Keystroke Analysis for Thumb-based Keyboards on Mobile Devices

Sevasti Karatzouni and Nathan Clarke
Network Research Group, University of Plymouth, Plymouth, PL4 8AA,
United Kingdom, nrg@plymouth.ac.uk
WWW home page: http://www.network-research-group

Abstract. The evolution of mobile networking has opened the door to a wide range of service opportunities for mobile devices, increasing at the same time the sensitivity of the information stored and access through them. Current PIN-based authentication has proved to be an insufficient and an inconvenient approach. Biometrics have proven to be a reliable approach to identity verification and can provide a more robust means of security, as they rely upon personal identifiers. Amongst various biometric techniques available, keystroke analysis combines features that can offer a cost effective, non-intrusive and continuous authentication solution for mobile devices. This research has been undertaken in order to investigate the performance of keystroke analysis on thumb-based keyboards that are being widely deployed upon PDA's and Smartphone devices. The investigation sought to authenticate users whilst typing text messages, using two keystroke characteristics, the inter-keystroke latency and hold-time. The results demonstrate the approach to be promising, achieving an average EER=12.2% with the inter-keystroke latency based upon 50 participants. Uniquely to this tactile environment however, the hold-time characteristic, did not prove to be a reliable feature to be utilised.

1 Introduction

The proliferation of mobile devices and mobile networking has introduced new challenges for the protection of the subscribers' assets. The security risks are no longer associated only with safeguarding the subscriber's account. With the introduction of 3rd generation mobile networks, the services and information accessible through mobile handsets has increased in sensitivity, as micro-payments, mobile banking and location-based services are all now a reality for the mobile world [1]. Statistics show that mobile theft in the UK accounts 45% of all theft [2], a fact,

Please use the following format when citing this chapter:

Karatzouni, S. and Clarke, N., 2007, in IFIP International Federation for Information Processing, Volume 232, New Approaches for Security, Privacy and Trust in Complex Environments, eds. Venter, H., Eloff, M., Labuschagne, L., Eloff, J., von Solms, R., (Boston: Springer), pp. 253–263.

which when combined with the information that can be stored on mobile handsets and the attraction that high-tech devices can pose, presents a further concern for enhanced security.

Current authentication, principally achieved by PINs, is not enough to substantially safeguard today's mobile handsets and the data accessed through them. As a secret knowledge technique it has several well established drawbacks, such as being shared, written down or kept at factory default settings [3]. Furthermore, as survey results demonstrate, subscribers consider it an inconvenient method and as such tend not to use it in the first place, leaving their device completely unprotected [4]. This is not only limited to the general public, as the Mobile Usage Survey 2005 reveals, only 2 thirds of the IT managers surveyed have enabled password security in their mobile devices, despite acknowledging the amount of sensitive business information that is stored upon them [5].

Of the two remaining authentication approaches - tokens and biometrics, the latter can offer a more viable approach. Token-based authentication implemented to date by SIM cards does not provide any protection for the user as it is unlikely to be ever removed from the device. Biometrics could provide an enhancement on the current security, as authentication is based upon a unique characteristic of a person. This fact introduces a unique level of security that other approaches are unable to accomplish, as it relates the process to a person and not to the possession of knowledge or a token. A biometric approach that can provide a cost-effective and a non-intrusive solution for mobile handset authentication is keystroke analysis, a technique which is based on the typing dynamics of a user.

The purpose of this research is to investigate the feasibility of keystroke analysis on thumb-based keyboards based on text messaging input, looking to apply this technique as an authentication method for mobile handsets that offer that unique tactile interface. The paper proceeds with section 2 describing the unique characteristics utilised in keystroke analysis and provides an overview of keystroke analysis studies to date. Sections 3 and 4 describe the methodology and results of the study. A discussion of the results, placing them in context and areas for future research are presented in Sections 5 and 6.

2 Keystroke analysis

Keystroke analysis is a behavioural biometric that attempts to verify identity based upon the typing pattern of a user, looking at certain physical characteristics of their interaction with a keyboard. Considerable research has been undertaken on the method since first suggested by Spillane [6] in 1975, with studies identifying two main characteristics that provide valuable discriminative information:

- Inter-keystroke latency, which is the interval between two successive keystrokes, and
- Hold-time, which is the interval between the pressing and releasing of a single key

The majority of the studies to date have investigated the feasibility of keystroke analysis on full QWERTY keyboards [7 – 10], showing good results for both of the characteristics mentioned. In general, the inter-keystroke latency has demonstrated better discriminatory characteristics for classification in comparison to hold-time.

As in all biometrics the method to assess the performance of keystroke analysis, is by using the False Acceptance Rate (FAR), which indicates the probability of an impostor being granted access to the system, and the False Rejection Rate (FRR), which represents the degree to which a legitimate user is rejected. A trade-off exists between these rates, in terms of increasing security (and therefore increasing user inconvenience) and increasing user convenience (and thus decreasing the security). The point at which those two rates cross is referred to as the Equal Error Rate (%) and is used as a more objective means of comparing the performance of different biometric techniques.

The underlying classification algorithms utilized in keystroke analysis were traditionally statistically based [7, 8, 10]. However, advancements in neural networks have shown this technique to be more successful. A summary of key literature and results within the domain of keystroke analysis on PC keyboards is illustrated in Table 1.

Table 1. A summary of literature & results on keystroke analysis on PC keyboards

Study	Users	Input	Inter-key	Hold-time	Approach	FAR (%)	FRR (%)
Umpress & Williams[7]	17	Alphabetic	●		Statistical	11.7	5.8
Joyce & Gupta [8]	23	Alphabetic	●		Statistical	0.3	16.4
Brown & Rogers [9]	25	Alphabetic	●	●	Neural N.	0	12
Obaidat & Sadoun [10]	15	Alphabetic	●	●	Neural N.	0	0
Ord & Furnell [11]	14	Numerical	●		Neural N.	9.9	30

Although continuous research on keystroke analysis has been conducted since the 1980's, it was not until more recently that the method was assessed on interfaces provided on mobile phones where the tactile environment considerably differs. A series of studies accessed the method on regular mobile phone keypads with promising outcomes, achieving an EER of 8% based on numerical input [12]. Nevertheless, the performance of keystroke analysis for other tactile environments such as thumb-based keyboards is undocumented. Thumb-based keyboards constitute an interesting gap in research as they provide the extensive interface of a PC keyboard and the thumb-based properties of a mobile phone.

3 Methodology

This study looked into the feasibility of authenticating a user whilst typing text messages. Two different types of analysis were conducted in the context of this research: static analysis utilising the inter-keystroke latency and pseudo-dynamic utilising the hold-time characteristic. A total of fifty participants took part in the study, involving the largest population of participants for a study such as this and enabling more statistically significant results to be concluded. The participants were asked to enter thirty messages, with each message specifically designed to ensure that certain requirements are met.

For the static analysis six varying sized keywords were included in the text messages providing a static classification component. The varying nature of the static keywords permitted an evaluation of the word length versus performance. Thirty repetitions of each keyword were included, to ensure enough data for classification. The words selected are listed in Table 2, along with the number of inter-keystroke latencies that they involve and the number of samples used for training and testing after outliers were removed (a standard procedure for keystroke analysis studies [7-15].

Table 2. Keywords used for inter-key latency

Keyword	# Inter-keystroke latencies	#Samples after outliers' removal	Training Set	Testing Set
everything	10	27	18	9
difficult	9	26	18	8
better	6	27	18	9
night	5	27	18	9
the	3	26	18	8
and	3	27	18	9

Fig. 1. An XDA IIs thumb-based keypad

Fig.2. Screenshot from experiment software

Literature has showed that attempts to perform dynamic analysis on keystroke dynamics [13, 14] did not yield satisfactory results. As such an attempt was made to utilize a static component – the recurrent letters, in a dynamic form of analysis. The

pseudo-dynamic analysis was based upon the hold-time of the six most recurrent letters in the English language – 'e', 't', 'a', 'o', 'n' and 'i' - an adequate number of repetitions of which were included within the messages.

The text messages were entered using an XDA IIs handset that deploys a representative example of today's thumb-based keyboards, as illustrated in Figure 1. In order to capture the keystroke data, appropriate software was developed using Microsoft's Visual Basic .NET, and deployed on the handset. A screenshot of the software is illustrated in Figure 2. As usual in keystroke analysis studies, corrections were not permitted in case the user misspelled a word as this would undesirably interfere with the data [7]. Instead, the whole word had to be retyped in the correct form. Although it would be preferred to collect the data during multiple sessions, as a more indicative typing profile of the users could be captured, the data collection was performed in a single session, to maximise the number of participants that completed the study.

4 Results

4.1 Inter-keystroke latency

An initial analysis of the input data showed a fairly large spread of values on the inter-keystroke latencies. Even though smaller keywords were expected to give a greater consistency in the typing pattern because of their length and commonality, that was not the case. Additionally, the difference between the values of the different users was not large. These factors put a burden on the classification algorithm, as they make the classification boundaries between users very difficult to establish successfully. Figure 3, illustrates the mean and standard deviation for the larger keyword 'everything' for all users as an example of the problem.

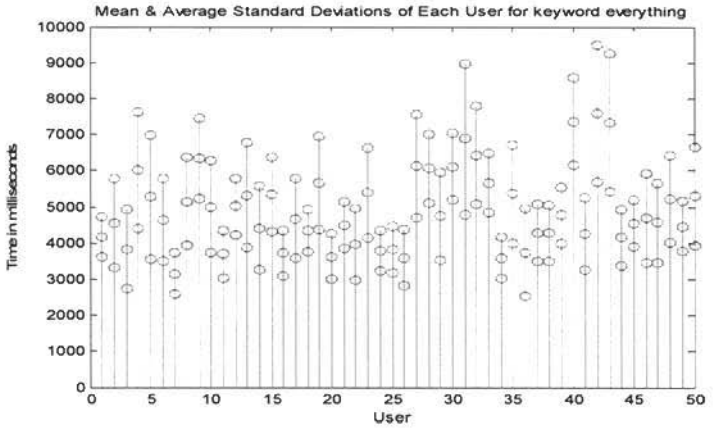

Fig. 3. Mean & Standard deviation for keyword everything

A number of analyses were undertaken, using Feed Forward Multilayer Perceptron Neural Network (FF-MLP) as it had demonstrated better performance in previous studies over other techniques [10, 12, 15, 16, 17]. Different network configurations were tested, looking for optimum performance. The best results achieved were for the keyword 'everything' with an EER of 23.4%. This was somewhat expected as the larger keywords contain more keystroke latencies and subsequently more discriminative information.

As illustrated in Table 3, the results show the FRR is much higher from the FAR which can be explained by the large number of impostors (49) extensively training the network versus the one authorised user. Furthermore, the number of samples assigned to the testing of the classification was small, resulting in the FRR encountering large steps in its transitions when being evaluated.

Although the error rate is fairly high, there were cases of users reaching an EER below 10% with the best case of user 1 achieving an EER of 0.3%, showing the ability to classify some users. The rest of the keywords resulted in higher error rates, with the error increasing as the length of the keyword was reducing. The best results for each keyword are illustrated in Table 3.

Table 3. Best results for each keyword

Keyword	FAR (%)	FRR (%)	EER(%)
everything	12.8	34.2	**23.5**
difficult	13.2	43.0	**28.1**
better	18.0	43.1	**30.5**
night	21.3	45.8	**33.5**
the	23.7	41.5	**32.6**
and	24.3	43.6	**33.9**

The average results of different networks showed minimal change in the EERs, although individual performances did vary. This suggests that the network does not

optimise for individual users but rather forces a standard training scheme upon the user. To overcome this problem a different approach was utilised by Clarke & Furnell [12], which provided an improvement in performance through optimising the number of training epochs. A gradual training technique was performed, training the network for an extensive number of epochs but periodically evaluating the performance. The results showed a noticeable decrease in the error rates with best case achieving an EER of 12.2% for the larger keyword. The summary of the gradual training results are presented in Table 4.

Table 4. Gradual training results for all keywords

Keyword	FAR (%)	FRR (%)	EER(%)
everything	15.8	9.1	**12.2**
difficult	16.8	12.0	**14.4**
better	23.5	14.4	**18.9**
night	24.2	14.4	**19.3**
the	29.3	19.5	**24.4**
and	28.7	17.6	**23.1**

Noticeably, for the keyword "everything", 20 users achieved an FRR of 0% with a respective FAR below 10%, with the best user achieving an FAR of 0.7% and FRR of 0%. The list of best and worst case users for all keywords are illustrated in Table 5. The results underline the requirement for different training intensiveness for each user, and that the inter-keystroke latency offers the discriminative data to classify users in the specific tactile interface.

Table 5. Best & worst case results from gradual training

Keyword	Best Case		Worst Case	
	User	EER (%)	User	EER (%)
everything	2	**0.4**	6	**32.4**
difficult	11	**1.3**	46	**34.1**
better	49	**1.6**	27	**34.2**
night	34	**2.3**	25	**40.5**
the	26	**6.4**	39	**45.8**
and	11	**5.4**	5	**49.4**

4.2 Hold-time

In contrary to the inter-keystroke latency investigation, the hold-time characteristic provided little discriminative information to classify users. A series of tests on different network configurations using all six letters (as to provide the largest possible input vector) resulted in an EER of around 50%, showing little classification performance. The same error rate was achieved using different size subsets of the letters with smaller input vectors (but with the advantage of more repetitions of each

letter) and also with a larger input vector of eight letters through the addition of letters 'r' and 's', as they appear next on the reoccurrence list.

In order to further assess the performance of hold-time, a group of only 20 users was utilised aiming to help the classification problem by reducing the amount and complexity of information presented to the network and thus assisting in the discrimination of authorised and unauthorised users. However, no change in the performance was experienced. Even when gradual training was applied, using the six letters set, no significant improvement was observed. Sample results from various tests are provided in Table 6. Even though there was a 10% decline on the EER using gradual training, the results are still too high to suggest that hold-time can offer any valuable discriminative information.

Table 6. Sample results from various tests on hold-time

Set	Training	Users	FAR (%)	FRR (%)	EER(%)
6 letters	normal	20	49.5	49.4	**49.5**
6 letters	normal	50	31.3	69.0	**50.2**
8 letters	normal	50	26.7	72.9	**49.8**
3 letters	normal	50	22.1	77.6	**49.9**
6 letters	gradual	50	34.2	36.8	**36.8**
6 letters	normal	20	49.5	49.4	**49.5**

5 Discussion

As the results showed the inter-keystroke latency can provide an effective means of differentiating between users. When based on a latency vector of 10, an EER of 12.2% was achieved with the gradual training approach. As was expected the use of smaller input vectors resulted in a corresponding increase in error rates, as the amount of unique discriminative information and feature space reduced.

With regards to the inter-keystroke latency, this study did not experience the very low rates in performance that have been found in previous studies based on regular keyboards. It is suggested that a number of aspects differentiate this investigation from previous studies. The keyboard utilised in this study provides a completely different tactile interface than traditional keyboards, with a more restricted keystroke interface, reduced distance between the keys and smaller key depth. In addition, the number of fingers utilised in typing has also been reduced from typically 10 fingers and thumbs to 2 thumbs. Both of these factors restrict the typing dynamics, as the combinations of the fingers in conjunction with the timing of the keystrokes and movement to achieve them, are reduced. This results in a smaller feature space for the keystrokes characteristics to reside in and subsequently making it more difficult to distinguish between them. Furthermore, although the layout was familiar to all users as it shares the same layout with a PC keyboard, some of the participants experienced difficulty in identifying the placement of the keys due to the different way of typing.

The hold-time characteristic did not provide any real evidence to suggest that it can be utilised in this specific typing interface, though there are a number of factors

that may explain the inability of the keystroke feature. Firstly, the keys that the thumb-based keyboard deploys are very small related to the chunky tactile environment that a normal keyboard offers, restricting the interval length between the pressing and releasing of a key and thus not providing much differentiation in values. Although the hold-time has performed well on regular keypads [12], the keys were larger than the keyboard used in this experiment and the method of calculating the hold time was different. In the study by Clarke & Furnell [12], the hold-time was defined by the first key press down until the last key release, increasing immediately the range of values and thus the feature space (for instance, for the character 'c' the number 1 button would need to be pressed three times).

Furthermore in a thumb-based keyboard, fingers stay almost static due to the limited area. As such, the hand movement which appears in PC keyboards and may affect the pressing of a key is unlikely to happen in this case. What must also be noticed is that some participants complained about the feedback from the keyboard, as they could not at all cases be sure if they had pressed a key, which might have further complicated matters.

6 Conclusions

This research conducted a feasibility study on the utilisation of keystroke analysis as an authentication method in devices that offer the tactile environment of a thumb-based keyboard. The results showed that from the two traditionally used keystroke characteristics, the inter-keystroke latency gave promising results in-line with previous studies undertaken. However, unusually the hold-time characteristic gave no promise of a potential use in this kind of keystroke interface, though further research must be undertaken to determine this conclusively.

Future research will be conducted looking to optimise network configurations for the inter-keystroke latency to take into account the bias towards the network responding in favour of the impostor. Furthermore, the use of different keywords will be investigated, as will the concurrent use of more than one keyword within a single authentication request, the latter aspect having the potential to substantially improve performance. In respect to hold-time, further tests are required before concluding to its ineffectiveness, exploring the use of longer input vectors and different letter subsets. A future experiment will also look to utilise different thumb-based keyboards that offer a slight different tactile environments than the one utilised in this study. Additionally, future work will seek to investigate the performance of the technique in environments representing more practical situations, thereby providing more balanced results. Factors such as the user's interaction with the handset whilst they are walking and their physical condition (e.g. tired or stressed) can be investigated for their impact upon performance.

This study has demonstrated promising results for the use of keystroke analysis, using a significantly large number of participants than previous studies. Although the accuracy of the method does not compete in distinctiveness with other biometrics such as fingerprints, the nature of keystroke analysis in that it can provide a monitoring authentication mechanism, transparent to the user (which is not feasible

for many other techniques) is a positive attribute. If used regularly and in conjunction with other transparent authentication techniques, keystroke analysis can be an effective means of providing a more enhanced level of security.

7 References

1. The UTMS Forum, Mobile Evolution – Shaping the future (August 1, 2003); http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/MultiMedia_PDFs_Papers_Paper-1-August-2003.pdf.

2. British Transport Police, Mobile phone theft (August 20, 2006); http://www.btp.police.uk/issues/mobile.htm.

3. R. Lemos, Passwords: The Weakest Link? Hackers can crack most in less than a minute, CNET.com, (2002), http://news.com.com/2009-1001-916719.html.

4. N. Clarke, S.M. Furnell, P.M. Rodwell, P.L. Reynolds, Acceptance of subscriber authentication method for mobile telephony devices, *Computers & Security*, 21(**3**), pp220-228, 2002.

5. Pointsec, IT professionals turn blind eye to mobile security as survey reveals sloppy handheld habits (November 17, 2005); http://www.pointsec.com/news/release.cfm?PressId=108.

6. R. Spillane, Keyboard Apparatus for personal identification, IBM Technical Disclosure Bulletin, 17(**3346**) (1975).

7. D. Umphress, G. Williams, Identity Verification through Keyboard Characteristics, *International Journal of Man-Machine Studies*, 23, pp. 263-273 1985.

8. R. Joyce, G. Gupta, Identity Authentication Based on Keystroke Latencies, *Communications of the ACM*, 39, pp 168-176 1990.

9. M. Brown, J. Rogers, User Identification via Keystroke Characteristics of Typed Names using Neural Networks, *International Journal of Man-Machine Studies*, 39, pp. 999-1014 (1993).

10. M. S. Obaidat, B. Sadoun, Verification of Computer User Using Keystroke Dynamics, *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 27(**2**), (1997).

11. T. Ord, User Authentication using Keystroke Analysis with a Numerical Keypad Approach, (MSc Thesis, University of Plymouth, UK, 1999).

12. NL. Clarke, S.M. Furnell, Authenticating Mobile Phone Users Using Keystroke Analysis, *International Journal of Information Security*, ISSN:1615-5262, (2006), pp.1-14.

13. G. Leggett, J. Williams, Verifying identity via keystroke characteristics, *International Journal of Man-Machine Studies*, Vol. 28(1), (1988), pp.67-76.

14. R. Napier, W. Laverty, D. Mahar, R. Henderson, M. Hiron, M. Wagner, Keyboard User Verification: Toward an accurate, Efficient and Ecological Valid Algorithm, *International Journal of Human-Computer Studies*, 43, pp.213-222 (1995).

15. S. Cho, C. Han, D. Han, H. Kin, Web Based Keystroke Dynamics Identity Verification Using Neural Networks, *Journal of Organizational Computing & Electronic Commerce*, 10, pp. 295-307 (2000).

16. S. Haykin, *Neural networks: A Comprehensive Foundation (2nd edition)*, (Prentice Hall, New Jersey, 1999).

17. M. Bishop, *Neural Networks for Pattern Classification*, (Oxford University Press, New York, 1995).