Keystroke Biometric Studies on Password and Numeric Keypad Input

Ned Bakelman, John V. Monaco, Sung-Hyuk Cha, and Charles C. Tappert
Seidenberg School of Computer Science and Information Systems
Pace University, White Plains, NY 10606, USA
{vinmonaco, nbakelman}@gmail.com, {scha, ctappert}@pace.edu

***Abstract*—The keystroke biometric classification system described in this study was evaluated on two types of short input – passwords and numeric keypad input. On the password input, the system outperforms 14 other systems evaluated in a previous study using the same raw input data. The three top performing systems in that study had equal error rates between 9.6% and 10.2%. With the classification system developed in this study, equal error rates of 8.7% were achieved on both the features from the previous study and on a new set of features. On the numeric keypad input, the system achieved an equal error rate of 10.5% on the features from the previous study and 6.1% on a new set of features.**

Keywords*—*pattern recognition, machine learning, biometrics, keystroke biometrics, user authentication

I. INTRODUCTION

Keystroke biometric systems measure typing characteristics believed to be unique to an individual and difficult to duplicate [5, 12]. The keystroke biometric is a behavioral biometric using a number of measurements or features to characterize an individual such as key press duration (dwell) times, transition (latency) times, and the identity of the keys pressed. Most previous keystroke biometric studies dealt with passwords or other short input [3, 10, 13, 15, 16, 17, 20, 22, 23], and there are now a number of "password hardening" commercial products, such as [1, 2, 4, 6, 11, 14]. Fewer studies have investigated long-text input [8, 10, 18, 19, 21, 24, 25]. The short and long input keystroke applications are considerably different and the short input is usually fixed in contrast to arbitrary text for long input.

Biometric analysis of short keystroke input has high security importance in two common applications – password input and number pad input. The password application is important because computer users frequently use passwords to logon to computers and to access email, bank, brokerage, and online store accounts. The number pad input application is important for similar reasons because digit only passcodes are used for access on automated teller machines (ATMs), on electronic security keypads for building and room access, and on mobile/digital phones. Passwords and digit-only passcodes are currently the only security measures employed in these access applications and the password/passcode information that we are required to remember is easily compromised. These security threat situations could be improved considerably if it were possible to authenticate the user more precisely, distinguishing between the genuine user and an imposter.

Two studies were conducted to measure the performance of the keystroke biometric authentication system in the password and numeric keypad input applications. Carnegie Mellon University (CMU) conducted similar studies [15, 17]. Using the passwords data from CMU [15], the first study consisted of experiments using two different feature sets – the features used in the CMU study and a new feature set created for this study. The second study on short numeric input used two different feature sets – the CMU features and a new feature sets created for the study.

The remainder of this paper is organized as follows: section 2 describes the classification system, section 3 the CMU data studies, section 4 the numeric keypad studies, section 5 the discussion, and section 6 the conclusion.

II. PACE UNIVERSITY CLASSIFICATION SYSTEM

This study evaluates the performance of short input on a recently developed, closed-system keystroke biometric classification system. [1] The key component of this system is the Pace University classification procedure, referred to here as the *Pace Classifier*, which is described briefly here and in more detail in the companion paper (see footnote). All of the experiments described in this paper used the *Pace Classifier*.

What was varied were the data and the feature sets, as described in the following sections. All the features used in these experiments were normalized into the range 0-1 by using a min/max of +/- 2 standard deviations of each feature obtained over the population.

The *Pace Classifier* is based on a vector-difference authentication model which transforms a multi-class problem into a two-class problem [19, 25]. The resulting two classes are *within-person* ("you are authenticated") and *between-person* ("you are not authenticated"). This dichotomy model is a strong inferential statistics method found to be particularly effective in large open biometric

[1] Monaco, et al., "Recent advances in the development of a long-text-input keystroke biometric authentication system for arbitrary text input", also submitted to this conference.

systems where it is not possible to train the system on all individuals in the population.

The application of interest here, however, involves closed populations of users where it is possible to train the system on all of the authorized users. Therefore, a more accurate "engineering" closed-system procedure was developed for these and similar applications. Two performance enhancing modifications were made in converting the open to the closed-system procedure. First, the new procedure matches the claimed user's sample against all the enrollment samples from that user for authentication rather than just one as in the open system. Second, the new procedure is user-focused in that only the claimed user's enrollment samples and their relationships to the other users' enrollment samples are utilized in the classification process.

In the simulated authentication process, a claimed user's keystroke sample requiring authentication is first converted into a feature vector. The differences between this feature vector and all the earlier-obtained enrollment feature vectors from this user are computed, and the resulting difference vectors are matched against the within-person training difference vectors for authentication or against the between-person difference vectors for non-authentication using the *k*-nearest-neighbor procedure. Thus, *differences of difference vectors are being calculated*.

Receiver operating characteristic (ROC) curves characterize the performance of biometric systems and show the trade-off between the False Accept Rate (FAR) and the False Reject Rate (FRR). In this study, the ROC curves were obtained by using a weighted procedure of the *k* nearest neighbors [25]. This procedure uses a linear rank weighting, assigning the first choice (nearest neighbor) a weight of *k*, second a weight of *k-1*, ... , and the k^{th} a weight of 1. The maximum score when all *k* choices are within-person is $k+(k-1)+...+1 = k(k+1)/2$, and the minimum score is *0*. Now, consider that a user is authenticated if the weighted-within-person choices are greater or equal to *l*, where *l* varies from *0* to $k(k+1)/2$, and compute the (FRR, FAR) pairs for each *l* to obtain an ROC curve. The Equal Error Rate (EER) is where FAR = FRR on the ROC curve. These experiments use $k = 21$ to provide weighted scores in the range 0-231 and thus 232 points on the ROC curve. This value of *k* was chosen to generate a reasonable number of points on the ROC curves. When deploying the system the value of *l* is chosen to establish an appropriate operating point trade-off between FAR and FRR on the ROC curve.

III. CMU Password Data Study

The data for the password study were made available by Killourhy and Maxion [15]. Two experiments were conducted on the CMU 51 subject data using the closed-system *Pace Classifier* described in the previous section.

Due to the large number of samples per user, a repeated random sub-sampling validation method was used to derive the ROC curve. Each experiment was repeated three times, each time randomly splitting the feature space into 380 reference samples and 20 query samples per user. The reference samples are further refined by k-means clustering in order to reduce the size of both the within and between spaces. For each user, this would be equivalent to an authentication attempt after having recorded 380 samples. The key independent variable was the feature sets.

The first experiment used the CMU 51 subject data and re-implemented the 31 CMU features [15]. From the 10-character password ".tie5Roanl" + Enter, the 31 CMU features were the 11 hold (dwell) times plus the 10 keydown-keydown and 10 keyup-keydown transition times.

The second experiment was similar to the first but used a set of 75 features designed for this study. The 75 features were time differences between atomic keystroke events, where each event is the action-key combination occurring instantaneously at time *t*. For example, 'press e' and 'release e' are considered as two separate events and are not necessary consecutive. The time differences of every 2-gram which occurs over the population was taken for each user. So although each feature vector contains 75 features, only 21 of these are non-zero, since there are only 22 events (21 pairs). The non-zero features are usually different for each user because some users consistently overlap certain keys while others do not. Thus, there are 75 features, but it depends on the user which features are non-zero.

A summary of the results is shown in Table 1, the ROC curves in Fig. 1, and the FAR/FRR versus parameter *L* curves in Fig. 2. Although the EER can be approximated from the ROC curve, it can be more accurately determined from the crossover point on the FAR/FRR versus *L* curve (*L* is used in place of *l* to avoid confusion with 1). Although *L* goes from 0-231, expanded FAR/FRR plots at low *L* values are shown here because the crossover points on the FAR/FRR versus *L* curves occur in that region.

Table 1. Experimental Results on CMU Password Data.

Experiment	Number of Subjects	Centroids per Subject	Number of Features	EER (%)
1	51	20	31	8.7
2	51	20	75 (21)	8.7

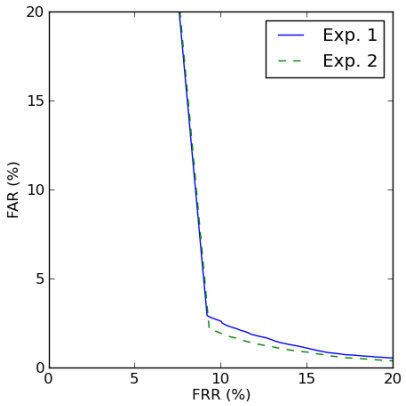

Fig. 1. ROC curves for the CMU data experiments.

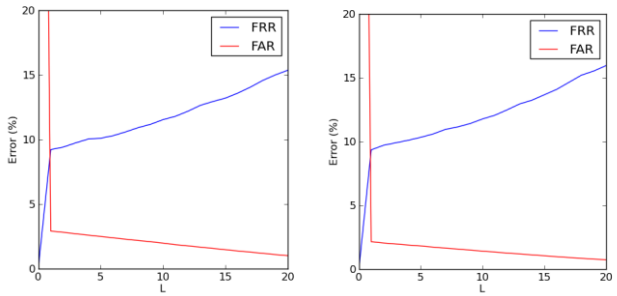

Fig. 2. FAR/FRR versus *L*: Exp. 1 (left) and Exp. 2 (right).

IV. Numeric Keypad Study

Two experiments were conducted on newly collected numeric keypad input to compare against the password performance above and the CMU number-pad study [17]. In this study, although it was not possible to compare the algorithms on the same input data as in the password study above because the CMU data were not available, new numeric keypad data were obtained in a manner similar to how the CMU data were collected.

A. Data Collection

The numeric keypad samples were captured using an open-source third-party keylogger designed by Fimbel [9]. This keylogger was originally developed for testing purposes and serves no malicious intent. It runs in the background of a computer system thus being unobtrusive to the user and allowing for the capture of keyboard and mouse input regardless of the application(s) running on a system.

The numeric keypad data were collected from 30 subjects over a four-day period with no more than 60 samples collected per subject per day. Each subject first practiced keying the input string several times before the samples were recorded. Each sample consisted of the numeric sequence **914 193 7761** (shown here in telephone number format) followed by the **Enter** key to provide a total of 11 keystrokes per sample. All samples were entered with the right hand as if entering a phone number on a digital phone or entering an ATM pin, and only correct numeric-sequence samples were accepted.

B. Experimental Design and Results

The first numeric keypad experiment used the same 31 CMU features as experiment 1 in the previous section. Because the numeric keypad data had only several instances of overlap between keys due to the use of only one hand, a feature set similar to the one used in experiment 2 in the previous section could not be used. Instead, the second keypad experiment used a feature set designed for arbitrary long-text input that consisted of 974 features of which only a portion were designed for the numeric keypad. The features with missing observations fall back to features with a higher frequency according to a fallback model [25]. All of the features are statistical in nature, comprised of averages and standard deviations of key press duration and digraph transition times. The numeric keypad duration features are shown in Fig. 3, transition features not shown. Because the input was restricted to a specific numeric sequence, only 44 of the numeric keypad features were actually used in this experiment.

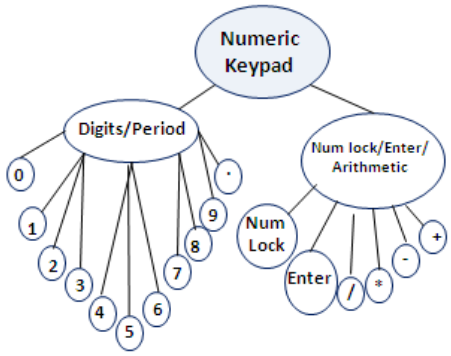

Fig. 3. Numeric keypad duration features.

The leave-one-out cross-validation (LOOCV) procedure was used to evaluate system performance. This procedure simulates many true users trying to get authenticated and many imposters trying to get authenticated as other users. For *n* users each supplying *m* samples, *m*n* positive (one for each sample) and *m*n*(n-1)* negative (each sample versus the other users) tests can be performed, for a total of *m*n*n* tests. Each experiment must be repeated only once since a full ROC curve as described in section II can be derived with this method.

The results are shown in Table 1, the ROC curves in Fig. 3, and the FAR/FRR versus parameter *L* curves in Fig. 4.

Table 2. Password and Numeric Keypad Results.

Experiment	Number of Subjects	Samples per Subject	Number of Features	EER (%)
Numeric Keypad 1	30	20	31	10.5
Numeric Keypad 2	30	20	44	6.1

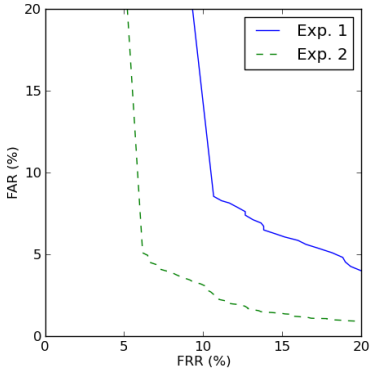

Fig. 3. Password and Numeric keypad ROC curves.

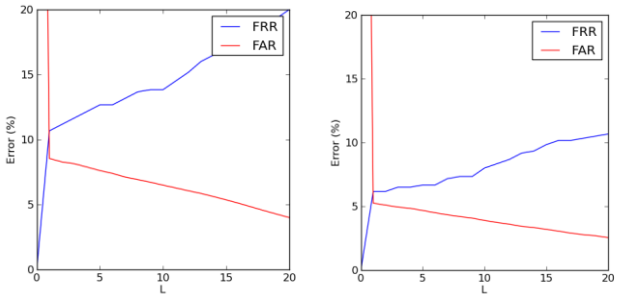

Fig. 4. FAR/FRR versus *L*: Exp. 1 (left) and Exp. 2 (right).

V. DISCUSSION

The performance of a behavioral biometric authentication system was evaluated on two different types of short input in which the sequence of keystrokes is fixed, password and numeric keypad sequences. In contrast to long text input applications, it is more feasible to obtain large numbers of enrollment samples for short input applications from the population of participants.

An advantage of the vector-difference model is that it operates efficiently with a small number of enrollment samples. Due to the large reference set in the first set of experiments, enrollment samples are reduced by k-means clustering, creating a smaller "ideal" reference set for each user. This model was then validated by a repeated-random subsampling procedure. With a smaller number of samples per subject in the second set of experiments, a leave-one-out validation method was used to obtain system performance.

Privacy is less of a concern for applications of interest which collect fixed-text short input, as opposed to those which require long-text arbitrary input. Many authorizing entities (ATM, security keypad, mobile phone) already have knowledge of the password or numeric sequence which must be entered. Others, such as a cash register, do not obtain any personal information from the user. Obtaining a metric on the keystroke dynamics of the fixed input does not require monitoring a session and possibly collecting personal information, as is the case in arbitrary input.

VI. CONCLUSION

The main contribution of this study was the development of an improved classification system and its performance evaluation. In the password study, on the same CMU input data, the *Pace Classifier* was compared against 14 other systems analyzed in a CMU study [15]. The three top performing systems in that study had EERs between 9.6% and 10.2%, while the EER achieved in this study was 8.7% on both the features from the previous study and on a new set of features.

In the numeric keypad study, the EERs achieved were 10.5% and 6.1%, which are comparable to those obtained in the password study described here and to the basic EER of 8.6% obtained in the CMU numeric keypad study [17].

REFERENCES

[1] AdmitOneSecurity. (Apr 2013). http://www.admitonesecurity.com/
[2] AuthenWare. (Apr 2013). http://www.authenware.com/
[3] S.S. Bender and H.J. Postley. "Key sequence rhythm recognition system and method." U.S. Patent 7,206,938, 2007.
[4] bioChec. (Apr 2013). http://www.biochec.com/
[5] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior, *Guide to biometrics*. New York: Springer, 2004.
[6] DeepnetSecurity. (Apr 2013). http://www.deepnetsecurity.com/
[7] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs, and wolves: A statistical analysis of speaker performance. *Proc. IC-SLD'98, NIST 1998 speaker recognition evaluation*, Sydney, Australia, 1998.
[8] J. Ferreira and H. Santos, "Keystroke dynamics for continuous access control enforcement," *Proc. Int.Conf. on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 216-223, 2012
[9] E.J. Fimbel, "Keylogger basic labbook", 2011. <http://sites.google.com/site/basiclabbook/keyloggerbasiclabbook>.
[10] R. Giot, M. El-Abed, and C. Rosenberger, "Keystroke dynamics with low constraints svm based passphrase enrollment," *IEEE Int. Conf. Biometrics (BTAS 2009)*, 2009.
[11] IDControl. (Apr 2013). http://www.idcontrol.com/
[12] L. Jin, X. Ke, R. Manual, and M. Wilkerson, "Keystroke dynamics: A software based biometric solution," *13th USENIX Sec. Sym.*, 2004.
[13] M. Karnan, M. Akila, and N. Krishnaraj, "Biometric personal authentication using keystroke dynamics: A review," *Applied Soft Computing J.*, vol. 11, 2011.
[14] KeyTrac. (Apr 2013). http://www.keytrac.de/
[15] K. Killourhy and R. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," *Int. Conf. Dependable Systems & Networks (DSN-09)*, Lisbon, 2009, pp. 125-134.
[16] Y. Li, B. Zhang, Y. Cao, S. Zhao, Y. Gao, and J. Liu, "Study on the BeiHang keystroke dynamics database," *Int. Joint Conf. Biometrics (IJCB 2011)*, Washington, D.C., 2011.
[17] R.A. Maxion, and K.S. Killourhy, "Keystroke biometrics with number-pad input." *Proc. IEEE/IFIP Int. Conf. Dependable Sys. & Netw. (DSN-10)*, pp. 201-210, 2010. IEEE Comp. Soc. Press, 2010.
[18] A. Messerman, T. Mustafic, S. Camtepe, and S. Albayrak, "Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics," *Proc. Int. Joint Conf. Biometrics (IJCB 2011)*, Wash. D.C., 2011.
[19] J.V. Monaco, N. Bakelman, S. Cha, and C.C. Tappert, "Developing a keystroke biometric system for continual authentication of computer users," *Proc. Euro. Intell. Sec. Inform. Conf.*, Denmark, 2012, pp. 210-216.
[20] F. Montrose, M.K. Reiter, and S. Wetzel. "Password hardening based on keystroke dynamics." *Int. J. Info. Security*, 1(2): 69-83, 2002.
[21] A. Peacock, X. Ke, and M. Wilkerson, "Typing patterns: A key to user identification," *IEEE Security & Privacy*, vol. 2, 2004, pp. 40-47.
[22] K. Revett, "Chapter 4: Keystroke dynamics," in *Behavioral biometrics: A remote access approach*: Wiley, 2008, pp. 73-136.
[23] R. N. Rodrigues, G.F.G. Yared, C.R. Costa, J.B.T. Yabu-Uti, F. Violaro, and L.L. Ling. "Biometric access control through numerical keyboards based on keystroke dynamics*." Lecture Notes Comp.Sci.*, 3832: 640-646, 2006.
[24] T. Shimshon, R. Moskovitch, L. Rokach, and Y. Elo,vici. "Continuous verification using keystroke dynamics." *Proc. Int. Conf. Comp. Intel. and Security* . IEEE Computer Soc., Washington, DC, 411-415, 2010.
[25] C.C. Tappert, S. Cha, M. Villani, & R. Zack, "A keystroke biometric system for long-text input," *Int. J. Info. Sec. Privacy*, 2010, pp. 32-60.