A Method for Preprocessing the Dwell Time and Flight Time of Biometric Keystroke Templates

Dr.G.Padmavathi[1],
ganapathi.padmavathi@gmail.com
Ms.D.Shanmugapriya[2]
ds_priyaa@rediffmail.com
Avinashilingam University for Women, Coimbatore – 43.

ABSTRACT

Securing the sensitive data and computer systems by allowing ease access to authenticated users and withstanding the attacks of imposters is one of the major challenges in the field of computer security. Traditionally, ID and password schemes are most 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. The Uniqueness of Biometrics for any specific human being provides more reliable and efficient means of authentication and verification. Keystroke Dynamics is one of the famous and inexpensive behavioural biometric technologies, which will try to identify the authenticity of a user when the user is working via a keyboard. The objective of the paper is to preprocess the obtained dwell time and flight time using Z-score in order to obtain the tolerable False Acceptance Rate and False Rejection rate.

Keywords: Biometrics, Entropy, False Acceptance Rate, False Rejection rate, Keystroke, Security.

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 the process of verifying claimed identity. The authentication is accomplished by matching some short-form indicator of identity, such as a shared secret that has been pre-arranged during enrollment or registration for authorized users. This is done for the purpose of performing trusted communications between parties for computing applications.

Conventionally, user authentication is categorized into three classes [17]:

Knowledge - based,
Object or Token - based,
Biometric - based.

The knowledge-based authentication is based on something one knows and is characterized by secrecy. The object-based authentication relies on something one has and is characterized by possession. The Biometric-based user authentication is based on something you are and depends on behavioural and physiological characteristics of individuals. There are many different examples of biometric authentication: it can be based on fingerprints, iris scanning, voice analysis, handwriting dynamics, keystroke dynamics and so on.

In knowledge-based and object-based approaches, passwords and tokens can be forgotten, lost or stolen. There are also usability limitations associated with them. For instance, managing multiple passwords / PINs, and memorizing and recalling strong passwords are not easy tasks. Biometric-based person recognition overcomes the above mentioned difficulties of knowledge-based and object based approaches. Biometric authentication is fundamentally different from the other two classes because it does not rely on secrets. Rather, it relies on registering and later matching what are believed to be distinguishing physical or behavioural characteristics of individuals.

Biometrics involves something a person is or does. These types of characteristics can be approximately divided int physiological and behavioural types [17]. Physiological characteristics refer to what the person is, or, in other words, the measure physical parameters of a certain part of the body. Some examples are: Fingerprint, Hand Geometry, Vein Checking, Iri Scanning, Retinal Scanning, Facial Recognition, Facial Thermo gram.

Behavioural characteristics are related to what a person does, or how the person uses the body. Some good examples of thi group are: Gait recognition, Signature Recognition, Mouse Dynamics, Keystroke Dynamics, Voice

1.1. Keystroke Dynamics as Biometric

Keystroke dynamics is considered as a strong behavioural biometric based authentication system [1]. It is a process of analyzin 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 is almost fre as the only hardware required is the keyboard. Some the features that are extracted using keystroke are- the way a person type the cumulative typing speed, time that elapses between consecutive strokes, time that each key is held down, frequency of th individual in using other keys on the keyboard, such as the number pad or function keys and the sequence that is utilized whe typing a capital letter for example, and whether the individual release the shift key or the letter key first. Using these features template is created forming a statistical profile of the individual's behavioural characteristics. The following Figure 1 shows th steps involved in keystroke authentication. The registration process captures the keystrokes and calculates the typing parameter and the verification process realizes the keystroke capture, the feature extraction and the comparison with the stored templat and authenticating the user.

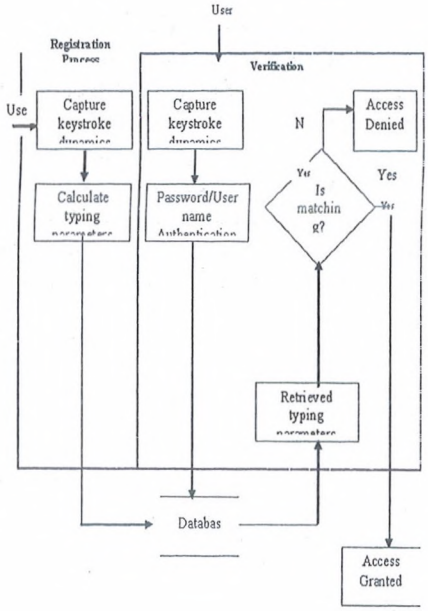

Fig. 1: Steps Involved in Keystroke Authentication

Keystroke dynamics system can run in two different modes [2] namely the Identification mode or Verification mod Identification is the process of trying to find out a person's identity by examining a biometric pattern calculated from the person' biometric features. A person's identity is checked in the verification case. The pattern that is verified is only compared with th person's individual template. There are two metrics used to verify the identity of a user—Duration and Latency (see Figure 2 Duration or dwell time measures how long a given key is pressed. Latency or flight time measures the time elapsed betwee releasing one key and pressing the next [5]. Analyzing the results of keystrokes statistically can produce surprisingly accurat results.

Fig. 2: Keystroke Duration and Latency

The rest of the paper is organized as follows: The section 2 summaries the background work done in this area. The methodology is explained in Section 3 and the section 4 concludes the paper.

2. Background

A number of studies [5,7,10,12,20-22,27,29] have been performed in the area of keystroke analysis since its inception. There are two main keystroke analysis approaches for the purposes of identity verification. They are statistical techniques and neural networks techniques. Some are the combinations of both the approaches. The basic idea of the statistical approach is to compare a reference set of typing characteristics of a certain user with a test set of typing characteristics of the same user or a test set of a hacker. The distance between these two sets (reference and test) should be below a certain threshold or else the user is recognized as a hacker. Neural Networks process first builds a prediction model from historical data, and then uses this model to predict the outcome of a new trial (or to classify a new observation). Although the studies tend to vary in approach from what keystroke information they utilize to the pattern classification techniques they employ, all have attempted to solve the problem of providing a robust and inexpensive authentication mechanism. Table 1 illustrates a summary of the main research approaches performed till date.

Study	Classification Technique		Users	FAR(%)	FRR(%)
Joyce & Gupta (1990) [16]	Static	Statistical	33	0.25	16.36
Leggett et al. (1991) [18]	Dynamic	Statistical	36	12.8	11.1
Brown & Rogers (1993) [6]	Static	Neural Network	25	0	12.0
Bleha & Obaidat (1993) [27]	Static	Neural Network	24	8	9
Napier et al (1995) [23]	Dynamic	Statistical	24	3.8 (combined)	
Obaidat & Sadoun (1997) [19]	Static	Statistical	15	0.7	1.9
		Neural Network		0	0
Monrose& Rubin (1999) [22]	Static	Statistical	63	7.9 (combined)	
Cho et al. (2000) [7]	Static	Neural Network	21	0	1
Ord & Furnell (2000) [25]	Static	Neural Network	14	9.9	30
Bergadano et al. (2002) [5]	Static	Statistical	154	0.01	4
Guven & Sogukpinar(2003) [13]	Static	Statistical	12	1	10.7
Sogukpinar & Yalcin(2004) [29]	Static	Statistical	40	0.6	60
Dowland & Furnell (2004) [9]	Dynamic	Neural Network	35	4.9	0
Yu & Cho (2004) [10]	Static	Neural Network	21	0	3.69
Gunetti & Picardi (2005) [12]	Static	Neural Network	205	0.005	5
Clarke & Furnell (2007) [8]	Static	Neural Network	32	5 (Equal Error Rate)	
Lee and Cho (2007) [14]	Static	Neural Network	21	0.43 (Average Integrated Erros)	
Pin shen The et al (2008) [27]	Static	Statistical	50	6.36 (Equal Error Rate)	
Hawang et al (2009) [28]	Static	Neural Network	25	4 (Equal Error Rate)	

Table 1: Approaches in Keystroke Analysis

3. Methodology

The proposed method consists of Enrollment module and preprocessing module. In the Enrollment module the user duration and latency is captured and in preprocessing module the captured parameters are analyzed by introducing the z-score.

3.1. Enrollment Module

This module captures and analyses the features of the users typing rhythm for template creation. To capture the keystroke 25 users are asked to type the weak password "computer science" 10 times. The 250 samples are got in a week period. The parameters duration or dwell time and latency or flight time are captured from all the users. The mean and standard deviation values are measured with the following equations.

$$\text{Mean } \mu(x_i) = (1/N)\sum x[i] \text{ where } i=1 \ldots N \quad (1)$$

$$\text{Standard deviation } \sigma_i = (1/N-1)\sum | x[i] - \mu[i] | \text{ where } i=1 \ldots N \quad (2)$$

3.2. Preprocessing Module

A degree of inconsistency exists in users typing rhythms. While some individuals may be highly consistent, others are not. This causes problems during classification due to large spurious outliers. In order to reduce these problems, the data is pre-processed before the template is created. To achieve this, the z-score values for each sample in the training set with respect to its class were calculated with the following equation.

$$\text{Z-score } Z_i = (x_i - \mu(x))/\sigma(x) \quad (3)$$

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.

Values that fall outside the neighbourhood of z are eliminated and replaced with the mean of the rest of the feature samples. Limit is placed on the allowable number of outliers for an ideal template. The following table 2 and table 2 gives the Effect of preprocessing Dwell time of sample keystroke obtained from a user.

Dwell Time before Preprocessing (ms)	Preprocessed Dwell Time (ms)	Z Score	Corrected Z Score
579	579	0.87	0.87
468	520	-0.76	0.00
499	520	-0.31	0.00
392	520	-1.88	0.00
485	520	-0.51	0.00
595	595	1.10	1.10
546	546	0.38	0.38
498	520	-0.32	0.00
624	624	1.53	1.53
516	520	-0.06	0.00

Table 2: Effect of Preprocessing Dwell Time of Sample Keystroke Obtained from a User

Flight Time before Preprocessing (ms)	Preprocessed Flight Time (ms)	Z Score	Corrected Z Score
4344	4344	0.12	0.12
1219	4159	-1.88	0.00
5156	5156	0.64	0.64

Flight Time before Preprocessing (ms)	Preprocessed Flight Time (ms)	Z Score	Corrected Z Score
3953	4159	-0.13	0.00
3625	4159	-0.34	0.00
3719	4159	0.28	0.00
4077	4159	-0.09	0.00
7594	4159	2.19	0.00
3922	4159	-0.15	0.00

Table 3: Effect of Preprocessing Flight Time of Sample Keystroke Obtained from a User

The following Figure 3 and 4. shows the distribution of a pre-processed and the otherwise keystroke. From the table 2 and from the figure 3, it is shown that after preprocessing the dwell time calculated using Z score is between 500 to 600. The table 3 and the Figure 4 shows that the flight time calculated using Z score is between 4100 to 5100. After preprocessing the obtained dwell time and the flight time are classified using the classifier and the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) can be calculated. The preprocessing of the parameters are very important in obtaining accurate FAR and FRR in keystroke dynamics.

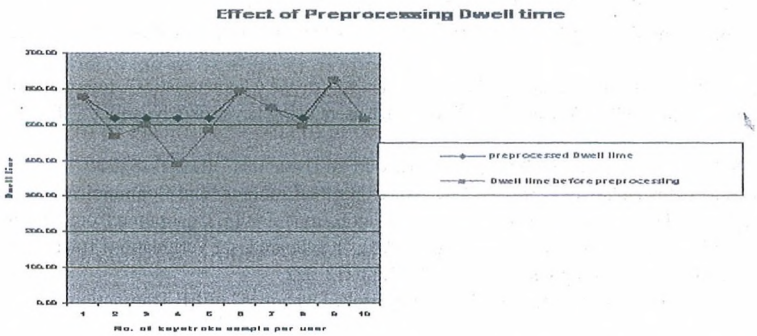

Fig. 3: Effect of Preprocessing Dwell Time (in Milli second) of Sample keystroke Obtained from a User

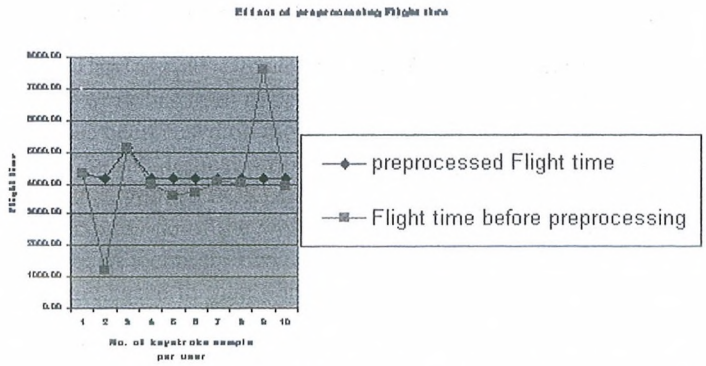

Fig. 4: Effect of Preprocessing Flight Time (in Milli second) of Sample keystroke Obtained from a User

4. Conclusion

Preprocessing of the Keystroke parameters are very important in obtaining accurate FAR and FRR in keystroke dynamics. In this paper, the parameters are preprocessed using Z-score and the values that fall outside the neighbourhood of Z-score are eliminated and replaced with the mean of the rest of the feature samples. The preprocessed parameters are classified using the classifier and the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) can be calculated.

REFERENCES

1. Ahmed Awad E. Ahmed, and Issa Traore (2005), Anomaly Intrusion Detection based on Biometrics, Proceedings of 6th IEEE Information Assurance Workshop: 452- 453.
2. Anil K. Jain, Arun Ross and Salil Prabhakar. January 2004."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
3. Attila Meszaros, Zoltan Banko, Laszlo Czuni. September 2007."Strengthening Passwords by Keystroke Dynamics", IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 6-8
4. Benny Pinkas (2002), Securing Passwords Against Dictionary Attacks, Proceedings of the 9th ACM conference on Computer and communications security: 161 - 170
5. Bergando et al, "User Authentication through keystroke Dynamics", ACM transaction on Information System Security" Vol.No. 5, Nov 2002, pg 367-397
6. Brown, M., Rogers, J. (1993). "User Identification via Keystroke Characteristics of Typed Names using Neural Networks". International Journal of Man-Machine Studies, vol. 39, pp. 999-1014
7. Cho et al. 2000. "Web based keystroke dynamics identity verification using neural network". Journal of organizational computing and electronic commerce, Vol. 10, No. 4, 295-307
8. Clarke, N. L. and Furnell, S.M. (2007) 'Authenticating mobile phone users using keystroke analysis' International Journal of Information Security, 6 (1): 1-14.
9. Downland, P. and Furnell, S. (2004). A long-term trail of keystroke profiling using digraph, trigraph and keyword latencies. In proceedings of IFIP/SEC 19th International Conference on Information Security pages 275-289
10. Enzhe Yu, Sungzoon Cho. (2004). "Keystroke dynamics identity verification and its problems and practical solutions", Computers & Security
11. Glaucya C. Boechat, Jeneffer C. Ferreira, and Edson C. B. Carvalho, Filho. December 2006. "Using the Keystrokes Dynamic for Systems of Personal Security", Proceedings Of World Academy Of Science, Engineering And Technology, Volume 18
12. Gunetti and Picardi. 2005 " Keystroke analysis of free text", ACM Transactions on Information and System Security, volume 8, pages 312–347.
13. Guven, A. and I. Sogukpinar (2003), Understanding users' keystroke patterns for computer access security, Computers & Security 22, 695–706.
14. Hyoungjoo Lee, Sungzoon Cho. (2007). Retraining a keystroke dynamics-based authenticator with impostor patterns. Computers & Security 26(4): 300-310
15. John A. Robinson, Vicky M. Liang, J. A. Michael Chambers, and Christine L. MacKenzie. March 1998. "Computer User verification Using Login String Keystroke Dynamics", IEEE transactions on systems, man, and cybernetics—part a: systems and humans, Vol. 28, No. 2
16. Joyce R., Gupta, G. (1990). Identity Authentication Based on Keystroke Latencies. Communications of the ACM, vol. 39; pp 168-176.
17. Lawrence O'Gorman. Dec. 2003. "Comparing Passwords, Tokens, and Biometrics for User Authentication", Proceedings of the IEEE, Vol. 91, No. 12, pp. 2019-2040
18. Leggett, J., Williams, G., Usnick, M. (1991). "Dynamic Identity Verification via Keystroke Characteristics". International Journal of Man-Machine Studies.
19. Mohammad S. Obaidat, Balqies Sadoun. April 1997. "Verification of computer users using keystroke dynamics", IEEE Transactions on Systems, Man, and Cybernetics, Part B 27(2): 261-269
20. Monrose, F., Reiter, M., Wetzel, S. 2001. "Password Hardening Based on Keystroke Dynamics", IIJS, 1-15 21. Monrose, F., Rubin, A. April 1997. "Authentication via Keystroke Dynamics", Proceedings of the 4th ACM Conference on Computer and Communications Security, p 48-56.
22. Monrose, R., Rubin, A. (1999). "Keystroke Dynamics as a Biometric for Authentication". Future Generation Computer Systems, 16(4) pp 351-359.
23. Napier, R., Laverty, W., Mahar, D., Henderson, R., Hiron, M., Wagner, M. (1995). "Keyboard User Verification: Toward an Accurate, Efficient and Ecological Valid Algorithm". International Journal of Human-Computer Studies, vol. 43, pp213-222.
24. Obaidat, M. S., Sadoun, B. (1997). "Verification of Computer User Using Keystroke Dynamics". IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics, Vol. 27, No.2.
25. Ord, T., Furnell, S. (2000). "User Authentication for Keypad-Based Devices using Keystroke Analysis". MSc Thesis, University of Plymouth, UK.
26. Pin Shen Teh Teoh, A. Thian Song Ong Han Foon Neo (2007). Statistical Fusion Approach on Keystroke Dynamics. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System.
27. S Bleha and M S Obaidat. May 1993. "Computer user verification using the perceptron," IEEE Trans. Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 900–902.
28. Seong-soeb Hwang, Sungzoon cho, Sunghoon park, "Keystroke dynamics based authentication for mobile phones", Computers & Security (2009), pages 85-93.
29. Sogukpinar. I, Yalcin. 2004. "User identification at logon via keystroke dynamics", Journal of Electrical and Electronics Engineering, Vol. 4, No. 1, 995-10