Statistical Median-Based Classifier Model for Keystroke Dynamics on Mobile Devices

Noor Mahmood Al-Obaidi
Dept. of Computer Science
Middle East University
Amman, Jordan
e-mail: noor_mah89@yahoo.com

Mudhafar M. Al-Jarrah
Dept. of Computer Science
Middle East University
Amman, Jordan
e-mail: maljarrah@yahoo.com

Abstract**—**User authentication on mobile devices, smartphones, and tablets have received considerable interest lately due to the increasing dependence on these devices for storing sensitive data. Keystroke dynamics is a biometrics method of authentication that has the advantage of being hardware independent. However, enhancements are needed to make it a viable security solution on mobile devices. This paper presents a statistical median-based classifier model to serve as an anomaly detector in keystroke dynamics authentication on mobile devices. The proposed classifier uses the timing features of keystroke dynamics as well as touch features of mobile devices. Formulation and evaluation of proposed classifier have been influenced by an empirical analysis of a public dataset of keystroke dynamics on Android platform. The classifier considers the distance from the median of a feature element as an indicator of whether it relates to a genuine user or an impostor, based on previous training data of genuine users. The formula for measuring the acceptable distance to the median is derived from the distance between the minimum value of a feature element and its median. The model was evaluated through comparison with previous detectors, using the same dataset, and the results have shown a significant reduction in the equal-error-rate.

Keywords—biometrics; keystroke dynamics; equal-error-rate; false-acceptance-rate; false-rejection-rate; distance-to-median; classifier; anomaly detector; pass-mark.

I. INTRODUCTION

The recent trend in using a mobile device (smartphone or tablet), as a personal computer and a data store, is putting at risk valuable personal and business data, in the case of hacking of the mobile device or if it is physically lost or stolen. Awareness of the sensitivity of data on mobile devices has drawn attention to the need for the development multi-level protection schemes for mobile devices, not only passwords, to prevent unauthorized access. One approach is to add fingerprint, iris and face detection sensors. Apart from the additional cost, the usefulness of these sensors is limited to static authentication phase, at the log-in phase. An alternative authentication approach relies on using behavioral biometrics [1], for either the static authentication or in the continuous authentication process. The keystroke dynamics is a behavioral authentication approach that has been investigated on desktop computers [2, 3, 4], and more recently on touch mobile devices that have additional touch and sensor features to enhance the authentication process [5 , 6, 7, 8, 9].

II. RELATED WORK

The work of Antal, et.al. [5, 6] at Sapientia University (SU) has provided a public benchmark dataset for keystroke dynamics on Android mobile devices which included touch screen features of pressure and size of finger area. The dataset collection process followed the line of research of Kilrouhy [2, 3] at Carnegie Mellon University (CMU), in particular by using the same password (".tie5Ronal"), which has become a standard password for comparative studies on keystroke dynamics. Also, the SU study used the new dataset in evaluating the Equal-Error-Rate (EER) of three verification models that were suggested in the CMU study.

Al-Ghamdi and Elrefaei [9] presented a keystroke dynamics system implemented on an Android phone which measured pressure and finger area, as well as the timing features, and the authentication is based on the statistical anomaly detector in [10]. The experimental work used a dynamic text chosen by each subject instead of a fixed password, an approach that can be extended to continuous authentication. The dynamic text was

ISBN: 978-1-4673-7504-7 ©2016 IEEE

decided to be of 11 characters so that the results can be compared with CMU and related experiments that used a 10-character password and the additional Enter key. The reported results included False-Acceptance-Rate (FAR), False-Rejection-Rate (FRR) and the EER metric.

Recent studies have also looked at new sensors data that are becoming available in mobile devices, such as accelerometer and gyroscope. In [8], results are presented on the effect of using sensor dynamics data to complement timing data of the keystroke dynamics, and the published results have shown lower EER rates and, therefore, better authentication.

The work in this paper follows the model of using the median of an authentication feature as the point of center, along the lines of the works in [10, 11, 12, 13], instead of the mean which can be influenced by outliers.

III. THE SAPIENTIA UNIVERSITY DATASET

The SU public dataset [5] is used in the present research, which has the following advantages:
1. It is available for download from the university's website.
2. The dataset is consistent and has been analyzed and verified using previously tested models.
3. The password that was used in the experiment is the CMU password (".tie5Roanl"), which has become a standard password for comparison in keystroke dynamics research.
4. The data was collected on mobile devices.

The SU dataset contains timing features (Hold, UD, DD) and additional features of touch mobile devices that are the pressure and size of the finger area when a key is pressed. In the SU experiment, the password consisted of 10 characters plus the Enter key, which resulted in 41 features for timing data only, and 71 features for timing, pressure, and finger area, as shown in tables1 and 2. The dataset contains typing records of 42 subjects; each subject has entered the same password 51 times (34 entries in the training session and 17 in the testing session). The dataset is divided into two sub-datasets, timing only sub-dataset and timing with pressure and finger area sub-dataset. The SU dataset was collected on Android devices, a tablet, and a mobile phone.

Table 1. SU dataset timing features

Feature name	Explanation	Number of features
Key Hold time (H)	Time between key press and release	14
Down-Down time (DD)	Time between consecutive key presses	13
Up-Down time (UD)	The time between key release and next key press	13
Average hold time (AH)	Average of key hold times	1
Total		**41**

Table 2. SU dataset timing features + touch screen features

Feature name	Explanation	Number of features
Key Hold time (H)	Time between key press and release	14
Down-Down time (DD)	Time between consecutive key presses	13
Up-Down time (UD)	The time between key release and next key press	13
Key hold Pressure (P)	Pressure at the moment of key press	14
Finger Area (FA)	Finger area at the moment of key press	14
Average hold time (AH)	Average of key hold times	1
Average Finger Area(AFA)	Average of key finger areas	1
Average Pressure (AP)	Average of key pressures	1
Total		**71**

The SU study [5] analyzed the dataset in terms of EER using the three verification models that were suggested in [3]. The analysis results are shown in table 3 for the 41 timing features and the 71 timing and touch features.

Table 3. EER analysis results of the SU dataset using three verification models

Detector	**H+DD+UD+AH** **(41 Features)**	**H+DD+UD+P+FA+AH+AP+AFA** **(71 Features)**
Euclidean	17.5%	15.7%
Manhattan	15.3%	12.9%
Mahalanobis	23.3%	16.6%

IV. THE PROPOSED AUTHNTICATION MODEL

The proposed keystroke dynamics authentication model is a statistical binary classifier that uses the median of a training set of feature values as the point of center against which testing (Authentication) feature values are evaluated. A testing value is considered acceptable (genuine) if it falls within upper and lower thresholds surrounding the median, obtained during the training phase.

A. Feature Set

The proposed model employs the same feature set as in the SU dataset, which consists of the three timing features that were suggested in the CMU study [3], as well as pressure and finger area touch features, as follows:

Hold (H): The elapsed time during key press, which is the difference between key-down and key-up timestamps, also referred to as the dwell time.

Up-Down (UD): The latency time between key-up timestamp of the first key in a typing sequence and key-down of the second key.

Down-Down (DD): The elapsed time between key-down of the first key and key-down of the second key, it is a composite feature of Hold of the first key and UD between the first and second keys.

.

Pressure (P): The pressure applied to the touch key area during a key-press.

Finger Area (FA): The occupied finger area during key- press on a touch key area.

In addition to the measurable features, the average of Hold, Pressure and Finger Area were included in the SU dataset, therefore we included them in our analysis.

B. The Med-Min-Diff classifier.

The proposed classifier is trained during the training session on the typing rhythm of a user by applying the following steps:

1. Collect keystrokes data of the feature set elements for some pre-determined repetitions.

2. Generate the training template which consists of two vectors representing the upper and lower thresholds of each feature set element. The two thresholds are calculated as follows:

Lower Threshold = Minimum value of a feature element obtained from training repetitions. Upper Threshold = Median of the feature element + Upper Distance to Median (UDM)

UDM = (Median – Minimum) x d, where d is a constant factor greater than 1, whose purpose is to allow the area between the Median and Upper Threshold to be wider than the area below.

C. Test-Score Calculation

During the testing phase (positive or negative tests) in which a testing vector of the feature set elements is collected from the typing attempt of the password, the following steps are used to classify the typing attempt as genuine or an impostor.

1. Compare the testing vector of feature elements with the corresponding elements of the upper and lower thresholds, and give a score of 1 if the feature element is within the thresholds, otherwise zero.

2. Sum-up the feature elements scores, where the sum is the test-score.

3. Compare the test-score with the pass-park; if the test-score is higher than or equal to the pass-mark, the typing attempt is classified as genuine, otherwise as an impostor.

D. Error Metrics calculation

The following error metrics are used as indicators of the authentication efficacy of a model:

- False Acceptance Rate (FAR): The ratio of false acceptance of the typing attempts of impostors.
- False Rejection Rate (FRR): The ratio of false rejection of the typing attempts of genuine users.
- Equal Error Rate (ERR): the value at which the FAR and FRR are equal.

The three error rates are extracted from the analysis of a set training and testing data; their purpose is to be as comparison metrics of various authentication models. The EER, FAR, and FRR metrics are generated from a training/testing collection of data as below:

1. Based on a set of typing attempts data from the training phase for a group of users, a template of upper and lower thresholds are generated for each user.

2. During the positive testing phase, a second set of typing attempts of each user is collected, and the testing vectors generated from the typing attempts are classified as genuine or impostor through comparison with users' training thresholds.

V. ANALYSIS OF THE SU DATASET USING THE PROPOSED CLASSIER

The proposed Median-Min-Diff classifier has been used in the EER analysis of the SU dataset, according to procedure below:

1. The 51 password typing records of each subject are divided into 34 training and 17 testing records.

2. An impostor set is created which contains the first five records of each subject, a total of 210 impostor records.

3. The 34 training records of each subject are used to generate a template of upper and lower thresholds for the subject, using the proposed model.

4. The 17 testing records of each subject are evaluated according to the subject's template and a score is given to each feature element.

5. A test-score is calculated for each record (typing attempt), which is the sum of individual feature element scores. The test result is determined as either genuine or impostor, depending on the pass-park.

6. The impostor set records are evaluated against each subject's template, and a test-score and the result of authentication are calculated for each impostor record, as in steps 4 and 5 (the actual number of impostor records against each subject is 205, which excludes the five records of the same subject).

7. For each subject, a pass-mark is tuned to get to the point of equal or near equal FRR and FAR, and the EER value is taken to be the average of FAR and FRR.

8. The overall EER for the population is calculated as the average of subjects' EER.

The EER analysis results of the proposed model, using the SU dataset, are shown in table 4.

Table 4. EER analysis of the SU dataset using the Median-Min-Diff classifier

H+DD+UD+AH (41 Features)	H+DD+UD+P+FA+AH+AP+AFA (71 Features)
8.53%	6.79%

The obtained results show a significant difference in EER between the 41 and 71 features datasets. Also, the proposed model has out-performed the three verification models in the EER results that are shown in table 3.

VI. COEFFICIENT OF VARIATION ANALYSIS OF THE SU DATASET

To study the behavior of individual features of the dataset, we performed the coefficient of variation analysis of each feature. The coefficient of variation is the ratio of standard deviation to the average, and it measures the degree of dispersion of a set of values. Table 5 shows the average coefficient of variation of the five features of the SU dataset. As each feature has several feature elements depending on the number of characters in the password, hence we have calculated the average of the coefficient of variation per feature.

Table 5. Coefficient of variation analysis of the SU dataset

Feature	Average of the Coefficient of variation
Hold	0.32
DD	1.33
UD	1.64
Pressure	0.42
Finger Area	0.39

It can be seen that the latency features of UD and DD have more variation than the Hold feature, which is due to the fact that transition time between keys is more variable than the key-press time. Also, the pressure feature has more dispersion than the finger area. This observation suggests that the more dispersed features can have higher effect in distinguishing between different typists.

VII. CONCLUSIONS

The inclusion of touch features of mobile devices in user authentication has resulted in lower EER, as reported in the SU study, and confirmed in the present work using the Med-Min-Diff classifier. This suggests that other mobile devices' related features can further reduce the error rate, leading to more accurate authentication. The differences in the reported coefficient of variation among features can be a useful guide in determining which features to be given more weight in the authentication process.

Also, the EER analysis results of the proposed classier have shown a significantly lower EER value compared with the three verification models, which should lead to further investigation, to enhance the median-based model as a classifier in keystroke dynamics authentication.

REFERENCES

[1] El-Abed M., Charrier C., and Rosenberger C., "Evaluation of Biometric Systems", New Trends and Developments in Biometrics, InTech, ISBN: 978-953-51-0859-7, DOI: 10.5772/52084, 2012.

[2] K. S. Killourhy, "A scientific understanding of keystroke dynamics", Carnegie Mellon University, PhD Thesis, No. CMU-CS-12-100, Department of Computer Science, 2012.

[3] K. Killourhy, and R. Maxion, "Comparing anomaly detectors for keystroke dynamics", Proceedings of the 39th Annual International Conference on Dependable Systems and Networks (DSN2009), June29-July2, pp.125–134, IEEE Computer Society Press, Los Alamitos, 2009.

[4] P. S. Teh, A. B. J. Teoh, & S. Yue, "A survey of keystroke dynamics biometrics", The Scientific World Journal, 2013.

[5] M. Antal, L. Z. Szabó, & I. László, "Keystroke dynamics on android platform" Procedia Technology, Vol. 19, pp.820-826. 2015.

[6] M. Antal, & L. Z. Szabó, "An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices", 2015.

[7] G. Kambourakis, D. Damopoulos, D. Papamartzivanos, & E. Pavlidakis, "Introducing touchstroke: keystroke-based authentication system for smartphones" Security and Communication Networks, 2014.

[8] V. Stanciu, R. Spolaor, and M. Conti, "On the Effectiveness of Sensor-enhanced Keystroke Dynamics Against Statistical Attacks", CODASPY '16 Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy, Pages 105-112, 2016.

[9] S. J. Alghamdi, & L. A. Elrefaei, "Dynamic User Verification Using Touch Keystroke Based on Medians Vector Proximity" IEEE, 7th International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), 2015.

[10] M. M. Al-Jarrah, "An anomaly detector for keystroke dynamics based on medians vector proximity", Journal of Emerging Trends in

Computing and Information Sciences, Vol. 3, No. 6, pp. 988-993, 2012.

[11] M. M. Al-Jarrah, “Multi-factor authentication scheme using keystroke dynamics and two-part passwords”, International Journal of Academic Research, Vo. 5, No. 3, 2013.

[12] A. O. Al-Rahmani, “An Enhanced Classifier for Authentication in Keystroke Dynamics Using Experimental Data”, Master dissertation, Middle East University, Jordan, 2014.

[13] S. A. Al-Robayei, “A Multi-Model Keystroke Dynamics Anomaly Detector for User Authentication”, Master dissertation, Middle East University, Jordan, 2016.