A Novel Mobilephone Application Authentication Approach based on Accelerometer and Gyroscope Data

Guoqiang Li[1], Patrick Bours[2]

Abstract: The advent of mobile phones have changed our daily life. We are heavily relying on various applications installed on mobilephones to communicate with other, to share personal information with them, and to access our bank account, etc. However, the security measurement in terms of accessing these applications is either omitted or user-hostile because of the burden of memorizing the PIN or password. In order to relieve people from such burden, we explore the possibility of developing a mobilephone application authentication approach by analyzing the accelerometer and gyroscope data collected from the first few seconds when the user opens an application. By evaluating several proposed authentication approaches on a dataset collected from a real-life scenario, we achieve the best EER at 22.72% by only using the data collected from first 3 seconds. We think integrating the proposed non-intrusive authentication approach into the mobilephone application as an alternative for PIN/password can provide a more user-friendly authentication mechanism.

Keywords: Behavioural biometrics, accelerometer, gyroscope, mobilephone, authentication.

1 Introduction

Mobile phones have become an essential part of our daily life. We are heavily relying on various applications installed on them to socialize and even do business. These applications either have a lightweight authentication mechanism or donot have any authentication measurement at all because of the priority from user friendliness aspect. For those applications with an authentication mechanism, using PIN (Personal Identification Number) or password is still dominant. In recent years, embedding a fingerprint sensor in mobile phone has become prevalent [Ga14]. However, fingerprint recognition system is vulnerable to the presentation attack, as our fingerprints can be easily obtained from the glass that we used, or the wall that we touched, and later can be used to attack our mobile phones. Many researchers have revealed the feasibility of such attacks [AJP17]. Even one fake fingerprint has the potential capability to hack into the fingerprint sensors from multiple mobile phones [Ya14], which makes fingerprint based authentication system less trustful.

Besides the fingerprint sensor, accelerometer sensor and gyroscope sensor are ubiquitously incorporated in the modern mobilephones. Accelerometer sensor and gyroscope sensor have been widely studied for implicit authentication which provides a user-friendly security mechanism for identifying the user. According to the definition proposed in the article

[1] Department of Information Security and Communication, Norwegian University of Science and Technology, Teknologivegen 22, 2815 Gjvik Norway, guoqiang.li@ntnu.no

[2] Department of Information Security and Communication, Norwegian University of Science and Technology, Teknologivegen 22, 2815 Gjvik Norway, patrick.bours@ntnu.no

[Bo12, TA12], they are two types of authentication models based on accelerometer sensor and gyroscope sensor: continuous authentication and static authentication, where static authentication is used as an alternative to substitute the password or PIN. Continuous authentication monitors the behaviour of a user in a constant manner in order to determine if the current user is still the genuine one that has been granted the access. Many articles have presented the approaches to analyze these behaviour characteristics, such as using accelerometer data for gait recognition [SJ15, De10], studying gyroscope data for continuous authentication on mobile phone [Ya14], and combining context information with sensors' data for continuous authentication [LL17] as well as detecting the physical activities [Si16]. Unlike a variety of approaches in the continuous authentication field, there is much less research work on static authentication based on accelerometer and gyroscope data. A paper [Li12] utilized the accelerometer data to study whether a user can be recognized by looking at the different orientation when the users rotate their mobilephone. Another work [Le17] analyzed the accelerometer and gyroscope data to authenticate the user when they pick up the mobile phone. However, there is a lack of research studying the feasibility of authenticating the user based on the accelerometer and gyroscope data when the user opens his/her mobile application. The way that the user holds the mobilephone (one hand holds or two hand hold) and the phone movement of clicking on the mobile application are discriminating from individual to individual. These difference can be reflected and measured by the accelerometer and gyroscope sensors. This is the main motivation that drives us to explore the feasibility of developing such mobilephone application authentication approach as an alternative to PIN / password when the user opens an application. This implicit authentication approach is more robust to the presentation attack than the fingerprint based authentication approach. By applying this novel authentication approach, we expect the user will have the chance to be automatically authenticated and log into the mobile phone App without asking PIN or Password.

The remainder of this paper is organized as follows: in Section 2, we describe the major component of the proposed approach, and a new weighting approach is also designed in this section; Section 3 reports the experimental results; Conclusions and future work are discussed in Section 4.

2 Authentication based on Accelerometer and Gyroscope Data

2.1 An Overview of the Proposed Approach

The proposed approach analyzes the first few seconds data collected from the accelrometer sensor and gyroscope sensor to authenticate the user when the user opens a mobilephone application. The proposed approach consists of four major components as briefly described below.

- **Interpolation:** due to the practical challenge, accelerometer and gyroscope data cannot be always collected at a uniformly manner from various mobilephones. An interpolation component is designed in our approach to construct the missing points when the sampling rate of capturing data is set a fixed value;

- **Feature extraction:** a number of features will be extracted from the accelerometer and gyroscope data for further processing;
- **Classification:** a Random Forest based classification component will recognize the features as a genuine one or a imposter one by outputting a similarity score that is defined in 1. Lets assume the probability of classifying the user as a genuine user is P_g, and the probability of classifying the user as an imposter user is P_i.
- **Fusion process from accelrometer and gyroscope data:** a feature-level fusion approach and a score-level fusion approach will be studied. The details will be given in 2.3.

$$similarity\_score = \begin{cases} P_g/P_i & \text{if } P_i > 0 \\ 1000 & \text{if } P_i = 0 \end{cases} \tag{1}$$

2.2 Feature Extraction from Accelerometer and Gyroscope Data

Given that accelerometer sensor and gyroscope sensor provide three values $(x,\ y,\ z)$ respectively from their three axes for each sampling. In our feature extraction method, we consider another value s which is the root-mean-square by combining $(x,\ y,\ z)$ as denoted in Equation: $s = \sqrt{\frac{x^2+y^2+z^2}{3}}$.

After calculating the root-mean-square values both for accelerometer data and gyroscope data, we assume $(x_a^i,\ y_a^i,\ z_a^i,\ s_a^i)$ represents the values for one session from accelerometer sensor, and $(x_g^i,\ y_g^i,\ z_g^i,\ s_g^i)$ represents the values for one session from gyroscope sensor, where $1 \leq\ i\ \leq M$, M is 50Hz (the sampling rate). In this work, one session is the data collected from the first few seconds when the user opens a mobilephone application. There are 12 features extracted from each session of accelerometer data:

- **Based on mean value:** mean(x_a^i), mean(y_a^i), mean(z_a^i), mean(s_a^i);
- **Based on median value:** median(x_a^i), median(y_a^i), median(z_a^i), median(s_a^i);
- **Based on standard deviation (std) value:** std(x_a^i), std(z_a^i), std(s_a^i);

There are 24 features extracted from each session of gyroscope data. They are:

- **Based on mean value:** mean(x_g^i), mean(y_g^i), mean(z_g^i), mean(s_g^i);
- **Based on median value:** median(x_g^i), median(y_g^i), median(z_g^i), median(s_g^i);
- **Based on standard deviation (std) value:** std(x_g^i), std(z_g^i), std(s_g^i);
- **Based on maximum value:** max(x_g^i), max(z_g^i), max(s_g^i);
- **Based on minimum value:** min(x_g^i), min(z_g^i), min(s_g^i);

- **Difference between maximum and minimum value:** max(x_g^i)-min(x_g^i), max(z_g^i)-min(z_g^i), max(s_g^i)-min(s_g^i);

2.3 Fusion Strategy

In order to leverage the combination from accelerometer data and gyroscope data, we study the fusion strategy at feature-level and score-level individually to evaluate the performance variation. The feature-level fusion mechanism is carried out by concatenating two feature vectors generated from accelerometer and gyroscope data into a combined feature vector with 36 values. Furthermore, this 36-D feature vector will be fed into the Random Forest classifier to authenticate the user.

With respect to the score-level fusion mechanism, the combining implementation is carried out after the accelerometer similarity score and gyroscope similarity score are generated from their classifiers. A common thought of fusing these two similarity scores is to set different weights in order to achieve the best performance. Several weighting approaches have been presented in the literature. We selected three recently published weighting approaches to find out the optimal weights. In addition, we also propose a weighting approach according to our own scenario. A short description of each weighting approaches is given as follows.

The Overlap Deviation Weighting (OLDW) approach was proposed by Daner et al. [DON14] who considers the overlap area between the imposter and genuine score distributions after removing the score outliers. The overall verification of performance of the biometric source (EER) is the second consideration of this weighting approach. The formula of OLDW is presented by Equation 2, where S^I and S^G are imposter scores and genuine scores. T is the score threshold corresponding to EER. N is the number of biometric modalities.

$$w_k = \frac{\frac{1}{OLD_k}}{\sum_{k=1}^{N} \frac{1}{OLD_k}},\ OLD_k = \sigma(S_k^I|S \geq T \cup S_k^G|S < T) \times EER \tag{2}$$

The Non-Confidence Width Weighting (NCW) approach was proposed by Chia et al. [CSN10] by simply considering the whole range of the overlapping area from the genuine and imposter scores. Equation 3 gives the formula of NCW weighting approach, where Max_k^I is the maximum imposter score, and Min_k^G is the minimum genuine score.

$$w_k = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^{N} \frac{1}{NCW_k}},\ NCW_k = Max_k^I - Min_k^G \tag{3}$$

The Fisher Discriminant Ratio Weighting approach (FDRW) was designed by Lorena et al. [LDC10] who considers the separability of the genuine and imposter scores. The formula of FDRW weighting approach is given in Equation 4, where μ_k^G and μ_k^I are mean values of genuine scores and imposter scores. σ_k^G and σ_k^I are the standard deviation of genuine scores and imposter scores respectively.

$$w_k = \frac{FDR_k}{\sum_{k=1}^{N} FDR_k}, \; FDR_k = \frac{(\mu_k^G - \mu_k^I)^2}{(\sigma_k^G)^2 + (\sigma_k^I)^2} \tag{4}$$

We extend the idea from FDRW approach by incorporating the overall verification performance EER and its corresponding threshold. The proposed weighting approach is presented in Equation 5. We think the lower EER and higher threshold shall lead to a higher weight. The reason of considering the threshold is that the higher threshold indicates that the generated Random Forest classifier has stronger capability to distinguish the genuine/imposter users, as our similarity scores is calculated by dividing the genuine user probability over the imposter user probability.

$$w_k = \frac{P_k}{\sum_{k=1}^{N} P_k}, \; P_k = \frac{(\mu_k^G - \mu_k^I)^2 \times T}{((\sigma_k^G)^2 + (\sigma_k^I)^2) \times EER} \tag{5}$$

3 Performance Evaluation

3.1 Dataset Preparation

The dataset used for evaluation was collected from a real-life scenario. A data collection component was integrated into a mobilephone application which was installed in the mobile phones from the customers of a mobile network operator. Reading the data from the accelerometer and gyroscope sensors starts once the customer opens the mobilephone application, and the data collecting session lasts until the customer closes this application or leaves it running in the background. By considering a user-friendly use case, the proposed approach will only use the first three second data to authenticate the user. We divide the whole dataset into two parts: one development dataset used for seeking the optimal weights, and one evaluation dataset used for generating the performance. Tab. 1 lists the number of subjects and the total number of sessions involved in the accelerometer data. Note that there is a same amount of gyroscope data for the analysis as well.

Dataset	Number of subjects	Total number of sessions for accelerometer data
Development dataset	142	8,785
Evaluation dataset	162	18,465

Tab. 1: Basic information of accelerometer datasets used for performance evaluation.

Approach	EER
Accelerometer based authentication	23.5%
Gyroscope based authentication	33.18%
Feature-level fusion based authentication	24.28%

Tab. 2: EER (Equal Error Rate) values for accelerometer based authentication, gyroscope based authentication and feature-level fusion based authentication.

Score-level fusion based authentication approach with different weighting method	EER
OLDW weighting method	23.34%
NCW weighting method	25.56%
FDRW weighting method	23%
Proposed weighting method	22.72%

Tab. 3: EER (Equal Error Rate) values for score-level fusion based authentication approach with different weighting methods.

3.2 Experimental results

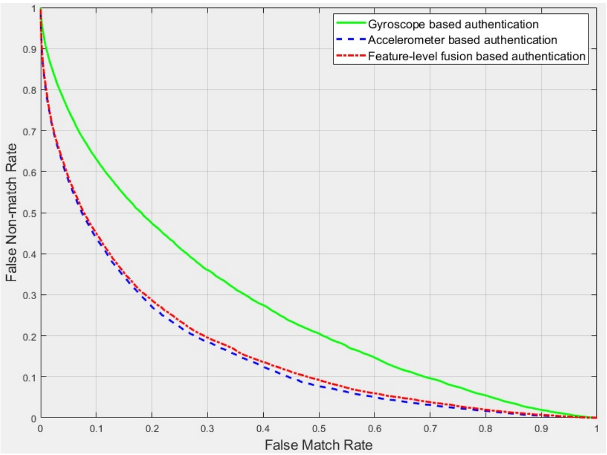

Fig. 1: DET (Detection Error Trade-off) curves for accelerometer based authentication, gyroscope based authentication and feature-level fusion based authentication.

In all experiments, the number of trees in the forest of Random Forest classification was set at 640. The number of genuine sessions for training the classifier is 50, and the number of imposter sessions for training the classifier is 100. After these setting, the number of genuine scores (both for accelerometer and gyroscope data) is 10366, and the number of imposter scores is 1,668,766. Tab. 2 lists the Equal Error Rate for the accelerometer based authentication, gyroscope based authentication and feature-level based authentication. The corresponding DET (Detection Error Trade-off) curves are illustrated in Fig. 1. As we can observe, the EER values from the accelerometer data is almost 10% lower than the gyroscope data, and the feature-level fusion approach is slightly higher than the accelerometer data. This implies that the accelerometer data is more distinguishable than the gyroscope data in such authentication scenario. Tab. 3 and Fig. 2 give the results for the score-level fusion approach with different weighting methods. As shown in the results, the proposed weighting approach is more appropriate for this behavioural dataset, and it further improves the performance comparing to the authentication approach that only uses the accelerometer data.

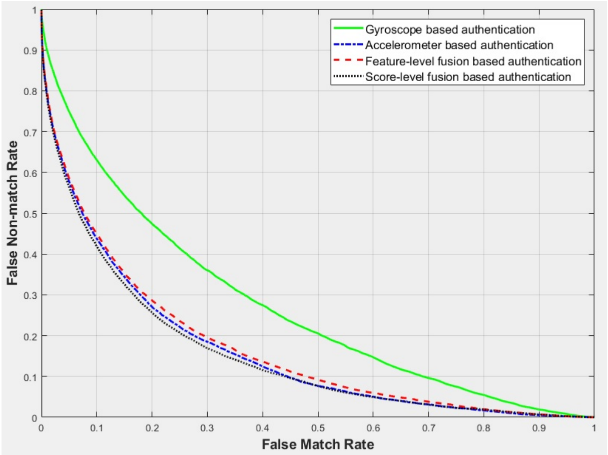

Fig. 2: DET (Detection Error Trade-off) curves for accelerometer based authentication, gyroscope based authentication, feature-level fusion based authentication and a score-level fusion based approach with the proposed weighting method.

4 Conclusions and Future work

Accelerometer data and gyroscope data have been widely studied for continuous authentication on mobile devices, yet they are rarely used for static authentication that can be an alternative for PIN or password. In this paper, we investigated the probability of using accelerometer data and gyroscope data for static authentication, and a new weighting approach is also developed in this work. By applying the proposed weighting approach and analyzing the 3 seconds data collected from accelerometer sensor and gyroscope sensor, we achieved the EER value at 22.72%. We think this performance could be further improved after we incorporate other source data from the mobilephone, such as WiFi and Bluetooth information.

5 Acknowledgment

This work was supported by the project “Awesome Possum: Data driven authentication”, led by Telenor Digital AS and financed by The Research Council of Norway under project number 256581.

References

[AJP17] Arora, Sunpreet S; Jain, Anil K; Paulter, Nicholas G: Gold fingers: 3d targets for evaluating capacitive readers. IEEE transactions on information forensics and security, 12(9):2067–2077, 2017.

[Bo12] Bours, Patrick: Continuous keystroke dynamics: A different perspective towards biometric evaluation. Information Security Technical Report, 17(1-2):36–43, 2012.

[CSN10] Chia, Chaw; Sherkat, Nasser; Nolle, Lars: Towards a best linear combination for multimodal biometric fusion. In: Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, pp. 1176–1179, 2010.

[De10] Derawi, Mohammad Omar; Nickel, Claudia; Bours, Patrick; Busch, Christoph: Unobtrusive user-authentication on mobile phones using biometric gait recognition. In: Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on. IEEE, pp. 306–311, 2010.

[DON14] Damer, Naser; Opel, Alexander; Nouak, Alexander: Biometric source weighting in multibiometric fusion: Towards a generalized and robust solution. In: Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European. IEEE, pp. 1382–1386, 2014.

[Ga14] Gao, Ming; Hu, Xihong; Cao, Bo; Li, Dianxin: Fingerprint sensors in mobile devices. In: Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on. IEEE, pp. 1437–1440, 2014.

[LDC10] Lorena, Ana Carolina; De Carvalho, André CPLF: Building binary-tree-based multiclass classifiers using separability measures. Neurocomputing, 73(16-18):2837–2845, 2010.

[Le17] Lee, Wei-Han; Liu, Xiaochen; Shen, Yilin; Jin, Hongxia; Lee, Ruby B: Secure pick up: Implicit authentication when you start using the smartphone. In: Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies. ACM, 2017.

[Li12] Lin, Chien-Cheng; Chang, Chin-Chun; Liang, Deron; Yang, Ching-Han: A new non-intrusive authentication method based on the orientation sensor for smartphone users. In: Software Security and Reliability (SERE), 2012 IEEE Sixth International Conference on. IEEE, pp. 245–252, 2012.

[LL17] Lee, Wei-Han; Lee, Ruby B: Implicit smartphone user authentication with sensors and contextual machine learning. In: Dependable Systems and Networks (DSN), 2017 47th Annual IEEE/IFIP International Conference on. IEEE, pp. 297–308, 2017.

[Si16] Sitová, Zdeňka; Šeděnka, Jaroslav; Yang, Qing; Peng, Ge; Zhou, Gang; Gasti, Paolo; Balagani, Kiran S: HMOG: New behavioral biometric features for continuous authentication of smartphone users. IEEE Transactions on Information Forensics and Security, 11(5):877–892, 2016.

[SJ15] Sprager, Sebastijan; Juric, Matjaz B: Inertial sensor-based gait recognition: a review. Sensors, 15(9):22089–22127, 2015.

[TA12] Traoré, Issa; Ahmed, Ahmed Awad E: Introduction to continuous authentication. In: Continuous Authentication Using Biometrics: Data, Models, and Metrics, pp. 1–22. IGI Global, 2012.

[Ya14] Yang, Qing; Peng, Ge; Nguyen, David T; Qi, Xin; Zhou, Gang; Sitová, Zdeňka; Gasti, Paolo; Balagani, Kiran S: A multimodal data set for evaluating continuous authentication performance in smartphones. In: Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems. ACM, pp. 358–359, 2014.