Subscriber Authentication for Mobile Phones using Keystroke Dynamics

N.L. Clarke[†], S.M. Furnell[†] B.M. Lines[†] and P.L. Reynolds[‡]

[†] Network Research Group, Department of Communication and Electronic Engineering,
University of Plymouth, Plymouth, United Kingdom.
[‡] Orange Personal Communications Services Ltd, Bradley Stoke, Bristol, United Kingdom.
Email: nrg@plymouth.ac.uk

Abstract

With the introduction of third generation phones, a technological transition is occurring in which the devices begin to have similar functionality to that of current personal digital assistants. The ability of these phones to store sensitive information, such as financial records, digital certificates and company records, makes them desirable targets for impostors. Current security for mobile phones is provided by the Personal Identification Number (PIN), which has weaknesses from both technological and end-user perspectives. As such, non-intrusive and stronger subscriber authentication techniques are required. This paper details the feasibility of one such technique, the use of keystroke dynamics. This feasibility study comprises a number of investigations into the ability of neural networks to authenticate users successfully based upon their interactions with a mobile phone keypad. The initial results are promising with individual users' classification performing as well as 0% false rejection and 1.3% false acceptance

1 Introduction

Mobile phones are becoming an ever-increasing part of our lives, with users becoming more reliant upon the services that they can provide. The evolution has been directed towards the provision of data services, by increasing data rates through technologies such as the General Packet Radio Service (GPRS) and the emerging third generation networks, which will enable a broadband service of up to 2Mbps (UTMS Forum 1998). With this increase in information capability, the mobile phone will begin to acquire many of the uses a personal computer has today. Access security currently takes the form of a Personal Identification Number (PIN), a secret-knowledge approach that relies heavily on the user to ensure its validity. For example, the user should not use the default factory settings, tell other people their PIN, or write it down. Apart from the technological arguments, a recent survey into attitudes and opinions of mobile phone customers found that 45% of respondents thought the PIN to be inconvenient and did not use the facility (Clarke et al. 2001). The findings also demonstrated the user's awareness of the security implications, with 81% of respondents overwhelmingly in supported for more security. Therefore the protection against unauthorised access and use of mobile phones is currently questionable – not because users do not want protection, but because they do not like the current method by which it is achieved.

It is clear that an alternative means of subscriber authentication is required to replace the PIN, but at the same time, other forms of secret knowledge-based approach are likely to be regarded as similarly inconvenient. It is, therefore, considered appropriate to examine the potential of a fundamentally different strategy. Amongst the most powerful approaches to facilitate this are biometrics, which are based not on what the user *knows*, but who the user *is*. Biometrics can include physiological characteristics, such as fingerprints and hand geometry, and behavioural traits, such as voice and signature. Another behavioural biometric is keystroke dynamics, which measures the typing pattern of a user. This paper presents the findings of an investigation into the feasibility of using keystroke dynamics to authenticate users on a mobile handset, according to the way in which they use the keypad.

2 Background Concepts

The principal concept behind keystroke dynamics is the ability of the system to recognise patterns, such as characteristic rhythms, during keyboard interactions. A significant amount of prior research has been conducted in this domain, dating back to the 1980s (Legett and Williams 1988; Joyce and Gupta 1990; Monrose and Rubin 1999). However, all of these studies, have focused upon alphabetic inputs from a standard PC keyboard. Little work to date has considered the feasibility of assessing numeric input as the basis for authentication (Ord 1999), and to the best knowledge of the authors, no work has evaluated the application of the technique to a context such as a telephony handset (although the idea was previously proposed by Furnell et al. (1996)).

The assessment of keystroke dynamics can be based upon the more traditional statistical analysis or relatively newer pattern recognition techniques, and previous published studies have incorporated both approaches. The results generally favour the effectiveness of the pattern recognition, with neural network approaches having been shown to perform well (Cho et al. 2000). The network configurations of particular interest are the *Feed-Forward Multi-Layered Perceptrons (MLP),* as they have particularly good pattern associative properties and provide the ability to solve complex non-linear problems (Bishop 1995).

The size of the neural network in terms of number of layers, and number of neurons per layer, plays a key role in the processing ability of the network. However, in the design of neural networks very few, if any, solid rules exist to govern the size of neural networks, with respect to problem complexity. As such, concerns over network size are solved in this study through an iterative process of review and modification. For more information about the design, structure, training and implementation of neural networks, see reference (Bishop 1995; Haykin 1999).

As with other biometric techniques, the performance of the neural network classification for keystroke dynamics is measured using two error rates, the False Acceptance Rate (FAR) and False Rejection Rate (FRR). The former represents the level to which impostors are authorised by the network, and the latter is the likelihood an authorised user being rejected. However with keystroke dynamics, as with all biometric techniques, a threshold must be chosen for the error rates. The trade-off exists between high security/low user acceptance (a threshold value that provides a low FAR and high FRR) and low security/high user acceptance (a threshold value that provides a high FAR and low FRR). It is generally held as being infeasible to simultaneously achieve zero as they share a mutually exclusive relationship (Cope 1990). The point at which the FAR and FRR errors coincide is termed as the Equal Error Rate (EER) (Ashbourn 2000) and is often used as a performance measure when comparing biometric techniques. These measures are used as the basis for evaluating the practical experiments discussed in this paper.

3 Experimental Procedure

The eventual application of keystroke dynamics to a mobile phone would ideally authenticate a user by monitoring his or her continuous 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 the feasibility of the technique rather than to provide a complete solution to the problem. As such, the initial study has been confined to two types of data, namely:

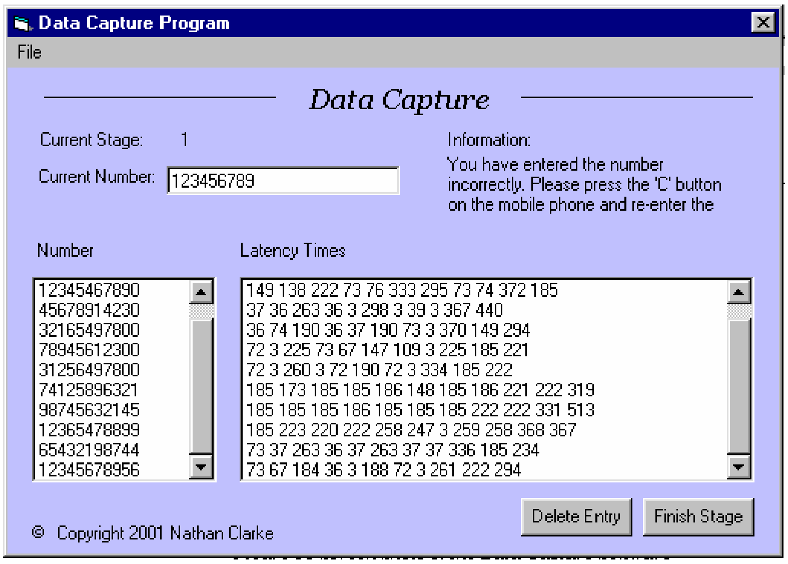
1. PIN code, representing a 4 digit number plus the enter key (i.e. 5 key presses in total).
2. Telephone Number, including area code, representing a 10/11 digit numerical number plus the call key (i.e.11/12 key presses in total).

From these sets of data, three investigations were designed, which sought to assess the ability of a neural network to classify users based upon:

1. Entry of a fixed four-digit number, analogous to the PINs used on many current systems. The users entered the same four-digit code thirty times. Twenty of these inputs were utilised in the training of the neural network, with the remaining ten used as validation samples.
2. Entry of a series of telephone numbers. Fifty mock telephone numbers are entered per user. The classification of inputs was expected to increase inter-sample variance, and thereby make it harder for the network to classify. Thirty samples were used in the training of the network, with the remaining twenty used as validation samples.
3. Entry of a fixed telephone number in order to facilitate a comparison against the results from the second experiment. As with the fixed four-digit investigation, there are thirty samples, twenty for training and ten for validation.

A total of sixteen test subjects provided the input data required for all three investigations. The neural networks in all investigations were trained 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. However, it was considered that the standard numerical keypad on a PC keyboard would not 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. Figure 1 shows a screenshot from the data capture software that was used.



Current Stage: 1

Current Number: 123456789

Information:
You have entered the number incorrectly. Please press the 'C' button on the mobile phone and re-enter the

Number	Latency Times
12345467890	149 138 222 73 76 333 295 73 74 372 185
45678914230	37 36 263 36 3 298 3 39 3 367 440
32165497800	36 74 190 36 37 190 73 3 370 149 294
78945612300	72 3 225 73 67 147 109 3 225 185 221
31256497800	72 3 260 3 72 190 72 3 334 185 222
74125896321	185 173 185 185 186 148 185 186 221 222 319
98745632145	185 185 185 186 185 185 185 222 222 331 513
12365478899	185 223 220 222 258 247 3 259 258 368 367
65432198744	73 37 263 36 37 263 37 37 336 185 234
12345678956	73 67 184 36 3 188 72 3 261 222 294

© Copyright 2001 Nathan Clarke

Figure 1 : Data Capture Software

Due to the limitations of data collection, the input data required for training and testing of the authentication system had to be collected in a single session. Ideally, the data would be collected over a period of time, in order to capture a truer representation of the users typing pattern. For example, by asking the user to type in 50 telephone numbers all at once, could result in an exaggerated learning curve.

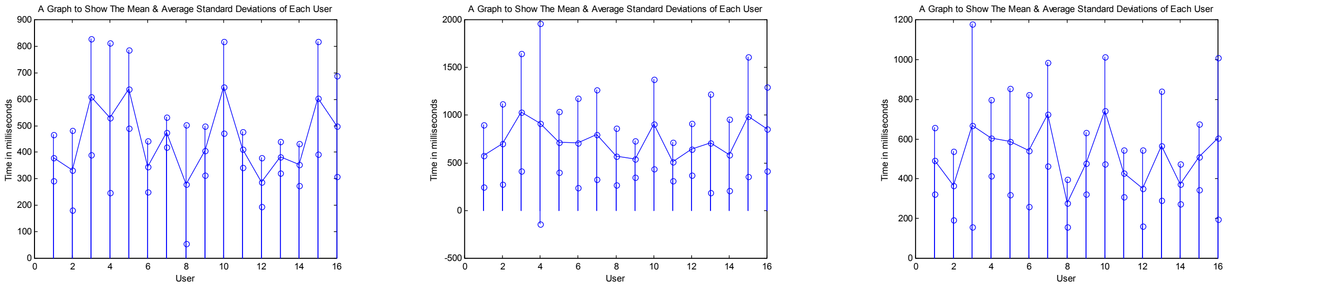
4 Results

The 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. This spread, otherwise known as variance, is likely to encompass input vectors that closely match other users. Because users' latency vectors do not exist on clearly definable classification regions, the problem is made that much more complex for the neural networks.

Two types of variance exist in the latency data:

- inter-sample variance, which ideally would be zero, so that every sample a user inputs would be identical and therefore easier to classify.
- inter-user variance, a measure of the spread of the input samples between users, which would be ideally as large as possible in order widen the boundaries between classification regions.

An initial analysis of the inter-sample variance indicates that they are not ideal by any means, however some users obviously have smaller inter-sample variances than others. The graphs in Figure 2 illustrate the inter-sample mean for each of the users in each investigation. Significant differences can be noted between the three sets of results, such as generally smaller standard deviations and the lower average latency for the fixed telephone number tests when compared to those from the test in which varying numbers were used. This was expected, in the sense that users would become used to entering the fixed telephone number, and therefore the inter-sample variation would progressively decrease. However, the 4-digit PIN investigation shows the lowest inter-sample variance, possibly indicating strong classifiable regions.



(a) (b) (c)

Figure 2 : Mean & Standard Deviation of User's for (a) 4-digit PIN, (b) varying telephone numbers, and (c) fixed telephone numbers

It is interesting to note that the inter-user variance is not considerably larger than the inter-sample variances, as would be favourable, indicating that less well defined classification boundaries exist.

Analysis of individual network performances shows unfavourably large error rates, with some users experiencing FAR/FRR pairs of 41%/20% and 37%/60% in the telephone number investigation. The large error rates suggests there are groups of users with more similar typing characteristics than others, thus making it difficult for the networks to classify them correctly. In particular, two groups of users were identified as having high false acceptances as each other. One such group is illustrated in figure 3(a), with figure 3(b) illustrating a group of dissimilar user inputs. However, in constrast, some users exhibited much more encouraging FAR/FRR figures, such as 1.3%/0% and 4%/10%, both of which were observed in the PIN code investigation. Results such as these suggest that keystroke characteristics can indeed be used to facilitate correct classification, but further development is required in improving network sensitivity and generalisation in other cases.

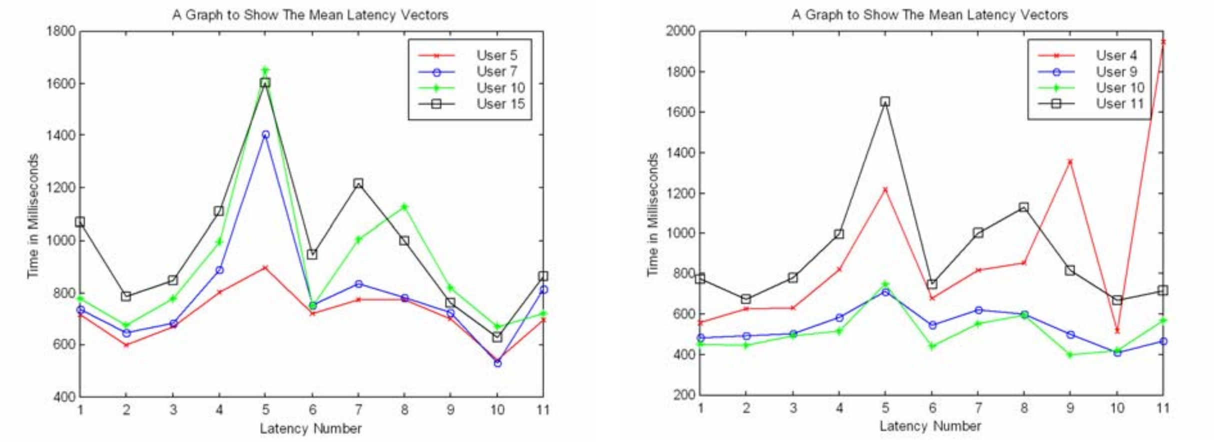


Figure 3 : (a) Similar User Input Latency Vectors (b) Dissimilar User Input Latency Vectors

The overall performances of the neural networks are illustrated in figures 4(a), (b) and (c). The optimum configurations for the MLP's were 11 inputs, 22 neurons for both the 1st and 2nd hidden layers, and 1 neuron in the output layer (i.e.11-22-22-1) for the telephone number based investigations, while the configuration for the PIN based investigation was 4-8-8-1. Unsurprisingly the fixed input networks of the PIN and fixed-telephone investigations performed substantially better than the pseudo-random telephone investigation. The difference between the two telephone investigations are an improvement in the FRR of over 50% and 35% in the FAR. Interestingly, the results indicate that the neural networks can classify the 4-digit PIN input at least as well as an 11-digit fixed telephone input. It would be normal to assume the more information a system has, the better it is able to classify the inputs.

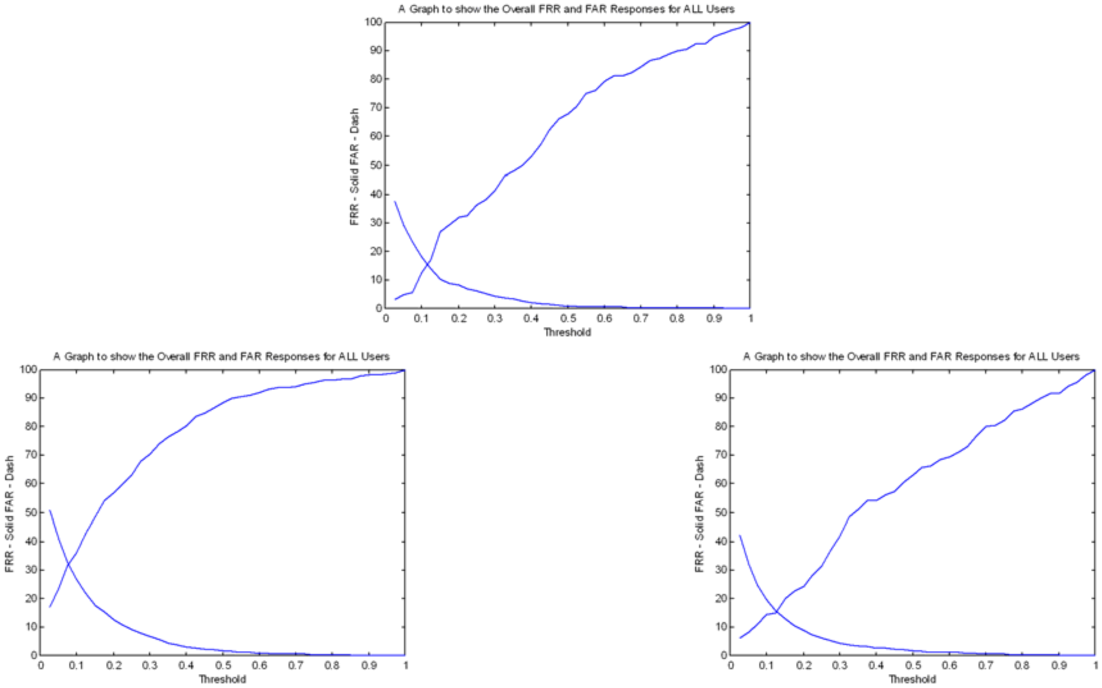


Figure 4 : Overall FRR & FAR for the (a) PIN Code Input Neural Network (b) Pseudo-Random Telephone Neural Network (c) Fixed Telephone Neural Network

The exclusivity of the FA and FR rates are clear, as one error rate decreases the other increases. The equal error rates (EER) for this study are shown in table 1. The threshold value assigned to the network is the level at which the network operator considers that the compromise between security and convenience has been established. For this study, the threshold level was kept constant throughout each of the networks per investigation to enable comparison. Both the PIN Code and mock telephone number networks were given static thresholds of 0.1, and the fixed telephone a threshold of 0.125. Tables of results for a static threshold level can be seen in table 1.

Investigation	FAR (%)	FRR (%)	EER (%)
PIN Code	18.1	12.5	15
Varying Telephone	36.3	24.3	32
Fixed Telephone	16	15	15

Table 1 : Investigation Results

5 Discussion

The investigations have shown the neural networks ability to classify valid and invalid users with a relative degree of success. The networks ability to classify users entering a varying series of telephone number was, as expected, the weakest of network configurations. The classification of fixed 4-digit input suggests that the entering of a PIN number has a quite unique dialling pattern of its own. The reason for this might lie in the fact that users become familiar with typing the 4 digit PIN quite quickly, enabling improved classification. The fixed telephone number has more digits, so while the entry is more consistent than for a variable series of phone numbers, it is not as fluid as the PIN code. However, more practice would probably improve this. Additionally, the 11-digit input has a longer feature set, making it more difficult for an impostor to duplicate.

From the two investigations surrounding the telephone number input, it can be seen that improvements in the inter-sample variance experienced between the varying and fixed telephone numbers has provided a proportionality higher improvement in network performance. However, it should also be noted that the inter-sample and inter-user variances are not the only relationships that determine the neural networks ability to classify users. For instance, the inter-sample variance of User 8 in the PIN investigation is one of the largest in the user group and covers the input latency range of other users, indicating a small inter-user variance. Yet, User 8 has the best FAR and FRR of all the users.

From the analysis of the individual network performances, it is clear that some networks perform far better than others. For instance User 9 in the PIN investigation has an FAR of 90% with Users 11 and 13. This could indicate one of two problems. Firstly, the typing patterns of those users are just too similar and no network would be able to successfully classify those users on a regular basis, or, more likely, it may be the case that the network is not sensitive enough to users data and through further training of the networks with those users with similar responses will help increase network sensitivity. Either way, this error rate is completely unsatisfactory and any further development will need to monitor the individual user performances, not just the average. Conversely, it is worth noting that individual networks performed as well as 0% FRR and 1.33% FAR, indicating that user typing patterns can be classifiable with a good degree of accuracy.

The FAR and FRR errors indicate how often a valid and invalid user will be authenticated onto the system. The trade off between the inconvenience of valid users not being accepted and invalid users being accepted means ideally a level has to be chosen at which these are both minimised. However the likelihood is one error rate will be minimised over another. From the results, if the FAR were to be set in the 2% range this would translate to having an FRR of approximately 55% for the PIN code and fixed telephone inputs. Inversely, setting the FRR in the 5% range (lowest FRR level) corresponds to approximately an FAR of 40%. It would be likely that a level in between theses extremes would be chosen by the network operator, to ensure the impact on legitimate users is minimised, but keeping a practical and useable level of security.

6 Conclusion

This paper has presented an investigation into the feasibilty of using keystroke analysis as a means of enhancing subscriber authentication on mobile handsets. Although the mis-authentications observed at this stage indicate that a practical implementation would prove too error prone, the nature of the investigtions, and the controlled environment in which they were carried out, are believed to be large contributing error factors (as well as actually being necessary to establish a worst case senerio).

The implementation employed in this study has adapted the neural network approach to determine the feasibility of a keystroke-based technique. As such, several areas for possible further research and experimentation can be identified. The first would be to obtain more representative input data, which would ideally incorporate all user input data from the mobile phone (including keystrokes relating to SMS text message entry and menu interactions), and be obtained over a reasonable period of time, in order to ensure a truer representation of users normal behaviour.

From an analysis perspective, further developments could include:

- Removal of outliers from the source input. A quick analysis of the user input data shows a small number of anomalies, which could be unfavourably biasing the network. This will have the effect of reducing the inter-sample variance.
- Increased network sensitivity by training the network using impostor input data that closely matches that of the authorised user, rather than training with all impostor input data.
- Use of generalisation techniques, such as early stopping and regularisation, to optimise the training of the network.
- Analysis of network structure, in terms of network interconnections and transfer functions. Although feed-forward backpropagation networks are amongst the best pattern associators at present, this need not be the case. A structure may exist that is better able to classify this particular problem.
- Updating network configuration over time through re-training.

However, no matter how accurate keystroke analysis becomes, the mutually exclusive relationship between false acceptance and false rejection rates would mean that it is unlikely that 0% can be achieved for both simultaneously. Therefore the study suggests the best implementation of a keystroke analysis authentication technique would be as part of a larger hybrid authentication algorithm, involving two or more non-intrusive biometric authentication techniques for normal authentication.

The technqiues discussed here will be the focus of futher research and practical experimentation by the authors.

References

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

Bishop, M. 1995. *Neural Networks for Pattern Recognition.* Oxford University Press.

Cho, S., Han, C., Han, D., Kim, H., 2000. *Web Based Keystroke Dynamics Identity Verification using Neural Networks.* Journal of Organisational Computing & Electronic Commerce, Vol. 10, No.4, pp 295-307.

Clarke, N., Furnell, S., Rodwell, P., Reynolds, P. 2001. *Acceptance of Subscriber Authentication for Mobile Telephony Devices*. Computers & Security. (In press)

Furnell, S.M.; Green, M.; Hope, S.; Morrissey, J.P. and Reynolds, P.L. 1996. *Non-Intrusive Security Arrangements to support Terminal and Personal Mobility.* Proceedings of EUROMEDIA 96, London, UK, 19-21 December 1996. pp167-171.

Haykin, S. 1999. *Neural Networks. A comprehensive Foundation. Second Edition.* Prentice Hall.
Cope, B. 1990. “Biometric Systems of Access Control”, *Electrotechnology*,
April/May: 71-74

Joyce, R., Gupta, G., 1990. *Identity Authorisation Based on Keystroke Latencies.* Communications of the ACM, 33(2): 168-176, February.

Legett, J., Williams, G., 1988. *Verifying User Identity via Keystroke Characteristics.* International Journal of Man-Machine Studies, Vol 28, pp 67-76.

Monrose, F., Rubin, A., 1999. *Keystroke Dynamics as a Biometric for Authentication.* Future Generation Computer Systems, 16(4) (2000) pp 351-359.

Ord, T. 1999. *User Authentication Using Keystroke Analysis with a Numerical Keypad Approach.* MSc Thesis, University Of Plymouth, UK.

UMTS Forum. 1998. *The Path Towards UMTS – Technologies for the Information Society.* Report Number 2. http://www.utms-forum.org/reports.html