How to improve performance of Neural Network in the hardened password mechanism

Narainsamy Pavaday, Insah Bhurtah and Dr. K.M.Sunjiv Soyjaudah, Member IEEE

Abstract— A wide variety of systems, ubiquitous in our daily activities, require personal identification schemes that verify the identity of individual requesting their services. A non exhaustive list of such application includes secure access to buildings, computer systems, cellular phones, ATMs, crossing of national borders, boarding of planes among others. In the absence of robust schemes, these systems are vulnerable to the wills of an impostor. Current systems are based on the three vertex of the authentication triangle which are, possession of the token, knowledge of a secret and possessing the required biometric. Due to weaknesses of the de facto password scheme, inclusion of its inherent keystroke rhythms, have been proposed and systems that implement such security measures are also on the market. This correspondence investigates possibility and ways for optimising performance of hardened password mechanism using the widely accepted Neural Network classifier. It represents continuation of a previous work in that direction.

Keywords— Biometric, Neural Networks, Perceptron and Fuzzy.

I. INTRODUCTION

It is fact that the first aspect to consider in computer system security is authentication as other components, such as access control, audit log and many others hinge on it. It guarantees that entities accessing system resources are what they claim to be. The most extended mechanism that covers the authentication process is the simple password and it is likely to stay for a long time ahead due to a number of reasons. It is straightforward to implement, easy to use and maintain, its precision can be adjusted by enforcing structure or hygiene policies such as inclusion of special characters, be of a minimum length and changed every week or so. Moreover it represents an inexpensive and scalable way of validating users, both locally and remotely, to all sorts of services [1, 2]. Username and password combinations have a fundamental flaw stemming from human psychology. Passwords should be easy to remember and provide swift authentication. Additionally in terms of security the password should be difficult to guess, changed from time to time, and unique to a single account [3]. These stringent requirements as well as the larger number of systems they access, make people adopt unsafe practices to remember their passwords. People record them on pieces of papers, near to the authentication devices and tell them to their confident among others. Furthermore, as technology increases, attacks targeting passwords (dictionary and brute force attacks) are becoming easier. Therefore, it is primordial to use alternative mechanisms to reinforce the former one. One such alternative is the exploitation of biometric features which are intimately linked to every individual. As human beings we have characteristics that help identify us from others. Our genetic code, fingerprints, handwriting, and ocular retinal pattern are examples of biometric features that make us unique and distinguishable as individuals. Society has and still relies on the written signature to verify the identity of an individual. The

Mr. N. Pavaday is now with the Computer Science University of Mauritius having previously done his training with the Biometric Lab, School of Industrial Technology, University of Purdue West Lafayette, Indiana, 47906 USA, (phone: +230-465-4780; e-mail: n.pavaday@uom.ac.mu).

Prof Dr. K.M.S.Soyjaudah is with the same university as the first author. He can also be contacted on the phone +230 4541041 ext 1367 (e-mail: ssoyjaudah@uom.ac.mu).

Ms Insah Bhurtah is a student reading for an MPhil/PhD in the field of communication at the University of Mauritius.

DOI: 10.5176_2010-2283_1.2.42

©2011 GSTF
dexterity and complexity of the human hand and its environment make written signatures highly characteristic and difficult to forge precisely. The handwritten signature has a parallel on the keyboard. The same neurophysiological factors that make a written signature unique are also exhibited in a user’s typing pattern as revealed by a number of researchers [4]. When a person types on a keyboard, he/she leaves a digital signature in the form of keystroke latencies. Latency between keystrokes, keystroke durations and finger positions can be used to uniquely identify a user. This soft biometric is not as precise as others in terms of entropy and classification power [5] but has the advantage of not requiring costly equipment and software to be implemented and can therefore be used to strengthen password-based authentication. As any other biometric it is free from unsafe storage, loss, forgery, cloning and associated memory problems.

The current paper started with an introduction to authentication with a focus on passwords schemes and the enhanced variant known as keystroke dynamics. After the motivation paragraph, the next section then concentrate on a study of the MLP/BP model and a review of previous work placing emphasis on the application of Neural Network to keyboard dynamics. Section 3 explains the different concepts behind the selected authentication mechanisms. Before discussing the results obtained with various system parameters, the methodology adopted for the experiment is detailed. Finally the paper ends with a conclusion as well as further work to be undertaken following the current one.

II. MOTIVATION

Previous works had some short comings in that they do not pay attention to the time required for training the model and the preprocessing required before the result becomes available. We should consent that such an authentication system has to be instantaneous and integrate seamlessly in existing passwords mechanism for it to achieve widespread acceptance and hence use. This work extends the previous work carried out in using Neural Network (NN) for enhancing the lifetime of passwords mechanism in a cheap and unobtrusive way even when the latter loses its secrecy [6]. The aim of this paper is to determine the parameters of the neural networks so that the best results can be obtained. Our focus is on the number of training data that yield the best performance, the transfer functions to be used as well as the number of neurons among others.

III. RELATED WORK

Since the uniqueness of a user’s typing pattern was first reported by Joyce and Gupta [7], work has progressed in using typing behavior as an authentication tool. Chronologically it kicked off with statistical classifier more particularly the T test by Gaines and his colleagues [8]. Statistical models and digraph latencies were the pioneers for some time and even had two patents issued [9]. The first approach to include the then new neural network (NN) was brought about by Brown and Rogers [10] where they used a simple multiple layer perceptron (MLP) with back propagation (BP) and they received a patent for their method [10,11]. Lin [12] extended the work involving MLP/BP by considering the variation on the structure and parameters of the neural network with a modified keystroke latency being used to compensate for cases when the second key is pressed before the first one is released, typical of experienced typists [12]. Obaidat and Sadoun performed a side-by-side comparison of five statistical classification methods and eight neural network paradigms [13]. Their results clearly favored neural networks over statistical methods and hence the current trend towards the use of the latter technology as the classifier. By altering the activation function for the hidden layer and the learning rule, Obaidat and Sadoun also achieved a near ideal performance with a Multiple Layer Perceptron/Back Propagation (MLP/BP) neural network using a sigmoid transfer function. The MLP/BP using a sine-delta transfer function was ranked as one of the worst classifiers, along with the Counter-Propagation neural network (CPNN) which had very high error rates. In a previous work dated 1994, Obaidat and Macchiaiolo [14], presented three different approaches using neural network based on the key interval. A multilayer feedforward network
trained using the BP algorithm, a sum of product (SOP) network trained with a modification of the back propagation and a hybrid of the two which achieved a performance of 97.5% for classification of users who were previously categorized. Similarly Wong compared the classification capabilities between ANN and KNN [15]. A preprocessing was first performed to remove unwanted data and noise. He showed that the neural network with the Hebbian Learning rule performed much better than the K nearest neighbor approach with k set to 1. Bleha tried using the single perceptron algorithm [16]. Mention on its limitations for access control systems due to high training requirements (e.g., time consuming) when a new user is added was revealed by Monrose [17]. A combination of the statistical, a neural network, and a fuzzy classifier were combined to achieve optimum performance in [18].

Concerning the variants of Neural Network a comparison between ADALINE (based on the single perceptron model) and the BP model, using both the hold time and digraph latency time, concluded that the BP surpass the ADALINE which was not capable of classifying patterns [18]. Capuano tried sorting the problem using the MLP with the transfer function based on the Radial Basis Function (RBF) instead of the sigmoid one previously used by others [19]. Obaidat and Sadoun, reported 0% error FRR (when FAR=0) using both the imposter and owner model and password of length of length 7 characters [13]. They considered both the hold time and interval between keystrokes as input to the system. The result was however verified by Bechtel et al who concluded that with 10% impostor rate obtained on the same ART2 neural network, the result is clearly impractical [9]. Another work worth mentioning is the one where the authors applied keystroke dynamics to strengthen a code of six digits using Neural Network with a multi-layer perceptron for each user with back-propagation [20]. This NN approach has also been used in java applet for secure web based transaction [21]. Recently a new NN variant of the auto associative neural network approach proposed by Cho [22] was revealed which use support vector machine (SVM) novelty detector with Genetic algorithm (GA) [21]. A table comparision of the performance of the different classifiers in different variants can be found in [23] while in [31] a keystroke dynamic sytem using Ant Colony Optimization and Neural Network in BP mode is detailed.

IV. BACKGROUND THEORY

Neural Network (also called Artificial Neural Networks) is a method of computation and information processing that takes advantage of today's technology by mimicking the architecture found in biological neurons. They are used in areas of pattern recognition, modelling and prediction to emulate the human brain which can perform tasks of high degree of complexity. Unlike mathematical models that require precise knowledge of all parameters and their interrelations, neural networks can provide an estimation of the parameters under various conditions without a precise knowledge of all contributing variables and their relations.

Artificial neural networks are constructed with artificial neurons that result in the formation of "layers" which are interconnected [24] as shown below (figure 1). A network consists of an input layer to receive inputs from the external environment and distribute them to the hidden layers. These hidden layers do all the necessary computations based on the

connection weights have to be set to indicate the effect that
this neuron exert on others connected to it. A NN learns by updating its architecture and connection weights iteratively from the supplied training patterns to achieve optimum performance. According to [25], an ANN learns by determining how close the output of the network is to the desired one in supervised learning. The Back Propagation learning process works in small iterative steps: one of the example cases (training sample) is applied to the network, and the network produces some output based on the current state of its synaptic weights. This output is compared to the known-good output, and usually the mean-squared error (MSE) signal is calculated. The error is propagated backwards into the network so as to alter the weights associated with each path so that the desired output is obtained in which case MSE is zero. The weight changes are calculated to reduce the error signal for the case in question. The whole process is repeated for each of the example cases and then back to the first case again, and so on. The cycle is repeated until the overall error value drops below some pre-determined threshold which ideally should be zero.

In the forward pass each hidden node receives a net input represented by the summation of all its input connection.

\[ X_j = \sum W_{jk} L_k \]  

(1)

Where \( j \) represents the node number and \( W_{jk} \) stands for weight \([\text{layer}]_{\text{number}} \) to \([\text{number}] \) connecting that particular node to all others.

This is repeated for all the hidden nodes (\( N_h \)) in that layer. Each node will then output:

\[ Y_j = f(X_j) \]  

(2)

Where function \( f(.) \) stands for the transfer function of that neuron.

Assuming one hidden layer as in most previous studies each output node thus receives

\[ X_{j+1} = f(\sum W_{jk} f(\sum W_{jk} I_k)) \]  

(3)

Consider the network shown in figure 1 with \( N_i \) input variable and one hidden layer of \( N_h \) nodes. This will imply that for each hidden node we have \( N_i \) computations to make and for each output \( N_o \). Thus for one such network of \( N_o \) outputs the required number increases to \( N_o N_i N_h \).

The difference between the desired output and the one obtained using different weight on each node is termed the error (E).

\[ E = \frac{1}{2} \sum \sum (d_i - Y_j) \]  

(4)

For all output units (i) over all input pattern (j)

In the backward pass the output unit error is used to alter weights as depicted in equation 4. The error at the hidden nodes is calculated (by back-propagating the error at the output units through the weights). The learning rate dictates the percentage of the error which is used to alter the weights. Clearly this will yield a total of \( 2 N_o N_i N_h \) computations for one epoch with \( N_i \) samples. The whole process starts again and again until the data can be correctly analyzed by the network after a long number of epochs. A thorough mathematical analysis of the model presented can be found in [26] and [27].

V. METHODOLOGY

For our experiment the sensor module that acquires biometric user data is the keyboard. Specifically two distinct variables are the hold and flight (dwell) times which are the amount of time you hold down a particular key and flight time which is the amount of time it takes a person to move between keys. For the system to work, it is primordial to obtain accurate timing information with sufficient resolution. The software based approach was therefore adopted for our experiments with a view of making it as simple as entering the password. A set of related programs was implemented in Microsoft Visual Basic and MatLab for analysis and investigation purposes. The nntool in MatLab was used for the NN part while a Visual Basic toolkit with a basic interface allowed user to enter a text string and recorded the timing information at the nearest millisecond in the background. A good tutorial on the use of Neural Network in Matlab can be found in [30]. The calculated
values as well as the captured one were stored in a text file for future use once the practical session was completed. The participants were informed the purpose of the experiment and given ample time to practice the desired password so as to emulate real world condition as far as possible. All intervals were captured with a counter monitoring the number times the password “Thurs1day” was entered correctly. Using the password mentioned above we obtained 8 keystrokes interval and 9 keystroke duration times neglecting the “Enter” key which was considered to be unstable.

To construct a reference signature or template for each user, we followed an approach similar to that used by the banks and other financial institutions. A new user goes through a session where he/she provides a number of digital signatures by typing in the same strings a number of times. For authentication purposes the person is allowed access if the correct password is typed and the captured keystroke is close to the reference signature stored based on the matching score obtained.

VI. RESULTS

For the MLP/BP approach the data collected was normalized as used in [28] and then passed to the NN discussed before and the parameters were varied and the observations noted as detailed.

The initial weights and bias were initialized randomly with the error level set to 0.01.

As mention in [6, 12] the transfer function used in the neural network were the sigmoid functions (tan-sigmoid and log-sigmoid functions) which achieved better performance than the sine delta function. Figure 2 shows the variation of the learning with varying number of neurons in 1 hidden layer. Figure 3 below shows the variation of the network with different values of the learning rate, proportion of the error which is propagated backward to alter the weights and bias of the nodes. The last line labeled ‘d.variable’ demonstrates how an increase in the number of inputs nodes improves the performance. From the graph about 20 epochs only is required to achieve the same performance reached previously by 55 epochs with learning rate at 0.8 with all other values remaining constant. It is to be noted that when changing the learning rate we reached a scenario where the learning became erratic. As demonstrated in [6], the increase in input nodes merely reflects the fact that enriching the network capability can only aid in the classification process as more components are involved in the classification tasks. The next step in the investigation phase was to vary the number of training data used to build the user template and find the optimum number of training attempts required to reach minimum MSE error. Ideally using a minimum number of training is aimed as it prevents the participant from getting bored and also minimizes processing time. In a typical password scheme the user is asked to type the password twice except if both attempts are not identical.
The graph in figure 4 demonstrates the results obtained using different numbers of trainings with a network consisting of two hidden layers. Our investigation has shown that the best performance is obtained with 7 trainings per user when using 20 neurons in each of the two hidden layers with a mean square error (MSE) of 0.0000000603. The line data of 7-t represents the situation where 7 training data per user (7-t) is being considered.

Close investigation of the different graphs however reveals that the 10-t graph has already reached that same level of MSE with about 14 neurons and at 20 neurons the performance is not optimized. The negative portions of the lines have been shown only for demonstrating the shapes of the graphs as these errors cannot have negative values. The graph in figure 5 shows the variation of time taken for 20 iterations when the NN is learning the training data. As expected the values shows that an increase in the number of training data together with an increase in number of neurons allows the network to learn faster. It is seen that time increases in a cubic manner with increasing number of neurons. With less training data the time taken is of the order of few second which is clearly impractical for authentication systems. With 5 to 10 training samples the value is close to 1 second as recommended in [29]. In figure 6 the number of iterations was increased up to 100 iterations to show variation of MSE in training process.

It is also observed that for a particular number of neurons, the best validation performance decreases exponentially with an increase in the number of neurons (figure 5). It has nearly the same shape of figure 2 which shows the number of incorrectly identified attempts as the number of iterations is increased.

Here the value of MSE is plotted with increasing iterations number, the network was trained 20 times, but the exponential decrease is more clearly seen with 100 trainings of the network. The figure pertains to variation for 7-t with 20 neurons. The table 1 below shows the best MSE reached by the network with different combinations of transfer functions in the two hidden layers.

Clearly with same transfer function in the two hidden layers of the neural network the optimum MSE reached is better and the best value is obtained with 20 neurons in each of the two hidden layers with tansig function. The value reached is much
lower than that achieved with 1 hidden layer and same number of neurons.

<table>
<thead>
<tr>
<th>TRANSFER FUNCTION</th>
<th>10 NEURONS</th>
<th>20 NEURONS</th>
<th>30 NEURONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tansig</td>
<td>0.0033302</td>
<td>0.0000000603</td>
<td>0.00019842</td>
</tr>
<tr>
<td>Logsig</td>
<td>0.013879</td>
<td>0.019841</td>
<td>0.011808</td>
</tr>
<tr>
<td>Tansig</td>
<td>0.016777</td>
<td>0.029154</td>
<td>0.011225</td>
</tr>
</tbody>
</table>

MSE for various combination of transfer function in layer 1 and 2

VII. CONCLUSION
In our study more than 100 students and staffs were involved. They were required to type the same word “Thurs1day” a number of times they wish. Ten users were selected to be authentic users and attempts collected after profile creation amounted to 5440. Some of these were genuine attempts while others impostor attempts. The important observation is that the sample size of 7 training was performing better than 10 training because the latter is not at its optimal operating point. A selection of seven combinations out of 10 combinations gave better performance than the whole lot taken together. As an example for one user the MSE was 6.0289 exp -08 for seven selected training. The same user had an MSE of 4.5044 exp -04 for 10 training data. Achieving a lower MSE with less training data remains an area that will get our focus in a forthcoming work. This was expected since the 20 neurons per layer had a lower value for MSE for the 10 training per user. This may be due to the latter exhibiting memory or overtraining.

Most of the works published in this field propose a neural network model that is first trained using the timing vectors of the owner’s keystroke dynamics and then used to solve the owner v/s impostor conflict. In an open system where users may join the system, one will have to be retrained each time a new user joins in placing a heavy burden on the system administrator. Given the inherent characteristic of Neural Network it will always match an impostor to one of the authentic users. Our work has shown that for a given network the parameters have to be varied in order to get the optimal configuration to use. Common sense would favour the use of maximum training captured to build the template but our results proves the contrary as we have already moved aside of the optimal point.

The other intriguing fact about the NN is that as the number of nodes in 1 hidden layer network was increased from 5 to 25, some users which were previously correctly identified were mismatched with higher probabilities, resulting in a degradation of performance. The time taken to train the network was not a major drawback as it was of the order of 1 second comparable to waiting time when assessing online services though the web. This clearly demonstrates the importance of a careful selection of system parameters for such an implementation. A cubic relation exists between time taken and number of neurons before reaching optimal point for MSE. This was expected as derived in previous equations.

As mentioned previously an intruder detection unit placed before the Neutral Neural network is primordial to enhance its usability and acceptability. By removing the intruder and presenting only authentic users to the neutral network a better performance can be achieved even with sample consisting of fewer attempts.

ACKNOWLEDGMENT

Once again the authors are grateful to the users have willingly participated in our experiment without whom our work would not have existed.

REFERENCES


