International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2015): 6.391

Keystroke Authentication with Back Propagation Neural Network

Qaswaa K. Aboud

[1]Lecturer, Computer Science Department, College of Science, University of Baghdad, Al-Jadriya, Baghdad, Iraq

Abstract: ***Security threats to computers and networks have increased in considerably because of the increasing use of computer systems and networks in almost every aspect of our daily lives; traditionally, key stroke dynamic authentication is vastly used to authenticate legitimate user in the current system but this method has many loop holes. The aim of this paper is to enhance the continuous authentication method by presenting a keystroke dynamics with back propagation neural network as a transparent layer of user authentication, this biometric technologies is of the most well known and not costly behavior. This paper utilizes keystroke features including Dwell time (DT), Flight time (FT), Up-Up time (UUT), Down-Down time (DDT) and a mixture of these features as strong new keystroke feature. The Back Propagation neural network is used with two cases: case-1, used the Back Propagation neural network with sigmoid function, the second case used the Back Propagation neural network with bipolar function, both of these two cases used with five different cases of features as mentioned above. These times are used to distinguish between the authentic users and impostors. Results of the experiments demonstrate that the Back Propagation network with mixed four features in case-1 with mixed of four features is more efficient comparable to case-2 and provide low False Accept Rate (FAR) and False Reject Rate (FRR) and high accuracy***

Keywords: Back propagation, Biometrics, Continuous Keystroke, Dwell Time, Flight Time, Up-Up Time, User Authentication.

1. Introduction

There is sharply increasable in number of computer users and so too has the use of internet applications such as online banking services.

All internet applications demand the user to use an authentication plan to make sure only the veritable individual can login to the application [1-3]. User authentication is the procedure of establishing claimed identity. This is done to perform trusted in communications between parts for applications compute. Biometrical mechanism is the robust way to authenticate people, that is, to verify their identity [4].Because biometric characteristics are exclusive to each person and could not be stolen, lost or forgotten. However, they usually require expensive hardware to support the dedicated function [5]. Biometrics can be classifies into two types, behavioral and physiological. The first type includes features that possible the people to learn it in a stable environment. Examples in this type are walking, and typing on a keyboard. The second type includes physically features that are related to a person, for examples, DNA, and finger print. The authentication via keystroke is based on the concept that each user has a keystroke latency pattern is diverging about others Keystroke dynamics, which fall into two types (static and continuous) [6].

This paper looks at continuous keystroke dynamics as a strategy for authentication offer. Keystroke dynamic is a process of dissecting keyboard typing properties or keyboard typing rhythms by observing keyboard inputs. In other words, the system verifies how a person types [7].This paper is systematic as follows. Section 2 exhibit related work. A section 3 crystallizes the features of key stroke dynamics. Section 4 demonstrates how the data was collected. The proposed approach presents in Section 5. Finally the results and the conclusion are given in sections 6 and 7 respectively.

2. Related Work

This section reviews some of the related work in area of keystroke dynamics.

Monrose et al. [8] confirmed that keystroke distinguishing depend on static text is more dependent than dynamic text. This is because of number of parameters for example: the parameters of environment un controlled, un constrained inputs, and the user was un cooperated. The data of keystroke were obtained from 63 users during 11 months. Two features of keystroke dynamics extracted duration and keystroke latency.

Cho et al. [9] a neural network was implemented to recognize between authorized and not authorized user. This work depends on two keystroke features: duration and key press time, in addition the experiment included 21 users.

Dowland et al. [10] obtained the typing style samples of 5 users by monitoring the activity of personal device, without any constraints imposed on the users, such as asking users to print predefined text. In this work select2-graphs features only, then compute the mean and standard deviation of 2-graphs latency for each user profile.

D'Souza [11] a statistical approach was used to identify user depend on her/him printing dynamics. The experiment included10users, for each user require to repeat the username, password and a predefine text for 10 times. The key press time and the flight time were considered.

Joshi [12] a neural network was used to classify authorized user and not authorized. The extract keystroke features was used for 43 users in their experiment.

Licensed Under Creative Commons Attribution CC BY

3. Keystroke Features

There are several different features which can be extracted when the user presses keys on as listed below [13-14]:

a) Dwell time (DT) or Duration: The interval time between a key pressed until it is released.
b) Flight time (FT) or latency: The interval time between a key release and the successive key press.
c) Up-Up time (UUT): The interval time between a key release and the successive key release.
d) Down-down time (DDT): the interval time between two successive key presses.
e) Pressure of keystroke used when hitting keys while typing.
f) Finger placement the place where the finger is placed on the key or even the tip of the finger when pressing the key (in this case a camera is recommended).
g) Finger choice which finger is used for which key of the keyboard.

4. Data Collection

A total of 15 users participated in the experiments of proposed system. The participated users are divided into two groups: the first one contains 10 users as authentic users while a second group contains 5 users' impostors. Initially, each participant had to register the determined paragraph three times continuously during a session to produce one paragraph. All participants were requested to enter the same paragraph three times for each one of five separated sessions, since the objective was to determine whether the proposed approach could identify and differentiate a particular user from the rest. Note that the participants were not informed of the data collection and analysis approach. During the data collection phase, a user typed the paragraph "***my name is suha , I'm in Baghdad university, college of science, computer science department***" during two weeks for three times within five session because there is a chance that when the user types down the same paragraph over and over again the typing speed may increase. Thus, a database of 15 user profiles, where each profile contained number of samples of keystroke features (timing vector) was measured in milliseconds. The entrants were asked to traineeship typing the paragraph beforehand. Moreover, the user typed the paragraph for 5 sessions during two weeks. The length of the timing vector is different and depending on the length of the paragraph and the type of feature used. For example, a part of selected paragraph "***my name***" which contains seven characters will result in seven DT, six FT, six UUT, and six DDT. Mostly, a paragraph with (n) character will donate (n) number of DT and (n – 1) number for FT, UUT, and DDT. Figure (1) illustrates BP Structure of DT.

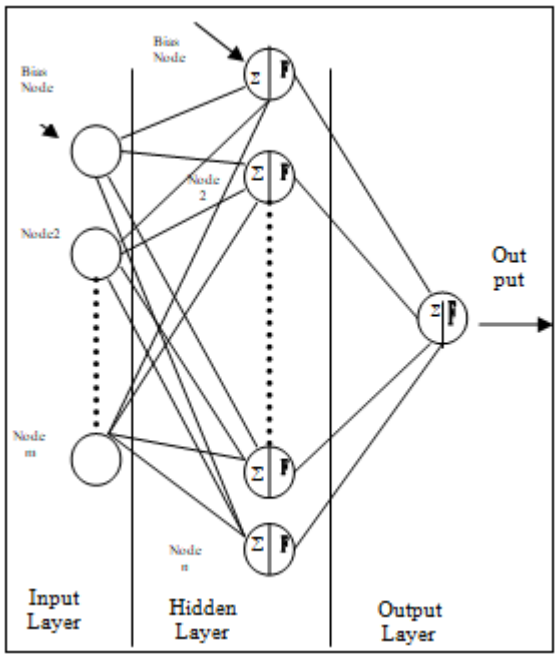

Figure 1 : BP Structure of DT

5. Proposed System

The authentication via continuous keystroke is based on the notion that each user posse's individual typing dynamics, this paper proposes a continuous keystroke dynamics user authentication by exploiting Back Propagation (BP) neural network algorithm with two different cases: case-1 using sigmoid function, and case-2 using bipolar function, and for each case using with two experiments for each one applied with five situations of features (DT, FT, DDT,UUT, and combination of 4 features). The following subsections will explain the basic steps of this proposal.

5.1 Preprocessing

Preprocessing step is done to obtain single template for each user that enable the proposed keystroke dynamics with neural network to identify between legitimate user and imposter one, with lower rate of error. First, compute the mean of timing vector for each user. Then, normalize the timing vector for each feature.

5.2 Back Propagation Neural Network (BP)

Classification is to find the best rating that is very close to the classified pattern. BP neural network algorithm is used to group the features in classification phase. A BP neural network is a form of supervised learning for Multi-Layer Perception (MLP). The MLP network consists of several layers of neurons; typically an input layer, hidden layers, and an output layer. Input layers take the input and distribute it to the hidden layers. These hidden layers do all the substantial calculations and output the results to the output layer, which redirects the data to the user as shown in figure (2).

Licensed Under Creative Commons Attribution CC BY

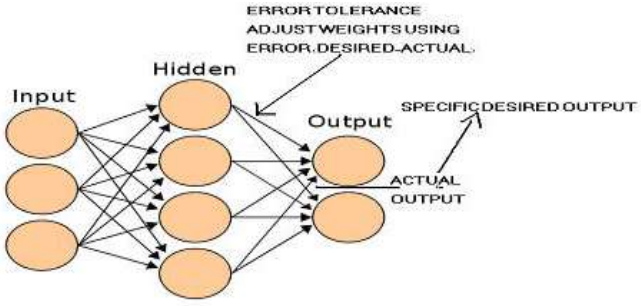

Figure 2: Structure of Back propagation Network

Error data was back propagated to earlier ones from the output layer, allowing incoming weights to these layers to be updated. The BP algorithm has been vastly used as a learning algorithm in feed forward multilayer neural networks [15]. A set of timing vectors from each user class was demand to train the neural network,. These timing vectors are composed for each user and stored in databases. The timing vector are averaged and normalized to form a set of patterns that will be passed to train the network. The number of patterns is 15. The training consists of taking timing vectors as a contribution, Compare the current output with the target output, and regulate the weight values according to the Bb training algorithm. Training was stopped when the error of the training vector set is reduced to a pre-defined threshold which is the total summed squared error less than or equal to threshold. Algorithm -1 illustrates BP steps.

Algorithm -1: BP

Input: Set of training timing information examples (X) and target vector (Y)

Learning Rate η

Momentum α

Stopping criterion ε

Output: Weight: W_1 and W_2

Step1: [Initialization]

Number of input layer node (I) according to length timing information.

Number of hidden layer nodes (H) to $2*I+1$

Number of output layer nodes (O) to 1.

Weights between input layer and hidden layer (W_1) to small random values.

Weights between hidden layer and output layer (W_2) to small random values.

Iteration Number t=0

Step2: [Compute activation function (Sigmoid) for all hidden nodes

$$h_j = \frac{1}{1 + e^{-\sum_{i=0}^{I} w1_{ij} * x_i}}, \quad for \quad j = 1 .. H$$

Step3: [Compute activation function (Sigmoid) for output nodes]

$$o_j = \frac{1}{1 + e^{-\sum_{i=0}^{H} w2_{ij} * h_i}}, \quad for \quad j = 1 .. O$$

Step4: [Compute errors in the output nodes]

$$\delta 2_j = o_j (1 - o_j)(y_j - o_j), \ for \ j = 1..O$$

Step5: [Compute errors in the hidden nodes]

$$\delta 1_j = h_j (1 - h_j) \sum_{i=1}^{O} \delta 2_i \times w2_{ij}, \ for \ j = 1..H$$

Step6: [Adjust weights between the hidden layer and the output layer

$$w2_{ij}^{t+1} = w2_{ij}^{t} + \eta * \delta 2_j * h_i + \alpha \Delta w_{ij}^{t}$$

Step7: [Adjust weights between the input layer and the hidden layer

$$w1_{ij}^{t+1} = w1_{ij}^{t} + \eta * \delta 1_j * x_i + \alpha \Delta w1_{ij}^{t}$$

Step8: [Stopping criterion]

$$\text{If } \frac{1}{2} \sum_j (y_j - t_j)^2 \prec = \varepsilon$$

Then **End**

Else increment iteration number $t = t + 1$ and GO TO **step 2**

The above algorithm BP in case-1 with sigmoid function, in case-2 the same algorithm with replace activation function to bipolar as shown in following equation:

$$o = (\frac{1}{1 + e^{-x}}) * 2 - 1$$

6. Experimental Result

This section investigates the performance of the proposed continuous keystroke dynamics with the most popular features DT, FT,DDT, UUT, and 4F(combination of four features) for satisfying user authentication based on keystroke dynamics by using BP algorithm with comparing between sigmoid and bipolar functions .Two experiments are performed independently on the paragraph writing three times at a session (***"my name is suha , I'm in Baghdad university, college of science, computer science department "***) five times during two weeks using BP with two functions with five features of keystroke dynamics to distinguish between authentic users and impostors with the parameters adjusted as follows:

a) Number of input layer nodes (IN) depends on the password length and the feature(s) that are used plus 1 node as bias.
b) Number of hidden layer nodes (HN) is set to. 1*2+IN
c) Number of output layer nodes (ON) is set to 1 since the output is either authentic users or impostors
d) Learning rate is set to 0.1 (η)
e) Momentum is set to 0.9 (α)

In the first experiment, the proposed approach was tested on the same patterns that neural net was trained on. The second experiment deals with test the proposed keystroke dynamic online. This experiment includes calculating the specific feature (s) for each key he/she typed then concern preprocessing on the computed vector time and finally test the approach The testing involves when the user types the paragraph incase-1(sigmoid function) and when the user types the paragraph in case-2 (bipolar function). For each case the system operates fifty second for each two minutes to check the authenticity of the current typing user, if he/she authorized continue else the system will be stopped. To evaluate the performance of the proposed continuous

Licensed Under Creative Commons Attribution CC BY

keystroke dynamics, three metrics are used including the false rejection rate (FRR) (i.e. the rate at which users are rejected when they could be authenticated), false acceptance rate (FAR) (i.e. the rate at which users are accepted when they should be rejected) and accuracy (i.e. the proportion of true results in the population). Tables (1- 2) in addition to the figures (3 -6) illustrate the results of the proposed continuous keystroke dynamics when each user types paragraph "***my name is suha , I'm in Baghdad university, college of science, computer science department***" three times for each one of five sessions during two weeks.

Table 1: Experiment-1

Metric	*Bipolar function*					*Sigmoid function*				
	DT	FT	DDT	UUT	All	DT	FT	DDT	UUT	*All*
FAR%	3	1	2	1	1	0	0	0	0	*0*
FRR%	1	3	1	2	1	1	1	0	0	*0*
ACC%	96	96	97	97	98	99	99	100	100	*100*

Table 2: Experiment-2

Metric	*Bipolar function*					*Sigmoid function*				
	DT	FT	DDT	UUT	All	DT	FT	DDT	UUT	All
FAR%	5	5	4	2	1	4	3	2	2	0
FRR%	7	4	3	2	2	3	2	1	1	1
ACC%	85	90	92	96	97	92	95	97	97	99

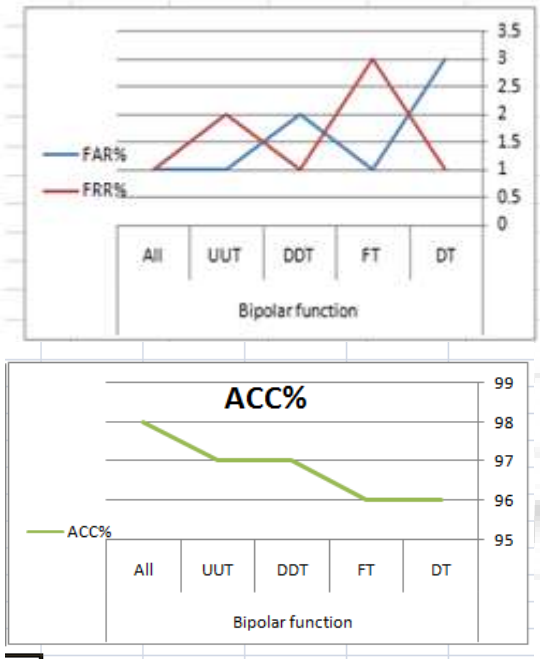

Figure 3: expermain1 case1

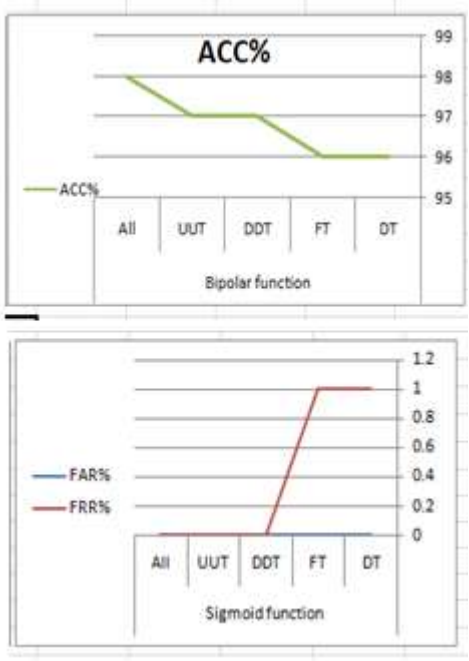

Figure 4: expermain1 case2

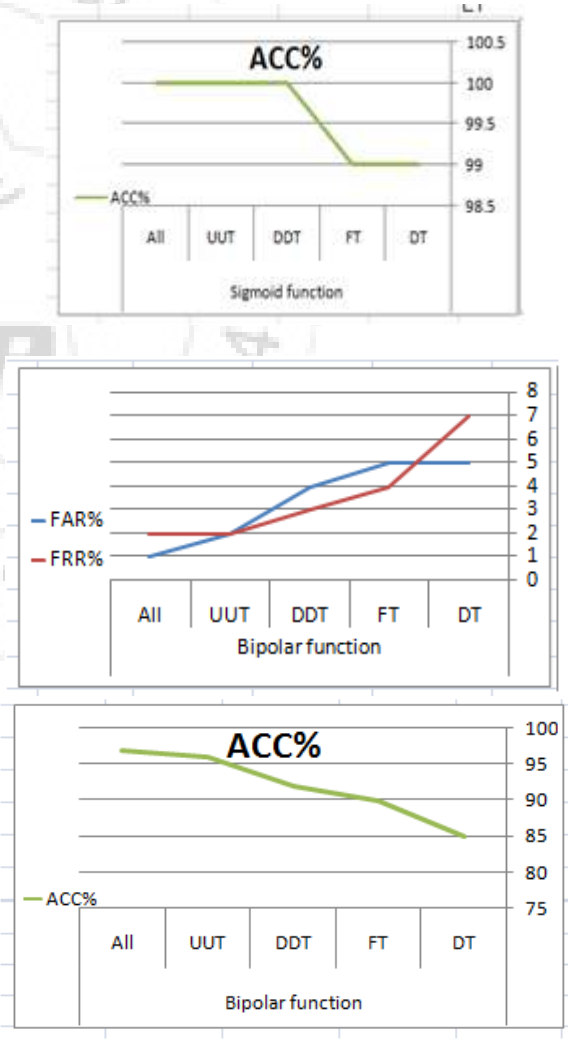

Figure 5: expermain2 case1

Licensed Under Creative Commons Attribution CC BY

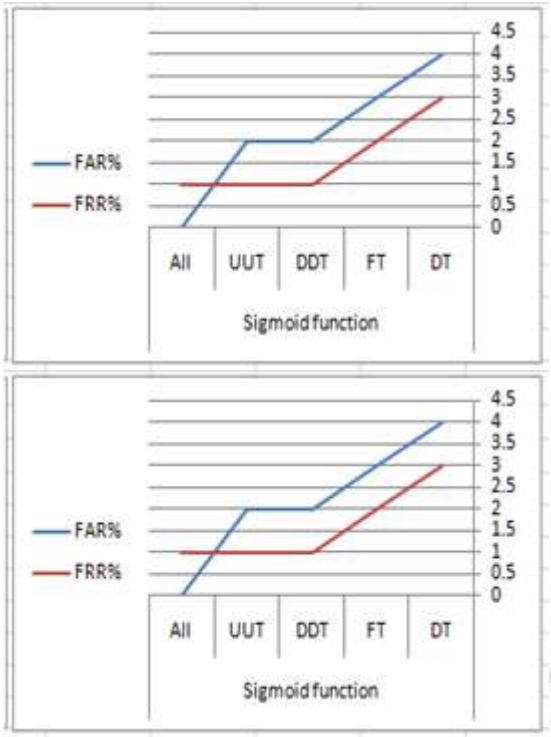

Figure 6: expermain2 case2

7. Conclusions

The primary goal of the proposed continuous keystroke dynamics is to prevent intruders from personifying legitimate users (a low FAR is desired), while also ensuring that legitimate users are not refused (a low FRR is also desired) and the accuracy of the system is high. This paper extracted the DT, FT, DDT, UUT, and combination of four features (4F) from participate users. The BP neural network was applied on different these situations of features to analysis the performance of these features with different cases (sigmoid and bipolar functions). The tested results demonstrate that the combination of four features with BP neural network with sigmoid function and bipolar function gives lower FAR, FRR and higher accuracy than the rest features. Also, using BP neural network in this field with sigmoid function is better than using bipolar function. In addition the UUT result is better than DT,FT, and DDT. On the other hand the result of combination of DT,FT, DDT and UUT features is best among the results. So, the testing results demonstrated that the BP neural network with sigmoid function was able to determine weights that distinguish the authentic users and impostors with low FAR, low FRR and high accuracy when using the combination of DT, FT, DDT, and UUT.

References

[1] A. Jain. and R. Jain, "Information Fusion in Biometrics", Pattern Recognition Letters, Vol. 24, No. 13, pp. 2115-2125, Elsevier, 2003
[2] U. Dieckmann and R.W. Frischholz "BioID: A Multimodal Biometric Identification Systems" IEEE Computer, Vol. 33, No. 2, pp. 64-68, 2000.
[3] Shanthi Bhatt and T.Santhanam."Keystroke dynamics for biometric authentication — A survey IEEE Pattern Recognition, Informatics and Mobile Engineering (PRIME),2013 International Conference on".
[4] S. Mandujano. and R. Soto " Deterring password sharing: User Authentication via fuzzy c-means clustering applied to keystroke profiles, Proceedings of the ENC International Conference on Computer Science (ENC '04), pp. 181-187, ed. IEEE Computer Society, Colima, Mexico, September 2004.
[5] Vishalim.Patel,.RemaChellappa,.DeepakChandra,.Brand onBarbello."Continuous User Authentication on Mobile Devices: Recent progress and remaining Challenges IEEE Signal Processing Magazine Volume: 33, Issue: 4, July 2016.
[6] G.A. Carpenter, M.A. Rubin and W.W. Streilein "ARTMAP-FD: Familiarity Discrimination Applied to Radar Target", Proceeding of the International Conference on Neural Networks, pp. 1459–1464, 1997.
[7] D. Davis and W. Price, "Security for Computer Networks", John Wiley & Sons, Inc., 1989.
[8] F. Monrose and A. D. Rubin "Keystroke Dynamics as a Biometric for Authentication", Future Gener Compute Syst Vol. 16m No.4, pp. 351–359, 2000.
[9] S. Cho,C. Han,D. H.Hee and H. Il Kim, "Web based Keystroke Dynamics Identity Verification Using Neural Network", Journal of organizational computing and electronic commerce, Vol. 10, No. 4, pp. 295-307, 2000.
[10] P. Dowland, S. Furnell, and M. Papadaki, "Keystroke Analysis as a Method of Advanced User Authentication and Response", Security in the Information Society: Visions and Perspectives, pp. 215, 2002.
[11] D. C. D' Souza. "Typing Dynamics Biometric Authentication", Bachelor engineering thesis, Faculty of Engineering and Physical Sciences, University of Queensland, Australia, 2002.
[12] D. Shanmugapriya and G. Padmavathi "Virtual Key Force - A New Feature for Keystroke Dynamics", International Journal of Engineering Science and technology, Vol. 13m No. 10, pp. 7738-7743, 2011.
[13] P.S. Tee, T.S. Ong and, A. B. J. Teoh "A Multilayer Layer Fusion approach on Keystroke Dynamics", Springer, Pattern Anal Applic, Vol. 14, pp. 23-36, 2011.
[14] H. Barghouthi "Keystroke Dynamics How Typing Characteristics Differ From One Application to Another", Master of Science in Information Security, Gjøvik University College, Norway, 2007.
[15] P. McCollum, "An Introduction to Back Propagation Neural Networks", the Newsletter of the Seattle Robotics Society, 1997.

Author Profile

Qaswaa K.Aboud received the B.Sc. in computer and control engineering from the University of Technology in 1989. She received her M.Sc.in computer engineering from the University of Baghdad in 2005. She received her scientific title "*lecturer*" in 2012. She is a faculty member in the Computer Science Department since 1989.

Licensed Under Creative Commons Attribution CC BY