A machine learning approach to keystroke dynamics based user authentication

Kenneth Revett*

Harrow School of Computer Science,
University of Westminster, London, UK
E-mail: revettk@westminster.ac.uk
*Corresponding author

Florin Gorunescu, Marina Gorunescu and Marius Ene

Department of Mathematics, Biostatistics and Computer Science,
University of Medicine and Pharmacy of Craiova,
Romania
E-mail: fgorun@umfcv.ro
E-mail: mgorun@umfcv.ro
E-mail: enem@umfcv.ro

Sérgio Tenreiro de Magalhães and Henrique M. Dinis Santos

Department of Information Systems,
Universidade do Minho,
Campus de Azurem,
Guimaraes 4800-058, Portugal
E-mail: psmagalhaes@dsi.uminho.pt
E-mail: hsantos@dsi.uminho.pt

Abstract: The majority of computer systems employ a login ID and password as the principal method for access security. In stand-alone situations, this level of security may be adequate, but when computers are connected to the internet, the vulnerability to a security breach is increased. In order to reduce vulnerability to attack, biometric solutions have been employed. In this paper, we investigate the use of a behavioural biometric based on keystroke dynamics. Although there are several implementations of keystroke dynamics available – their effectiveness is variable and dependent on the data sample and its acquisition methodology. The results from this study indicate that the Equal Error Rate (EER) is significantly influenced by the attribute selection process and to a lesser extent on the authentication algorithm employed. Our results also provide evidence that a Probabilistic Neural Network (PNN) can be superior in terms of reduced training time and classification accuracy when compared with a typical MLFN back-propagation trained neural network.

Keywords: biometrics; equal error rate; EER; keystroke dynamics; probabilistic neural networks; PNNs.

Reference to this paper should be made as follows: Revett, K., Gorunescu, F., Gorunescu, M., Ene, M., de Magalhães, S.T. and Santos, H.M.D. (2007) 'A machine learning approach to keystroke dynamics based user authentication', *Int. J. Electronic Security and Digital Forensics*, Vol. 1, No. 1, pp.55–70.

Biographical notes: Kenneth Revett is a Senior Lecturer at the University of Westminster, Harrow School of Computer Science. One of his principal research areas is biometric security. He has written more than a dozen papers on the topic of keystroke dynamics. Much of the research in this area entails the deployment of novel computational algorithms such as artificial immune systems and bioinformatics based classification techniques. In addition, he holds a patent on the implementation of keystroke dynamics called KBD-(Secure). He is on the editorial board of a new journal dedicated to electronic security – the *International Journal of Electronic Security and Digital Forensics*.

Florin Gorunescu received a PhD in Statistics from the University of Bucharest in 1979. Currently, he is a Professor of Mathematics, Biostatistics and Computer Science at the University of Medicine and Pharmacy of Craiova, Romania. He has published numerous technical papers on applied statistics in healthcare, stochastic modelling, data mining and natural computation.

Marina Gorunescu received a PhD in Functional Analysis from the University of Craiova in 1984. Currently, she is Lecturer of Scientific Calculus, Classification and Prognosis within the University of Craiova, Romania. She has published numerous technical papers on stochastic modelling and data mining.

Marius Ene is a PhD student in Artificial Intelligence at the University of Pitesti. Currently, he is an Assistant Professor at the University of Medicine and Pharmacy of Craiova, Romania.

Sérgio Tenreiro de Magalhães is an Invited Teacher in the University of Minho, Portugal, Lecturing Computing Systems and Computer Technology. In the last few years he has developed a research in of Information Services (especially in information retrieval) and Information Security, more especially in what concerns to behavioural and/or stealth biometric technologies. He is the author of several papers published in international journals and in international conference proceedings. He is also a Member of the Programme Committee of several workshops and conferences in his areas of knowledge.

Henrique M. Dinis Santos received his first degree in Electric and Electronic Engineering, University of Coimbra, Portugal, in 1984. In 1996, he received a PhD in Computer Engineering, at the University of the Minho, Portugal. Currently, he is an Associate Professor at the Information Technology and Communications group, at the University of Minho, being responsible for several graduate and post-graduate courses, as well as the supervision of several final year projects. He is also the President of the ALGORITMI Research Centre, University of Minho and President of a National Technical Committee (CT 136) related to information system security standards.

1 Introduction

User access to most computer systems is secured through possession of a login ID and password combination (i.e. C2 security level). Once the login details have been exposed

to an unauthorised user – they have complete access to the computer system in a transparent manner/such security breaches may result in direct financial loss and information security leaks. Such breaches produce an increased perception of the vulnerability of computer systems as portals for computer based transactions. This apprehension is exacerbated when transactions occur in a publicly accessible medium such as the internet. In response to public awareness of perceived and real threats, researchers in academia and industry alike have sought ways to enhance computer security. These research efforts have spawned a new industry – with the sole purpose of providing solutions to enhance computer security – the biometrics industry.

Currently, there are two major forms of biometrics: those based on physiological attributes and those based on behavioural attributes (this of course excludes ID cards and *what-we-possess* mechanisms). Physiological biometrics integrate a measurement of some physiological feature such as fingerprints, retinal blood vessel patterns and iris patterns into an automated authentication schema. Behavioural biometrics on the other hand extract and integrate information about human behaviour such as variations in our speech pattern, gait, signature and the way we type into the authentication schema (see Jain and Sharath, 2003; Peacock, 2004).

Each major class of biometrics has their pros and cons. Physiological biometrics is generally considered/perceived to be extremely robust and hence more secure. For instance, we each possess unique fingerprints (even identical twins differ in their fingerprint patterns) and such measures of identity are thought to be foolproof. But current literature reports indicate that fingerprints can be spoofed (Jain and Sharath, 2003; Peacock, 2004). In addition, even though fingerprint scanners are becoming more reliable and cheaper to acquire, they still are subject to noise and wear and tear. They require replacement approximately once a year and are difficult to deploy on remote access systems, such as a home computer used in a credit card transaction over the internet. Iris scanners and more noise tolerant and are highly accurate, but are certainly more expensive than fingerprint scanners. In addition, they are (or at least appear to be) more intrusive – a very important factor in a biometric. Any biometric solution that is to be used on the internet (and therefore accessed by potentially 100s of millions of users) must be effective and yet very unobtrusive. Behavioural biometrics on the other hand are unobtrusive – but are considered/perceived to be more fallible than physiological techniques. Signatures can be forged, speech can be replicated through moderately sophisticated speech synthesis machinery, etc. Which class of biometrics one selects will depend on the need at hand. It is the contention in this, paper that behavioural biometrics – and keystroke dynamics in particular affords a high level of security – on par with fingerprint systems – without compromising usability and the need for expensive hardware.

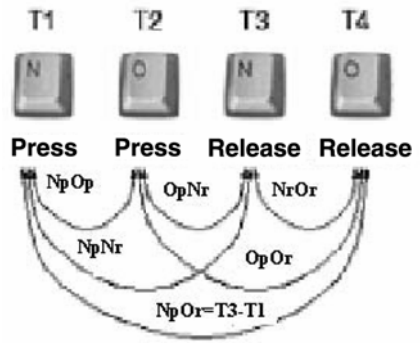
2 Keystroke dynamics

Keystroke dynamics is a class of behavioural biometrics that captures the typing style of a user. Typing style includes such factors as the length of time it takes to type the login id/password, how long we depress a key and how long we take to type successive keys.

Figure 1 illustrates the classic example of the data that can be extracted by entering two keys on a standard keyboard. By collecting all possible digraphs (two-letter combinations) from the login Id/password – one can develop a model of how the person

types these credentials for example. In addition to this static information, one can investigate how a person's typing style evolves with continued practice. This practice effect – or learning curve – can be quantified and used as a metric directly. In addition, any attributes collected for the authentication process must be updated over time. In addition to the static direct attributes mentioned above – secondary or derived attributes should be acquired. These include typing speed, edit distance and entropy to name a few. These attributes provide at the very least an additional range of attributes that can be used in the classification process. In addition, they may provide useful classification information not found in primary attributes. We provide information on the role of primary and secondary attributes in the authentication process later in this paper. In addition to the attributes one collects, there must be some objective function that can be used to measure the accuracy of the authentication process.

Figure 1 The concept of a digraph – and the various combinations that can be extracted and used for biometric authentication. In this particular example, the digraph is based on the character sequence 'no'. Note that the subscript 'r' = indicates release and the subscript 'p' = press



Source: Peacock (2000).

In the biometrics literature – there are two primary objective metrics used to quantify the efficacy of the authentication process: False Rejection Rate (FRR) and False Acceptance Rate (FAR). The former is usually reported as a measure of false rejection – a type I error and the later a false acceptance or type II error. Another measure – called the Cross-over Error Rate (CER) – sometimes referred to as the Equal Error Rate (EER) is also reported – they provide a measure of how sensitive the biometric is at balancing ease of use for the authentic user while at the same time reducing the imposter access rate. All extant biometric systems yield a trade-off between these two measures – those that reject imposters effectively (low FAR) are usually accompanied by a high FRR and vice versa. Figure 2 depicts a typical plot of FAR/FRR and indicates the CER point – where the two plots intersect. With the attributes at hand – and suitable metrics for quantifying the error level in the authentication process, the last phase entails developing the operational aspects of the authentication process. In most behavioural based biometrics – this is a two-stage process involving an enrollment process and a subsequent authentication process.

Enrollment is the process whereby users are asked to present their login details repeatedly in order for the system to be able to extract a statistically significant sample of the users typing style. In the background of course the biometric system is extracting

both primary and secondary attributes to be used in the authentication stage. There are several different methodologies employed in the enrollment process. Some require entering a string of text of at least 400–1500 characters, depending on the required level of security (http://www.psylock.com). Others require that a user enter their login ID/password multiple times (typically 10–15) such as the case with the commercial product Biopassword (http://www.biopassword.com). Lastly, some methods employ a combination of both strategies: with an initial Biopassword like enrollment followed by subsequent monitoring of keystrokes at periodic intervals. Whichever method is employed, a fine balance has to be achieved during the enrollment procedure. If it is too lengthy, then users will consider it a nuisance and if it is too short, will result in a classifier with reduced accuracy. Many surveyed users (personal observations) claim that periodic checking of their typing style is obtrusive and considered as an unacceptable invasion of their privacy.

Table 1 commonly reported attributes collected during and/or generated from the enrollment process.

Figure 2 The CER is indicated as the intersection between the FAR and FRR – when measured against a changing threshold

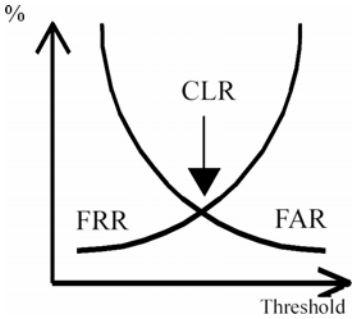


Table 1 Summary of the attributes that were extracted from both enrollment and authentication attempts

Digraphs	10–28 nodes
Trigraphs	8–12 nodes
Total username time (∗)	1 node
Total password time (∗)	1 node
Total entry time (∗)	1 node
Speed	1 node
Scan code (∗)	1 node
Edit distance	1 node

Note:

1 The attributes were utilised both in the PNN and the back-propagation neural network authenticators.

2 All time related attributes are recorded in milliseconds with an accuracy of ±1 mS. Items in the first column with an '∗' are considered primary attributes in this work.

Once this data has been collected, a reference 'signature' is obtained for this user. The reference is then used on subsequent login attempts – a user with that particular login id/password combination has their keystroke dynamics extracted and then compared with a stored reference value. If they are within a prescribed tolerance limit – the user is authenticated. If not – then the system can decide whether to lock up the workstation – or take some other suitable action. When devising such a biometric solution – there is always a trade – off between being overly stringent – rejecting every attempt to login in and being overly lenient – allowing imposters to access the computer. This balance is reflected in the resultant EER from pilot studies and is used to help tweak the system appropriately. How this is done is still a key research area in biometrics.

In order to evaluate keystroke dynamics as a suitable methodology for user authentication, the following questions must be addressed:

1 what attributes can be extracted from the user based on their typing styles and

2 what authentication algorithm(s) can be employed to maximise subtle differences between the typing styles of individuals.

The first question is intimately associated with the individuality of typing – is it as unique as a fingerprint for instance? In order to address this issue, one must examine the various attributes that can be extracted (with appropriate accuracy) from the typing pattern of a user. These attributes are the summation of a number of unique measurements which are recorded when a user types. For instance, how quickly they type, the duration of keypresses (dwell time), the latency between successive keypresses (time of flight), the keypress ordering, their usage of the shift keys and how their typing style evolves with practise – referred to as a practise effect (Card et al., 1980). These attributes must be collected for each user and stored for future use in the authentication process. In order to acquire a sufficient sample size for each the attributes, most keystroke based biometric systems require an initial enrollment phase.

With the enrollment data in hand, one can build a model of the typing style of a given user. The parameters of this model then become the inputs to a given classifier. Extracting the correct parameters is undoubtedly *the* critical issue with regards to keystroke dynamics based biometrics. What information can be extracted from typing? Table 1 lists the major attributes that have been reported in the literature in the context of keystroke dynamics (see Jain and Sharath, 2003; Obaidat and Macchairolo, 1994; Peacock, 2000). The attributes are generally collected during some form of enrollment process. The attributes can be categorised into primary and derived attributes. By primary, we refer to attributes such as duration, latency and scan code – attributes that can be extracted while the user is typing. From this set of primary attributes, we can derive a variety of second order attributes such as digraph/trigraph latencies, entropy and edit distance and speed (Bergadano et al., 2002; Leggett et al., 1991; Magalhães and Santos, 2005; Mahar et al., 1995; Obaidat and Sadoun, 1997b). The choice of attributes is an active area of research and we present some results on this issue in this paper. Once a suitable choice of attributes has been selected, a model is generated and the classification task is performed. It should be noted that most biometric systems employing keystroke dynamics, the classification task is reduced to one of authentication – a much simpler task than identification. There are a number of reported authentication algorithms that have been employed in keystroke dynamics. In the next section we provide a survey of the major research efforts which have focused on

authentication algorithms. Lastly, we describe our approach and present some data that indicate the Probabilistic Neural Network (PNN) approach to keystroke dynamics based authentication is indeed a reasonable one.

2.1 Classification algorithms

In the context of keystroke dynamics based security enhancement, the classification algorithm is really an authentication algorithm. It is usually described as a 1:1 mapping. A table of some sort is maintained that contains a user's details along with associated discrimination data collected during the enrollment process. When those access details are entered, the system looks up the respective details and performs a similarity measure of some sort. Identification is not prevalent in this domain primarily because of the possible time lag that might be involved if this system were deployed for a large number of users. Therefore, classification is synonymous with authentication in this present paper. Another important issue in keystroke dynamics based authentication algorithms is whether one has both legitimate and imposter data available. If the classification method is supervised, then the system must be trained on legitimate and imposter data samples where does one acquire imposter data? There are a couple of solutions to this problem. One solution is to select a random set of legitimate data samples and add a variable amount of noise. This will produce two sets of training data that can be used to train a supervised authentication scheme such neural network paradigms (MLFNs) trained with backpropagation. These issues will be discussed in turn as we enumerate some of the major research efforts in this domain next.

In Gaines et al. (1980) presented a report of his work to study the typing patterns of seven professional typists. The small number of volunteers and the fact that the algorithm is deduced from their data and not tested in other people later, results on a lower confidence on the FAR and FRR values presented. But the method used to establish a pattern was a breakthrough: a study of the time spent to type the same two letters (digraph), when together in the text. Since then, many algorithms based on Algebra and on Probability and Statistics have been presented. Joyce and Gupta (1990) presented in 1990 an algorithm to calculate a value that represents the distance between acquired keystroke latency times and correspondent times previously stored. In Monrose and Rubin (1997) use the Euclidean Distance and probabilistic calculations based on the assumption that the latency times for one-digraph exhibits a Normal Distribution. Later, in 2000, they also present an algorithm for identification, based on the similarity models of Bayes and in 2001 they present an algorithm that uses polynomials and vector spaces to generate complex passwords from a simple one, using the keystroke pattern (Monrose et al., 2001).

Various fuzzy logic algorithms have been applied – mapping the variability in ones typing patterns to a fuzzy concept. For instance, Hussien et al. (1989) and de Ru and Eloff (1997) use a combination of fuzzy clustering algorithms – obtaining an error rate of approximately 5–10% – depending on the number of samples they acquired per login Id/password combination. Another study (Tapiador and Siguenza, 1999) employed a fuzzy rule set in order to classify login Id/password combinations with somewhat better success than Hussein – although they report only their preliminary results.

Techniques based on neural networks have been explored – focusing on ART-2 and multilayer perceptrons trained with the backpropagation algorithm. For instance, Obaidat provides data that suggests that the error rate can be reduced to approximately 2.4–4.2%,

depending on the exact preprocessing performed using a non-standard neural network (Obaidat and Sadoun, 1997a), has also applied neural networks (using standard backpropagation) to keystroke dynamics, generating error rates on the order of 2–4% (Bleha et al., 2002; Brown and Rogers, 1993).

Other machine learning approaches, based on Support Vector Machines (SVM) have been used to address the classification problem presented by keystroke dynamics. de Oliveira et al. (2005), Sang et al. (2004) and Sung and Cho (2006) have applied SVM to a small keystroke dataset and compare their results to standard neural network technology. The authors claim that the SVM classifier is more efficient and at least as accurate as neural network technologies, have also applied SVM to this domain, reporting an error rate of approximately 8–10% (Sung and Cho, 2006).

Lastly, Revett et al. have used the rough sets induction algorithm to extract rules that form models for predicting the validity of a login Id/password attempt (Revett et al., 2005b). The results indicate that the error rate can be as low as 2–4% in many cases.

There is a substantial body of literature accruing focused on the use of graphical based authentication schemes (see Davis et al., 2004; Jermyn et al., 1999). Although the results look fairly promising, we will not discuss this research stream in this paper. It should be noted though that the approach taken in this paper does not preclude using graphical authentication derived attributes – and research in this direction is already underway.

The algorithms cited are a small example of the many approaches used to find adequate keystroke dynamics algorithms with a reasonable CER. Many others could also be cited, all with different evaluation methods, different number of subjects, different number of keystrokes required to enroll the system and different number of repetitive operations required to authenticate and/or identify the user. This diversity in the algorithm parameters and in the evaluation method makes the task of comparing their results a very difficult one. Furthermore, there is, in this subject, no concept of what is a representative data sample. The same algorithm presents different results when tested with different volunteer groups (datasets). The best way to meaningfully compare two algorithms is to test it against the same group.

In this study, we have deployed the use of a PNN as the authentication technique. The reason for this is that we have some expertise in this area (Gorunescu et al., 2005a,b; Revett et al., 2005b) and that it is a novel approach to the problem. To our knowledge, no other author(s) have employed a PNN to this domain. A PNN operates in a supervised fashion – so we collected data with respect to FRR and FAR. The details of the data collection are described in the methods section. We also compared our results with that obtained by using a standard three layer MLP and the vanilla back-propagation algorithm. Another justification for applying a PNN to this domain is the efficiency of training relative to a back-propagation based NN. Consider the deployment of a keystroke dynamics based authentication system on the internet. One would expect that users will be added to the system constantly – if you are using standard backprop – the training time will increase almost exponentially. With PNN however – training time is minimised – provided the system has sufficient memory to house all objects in memory. Lastly, we used the authentication algorithm we have developed to examine the relative importance of the attributes that are collected during enrollment and subsequent data processing. We present this data after a brief discussion of the PNN algorithm.

2.2 Probabilistic neural networks

The PNN is essentially a classifier implemented as a neural network version of a Bayes-Parzen classifier (Specht, 1988, 1990). The general classification problem is to determine the category membership of a multivariate sample data (i.e. a p-dimensional random vector $\mathbf{x}$) into one of q possible groups Ω_i, $i = 1, 2,\ldots, q$, based on a set of measurements. If we know the probability density functions (p.d.f.) $f_i(\mathbf{x})$, usually the Parzen-Cacoulos or Parzen like p.d.f. classifiers:

$$f_i(x) = \frac{1}{(2\pi)^{p/2}\sigma^p}\frac{1}{m_i}\sum_{j=1}^{m_i}\exp\left(-\frac{\left\|x - x_j\right\|^2}{2\sigma^2}\right) \tag{1}$$

the *a priori* probabilities $h_i = P(\Omega_i)$ of occurrence of patterns from categories Ω_i and the *loss* (or *cost*) parameters l_i associated with all incorrect decisions given $\Omega = \Omega_i$, then, according to the Bayesian decision rule, we classify $\mathbf{x}$ into the category Ω_i if the inequality $l_i\ h_i\ f_i(\mathbf{x}) > l_j\ h_j\ f_j(\mathbf{x})$ holds true. The standard training procedure for PNN requires a single pass over all the training patterns, giving them the advantage of being faster than the feed-forward neural networks (Specht, 1988, 1990).

Basically, the architecture of PNN is limited to three layers: the *input/pattern layer*, the *summation layer* and the *output layer*. Each input/pattern node forms a product of the input pattern vector $\mathbf{x}$ with a weight vector W_i and then perform a non-linear operation, that is $\exp[-(W_i - x)^\tau (W_i - x)/(2\sigma^2)]$ (assuming that both $\mathbf{x}$ and W_i are normalised to unit length), before outputting its activation level to the summation node. Each summation node receives the outputs from the input/pattern nodes associated with a given class and simply sums the inputs from the pattern units that correspond to the category from which the training pattern was selected, $\sum_i \exp[-(W_i - x)^\tau (W_i - x)/(2\sigma^2)]$. The output nodes produce binary outputs by using the inequality:

$$\sum_i \exp\left[\frac{-\left(W_i - x\right)^\tau \left(W_i - x\right)}{\left(2\sigma^2\right)}\right] > \sum_j \exp\left[\frac{-\left(W_j - x\right)^\tau \left(W_j - x\right)}{\left(2\sigma^2\right)}\right] \tag{2}$$

related to two different categories Ω_i and Ω_j.

The key to obtaining a good classification using PNN is to optimally estimate the two parameters of the Bayes decision rule, the misclassification costs and the prior probabilities. In our practical experiment we have estimate them heuristically. Thus, as concerns the costs parameters, we have considered them depending on the average distances D_i, inversely proportional, that is $l_i = 1/D_i$. As concerns the prior probabilities, they measure the membership probability in each group and thus, we have considered them equal to each group size, that is $h_i = m_i$. As in our previous work, we employed an evolutionary technique based on the genetic algorithm to find the smoothing parameters (see Gorunescu et al., 2005a,b for implementation details). In the next section, we describe the experimental methods, with a brief description of the dataset.

3 Methods

The dataset we examined consisted of a group of 50 subjects (all university students in a computer science department) – 20 acting as authentic users and the balance (30) acting as imposters. We asked the authentic user group to enter a login Id/password of their choice (minimum of 6 characters each, with a maximum limit of 15 characters for each). This was immediately followed by an enrollment period that consisted of entering their selected user Id/password for ten trials. We collected a series of attributes (see Table 1 for a complete listing) which were to be used during the authentication process. The data samples were collected over a 14–day period, throughout specified periods of the day. We requested that the participants login during a morning, midday and late afternoon session in order to replicate the average login times during the course of a normal working day. We maintained a running average of the primary and derived attributes – where the oldest sample of ten was replaced and all derived attributes were recalculated. We invited the imposter group (30 participants) to 'hack' into all of the legitimate accounts after providing them with the account holders' login Id/passwords. They were given 1 week to log into all 20 authentic accounts approximately 100 times each (total of 2000 attempts) and the success/failure rates were recorded. More specifically, each participant of the imposter group attacked each account four times – for a total of 80 login attempts for each imposter. Therefore, each account will be attacked 120 times. We randomly selected 100 imposter login attempts for each account and used these values in all subsequent calculations in order to keep the numbers in multiples of 100. This was used to estimate the average FAR for the user group. In addition, the authentic users were asked to log into their own accounts 100 times during the same period. This data was to be used for estimating the average FRR for the user group. The resultant data will contain 2000 FRR attempts and 2000 imposter login attempts. We then used this data to train our PNN algorithm to perform the required class discrimination task. We cross-validated our results and we report the average results from these experiments. The particular version of the PNN we employed in this paper was the same as that employed in previous work (Gorunescu et al., 2005a,b; Revett et al., 2005b). We also applied a modified version of our PNN algorithm, that used separate smoothing factors for each class (authentic and imposter). We report both results in this work – and found that using a separate smoothing factor provided consistently better classification results. To provide a direct comparison of the PNN results with another recognised classification technique, we developed a three layer multilayer perceptron neural network trained with back-propagation. The following parameters were used for the MLFN the input layer was contained the minimal number of nodes 23, the hidden layer had 14 nodes and the output layer 2 nodes (corresponding to the two decision classes). Please note the actual number of potential input nodes really ranges from 23 to 45 in this particular dataset. We therefore used the minimal number of digraphs (10) and trigraphs (8) as per Table 2. The learning parameter $\eta = 0.2$ without a momentum term. The acceptable error rate was set to 0.01. The actual data (extracted attributes) that was presented to both authentication systems is summarised in Table 2. Note the difficulty one encounters when using a fixed architecture like an MLP when the number of digraphs varies as is the case with a variable length user Id/password. The results were assessed using 10-fold cross validation and the results presented in this paper are the average values for each network (PNN or BP).

Table 2 FAR/FRR values as a function of the division level

Division points	*False acceptance*	*False rejection*
Panel A		
10	0.0483	0.0481
20	0.0192	0.0197
30	0.0576	0.0376
40	0.0576	0.0566
50	0.0576	0.0483
60	0.0001	0.0021
70	0.0576	0.0598
80	0.0481	0.0483
90	0.0288	0.0312
100	0.0480	0.0427
	0.0422	0.0394
Panel B		
10	0.0583	0.0481
20	0.0692	0.0997
30	0.0976	0.0876
40	0.0741	0.0566
50	0.0716	0.0983
60	0.0411	0.0521
70	0.0576	0.0598
80	0.0481	0.0483
90	0.0588	0.0912
100	0.0638	0.0727
	0.06402	0.07144

Note: The values in the last row of the right-most columns are the averages of their respective columns. In Panel A, the results were obtained using a different smoothing factor for training/testing and Panel B using the same smoothing factor for each.

In addition to the classification task per se, we also sought to investigate the information content of each of the attributes that were acquired during the enrollment phase. We therefore tested a variety of combination of attributes (see Table 4 for a summary of the collected attributes). The digraphs were recorded during as the time in mS between the release of two successive keystrokes, with an accuracy of 1 mS. The trigraphs were recorded in the same way, based on the release time between the first and third keys. The speed was measured as the total time taken to enter the login ID and password divided by the total number of characters, excluding the return key and the interval between entering the login ID and password. The edit distance was recorded as per (Bergadano et al., 2002) using trigraphs only. Briefly, the edit distance is an indication of the entropy different between two typing samples. The trigraphs entered which span across the login ID/password boundary, are arranged in order of ascending order of time. Then the

number of rearrangements required to order the trigraphs is measured and divided by the total number of trigraphs that are available for a string of a given length. The specific details of which attributes and their combinations that were tested is described in detail in the results section. These attribute combinations were used for classification purposes with both versions of the PNN and the MLFN as indicated. The result of these experiments is presented in the next section.

4 Results

We first describe an experiment where we examined which division used in the PNN and whether a single or separate smoothing factor gave us the best classification accuracy. For this experiment, we used the full set of attributes (see Table 4 for details). We selected random samples for training and testing (50/50 in this case) and applied our PNN algorithm to these random samples. More specifically, we selected 200 samples of FAR and FRR data for the training/testing purposes, repeating this process until all samples were utilised (10 trials). The data in Tables 2 indicate that the classification accuracy was essentially independent on the number of divisions employed for this dataset. Please note that the modified PNN algorithm (separate smoothing factors for each category) yielded consistently higher results than one that employed the same smoothing factor for both classes.

The results presented in Table 4 indicate that our FAR/FRR is on the order of 4% – and the data in Table 4 indicate a slightly higher error rate of approximately 6.8% (sum of the FAR and FRR errors) for the MLFN.

Table 3 Summary of the authentication accuracy using the back-propagation algorithm summarised as a series of confusion matrices

Trial 1	*Legit*	*Impost*	
Legit	*89*	*11*	0.89
Impost	*10*	*90*	0.90
	0.89	0.89	*0.895*
Trial 2			
Legit	*93*	*7*	0.93
Impost	*8*	*92*	0.92
	0.92	0.93	*0.925*
Trial 3			
Legit	*91*	*9*	0.91
Impost	*7*	*93*	0.93
	0.93	0.91	*0.920*

Note:

1 The values in the right hand column (in italics) are the accuracy for the particular classification run.

2 'Legit' is short for legitimate and 'impost' is short for imposter.

We also investigated the ability of a multilayered neural network running the back-propagation learning algorithm on the data that was collected and previously analysed with the PNN algorithm. The data extracted from the user input was summarised in Table 1. The dataset was sampled 50/50 training/testing and repeated using *n*-fold validation and the results are reported as the average values. The data for the MLFN classification experiment is presented as in Table 3. Note that the same set of subsamples was used in both experiments. Note that the average accuracy from the samples in Table 3 yield a value of approximately 91%. This value is approximately the same for the average from resampling the entire dataset without replacement (92.1%). We then repeated the comparison experiments (using the dual smoothing factors only for the PNN) but varied the particular attributes that were used. In particular, we were interested in what effect the various attributes contributed towards the classification accuracy and whether the particular classification technique would be differentially influenced by the attribute selection process. Table 4 presents a summary of the various attributes that were tested, along with the classification accuracy (reported as the average of the FRR/FAR results) for the PNN and the MLFN, both trained as per the previous results.

Table 4 Summary of the classification accuracy for the PNN and MLFN classifiers when using various combinations of primary and/or derived attributes, values are the sum of the FAR and FRR results

Attributes	*PNN (%)*	*MLFN (%)*
All	3.9	5.7
Primary only	5.2	6.5
Derived only	4.2	6.2
DG + primary	4.4	5.3
TG + primary	4.0	5.8
Edit distance only	3.7	5.0

Note: See Table 1 for details on which are primary and derived attributes.

Finally, there is the issue of computational time – both for training and the classification tasks. Generally speaking, the PNN training time was significantly lower than that for the MLFN – 1.4 min versus 5.6 min for training on 400 objects – 200 for each class. The classification task was approximately the same for each classifier – with an average value of 8 sec.

5 Conclusions

We have successfully applied our modified Specht PNN to difficult biomedical datasets and have obtained accuracy levels comparable to other more traditional methods (Gorunescu et al., 2005a,b; Revett et al., 2005a,b). In this study, we have employed our modified PNN to a small dataset of login Id/password samples. The modified classifier performed better than the standard PNN algorithm by a considerable margin (4% versus 8% approximately). These results are comparable to traditional neural network approaches as well as more ‘modern’ approaches such as SVM. We also used the same

dataset to test the classification accuracy of standard implementation of the MLFN trained with backpropagation. The results from this study indicate that the PNN is superior to the MLFN with respect to the classification accuracy and training time. It must be noted that these results were obtained without any data preprocessing. We simply collected the data, selected a random subset for training 50% and 50% for testing. This algorithm is time efficient when login id/password credentials are used for authentication purposes. It is a well known fact that the training phase of the PNN algorithm begins to degrade in terms of time efficiency when the sample numbers are large. But in this area of application, where we have a relatively small number of samples for training (on the order of 100–200) – and can select an equal number of testing samples, training performance is not an issue. This is in contrast to other techniques such as the backpropagation algorithm that requires a substantial number of training data in order to generate accurate classification. These advantages make the PNN a very suitable candidate for a novel machine learning algorithm in the context of keystroke dynamics authentication.

With regards to the attribute used in this study we found that the derived attributes such as digraph/trigraph times, speed and edit distance were more effective compared to primary attributes such as dwell time, time of flight and scan codes. The edit distance attribute produced results with the lowest error rate for both the PNN and MLFN. Although not statistically different from using all attributes – the data from this study suggest that this attribute is quite important in the classification task when using these types of classifiers. This data is consistent with the results published by Bergadano. Clearly there is a trend in the results from this study to investigate a variety of first and second order attributes before selecting a classifier tool. Not all attributes produce the same classification accuracy over different techniques.

Another aspect of keystroke dynamics based biometrics entails how to collect samples for classification. In this study, we attempted to collect data from users in a manner consistent with normal daily computer usage. It may be fairly unrealistic to ask imposters to hack into a system for 100 consecutive attempts. Most computer systems turn of access opportunities after three failed login attempts – so attempting 100 is an unreasonable situation. In addition, when does one count a login attempt as a failure? Is it after three continuously failed attempts or after each single failed attempt? In this study – we used single failed attempts. If one counts only three failed attempts before considering the attempt as a failure – then both FAR/FRR rates will be significantly different from what we report. As long as the author makes it clear what their criterion is – then different laboratories can meaningfully compare their results.

This preliminary study has yielded promising results with respect to the use of a PNN for keystroke dynamics based authentication. One advantage of the PNN is that it can work with missing and mixed datasets. The amount of preprocessing required is minimal which is critical with a system that must be automated and work in an online basis. We will continue to variations on the PNN architecture as originally proposed by Specht. In particular, hybrid systems employing evolutionary techniques may enhance the classification accuracy through the discovery of better parameters for the model (Gorunescu et al., 2005a). In addition, further variations and combinations in attribute selection may enhance the classification performance as well. Finally, the results of this work provide another example of how behavioural biometrics can produce an accurate authentication system that is both robust with respect to the parameter selection process, while maintaining the usability advantage over current physiological based biometrics.

References

Bergadano, F., Gunetti, D. and Picardi, C. (2002) 'User authentication through keystroke dynamics', *ACM Transactions on Information and System Security*, Vol. 5, No. 4, pp.367–397.

Bleha, S.A., Knopp, J. and Obadiat, M.S. (2002) 'Performance of the perceptron algorithm for the classification of computer users', *Proceedings of the ACM/SIGAPP Symposium on Applied Computing*, New York Press.

Brown, M. and Rogers, S.J. (1993) 'User identification via keystroke characteristics of typed names using neural networks', *International Journal of Man-Machine Studies*, Vol. 39, pp.999–1014.

Card, S.K., Moran, T.P. and Newell, S. (1980) 'The keystroke-level; model for user performance time with interactive systems', *Communications of the ACM*, Vol. 23, No. 7, pp.396–410.

Davis, D., Monrose, F. and Reiter, M.K. (2004) 'On user choice in graphical password schemes', *Thirteenth USENIX Security Symposium*.

de Oliveira, M., Kinto, V.S.E., Hernandez, E.D.M. and de Carvalho, T.C. (2005) *User Authentication Based on Human Typing Patterns with Artificial Neural Networks and Support Vector Machines*, SBC.

de Ru, W.G. and Eloff, J. (1997) 'Enhanced password authentication through fuzzy logic', *IEEE Expert*, Vol. 12, No. 6, pp.38–45.

Gaines, R., et al. (1980) *Authentication by Keystroke Timing: Some Preliminary Results*, Rand Report R-256-NSF, Rand Corp.

Gorunescu, F., Gorunescu, F., El-Darzi, E., Gorunescu, S. and Revett, K. (2005a) 'A cancer diagnosis system based on rough sets and probabilistic neural networks', *First European Conference on Health care Modelling and Computation*, University of Medicine and Pharmacy of Craiova, pp.149–159.

Gorunescu, F., Gorunescu, M., El-Darzi, E. and Gorunescu, S. (2005b) 'An evolutionary computational approach to probabilistic neural network with application to hepatic cancer diagnosis, CBMS', *Eighteenth IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*, pp.461–466.

Hussien, B., Bleha, S. and McLaren, R. (1989) 'An application of fuzzy algorithms in a computer access security system', *Pattern Recognition Letters*, Vol. 9, pp.39–43.

Jain, R.B. and Sharath, P. (2003) 'Introduction to biometrics', in A. Jain, R. Bolle and S. Pankanti (Eds). *Biometrics. Personal Identification in Networked Society*, Kluwer Academic Publishers, pp.1–41.

Jermyn, I., Mayer, A., Monrose, F., Reiter, M. and Rubin, A. (1999) 'The design and analysis of graphical passwords', *Eighth USENIX Security Symposium*.

Joyce, R. and Gupta, G. (1990) 'Identity authorization based on keystroke latencies', *Communications of the ACM*, Vol. 33, No. 2, pp.168–176.

Leggett, J., Williams, G., Usnick, M. and Longnecker, M. (1991) 'Dynamic identity verification via keystroke characteristics', *International Journal of Man-Machine Studies*, Vol. 35, pp.859–870.

Magalhães, S.T. and Santos, H.D. (2005) 'An improved statistical keystroke dynamics algorithm', *Proceedings of the IADIS MCCSIS 2005*.

Mahar, D., Napier, R., Wagner, M., Henderson, R.D. and Hiron, M. (1995) 'Optimizing digraph-latency based biometric typist verification systems: inter and intra typist differences in digraph latency distributions', *International Journal of Human-Computer Studies*, Vol. 43, No. 4, pp.579–592.

Monrose, F., Reiter, M.K. and Wetzel, S. (2001) 'Password hardening based on keystroke dynamics', *International Journal of Information Security*.

Monrose, F. and Rubin, A.D. (1997) 'Authentication via keystroke dynamics', *Proceedings of the Fourth ACM Conference on Computer and Communication Security*, Zurich, Switzerland.

Obaidat, M. and Sadoun, S. (1997a) 'Verification of computer users using keystroke dynamics', *IEEE Transactions on Systems, Man and Cybernetics, part B: Cybernetics*, Vol. 27, No. 2, pp.261–269.

Obaidat, M.S. and Macchairolo, D.T. (1994) 'A multilayer neural system for computer access security', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 24, No. 5, pp.803–816.

Obaidat, M.S. and Sadoun, S. (1997b) 'A simulation evaluation study of neural network techniques to computer user identification', *Information Sciences*, Vol. 102, pp.239–258.

Peacock, A. (2000) 'Learning user keystroke latency patterns', Available at: http://pel.cs.byu.edu/~alen/personal/CourseWork/cs572/Keystrokepaper/index.html.

Peacock, A. (2004) 'Typing patterns: a key to user identification', *IEEE Security and Privacy*, Vol. 2, No. 5, pp.40–47.

Revett, K., Gorunescu, F., Gorunescu, M., El-Darzi, E. and Ene, M. (2005a) 'A breast cancer diagnosis system: a combined approach using rough sets and probabilistic neural networks', *Proceedings Eurocon 2005 – IEEE International Conference on 'Computer as a Tool'*, 21–24 November, Belgrade, Serbia, pp.1124–1127.

Revett, K., Magalhaes, S. and Santos, H. (2005b) 'Developing a keystroke dynamics based agent using rough sets', *The 2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology Compiegne*, pp.56–61.

Sang, Y., Shen, H. and Fan, P. (2004) *Novel Imposters Detection in Keystroke Dynamics Using Support Vector Machines*, LNCS Springer Berlin/Heidelberg, pp.666–669, ISSN 0302-9743.

Specht, D.F. (1988) 'Probabilistic neural networks for classification mapping or associative memory', *Proceedings IEEE International Conference on Neural Networks*, Vol. 1, pp.525–532.

Specht, D.F. (1990) 'Probabilistic neural networks', *Neural Networks*, Vol. 3, pp.110–118.

Sung, K.S. and Cho, S. (2006) 'GA SVM wrapper ensemble for keystroke dynamics authentication', *International Conference on Biometrics*, Hong Kong, pp.654–660.

Tapiador, M. and Siguenza, J.A. (1999) 'Fuzzy keystroke biometrics on web security', *AutoID '99 Proceedings Workshop on Automatic Identification Advanced Technologies. IEEE*, pp.133–136.