# Using Keystroke Analysis as a mechanism for Subscriber Authentication on Mobile Handsets

N.L. CLARKE*, S.M. FURNELL*, B. M. LINES* and P.L. REYNOLDS**

*info@network-research-group.org
*Network Research Group*
*Department of Communication and Electronic Engineering*
*University of Plymouth*
*PLYMOUTH*
*PL4 8AA*
*United Kingdom*
*Tel: +44 1752-233520 Fax: +44 1752-233520*

** *Orange Personal Communications Services*
*Bradley Stoke*
*BRISTOL*
*United Kingdom*

Key words: Keystroke Analysis, User Authentication, Biometrics, Mobility.

Abstract: The mobile communications industry will experience an evolutionary step within the next two years with the introduction of third generation mobile networks, completing the handset transition from a purely telephony device of the first generation analogue networks into a multimedia multi-purpose mobile communications tool. The ability of these new handsets to store and access sensitive information such as financial records, digital certificates and company records in association with a large handset penetration (864 million subscribers) makes them a desirable target for impostors. The authentication technique for current mobile phones has many weaknesses from a technological and subscriber perspective, and as such non-intrusive and stronger subscriber authentication techniques are required. This study investigates the plausibility of one such technique that of keystroke analysis, comparing and contrasting a number of pattern recognition and neural network based approaches to classification. It was found that neural network-based

approaches performed substantially better than the pattern recognition-based approaches with false acceptance and false rejection rates of 3.2%.

# 1. INTRODUCTION

The mobile communications sector has witnessed substantial growth in recent years with global mobile subscribers forecasted to rise from 864m in 2002 [1] to 1,848m by 2004 [2]. However, in parallel with this rise in ownership there has been a rise in mobile related abuse, with over 700,000 handsets stolen from subscribers in 2001, in the UK [3]. It can be conjectured that the more advanced capabilities of third generation handsets with their ability to pay for products using micro-payments and digital money, surf the internet, buy and sell stocks, transfer money and manage bank accounts will make the handsets even more desirable targets. Current authentication for handsets is achieved through a PIN (Personal Identification Number) approach, which relies heavily on the user to ensure its validity. For example, the subscriber should not use the default setting, tell other people, or write it down. Apart from the technological arguments, a recent survey into attitudes and opinions of mobile phone subscribers found that 45% of respondents thought the PIN to be inconvenient and did not use the facility [4]. The findings also demonstrated the user's awareness of the security implications, with 81% of respondents in support for more security.

Approaches to the verification of an identity can be achieved in one of three ways. Something the user *knows, has* or *is* [5]. The first approach is a secret-knowledge technique identical to the PIN and will therefore be just as inconvenient. The second is based on the user having to carry a token. However, due to the very nature of a mobile handset it is likely to remain within the handset permanently and thus diminished any security the token would provide (for example, subscriber's use of the SIM). The last approach, commonly termed as biometrics, is based on some unique characteristic feature of a person and includes physiological characteristics such as, fingerprints and hand geometry and behavioural traits such as a person's voice and signature. Another behavioural biometric is keystroke analysis which measures the typing characteristic of a user. In this context it has a number of advantages including a keypad that already resides on the device and the possible non-intrusive application of the technique thereby reducing user inconvenience. This paper will compare and contrast the performance of a number of pattern recognition and neural network approaches to solving the problem of keystroke analysis on a mobile handset keypad.

## 2. KEYSTROKE ANALYSIS

The principal concept behind keystroke analysis is the ability of the system to recognise patterns, such as characteristic rhythms, during keyboard interactions. In particular, this study utilises the time between two successive keystrokes (known as a digraph pair) and is referred to as the inter-keystroke latency. Classification is achieved by comparing an input sample against a reference template for the claimed user and given sufficient similarity the input sample is deemed to have come from the authorised user. The reference template is securely acquisitioned from the user when they enrolled on the system initially. However this template matching process gives rise to a characteristic performance plot between the two main error rates governing biometrics, the False Acceptance Rate (FAR), or the rate at which an impostor is accepted by the system, and the False Rejection Rate (FRR), or the rate at which the authorised user is rejected from the system. A third error rate known as the Equal Error Rate (EER) is used as a comparative measure between biometric techniques and equates to the mean value of the FAR and FRR [6].

A significant amount of prior research has been conducted in this domain, dating back to the 1980s. However, all of these studies have focused upon alphabetic inputs from a standard PC keyboard. Little work to date has considered the application of keystroke analysis to a mobile handset keypad, which has obvious tactile and interoperability differences. A previous feasibility study by the authors [7] has demonstrated promising results. However, the classification algorithm was an un-optimised neural network classifier. It is the aim of this paper to present a number of classification algorithms, including optimised neural network configurations, in order to define the most appropriate classifier for this particular problem.

## 3. CLASSIFICATION ALGORITHMS

Previous researchers have utilised a number of pattern recognition approaches such as linear and non-linear distance techniques [8], z-tests [12] and Bayesian classifiers [16], with more recent research efforts focussing on the use of neural network approaches [14-16]. A number of these techniques were selected on the basis of providing a broad range of classification techniques and this section will give a brief outline of them. For more detailed information and analysis of the techniques refer to references [8, 12, 18-22].

- *Mean & Standard Deviation Algorithm.*

  This is a traditional pattern recognition algorithm [8], based on the assumption that users keystroke latencies for a given digraph pair will be similar within an acceptable tolerance. A mean and standard deviation is calculated from the user's reference profile for the most regular digraph pairs which is used in comparing against an unseen input vector. If a sufficient number of digraph pair keystroke latencies reside within the tolerance envelope then the user is deemed to be the authorised user, if not, then an impostor.

- *Z-Test*

  The z-test is a statistical hypothesis test which can be used to establish whether an input vector comes from a particular sample population or not. The test assumes the sample size is large, so that the central limit theorem applies and we can use the normal distribution and assume that the sample standard deviation is an estimate of the population standard deviation. The null and alternative hypotheses are defined as:

  *Null Hypothesis,* $H_0$*:*
  *Reference Profile Mean,* $\mu_x$ *= Input Vector Mean,* $\mu_o$
  *Alternative Hypothesis,*$H_1$*:*
  *Reference Profile Mean,* $\mu_x \neq$ *Input Vector Mean,* $\mu_o$

  The z-test investigation is a two-tailed test and will vary the level of significance, α in order to determine the most efficient level in terms of the performance rates.

- *Euclidean Distance Algorithm*

  The Euclidean distance algorithm is a linear minimum distance technique that computes the Euclidean distance between an input vector and reference profile. If the distance is within a predefined tolerance level then the input vector is deemed to have come from the authorised user, if not then an impostor. The Euclidean distance is calculated using:

  $|| x - m_k ||$ Where x = input vector; $m_k$ = reference profile

  Here || u || is called the norm of the vector u and corresponds to different ways of measuring distance. The Euclidean metric is calculated using:

  $$|| u || = (u_1^2 + u_2^2 + u_3^2 + ...u_d^2)^{\frac{1}{2}}$$

- *Mahalanobis Distance Algorithm*

  The Mahalanobis distance algorithm is a non-linear minimum distance technique which uses the same mechanism as the Euclidean algorithm but with a difference technique for measuring the distance. The Mahalanobis metric is calculated using:

$$r^2 = (x - m_x)^T C^{-1} (x - m_x)$$

  In principal the non-linear problem solving abilities of the Mahalanobis classifier should provide better decision boundaries and improve the performance over the Euclidean algorithm.

- *Feed-Forward Multi-Layered Perceptron (FF MLP) Neural Network*

  A FF MLP utilising a backpropagation training algorithm are best known for their pattern associative properties. Patten associative networks work by training the network to respond in a certain way given a certain input sample and backpropagation training is mathematically proven to converge towards the most optimal results [20]. However great care needs to be taken to ensure the training data is representational of the problem the network is to solve.

  Unfortunately FF MLP networks have no rules governing what the network parameters need to be given a certain complexity of classification problem. As such trial and error approaches are often utilised in order to achieve the most desirable performance rates.

- *Radial Basis Function (RBF) Neural Network*

  RBF networks are very similar mathematically to MLP networks in that they both provide techniques for approximating arbitrary non-linear functional mappings between multi-dimensional spaces. An advantage of RBF over FF MLP is their ease of implementation with only two network parameters to define, i.e., the mean sum-squared error and the spread of the radial basis neurons. The network works on the principle of transforming the input space into a higher dimensionality in the likelihood that it will be more linearly separable [20].

- *Generalised Regression Neural Network (GRNN)*

  GRNNs are another network topology often used for function approximation tasks and have a similar network paradigm to the RBF networks. GRNNs are again useful because of their ease of implementation and in particular their speed of training. A potential disadvantage is the one-to-one mapping of training vectors to radial basis neurons resulting in a large and computationally complex network with large training datasets.

- *Learning Vector Quantisation (LVQ) Neural Network*

  LVQ networks are a supervised version of vector quantisation [22] designed for adaptive pattern classification. The network paradigm utilises a competitive layer which will automatically learn to classify input samples similar to an unsupervised clustering technique, however the LVQ network also has a mechanism to transform the competitive layer classes into target classifications defined by the user. This technique was used with notable success in [16].

Each of the classification techniques implements an identical mechanism for the evaluation of valid and impostor input samples, so that a fair comparison of the approaches can be achieved. The correct or false acceptance of a user is based not on the success or failure of individual digraphs but on a complete input sample.

## 4. EXPERIMENTAL PROCEDURE

The eventual application of keystroke analysis to a mobile phone would ideally authenticate a user by monitoring their use of the phone, during activities such as the entry of telephone numbers, use of the menu system, and composition of text messages. However, the objective at this stage is to investigate a number of classification techniques rather than to provide a complete solution to the problem. As such, the initial study has been confined to two types of data, namely:

1. Entry of a fixed four-digit number, analogous to the PINs used on many current systems.
2. Entry of a fixed eleven-digit number, analogous to the telephone numbers in which you would enter on a handset.

A total of sixteen test subjects were asked to enter the data for both sets of data thirty times. Twenty of these inputs were utilised in the generation of the reference profile, with the remaining ten used as validation samples. The pattern classification tests were performed with one user acting as the valid authorised user, whist all the other users are acting as impostors. A specially written application was used to collect the sample data.

A standard numerical keypad on a PC keyboard was not deemed to be an appropriate means of data entry, as it differs from a mobile handset in terms of both feel and layout, and users would be likely to exhibit a markedly different style when entering the data. As such, the data capture was

performed using a modified mobile phone handset, interfaced to a PC through the keyboard connection.

## 5. RESULTS

An analysis of the input data allows an insight into the complexities of successfully authenticating a person from a single input vector of latency values. The problem is that latency vectors observed from a single user may incorporate a fairly large spread of values and as such do not exist on clearly definable classification regions. Figure 1 illustrates some similar and dissimilar input vectors as an indication of the difficulties the pattern classification techniques have in discriminating between users.

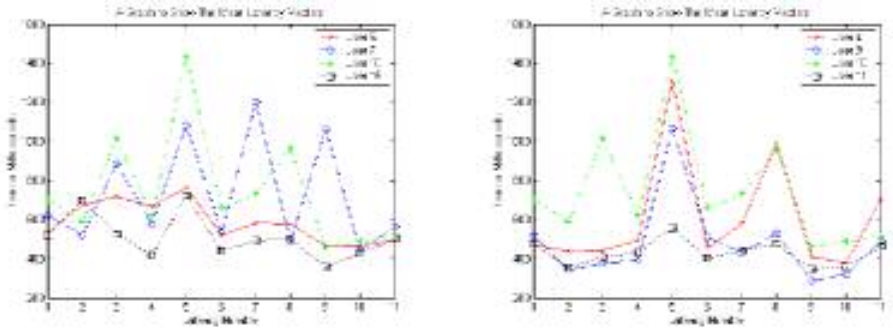

Figure 1: (a) Dissimilar User Input Latency Vectors (b) Similar User Input Latency Vectors

In order to help improve the boundaries between user's responses any input vectors three standard deviations away from the users mean latency value were removed. Table 1 illustrates the effect upon the dataset sizes.

| Input | Original # of Samples | Modified # of Samples | Training Dataset Size | Validation Dataset Size |
|---|---|---|---|---|
| 4-Digit | 30 | 26 | 18 | 8 |
| 11-Digit | 30 | 21 | 14 | 7 |

Table 1: Training & Validation Dataset

Figures 2 and 3 illustrate the results of the comparison of the various classification techniques. The most successful pattern recognition technique with the 4-digit input was the Euclidean technique, with an EER of 14.2%, followed by the mean and standard deviation algorithm, with an EER of 17.7%. Conversely, with the 11-digit input, the mean and standard deviation algorithm performed most successfully with an EER of 17.9%, followed by

the Euclidean technique with an EER of 21.2%. The worst classifier with both the 4-digit and 11-digit inputs was the Mahalanobis algorithm, with EERs of 19.3% and 28.7% respectively. This is somewhat unusual as the Mahalanobis distance algorithms performance was significantly inferior to its linear distance (Euclidean) counterpart. Re-testing both the Euclidean and Mahalanobis algorithms with the validation dataset replaced with the training dataset found that the performance of the Mahalanobis algorithm superseded the Euclidean as would be expect due to the non-linear boundaries it can form. The fact it does not perform as well using the validation dataset would suggest the more general boundaries produced by the Euclidean technique are more appropriate to the complete dataset, indicating the training dataset may not be as representative as required.

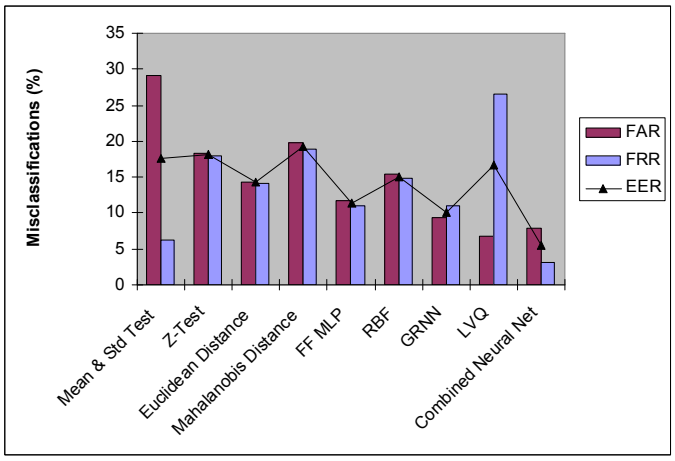


Figure 2: Classification Results for the 4-Digit Input

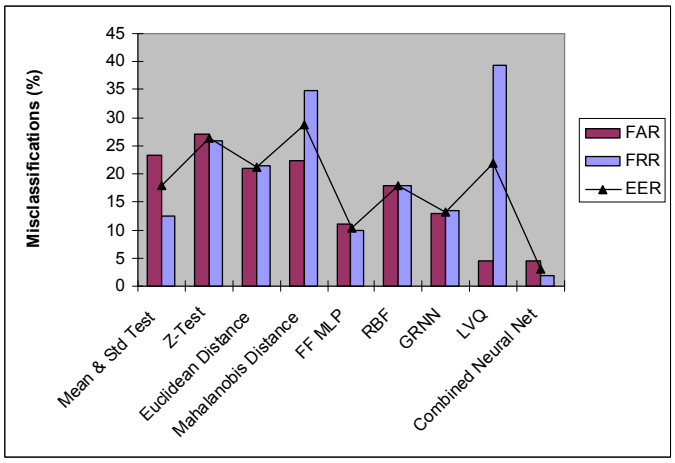


Figure 3: Classification Results for the 11-Digit Input

Overall it was the neural network based approaches that performed best, significantly improving the performance rates. The GRNN was most successful with the 4-digit input achieving a EER of 10.1%, followed by the FF MLP with an EER of 11.3%. Conversely again this role was reversed with the 11-digit input, as the FF MLP was the most successful, obtaining an EER of 10.4%, with the GRNN achieving an EER of 13.1%.

Analysis showed that individual user's performance varied with the different neural network techniques, with some users performing better on one than another. As such a combined neural network technique was created using the best result achieved by each user in any of the neural based techniques, resulting in the classifier achieving an EER of 5.5% and 3.2% with the 4-digit and 11-digit inputs respectively – by far the best result achieved thus far.

As both the 4-digit and 11-digit inputs represent a static keystroke analysis approach, in that they discriminate users based on a identical input string, an extension to the investigation was sought that provided a dynamic approach, in order to gauge the viability. As such 16 participants entered 50 random telephone numbers, which after outliers were removed decreased to 38 (26 for training and 12 for validation). A larger dataset was utilised so that more training data was available due to the more difficult task of discriminating users based on varying input vectors. The results, as would be expected, show the performance of the classification algorithms to be far poorer than the static-based techniques, with the best performance being achieved by the feed-forward MLP with an EER of 24.8%. Using the combined neural network technique the EER reduces to 16.1%.

## 6. DISCUSSION

The investigations have shown the ability for classification algorithms to correctly discriminate between users with a relatively good degree of success, with neural network approaches performing significantly better than their pattern recognition counterparts. The general performance of the 4-digit input vector, (analogous to the PIN) suggests that the entering of a PIN number on a mobile handset has a quite unique dialling pattern, perhaps due to user's previous experience of having to enter such short sequences on a regular basis. Classification algorithms typically find verification simpler when the input vector is larger, as it will typically contain more discriminative information. Overall, this was not the case, with the 4-digit input outperforming the 11-digit input, suggesting that although the user's entry of a fixed 11-digit number is more consistent than for a variable series of numbers, it is not as fluid as the 4-digit PIN input.

From an analysis of the classification algorithms, it is clear that some of the individual network performances experienced 40%+ false acceptance and false rejection rates. This would indicate two problems. Firstly, a user's input varies too considerably from input to input for even a static keystroke analysis technique to prove useful, or secondly, the classifier may not be sensitive enough to the users' data. Both of these problems could be counteracted as the user enters more and more data to the classification engine. However, this error rate is currently unsatisfactory in a mobile operator context and further developments will need to monitor the individual error rates not only the average. Conversely, a number of classifiers (particularly the neural network based techniques) experienced a number of users achieving an FAR and FRR of 0%, reiterating the ability for user's keypad interactions to be discriminative.

Although the static-based classifiers were relatively successful, any effective implementation of keystroke analysis would depend on the ability to provide dynamic-based classification in order to provide non-intrusive, ad hoc authentication. Although the results for dynamic-based classification have been poor in comparison with a static approach they are nevertheless encouraging, especially considering the small datasets with which the classifier was trained and validated. It may also be possible to improve dynamic authentication performance by utilising the more static elements of a varying input such as the area code of a telephone number, thereby reducing the number of varying telephone numbers that could be entered.

The mutually exclusive relationship that exists between the false acceptance and false rejection mean that it would be unlikely for both error rates to achieve near 0% simultaneously [23]. Therefore, the study suggests the best implementation of a keystroke analysis authentication technique would be as a larger hybrid authentication algorithm, involving two or more non-intrusive biometric authentication techniques, such as utilising voice recognition during a voice call, and facial recognition during a video conference.

## 7. CONCLUSIONS & FUTURE WORK

The implementation employed in this study has focussed on the feasibility of a keystroke analysis technique using a number of classification algorithms. Although neural network approaches have clearly outperformed the traditional pattern recognition techniques, the variability of the results in and between the neural network approaches means that much scope remains for fine tuning the networks – especially if larger and more representative input data were made available, with a larger group of participants. In

particular, the most successful algorithm implemented in this study was the combined neural network, but the use of such a technique in practicality is difficult, as training the user iteratively on a wide spread of networks is computational intense and time consuming. In order for this technique to be of any practical relevance it would be necessary to develop an algorithm for analysing a user's input data and (dependant on its complexity) decide which network was most relevant.

This study used keypad interactions exclusively linked with dialling numbers. However the use of mobile handsets for data services such as SMS (Short Message Service) messaging, and other mobile related interactions such as the menu system, opens the possibility of authenticating a user by other means such as the way in which they type words and characters. Additionally, recent research has shown that using classification algorithms that utilise both the inter-keystroke time and hold-time (the time taken to press and release a single key) has more distinct and thus discriminative information than the traditional inter-keystroke timings used in this study [16, 24].

## 8. REFERENCES


[1] Cellular Online. September 2002. www.cellular.co.za

[2] Giussani, B. 2001. Roam – Making Sense of the Wireless Internet. Random House Business Books, London, UK.

[3] BBC. 2002. "Huge surge in mobile phone thefts", BBC News Report, 8th January 2002. http://news/bbc.co.uk/hi/english/uk/newsid_1748000/1748258.htm

[4] Clarke, N.L., Furnell, S.M., Rodwell, P.M. and Reynolds, P.L. 2002. "Acceptance of subscriber authentication methods for mobile telephony devices". Computers & Security, vol. 21, no.3, pp. 220-228.

[5] Wood, H.M. 1978. "The Use of Passwords for controlling the Access to Remote Computer Systems and Services". Computers & Security, vol. 3.

[6] Ashbourn, J. 2000. *Biometric. Advanced Identity Verification. The Complete Guide*. Springer.

[7] Clarke, N.L., Furnell, S.M., Lines, B., Reynolds, P.L. 2002. "Subscriber Authentication for Mobile Phones using Keystroke Dynamics". Proceedings of the Third International Network Conference (INC 2002), Plymouth, UK. pp.347-355.

[8] Umphress D., Williams, G. 1985. "Identity Verification through Keyboard Characteristics". International Journal of Man-Machine Studies, vol. 23, pp. 263-273.

[9] Leggett, J., Williams, G. 1987. “Verifying Identity via Keystroke Dynamics”. International Journal of Man-Machine Studies, vol. 28, pp 67-76.
[10] Joyce, R., Gupta, G. 1990. “Identity Authorisation Based on Keystroke Latencies”. Communications of the ACM, 33(2): 168-176.
[11] Brown, M., Rogers, J. 1993. “User Identification via Keystroke Characteristics of Typed Names using Neural Networks”. International Journal of Man-Machine Studies, vol. 39, pp 999-1014.
[12] Napier, R., Laverty, W., Mahar, D., Henderson, R., Hiron, M., Wagner, M. 1995. “Keyboard User Verification: Toward an Accurate, Efficient and Ecologically Valid Algorithm”. International Journal of Human-Computer Studies, vol.43, pp 213-222.
[13] Monrose, R., Rubin, A. 1999. “Keystroke Dynamics as a Biometric for Authentication”. Future Generation Computer Systems, 16(4) pp 351-359.
[14] Cho, S., Han, C., Han D., Kin, H. 2000. “Web Based Keystroke Dynamics Identity Verification Using Neural Networks”. Journal of Organisational Computing & Electronic Commerce, vol. 10, pp 295-307.
[15] Obaidat, M., Macchairolo, D. 1994. “A Multilayer Neural Network System for Computer Access Security”. IEEE Transactions on Systems, Man, and Cybernetics, vol. 24, no. 5, pp. 806-813.
[16] Obaidat M., Sadoun, B. 1997. “Verification of Computer Uses Using Keystroke Dynamics”. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 27, no. 2, pp.261-269.
[17] Ord, T., Furnell, S. 2000. “User Authentication for Keypad-Based Devices using Keystroke Analysis”. MSc Thesis, University of Plymouth, UK.
[18] Triola, M. 1998. *Elementary Statistics (7th Edition)*. Addison Wesley.
[19] Hogg, R., Ledolter, J. 1989. *Engineering Statistics*. Macmillan Publishing.
[20] Bishop, M. 1995. *Neural Networks for Pattern Classification*. Oxford University Press.
[21] Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation (2nd Edition)*. Prentice Hall.
[22] Kohonen, T. 1997. *Self Organising Maps*. Springer.
[23] Cope, B. 1990. “Biometric Systems of Access Control”. Electrotechnology, April/May: 71-74.
[24] Robinson, J., Liang, V., Chambers, J., MacKenzie, C. 1998. “Computer User Verification Using Login String Keystroke Dynamics”. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, vol. 28, no.2, pp. 236-241.