering techniques and neural network approaches

D. Shamugapriya
Lecturer, Dept. of Information Technology
Avinashilingam Deemed University for women
Coimbatore, Tamilnadu, India
ds_priyaa@rediffmail.com

Dr. G. Padmavathi
Prof and Head, Dept. of Computer science,
Avinashilingam Deemed University for women,
Coimbatore, India,
ganapahti.padmavathi@gmail.com

ABSTRACT

Securing the secret data and computer systems by allowing ease access only to the authenticated users and enduring the attacks of imposters is one of the major challenges in the field of computer security. Traditionally, user name and password schemes are widely used for controlling the access to computer systems. But, this scheme has many flaws such as Password sharing, Shoulder surfing, Brute force attack, Dictionary attack, Guessing, Phishing and many more. Biometrics technologies provide more reliable and efficient means of authentication and verification. Keystroke Dynamics is one of the famous biometric technologies, which will try to identify the authentication of a user when the user is working with a keyboard. In this paper, neural network approaches with keystrokes for three different passwords namely weak, medium and strong passwords are taken into consideration. Neural Network algorithms are extended by applying normalization techniques for data filtering before performing the classification on the datasets. The performance of normalization based neural network algorithms is compared against neural network algorithms with all different category of passwords and the accuracy obtained is compared.

Keywords

Biometrics, Keystroke dynamics, Back propagation neural network, cascade forward back propagation neural network, Radial basis function, Min-max, Z-score, Tanh

1. INTRODUCTION

Almost all the people rely on computers at certain level in day today life. Many of these systems store highly sensitive, personal, commercial, confidential or financial data. Unauthorized access to such data will lead to loss of money or unwanted disclosure of highly confidential data by threatening the Information security. The first and foremost step in preventing unauthorized access of information for providing information security is user Authentication.

User authentication is categorized into three classes [21]: Knowledge - based, Object or Token – based and Biometric - based. The knowledge-based authentication is based on something one knows. The object-based authentication relies on something one has and the Biometric-based user authentication is based on something you are and depends on behavioral and physiological characteristics of individuals. The first two methods are more used but are very vulnerable.

In the first method the person can forget and can share their data. In the second method the object user possess can be lost or be stolen and in the third method the person presents a characteristic that cannot be forged and nor be forgotten. Biometrics involves something a person is or does and depends on the characteristics of the person. Biometrics is classified into physiological and behavioral biometrics [21]. Physiological biometric refers to what the person is and the Behavioral types are related to what a person does, or how the person uses the body. The figure 1. shows the classification of user authentication. Keystroke dynamics is considered as a strong behavioral biometric based authentication system [1]. It is a process of analyzing the way a user types at a terminal by monitoring the keyboard in order to identify the users based on habitual typing rhythm patterns. Moreover, unlike other biometric systems, which may be expensive to implement, keystroke dynamics does not require any sophisticated hardware as the only hardware required is the keyboard, which is universally available. Keystroke analysis contains two approaches: static and dynamic [14]. In static approach, the system checks the user one time that is at authentication time or login time. In the dynamic approach, the system checks the user continuously throughout the session. The approach used here is static since the authentication is done only during login time.

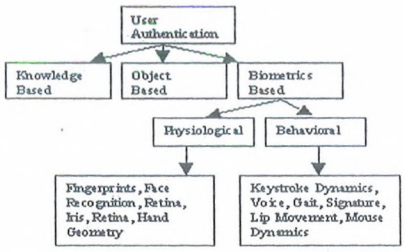

Figure 1. User Authentication classification.

The remaining of this paper is organized in four sections. Section 2 gives the works published in the area. In the section 3 the methodology is discussed. The experiments are presented and discussed in the section 4, and finally the conclusions are found in the section 5.

2. LITERATURE REVIEW

A number of studies [7,10,18-19,22-23,25-28,30] have been performed in the area of keystroke analysis since its inception. Table 1. illustrates a summary of the main research approaches performed in till date using static and neural network approaches.

Table 1. Keystroke approaches

Study	Classification Technique		Users	FAR %)	FRR (%)
Brown & Rogers [7]	Static	Neural network	25	0	12.0
Bleha & Obaidat) [31]	Static	Neural network	24	8	9
Obaidat & Sadoun [23]	Static	Neural network	15	0	0
Cho et al. [8]	Static	Neural network	21	0	1
Ord & Furnell [29]	Static	Neural network	14	9.9	30
Yu & Cho [12]	Static	Neural network	21	0	3.69
Gunetti & Picardi) [14]	Static	Neural network	205	0.005	5
Clarke & Furnell [9]	Static	Neural network	32	5 (Equal error rate)	
Lee and Cho [17]	Static	Neural network	25	0.43 (Average Integrated Error)	
Hawang et al [32]	Static	Neural network	25	4 (Equal error rate)	

3. PROPOSED METHODOLOGY

In the proposed methodology, there are three important phases involved in keystroke dynamics. First, a user registers or enrolls his/her timing Vector patterns. Second, a preprocessing is done. Third, neural network classifier is built using the timing vector patterns to measure the accuracy.

3.1 Registration or Enrollment

During the registration phase 26 users were asked to type three different passwords. Each user typed each password 10 times. Totally 780 samples were collected within a week time. Age group of users is between 18-2[illegible]. The three passwords used are 'pass_tie.R', '.tie5Roanl' and 'nopassword'. The password-strength checker [16] rates the above passwords as strong, medium and weak respectively.

3.2 Feature Extraction

The main function of feature extraction is to extract important features from the collected f raw keystroke data for template generation. There are many types of features that can be extracted from a human keystroke such as Duration, Latency, Digraph, Tri-graph, Pressure of keystroke, Force of Keystroke, Difficulties of typing text, Frequency of word errors, Typing rate, etc. However, not all kinds of the above-mentioned features are useful and widely used [30]. In order to measure keystroke pressure, a special type of pressure sensitive keyboard needs to be used. To measure Keystroke force, a special force sensitive resistor keyboard to be used. However, frequency of word errors, typing rate, and difficulties of typing text is only useful on long text. Since user will be providing only username and password, the features such as difficulties of typing text, frequency of word errors, typing rate, etc. are not suitable for the proposed work. Many works that were done in this area have measured only the Duration and Interval. In the proposed experimentation all the timing features such as Dwell time or Duration, Flight time or Latency, Digraph and Tri-graph are measured. The following Figure.3 shows the possible timing features that can be extracted for an 8-character password.

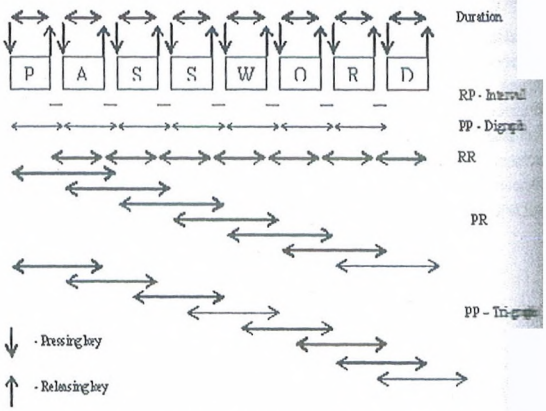

Figure 3. Measurement units of Keystrokes

3.3 Preprocessing

The typing pattern of the user varies from time to time even for the same user. The figure 4 shows the difference in typing pattern of same user. Four sample-typing patterns of same password by same user is shown.

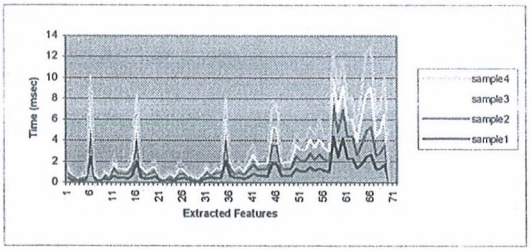

Figure 4. Difference in typing pattern of a User

The typing range of four sample-typing patterns which is shown is different and may not grant access to genuine user and provides low accuracy. Data filtering methods such as normalization may improve the accuracy and efficiency of algorithms involving neural networks classifiers. Such methods provide better results if the data to be analyzed have been normalized to specific ranges such as [0.0, 1.0]. Many normalization techniques have been proposed in the literature [2]. It includes Min–max, Decimal scaling, z-score, Median and MAD, Double sigmoid, Tanh-estimators, Biweight estimators and were tested with Face, Hand Geometry images and Finger print databases. For Keystroke dynamics, the previous works that involves filtering techniques has been tabulated in [21]. Joyce & Gupta [19] filtered the users whose typing times were highly variable or inconsistent during data collection and excluded from the study. Cho et al. [8] processed the collected timing data with an outlier-handling procedure to remove extreme values and also excluded the users from study whose typing times were highly variable or inconsistent. In this study, three different range normalization methods are considered: z-score normalization, min-max normalization, Tanh normalization.

3.3.1. Z-Score Normalization

The Z-Score normalization technique uses the mean and standard deviation for each feature across a set of training data to normalize each input feature vector. This method transforms the scores having some Gaussian distribution to a standard Gaussian distributional form. The mean and standard deviation are computed for each feature and then the normalization is done as given in eq.1

$$X_i' = \frac{(X_i - \mu(x))}{\sigma(x)} \text{ -------(1)}$$

where $\mu(x)$ and $\sigma(x)$ are the mean and standard deviations of the feature and x_i is the ith sample of the feature.

3.3.1 Min-Max Normalization

There may be a need in neural network to constrain the range of each input feature or each output. This is done by rescaling the features of output from one range of values to a new range of values. Most often the features are rescaled to lie within a range of 0 to 1 or from –1 to 1. The rescaling is accomplished by using min max formula as given in equation 2.

$$X_i' = \frac{X_i - \min}{\max - \min} \text{ ——(2)}$$

3.3.2 Tanh Normalization

This method is one of the robust statistical techniques. It maps the raw scores to the (0, 1) range using the equation (3).

$$n = \frac{1}{2}\left[\tanh\left(0.01\frac{(X_i - \mu(x))}{\sigma(x)}\right) + 1\right] \text{ ——(3)}$$

The effect of preprocessing of sample of a user using the week password is shown in figure 5.From the above graph it is shown clearly that the preprocessing brings the score range between 0 and 1, which remove the ambiguity of the obtained scores.

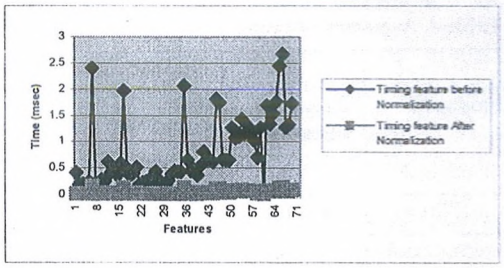

Figure 5. Effect of Min-max normalization

3.4 Classification

The preprocessed timing vectors are classified using three different neural network approaches namely back propagation neural network (BPN), Cascade forward back propagation neural network(CFBPN) and radial basis function (RBF) and the obtained classification accuracy are compared.

4. EXPERIMENTAL RESULTS

The collected samples of all the three passwords are preprocessed using three different normalization methods and feed into BPN, CFBP and RBF for classification. The following Table 2, 3 and 4. shows the obtained accuracy using BPN, CFBP and RBF respectively

Table 2. Accuracy obtained using BPN

		Normalization Methods		
Password	**Password category**	**Z-Score**	**Min-max**	**Tanh**
Pass_tie.R	Strong	78	71	72
.tie5Roanl	Medium	77	69	69
nopassword	Weak	69	69	69

Table 3. Accuracy obtained using CFBPN

		Normalization Methods		
Password	**Password category**	**Z-Score**	**Min-max**	**Tanh**
Pass_tie.R	Strong	69	68	71
.tie5Roanl	Medium	68	73	71
nopassword	Weak	61	67	67

Table 4. Accuracy obtained using RBF

		Normalization Methods		
Password	**Password category**	**Z-Score**	**Min-max**	**Tanh**
Pass_tie.R	Strong	71	70	70
.tie5Roanl	Medium	63	69	76
nopassword	Weak	65	63	65

From the above table it is shown that BPN with z-score normalization and with strong password provides higher accuracy i.e 78% than the other methods. The works that were done in this area so far has used only Duration or Interval as features but the proposed method uses all the timing measurements of the obtained Keystroke. The accuracy of the proposed work can be improved by applying feature reduction techniques like Genetic Algorithm. A single password or any kind of password typed by the user was used in the previous studies. In this experiment passwords are rated as strong, medium and weak and tested for accuracy.

5. CONCLUSION

A system is designed for user multifactor authentication combining the traditional password with the keystroke dynamics. Keystroke of strong, medium and weak passwords are measured and the accuracy is calculated using neural network approaches by preprocessing the captured data using different data filtering methods and results obtained show that keystroke in combination with strong password, Z-score normalization Back propagation neural network gives better accuracy. Since the strong passwords are hard to remember, user may store it in a database or write down the passwords, which may lead to dictionary attacks when hacked by imposters. A password combined with keystroke dynamics provides more security. Even when the other person knows the passwords, he cannot steal the typing rhyme of the user, which adds more security.

6. REFERENCES

[1] Ahmed Awad E. Ahmed, and Issa Traore, Anomaly Intrusion Detection based on Biometrics, Proceedings of 6th IEEE Information Assurance Workshop, (2005), pp.452- 453.

[2] Anil jain, Karthik Nandakumar, Arun Ross, Score normalization in multimodal biometric systems, pattern recognition, 38, (2005.), pp 2270-2285.

[3] Anil K. Jain, Arun Ross and Salil 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, 2004.

[4] Attila Meszaros, Zoltan Banko, Laszlo Czuni Strengthening Passwords by Keystroke Dynamics, IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications , (2007), pp 6-8.

[5] Benny Pinkas, Securing Passwords against Dictionary Attacks, Proceedings of the 9th ACM conference on Computer and communications security, (2002), pp 161 – 170.

[6] Bergando et al., User Authentication through keystroke Dynamics, ACM transaction on Information System Security, Vol.No. 5, (2002.),pp. 367-397.

[7] Brown, M., Rogers, J, User Identification via Keystroke Characteristics of Typed Names using Neural Networks, International Journal of Man-Machine Studies, vol. 39,(1993), pp. 999-1014.

[8] Cho et al., Web based keystroke dynamics identity verification using neural network, Journal of organizational computing and electronic commerce, Vol. 10, No. 4, (2000), 295-307.

[9] Clarke, N. L. and Furnell, S.M., Authenticating mobile phone users using keystroke , International Journal of Information Security, 6 (1), (2007), pp.1-14.

[10] D.-T. Lin, Computer-access authentication with neural network based keystroke identity verification, Proceedings of IEEE Int. Conf. On Neural Networks, vol. 1, (1997), pp. 174-178.

[11] Downland, P. and Furnell, S., A long-term trial keystroke profiling using digraph, trigraph and keyboard latencies, Proceedings of IFIP/SEC 19th International Conference on Information Security, (2004), pp 289.

[12] Enzhe Yu, Sungzoon Cho, Keystroke dynamics identity verification and its problems and practical solutions, Computers & Security, (2004.).

[13] Glaucya C. Boechat et al., Using the Keystrokes Dynamic for Systems of Personal Security, Proceedings Of World Academy Of Science, Engineering And Technology, Vol. 18, (2006).

[14] Gunetti and Picardi, Keystroke analysis of free text, ACM Transactions on Information and System Security, Vol 8, (2005), pp. 312–347.

[15] Guven, A. and I. Sogukpinar, Understanding users' keystroke patterns for computer access security, Computers & Security, Vol 22, (2003), pp. 695–706.

[16] http://www.microsoft.com/protect/yourself/password/checker.mspx.

[17] Hyoungjoo Lee, Sungzoon Cho, Retraining a keystroke dynamics-based authenticator with impostor patterns, Computers & Security, 26(4), (2007), pp. 300-310.

[18] John A. Robinson, Vicky M. Liang, J. A. Michael Chambers, and Christine L. MacKenzie, Computer User verification Using Login String Keystroke Dynamics, IEEE transactions on systems, man, and cybernetics—part a: systems and humans, Vol. 28, No. 2, (1998).

[19] Joyce R., Gupta, G., Identity Authentication Based on Keystroke Latencies. Communications of the ACM, Vol. 39, (1990), pp 168-176.

[20] Kevin S. Killourhy and Roy A. Maxion, Comparing Anomaly Detectors for Keystroke Dynamics, Proceedings of the 39th Annual International Conference on Dependable Systems and Networks, IEEE Computer Society Press (2009), pp. 125-134.

[21] Lawrence O'Gorman, Comparing Passwords, Tokens, and Biometrics for User Authentication, Proceedings of the IEEE, Vol. 91, No. 12, (2003), pp. 2019-2040.

[22] Leggett, J., Williams, G., Usnick, M., Dynamic Identity Verification via Keystroke Characteristics, International Journal of Man-Machine Studies, (1991).

[23] Mohammad S. Obaidat, Balqies Sadoun, Verification of computer users using keystroke dynamics, IEEE Transactions on Systems, Man, and Cybernetics, Part B 27(2), (1997), pp. 261-269.

[24] Monrose, F., Reiter, M., Wetzel, S., Password Hardening Based on Keystroke Dynamics, International journal of Information Security, (2001), pp.1-15.

[25] Monrose, F., Rubin, A., Authentication via Keystroke Dynamics, Proceedings of the 4th ACM Conference on Computer and Communications Security, (1997), pp. 48-56.

[26] Monrose, R., Rubin, A.,Keystroke Dynamics as a Biometric for Authentication, Future Generation Computer Systems, 16(4), (1999), pp 351-359.

[27] Napier, R., Laverty, W., Mahar, D., Henderson, R., Hiron, M., Wagner, M. , Keyboard User Verification: Toward an Accurate, Efficient and Ecological Valid Algorithm, International Journal of Human-Computer Studies, vol. 43, (1995), pp213-222.

[28] Obaidat, M. S., Sadoun, B., Verification of Computer User Using Keystroke Dynamics, IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics, Vol. 27, No.2, (1997).

[29] Ord, T., Furnell, S., User Authentication for Keypad-Based Devices using Keystroke Analysis, MSc Thesis, University of Plymouth, UK., (2000).

[30] Pin Shen Teh Teoh, et al., Statistical Fusion Approach on Keystroke Dynamics, Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, (2007).

[31] S Bleha and M S Obaidat, Computer user verification using the perceptron, IEEE Trans. Systems, Man, and Cybernetics, vol. 23, No. 3, 900–902(1993)

[32] Seong-soeb Hwang, Sungzoon cho, Sunghoon park, Keystroke dynamics based authentication for mobile phones, Computers & Securit, 85-93, (2009)

AUTHORS

Shanmugapriya. D, received the B.Sc. and M.Sc. degrees in Computer Science from Avinashilingam University for Women, Coimbatore in 1999 and 2001 respectively. And, she received the M.Phil degree in Computer Science from Manonmaniam Sundaranar University, Thirunelveli in 2003 and pursuing her PhD at Avinashilingam University for Women. She is currently working as a Lecturer in Information Technology in the same University and has nine years of teaching experience. Her research interests are Biometrics, Network Security and System Security.

Dr. Padmavathi Ganapathi is the professor and head of Department of Computer Science, Avinashilingam University for Women, Coimbatore. She has 22 years of teaching experience and one year Industrial experience. Her areas of interest include Network security and Cryptography and real time communication. She has more than 100 publications at national and International level. She is a life member of many professional organizations like CSI, ISTE, AACE, WSEAS, ISCA, and UWA.