2012 2nd International Conference on Information Communication and Management (ICICM 2012)
IPCSIT vol. 55 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V55.7

Personalized Rhythm Click Based Authentication System Improvement using a Statistical Classifier

Ting-Yi Chang[1 +], Chun-Cheng Peng[2], Cheng-Jung Tsai[3], Yen-Lin Chen[4],Pei-Cheng Cheng[5]

[1]Department of Industrial Education and Technology, National Changhua University of Education
tychang@cc.ncue.edu.tw
[2]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology
[3]Department of Mathematics, Graduate Institute of Statistics and Information Science, National Changhua University of Education
[4]Department of Computer Science and Information Engineering, National Taipei University of Technology
[5]Department of Information Management, Chien Hsin University

Abstract. Chang et al. recently proposed a personalized rhythm click-based authentication system implemented using a neural network classifier. Unfortunately, the neural network classifier requires impostor patterns as training samples to train the network. It is impractical to collect impostor patterns in the real world. This paper presents a statistical classifier that solves these problems. The proposed system does not require impostor patterns to bulid the classifier and the computation is efficient. With the same benchmark dataset, FAR=2.46% and FRR=29.2%, in Chang et al.'s system is reduced to FAR=0.00% and FRR=0.06% in the proposed system.

Keywords: user authentication, biometric, keystroke feature,keystroke dynamics authentication.

1. Introduction

Biometric keystroke dynamics methods have been used recently to enhance identification security[4][5][6][14], which do not require extra special tools. A user types his/her password on a QWERTY keyboard at the enrollment phase, the keystroke authentication system records the corresponding keystroke features and builds a classifier to ascertain personal identity. It is worth mentioning that the keystroke dynamic authentication system has great convenience by its user-friendly features since users do not need any additional equipment or process. In the *Keystroke Dynamics Authentication* (KDA) system, both the password correctness and its keystroke features are verified. Keystroke dynamic authentication can be easily applied into password mechanisms via its unobtrusive software.Many KDA researches based on the QWERTY keyboard have been done in the last several years[1][8][10]. With the popularity of smart mobile devices, these researches originally designed for the personal computer are no longer suitable for mobile devices. Different input devices produceinconsistent typing results against QWERTY keyboards. The inconsistencies between devices such as the location and size of the keyboardkeys change the keystroke results and decrease the authentication accuracy. In addition, KDA accuracy depends on using a classifier able to verify the identity of unknown users based only on keystroke samples extracted from the user during their login sessions.Researchers therefore proposed various algorithms to designclassifiersin last decades.(e.g. statistical method method[2[13][14], neural network[2][11], degree of disorder[6], fuzzy logic[7], etc.)

A convincible criterion is required to compare the utility between the different proposed methods. All participants, especially those who act as an impostor in the experiment, knew the password content during the assessment. In this context, *False Rejection Rate* (FRR) and*False Acceptance Rate* (FAR) are used to evaluate the system accuracy and error rate. However, the FAR and FRR values are evaluated separately using only the number of impostors and legitimate users. A system is evaluated as efficient if the values of these criteria incline towards zero, making FAR inversely proportional to FRR. These two rates vary according to the sensitivity level of the algorithm applied to the system. Therefore, a threshold value is often used as a system security tradeoff. The data quality is also an important factor in adopting keystroke dynamics. Because human behavior is a complex andunpredictable system of attributes, Hwang et al. [10] proposed three primary measures of data qualityfor consideration before adopting keystroke dynamics. Their research applied artificial rhythms and cues to assist user typing to improve discriminability. They used metronome audio to make artificial rhythms to improve the uniqueness, as well as asking the user to pause during their inputs to improve the consistency. They also proposed five typing strategies to improve the data quality[10].The sample made by the five typing strategies produced higher quality than the user's natural rhythm which did not apply any pauses or cues. According to Hwang's experiments, using pauses with cues produced the best performance. These typing strategies are more effective in those whose typing abilities are poor.

Although these typing strategies improved the data quality with so many benefits, here are several problems that should be considered. First, typing strategies are not a user-friendly design. For example, users need to remember not only the password but also the locations and lengths of the pauses. Secondly, musical rhythm and slow tempo are implemented by a metronome which should be prepared before adopting. Although today's hardware devices are sufficient to offer software with a similar function as an alternative, it should still be considered that some places do not allow voices. Finally, different input devices cause typing pattern inconsistencies. For instance, producing a typing pattern on a conventional keyboard or on handheld mobile device will produce different results by the same person. These different results are affected by the size of the keys on keyboard.Motivated to correct these weakness, Chang et al. [2] proposed a personalized rhythm click-based authentication system. Unlike Hwang et al.'s method, Chang et al. purposed the click feature method to solve the inconsistency issue between various devices. This click feature could be effective on a conventional keyboard, mouse or touch screen of a handheld mobile device such as a personal digital assistant or smart phone. This method was implemented by asking every participant to click the same private rhythm named "Encourage with Love". The click features are extracted from the user as his/her feature sample.Although this personalized rhythm click method can improve the consistency between samples and solve device inconsistency, Chang et al.'s method has two main drawbacks. First, it is impractical to collect impostor patterns in the real world. Secondly,their system requires complex computation for training the neural network.

This paper proposes a statistical method to build a classifier that requires only legitimate user training samples to construct the classifier. The computing complexity of the proposed system is lower than [2][11][12] by adopting the personalized statistical template method established in the training phase. This method requires data only from the legitimate user to build the classifier. When a user enrolls the system only his/her statistical template will be constructed without the need for data from other users' in the system. Finally, in the experiment results, FAR=2.46% and FRR=29.2% in Chang et al.'s system are reduced to FAR=0.00% and FRR=0.06% in the proposed system.

2. Methodology

Enrolment phase: In the enrolment phase, a user registers his/her personalized private rhythm. For evaluating the system utility, all users click the same target rhythm. The system extracts the click rhythm features to form feature samples. There are four combinations of up and down events for click rhythm features comprised as follows:

- Down-up (DU): the time interval between the pressure and release of the same click event on a mouse button.
- Down-down (DD): the time interval between two successive pressure events.
- Down-up (UD): the time interval between the mouse button releases to the next mouse button being pressed.

DU, DD and UD combination is used to form the feature samples. Let $\mathrm{x}_{\mathrm{m,i}}$ represent the feature value of *i*th click of the *m*th sample made by user *x*. For instance, "Encourage with Love" includes 11 clicks. Values of 11 *DU*, 10 *DD* and 10 *UD* are involved in the sample of user *x*. The click rhythm features adopted in this paper are given by$DU_{x_{m,i}}$, for $m = 1$ to $n, i = 1$ to $k$and $DD_{x_{m,i}}, UD_{x_{m,i}}$, for $m = 1$ to $n, i = 1$ to $k - 1$, where k is the number of clicks and *n* denotes the number of training samples. These click features formed a feature sample for each user entry, including 3*k*–2 =3×11–2=31 feature values.

Training phase: After collecting the data in the enrolment phase, these feature samples are applied to train a classifier. A novel statistical classifier method is proposed to deal with impostor samples in the real world and decrease the number of computations by adopting a personalized statistical template. With this method, only the feature samples of a legitimate user are collected to build his/her template. The template is used to compute a score with samples generated from each login attempt. A most applicable threshold value α applied to our system is measured via Receiver Operating Characteristic curve [9]. The value α is a benchmark for judging the user identity. Furthermore, the personalized statistical template is built according to the legitimate users' samples by computing the following values:

- **Step 1:** Compute the maximum value of *DU*, *DD* and *UD* for the *m* samples of user ***x*** from Eq. (1), minimum value from Eq. (2) and mean value from Eq. (3).
- **Step 2:** Compute the standard deviation of *DU*, *DD* and *UD* from Eq. (4).

$$\begin{cases} DU_{max_i} = \max\limits_{m=1 \text{ to } n} \{DU_{x_{m,i}}\}, \text{for } i = 1 \text{ to } k \\ DD_{max_i} = \max\limits_{m=1 \text{ to } n} \{DD_{x_{m,i}}\}, \text{for } i = 1 \text{ to } k-1 \\ UD_{max_i} = \max\limits_{m=1 \text{ to } n} \{UD_{x_{m,i}}\}, \text{for } i = 1 \text{ to } k-1 \end{cases} \quad (1)$$

$$\begin{cases} DU_{min_i} = \min\limits_{m=1 \text{ to } n} \{DU_{x_{m,i}}\}, \text{for } i = 1 \text{ to } k \\ DD_{min_i} = \min\limits_{m=1 \text{ to } n} \{DD_{x_{m,i}}\}, \text{for } i = 1 \text{ to } k-1 \\ UD_{min_i} = \min\limits_{m=1 \text{ to } n} \{UD_{x_{m,i}}\}, \text{for } i = 1 \text{ to } k-1 \end{cases} \quad (2)$$

$$\begin{cases} DU_{avg_i} = \frac{1}{m}\sum\limits_{m=1}^{n} DU_{x_{m,i}}, \text{for } i = 1 \text{ to } k \\ DD_{avg_i} = \frac{1}{m}\sum\limits_{m=1}^{n} DD_{x_{m,i}}, \text{for } i = 1 \text{ to } k-1 \\ UD_{avg_i} = \frac{1}{m}\sum\limits_{m=1}^{n} UD_{x_{m,i}}, \text{for } i = 1 \text{ to } k-1 \end{cases} \quad (3)$$

$$\begin{cases} DU_{sd_i} = \sqrt{\frac{1}{n}\sum\limits_{m=1}^{n} \left(DU_{x_{m,i}} - DU_{avg_i}\right)}, \text{for } i = 1 \text{ to } k \\ DD_{sd_i} = \sqrt{\frac{1}{n}\sum\limits_{m=1}^{n} \left(DD_{x_{m,i}} - DD_{avg_i}\right)}, \text{for } i = 1 \text{ to } k-1 \\ UD_{sd_i} = \sqrt{\frac{1}{n}\sum\limits_{m=1}^{n} \left(UD_{x_{m,i}} - UD_{avg_i}\right)}, \text{for } i = 1 \text{ to } k-1 \end{cases} \quad (4)$$

- **Step 3:** Generate the personalized statistical template composed of function f_{DU}, f_{DD}, andf_{UD} using the results from Steps 1 and 2. $f_{DU}$is generated by *i*th*DU* of every samples of user ***x***, as shown in Eq. (5). Likewise, based on the results from Steps 1 and 2,$f_{DD}$and$f_{UD}$are formed (DD and UD replace DU in Eq. (6) respectively) asEqs. (6) and (7).

$$f_{DU}\left(DU_{x_i}\right) = \begin{cases} 0 & , \text{for } DU_{x_i} \leq DU_{min_i} + 3 \times DU_{sd_i} \\ \dfrac{DU_{x_i} - \left(DU_{min_i} + 3 \times DU_{sd_i}\right)}{\left(DU_{avg_i} - 1 \times DU_{sd_i}\right) - \left(DU_{min_i} + 3 \times DU_{sd_i}\right)} & , \text{for } DU_{min_i} + 3 \times DU_{sd_i} \leq DU_{x_i} \leq DU_{avg_i} - 1 \times DU_{sd_i} \\ 1 & , \text{for } DU_{avg_i} - 1 \times DU_{sd_i} \leq DU_{x_i} \leq DU_{avg_i} + 1 \times DU_{sd_i} \\ \dfrac{\left(DU_{max_i} + 3 \times DU_{sd_i}\right) - DU_{x_i}}{\left(DU_{max_i} + 3 \times DU_{sd_i}\right) - \left(DU_{avg_i} + 1 \times DU_{sd_i}\right)} & , \text{for } DU_{avg_i} + 1 \times DU_{sd_i} \leq DU_{x_i} \leq DU_{max_i} + 3 \times DU_{sd_i} \\ 0 & , \text{for } DU_{max_i} + 3 \times DU_{sd_i} \leq DU_{x_i} \end{cases} \quad (5)$$

As a result, if the private rhythm includes *k* clicks, the system will generate 3*k*-2 functions to form a personal statistical template for each user. The system would store all users' personalized statistical templates using Step 1 to Step 3.

Authentication phase:Once an unknown user attempts to login as user *x*, the system would extract his/her click rhythm features from the sample that he/she provided. The system then verifies the clicked target rhythm by examining the number of features. As described in the enrolment phase, a legal click rhythm should have 11 *DU*, 10 *DD* and 10 *UD* involved. If the unknown user failed in this verification, the system would offer a second chance. Each user has two opportunities to provide his/her click rhythm for target rhythm verification. After the click rhythm verification, the system would continuously authenticate the unknown user by verifying the extracted click features through his/her personalized statistical template which acquired on training phase.

Let $DU_{x_i}^* = \{DU_{x_1}^*, DU_{x_2}^*, \ldots, DU_{x_k}^*\}$, $DD_{x_i}^* = \{DD_{x_1}^*, DD_{x_2}^*, \ldots, DD_{x_{k-1}}^*\}$ and $UD_{x_i}^* = \{UD_{x_1}^*, UD_{x_2}^*, \ldots, UD_{x_{k-1}}^*\}$ beth e click features extracted from the unknown user sample. These click features would be substituted into Eq. (8) to obtain a score as the benchmark for verifying the user.Finally we include a constant *α* for the score threshold using ROC curve. If the score = (DU_{sum} +DD_{sum} + UD_{sum}) / 3 is smaller than the value α, the system will offer a second chance for the unknown user to reenter again. Once the user fails on his/her second attempt, the system denies access from that user and labels him/her an impostor. Otherwise, the unknown user is judged a legitimate user.

$$DU_{sum} = \sum_{i=1}^{k} f_{DU}\left(DU^*_{x_i}\right), \text{for } i = 1 \text{ to } k, DD_{sum} = \sum_{i=1}^{k-1} f_{DD}\left(DD^*_{x_i}\right), \text{for } i = 1 \text{ to } k-1, UD_{sum} = \sum_{i=1}^{k-1} f_{UD}\left(UD^*_{x_i}\right), \text{for } i = 1 \text{ to } k-1 \quad (8$$

3. Experiment results

Comparing our method to others, the same dataset is used to calculate FAR and FRR. We give special thanks to Chang et al. for providing the raw dataset from their experiment. There were 25 participants, with each clicking the private rhythm 30 times to generate 30 samples, producing 25×30 samples overall. For every sample set from each user, 10 of 30 samples were randomly captured as the training sample to form the personalized statistical template. The last 20 samples (25×20=500 legitimate samples) were used to examine the system to obtain FRR. Each user was attacked by all others. These test samples were applied to calculate FAR, in addition to the samples against user's own (24×20=480 impostor samples). The result shows that FRR and FAR achieved 0.06% and 0.00%. Table 1 shows a comparison between this paper and Chang et al.'s [4] results based on the same dataset.

In Table 1, with the same dataset employed, our results are better than Chang et al.'s for FRR or FAR.

Table 1.A comparison sheet of classifier performance with the same dataset

Classifier	Statistic(this paper)	Neuralnetwork [2]
FAR (%)	0.00	0.06
FRR (%)	2.46	29.2
Numbers of training sample	250	490
System Overall threshold (α)	0.6	0.03

4. Conclusion

The proposed approach requires only legitimate samples into account, calculating the maximum, minimum, mean and standard deviation of registered samples to build the classifier. Finally, FAR and FRR are decreased to 0.00% and 0.06%, respectively. The proposed statistical algorithm has lower computational load which can enable this rhythm click-based authentication system to work on low computational handheld devices. to the proposed method improves the synergistic reinforcement of password mechanisms with rhythm click-based authenticaiton on various devices perfectly.

5. Acknowledgment

This research was partially supported by the National Science Council, Taiwan, ROC, under contract nos.: NSC100-2221-E-018-025 and NSC100-2221-E-018-034.

6. References

[1] Campisi P., Maiorana E., Bosco M. L. and Neri A., "User authentication using keystroke dynamics for cellular phones," *IET Signal Processing*, Vol. 3, No. 4, pp. 333-341, 2009.

[2] Liu C. L., Chang T. Y., Chiang M. L. and Tsai C. J., "A Simple Keystroke Dynamics-Based Authentication System using Means and Standard Deviations," *Journal of Internet Technology*, vol. 13, no. 3, pp. 439-444, 2012.

[3] Chang T. Y., Yang Y. J. and Peng C. C., "A personalized rhythm click-based authentication system," *Information Management & Computer Security*, Vol. 18, No. 2, pp. 72-85, 2010.

[4] Chang T. Y., Tsai C. J. and Lin J. H., "A Graphical-based Password Keystroke Dynamic Authentication System for touch Screen Handheld Mobile Devices," *Journal of Systems and Software*, Vol. 85, No. 5, pp. 1157-1167, 2012.

[5] Cheng P. C., Chang T. Y., Tsai C. J., Li J. W. and Wu C. S., "A Novel and Simple Statistical Fusion Method for User Authentication through Keystroke Features," *Convergence Information Technology*, Vol. 6, No. 2, pp. 347-356, 2011.

[6] Choras M. and Mroczkowski P., "Keystroke Dynamics for Biometrics Identification," *Adaptive and Natural Computing Algorithm*, Vol. 4432, pp. 424-431, 2007.

[7] De Ru W. G. and Eloff J. H. P., "Enhanced password authentication through fuzzy logic," *IEEE Expert: Intelligent Systems and their Applications*, Vol. 12, No. 6, pp. 38-45, 1997.

[8] Enzhe Y. and Sungzoon C., "Keystroke dynamics identity verification--its problems and practical solutions," *Computers & Security*, Vol. 23, No. 5, pp. 428-440, 2004.

[9] Fawcett T., "An introduction to ROC analysis," *Pattern Recognition Letters*, Vol. 27, No. 8, pp. 861-874, 2006.

[10] Hwang S. S., Lee H. J. and Cho S., "Improving authentication accuracy using artificial rhythms and cues for keystroke dynamics-based authentication," *Expert Systems with Applications*, Vol. 36, No. 7, pp. 10649-10656, 2009.

[11] Ku W. C., "Weaknesses and drawbacks of a password authentication scheme using neural networks for multiserver architecture," *IEEE Transactions on Neural Networks*, Vol. 16, No. 4, pp. 1002-1005, 2005.

[12] Lee H. J. and Cho S., "Retraining a keystroke dynamics-based authenticator with impostor patterns," *Computers & Security*, Vol. 26, No. 4, pp. 300-310, 2007.

[13] Teh P. S., Teoh A. B. J., Ong T. S. and Neo H. F., "Statistical fusion approach on keystroke dynamics," *IEEE Conference on Signal-image Technologies and Internet-based System*, pp. 918-923, 2008.

[14] Zhang Y., Chang G., Liu L. and Jia J., "Authenticating User's Keystroke Based on Statistical Models," *Genetic and Evolutionary Computing*, pp. 578-581, 2010.

Ting-Yi Chang received his MS from the Graduate Institute of Computer Science and Information Engineering at Chaoyang University of Technology, and his PhD in the Department of Computer Science at National Chiao Tung University (2003-2006). Currently, he is an associate professor with the Department of Industrial Education and Technology, National Changhua University, Taiwan. His current research interests include artificial intelligence, e-Learning, information security, cryptography, and mobile communications.

Chun-Cheng Peng has a Database Development from the Computer Science and Information Management Department of the Providence University, Taiwan, a MSc (DNA Sequence Prediction) from the Computer Science and Information Engineering Department of the Chaoyang University of Technology, Taiwan, and a PhD (Neural Networks Learning) from the Computer Science and Information Systems Department of the Birkbeck College, University of London, England. His research interests are mainly in nonmonotone learning, unconstrained optimization, recurrent neural networks and artificial intelligence applications.

Cheng-Jung Tsai received the BS degree in mathematics and science education from National Ping Tung University of Education in 1995, the MS degree in information education from National University of Tainan in 2000, and the PhD degree in computer science and information engineering from National Chiao Tung University in 2008. Currently, he is an Assistant Professor in the Graduate Institute of Statistics and Information Sciencethe, National Changhua University, Taiwan. His current research interests include data mining, information security, and e-learning.

Yen-Lin Chen received the B.S. and Ph.D. degrees in electrical and control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2000 and 2006, respectively. He is now an Associate Professor at the Dept. of Computer Science and Information Engineering, National Taipei University of Technology, Taipei, Taiwan. His research interests include image and video processing, pattern recognition, embedded systems, document image analysis, and intelligent transportation systems.

Pei Cheng Cheng received the Ph.D. degree in Computer Science and Engineering from NationalChiao Tung University (NCTU), HsinChu, Taiwan in 2006. Now he is an assistant professorof the Department of Information Management, ChienHsin University, Taoyuan, Taiwan. Hiscurrent research interests include content-based image retrieval, machine learning, knowledgemanagement, and artificial intelligence.