Evaluation of Motion-based Touch-typing Biometrics in Online Financial Environments

Attaullah Buriro[1], Sandeep Gupta[1], Bruno Crispo[1,2]

Abstract: This paper presents a bimodal scheme, the mechanism which contemplates the way a user enters an 8-digit PIN/password and the phone-movements while doing so, for user authentication in mobile banking/financial applications (apps). The scheme authenticates the user based on the timing differences of the entered strokes. Additionally, it enhances the security by introducing a transparent layer utilizing the phone-movements made by the user. The scheme is assumed to be highly secure as mimicking the invisible touch-timings and the phone-movements could be extremely onerous. Our analysis is based on 2850 samples collected from 95 users through a 3-day unsupervised field experiment and using 3 multi-class classifiers. Random Forest (RF) classifier out-performed other two classifiers and provided a True Acceptance Rate (TAR) of 96%.

Keywords: Smartphones, Biometric Authentication, Human-Computer Interaction

1 Introduction

Mobile banking is among the most sensitive activity a user performs on the Internet. Almost every bank now offers mobile banking through their dedicated apps. Thus, increasing number of smartphone users carry their banks around in their pocket rather than limiting themselves to just desktops or laptops. Recent research revealed that more than 82% of the teenage users (between 25 to 35 years) and 70% of the household users use online banking from their smartphones[3].

Mobile banking apps perform remote authentication requiring user credentials, as proof of identities, over the network. The credentials include `user-name` and the `password` (given by the bank or chosen by the user). The entered credentials are matched with the bank's database, and if found correct, the identity is confirmed. Since, they are open (exposed to view), uncontrolled and unsupervised, they pose several security challenges. Banks are shown to be reluctant replacing completely these schemes with the newer ones because there are no extensive data on their security.

[1] Department of computer Science and Information Engineering (DISI), University of Trento, Via Sommarive 5, Povo, Trento, Email: {attaullah.buriro, sandeep.gupta, bruno.crispo}@unitn.it

[2] DistriNET, KULeuven, Belgium, bruno.crispo@cs.kuleuven.be

[3] https://thefinancialbrand.com/62013/mobile-online-banking-payments-billpay/

This paper proposes motion-assisted touch-typing biometrics - a method to overcome the limitations of PINs/password, for the users of mobile banking apps. The scheme leverages two common human behaviors, i.e., touch-typing and phone-movements by the user. It identifies a user based on the timing differences of the entered strokes and the phone-movements made during the period of text entry. The user is authenticated on the basis of what and how she entered the text. In the case an adversary finds what is being entered, the access will still be denied because of the presence of the two invisible and inherently secure behaviors, i.e., touch-type timing differences and phone-movement. What we propose is also a effortless way to adopt behavioral biometrics, which not only complements the existing traditional methods but also keeps collecting data and security incidents with respect to time to evaluate dynamically the use of behavioral biometric only or both. We did not have time in this paper to collect historical data to say something specific about the choice, but the scheme offers gradual enrollment strategy.

The proposed scheme is fully transparent as it does not require any additional input from a user besides entering the credentials that makes it not only more usable but augments an additional layer enhancing the security of PINs/passwords as mimicking the person-specific movements are extremely onerous. Our scheme utilizes the built-in hardware, i.e., 3-dimensional sensors, to register user-generated phone-movements, and touchscreen to obtain touch-strokes. The sensors are started on the first touch and stopped on the last (the 8th). We evaluated our scheme on our collected dataset of 95 users by applying multi-class classification approach replicating the banking scenario. The main contributions of this paper are listed below:

- The proposal of a secure and usable behavioral-biometric-based authentication solution, for mobile banking. The scheme contemplates the touch-strokes timing-differences and the phone-movements during the process of entering PIN/password.
- Proof-of-concept prototype Android application of a proposed scheme for smartphones.
- Collection and sharing (in the due course of time) of the collected dataset of 95 users.

2 Related Work

Since the behavioral patterns can be collected unobtrusively, behavioral-biometric-based schemes are widely being researched for smartphone user authentication, these days. The search of new human behaviors, profiled through mobile sensors, have gained significant focus these days. Among all the researched schemes, i.e., the way a user walks (gait) [NWB12, De10] and they way a user types/enters any text (touchstroke)[Gi14, Bu15b, Bu16] are very popular.

Our scheme is a bimodal system which leverages the timing-differences from the entered 8-digit secret and the phone-movements while the user enters the text to login to the banking app, we compare our work with the closely related work proposed in the literature, i.e., [Gi14, Bu15b, Bu16] .

Giuffrida et al., [Gi14], proposed sensor enhanced keystroke based scheme for user authentication on Android smartphones. They reported an Equal Error Rate (EER) of 4.97% and 0.08% on fixed-text passwords (keystroke) and on sensory data, respectively, on a dataset of 20 users. Later, Buriro et. al., [Bu15b, Bu16] modeled sensory readings as hold behavior and introduced free-text secret the user needs to enter or writes on the touchscreen. They reported 1% EER on a dataset of 12 users for touch-typing [Bu15b] and ≈ 95% TAR at 3.1% False Acceptance Rate (FAR) on the dataset of 30 users [Bu16].

Our scheme is different from the previously proposed schemes in at least two ways: (i) all these papers performed in-lab supervised experiments and their analysis was based on a small number of users, i.e., just 12 [Bu15b], 20 [Gi14], and 30 [Bu16]. We evaluated our scheme on a comparatively larger dataset of 95 users collected in the wild. Since the number of users in previous studies was less and data was collected in the lab settings, it is difficult to examine how their achieved error would have varied if the number of users was more and data was collected in the wild. (ii) All of the papers evaluated their data either using one class or binary class classification [Bu15b] - replicating authentication on their smartphones [Bu16, Si15], but we have evaluated our data by applying multi-class classification replicating server based remote client authentication.

3 Motion-based Touch-typing Biometrics

To perform an online transaction, the user is required to login to the banking app, which is generally performed by entering the credentials i.e., email, customer-id, and 8-digit PIN/password. Banking server matches the credentials and decides accordingly. Hence, the user is authenticated on the basis of entered text and one who enters the correct pre-stored credentials is treated as the genuine customer/user. Since the password is vulnerable to spoofing, this mechanism poses a threat to the customer privacy.

Our scheme authenticate the user based on what and how she enters the text. Our scheme computes the key-hold and inter-stroke timings from the entered 8-digit secret and extracts the statistical features from different 3-dimensional sensors, for the entire duration of input, to profile the genuine user. In this way, it provides both usability (because the authentication mechanism is hidden from the user/customer), and security (because it is very difficult to impersonate the two inherently secure invisible human behaviors). Thus, the attacker needs to successfully mimic both invisible and person-specific characteristics to get access.

Figure 1 illustrates our approach. In enrollment phase, the banking server collects all the required features from the entered text and phone-movements to form a feature vector. Then it applies feature selection scheme to find out the most productive subset and calls it final feature. This final feature vector is saved in the bank's database under a particular label (i.e., user id, etc.)

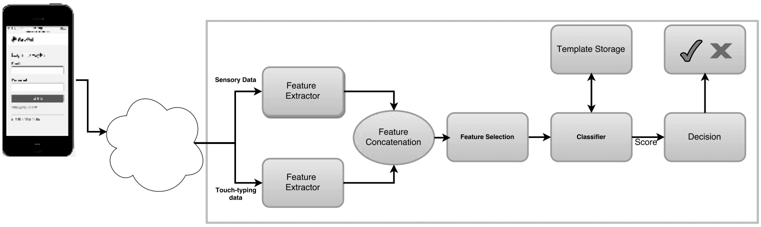


Fig. 1: Model diagram

In verification phase, the user is required to enter the label and the 8-digit password. The banking server picks the earlier pre-selected features from the features of the entered sample, and forms the query feature vector. Later, it compares this feature vector with all the pre-stored feature vectors under that label to find similarity, and authenticates the user, accordingly.

It is our assumption that the users would happily provide these much number of samples for enrollment in final systems, we evaluated our approach for different number of samples, i.e., 5, 10 and 15. However, we consider these samples too few to train the advance machine learning algorithms, i.e., deep learning, we chose simple but effective classifiers, which can perform pretty well even on less training-samples, for our evaluation.

4 Evaluation

4.1 Dataset

We collaborated with "UBERTESTERS a crowd sourcing platform to test the application, involving 95 users. We prototyped an Android application, namely, *PIN&WIN* to collect data. Our application can be installed on any Android device running 4.4.x OS or higher. We setup a web page with the complete explanation of *PIN&WIN*, i.e., the user consent, the procedure to install/uninstall the application. The testers had to agree to the consent form in order to download the app and to participate to the experiment. Then, they had to install the application, answer to the demographic questions, enter $8 - digit$ touch-types and keep the application running for at least 3 days. *PIN&WIN* required user's interaction in 3 sessions in 3 days. *PIN&WIN* required 30-minutes of user interaction on the first day, after installation, and 15 minutes of interaction on the following two days. In this manner, each user had to test the application for 1 hour, however, they needed to keep the application installed for $3 - days$. We collected 30 samples from each participant (in total 2850).

Tab. 1: User demographics (M = Male, F = Female, R = Right, L = Left)

Information	Description	Information	Description
No. of Users	95	Gender	75(m), 20(f)
Sample Size	2, 850 (30 X 95)	Password	8-digit
Devices	Android Smartphones with atleast `4.4.x` version	Handedness	89(R), 6(L)
No. of Sessions	3	Age Groups	90 (20 – 40), 5 (41 – 60)

4.2 Features Extraction & Selection

Our solution leverages all the 3-dimensional sensors i.e., the accelerometer, the orientation, the gravity sensor, the magnetometer and the gyroscope besides the touchscreen. Additionally, it also derives two other sensory readings by applying two filters, i.e., Low-Pass Filter (LPF) and high-Pass Filter (HPF). The value of $\alpha = 0.1$ was computed dynamically[4] to apply to these filters [Bu17]. Our solution leverages sensory readings from all 7 (3-dimensional) sensors in addition to the touchscreen data.

We gathered 4 datastreams from 3-dimensional sensors. Additionally, we computed 4th dimension for all the sensors, and called it magnitude, like in the previous studies [Bu15b][Zh14][Si15] .

We extracted 4 statistical features, namely mean, standard deviation, skewness, and kurtosis, from every data stream [Bu15b, Bu16, BCZ17]. Data from every sensor was transformed into a 4 by 4 features matrix. In total, we obtained 16 features from all four dimensions of each sensor. So the final feature vector for phone-movement behavior, from 7 physical sensors, becomes 112 features long. Similarly, the touch-typing feature vector is 30 features long extracted from the 8-digit password (similar to [Bu15b]). Hence, the final feature vector after concatenation becomes 142 features long.

The primary purpose of any feature selection scheme is to filter out the redundant and less productive features and feed the classifier with the most productive ones. Additionally, this helps also in decreasing the computational cost, i.e., processing smaller feature vectors would take less time. We applied Information Gain Attribute Evaluator[5](IGAE)- a Weka[6] implemented Information Gain based feature selection scheme. This scheme evaluates the worth of a feature by computing its information gain with respect to the class [BCZ17]. We obtained the threshold for feature selection by dividing the number of users (95) by the total number of features (142). The feature with higher weight was picked for further analysis.

[4] https://developer.android.com/reference/android/hardware/SensorEvent.html

[5] http://weka.sourceforge.net/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html

[6] http://www.cs.waikato.ac.nz/ml/weka/downloading.html

4.3 Classifiers

The classifier selection depends upon various parameters, i.e., data size, nature of the data, training time, etc. Our classification toolbox consists of simple but effective state-of-the-art classifiers: Naive Bayes(NB), NeuralNet(NN), and RF classifiers. All these classifiers are considered useful for smaller datasets and found useful in recent studies. We used PRTools[7], a matlab-based toolbox, for all the adopted classifiers and applied all of them in their default settings.

4.4 Experimental Protocol

As we collected 30 observations from each user, we picked first 5, 10 and 15 training samples for simultaneous training of the classifier and used the remaining samples for testing. The training with prior samples looks justified because after repetition the behavior becomes consistent and might show some biased results, i.e, training with prior samples and testing with remaining samples provides comparatively less accuracy.

4.5 Results

We report our obtained results in terms of True Acceptance Rate (TAR), False Reject Rate (FRR), False Acceptance Rate (FAR), True Reject Rate (TRR) and Receiver Operating Characteristics (ROC) curves. In particular, TAR, FAR, FRR and TRR can be defined as the fraction of the genuine samples correctly classified as genuine, the impostor samples incorrectly classified as genuine, the genuine samples incorrectly classified as impostors, and the impostor samples correctly classified as impostor, respectively. Since the FRR and TRR can be estimated by computing $1 - TAR$ and $1 - FAR$, respectively, we show TAR and FAR to avoid redundancy.

In Table 2, we show the TAR and FAR of our chosen classifiers on full and IGAE feature sets. It is evident that the TAR of all the classifiers increased on IGAE features, i.e., for RF classifier, it was 80.51% on full features and it increased upto 89.09% on selected features, for 5 training samples. Similarly, the TAR improved, significantly, as the number of training samples increased, i.e., from 80.51% to 89.09%, from 87.87% to 95.15%, and from 91.79% to 96.00%, for 5, 10, and 15 samples, respectively. The maximum TAR obtained by RF classifier is 96% on 15 training samples.

ROC curves are typically plotted between TAR on the y-axis and FAR on the x-axis. The curve starts from (0,0) and ends at (1,1) coordinates. The curve closer to (0,1) shows the better performance. We show an average ROC of all the users obtained through Vertical

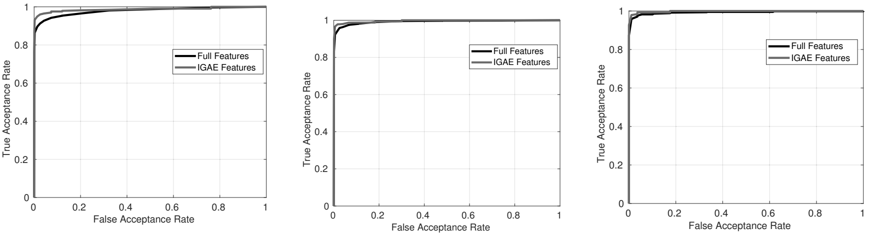
[7] http://prtools.org/

Averaging (VA)[Fa04] in Figure 2. In this averaging, the averages of the TAR rates is plotted against the researcher-defined fixed FAR. Due to the space limitations, we illustrate ROC curves for best performing classifier, i.e., for RF.

RF classifier outperformed both NB and NN classifier because of its ability to reduce the variances and its most unlikeliness of overfitting. NB classifier requires Gaussian distributed data, which might not be true in the dataset, hence it failed to address the problem of concept-drift. The NN classifier failed because of the limited number of training samples. It generally requires higher number of training samples to learn well.

Tab. 2: Results of different classifiers (averaged over all 95 users) on full and IGAE features.

	5				*10*				*15*			
	Full		*IGAE*		*Full*		*IGAE*		*Full*		*IGAE*	
Classifiers	TAR	FAR	TAR	FAR	TAR	FAR	TAR	FAR	TAR	FAR	TAR	FAR
NB	72.72	0.24	79.16	0.19	83.66	0.12	85.11	0.11	87.58	0.07	86.88	0.07
NN	57.81	0.37	77.26	0.20	63.61	0.27	84.51	0.11	70.53	0.16	85.89	0.08
RF	**80.51**	**0.17**	**89.09**	**0.09**	**87.87**	**0.09**	**95.19**	**0.04**	**91.79**	**0.04**	**96.00**	**0.01**



(a) 5 training samples (b) 10 training samples (c) 15 training sample

Fig. 2: ROC curves of RF classifier for (a) 5, (b) 10, and (c) 15 training sample scenarios.

5 Conclusion & Future Work

We have proposed a simple, effective and user-friendly, behavioral biometric-based remote user authentication solution for financial sector. The paper targets the users of mobile banking apps and helps the bank server in identifying the genuine user from the timing-differences of the entered strokes and the movements the user makes while entering the 8-digit secret. Our schemes is user-friendly, as it does not require any extra action for authentication. The transparent additional security layer based on phone motion enhances the security of the scheme, as mimicking simultaneously the two invisible and inherently secure human behaviors is very difficult, if not impossible.

We tried three different classification techniques and RF outperformed the other two. With RF as classifier, we obtained as high as 96% TAR on 15 training samples.

As some papers show [Bu16, Bu15a, Si15] that the behavioral patterns vary in different situations, so it will be interesting to test the scheme in different situations. We have already prototyped the proof-the-concept app based on our findings, however, its evaluation in terms of usability, and robustness against attacks, is a subject of future work. Additionally, its performance evaluation in terms of power consumption, computational constraints, i.e., CPU and memory overhead, and the sample acquisition time and decision time will be investigated as well.

References

[BCZ17] Buriro, A.; Crispo, B.; Zhauniarovich, Y.: Please Hold On: Unobtrusive User Authentication using Smartphone's built-in Sensors. In: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA-2017). 2017.

[Bu15a] Buriro, A.; Crispo, B.; Del Frari, F.; Klardie, J.; Wrona, K.: Itsme: Multi-modal and unobtrusive behavioural user authentication for smartphones. In: International Conference on Passwords. Springer, pp. 45–61, 2015.

[Bu15b] Buriro, A.; Crispo, B.; Del Frari, F.; Wrona, K.: Touchstroke: Smartphone User Authentication Based on Touch-Typing Biometrics. In: proceedings of the New Trends in Image Analysis and Processing– ICIAP 2015 Workshops. Springer, pp. 27–34, 2015.

[Bu16] Buriro, A.; Crispo, B.; Del Frari, F.; Wrona, K.: Hold and Sign: A Novel Behavioral Biometrics for Smartphone User Authentication. In: IEEE Security and Privacy Workshops (SPW). pp. 276–285, 2016.

[Bu17] Buriro, A.: Behavioral Biometrics for Smartphone User Authentication. PhD thesis, University of Trento, 2017.

[De10] Derawi, M.O.; Nickel, C.; Bours, P.; Busch, C.: Unobtrusive user-authentication on mobile phones using biometric gait recognition. In: IEEE 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). pp. 306–311, 2010.

[Fa04] Fawcett, Tom: ROC graphs: Notes and practical considerations for researchers. Machine learning, 31(1):1–38, 2004.

[Gi14] Giuffrida, C.; Majdanik, K.; Conti, M.; Bos, H.: I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics. In: International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. Springer, pp. 92–111, 2014.

[NWB12] Nickel, C.; Wirtl, T.; Busch, C.: Authentication of smartphone users based on the way they walk using k-NN algorithm. In: IEEE 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). pp. 16–20, 2012.

[Si15] Sitova, Z.; Sedenka, J.; Yang, Q.; Peng, G.; Zhou, G.; Gasti, P.; Balagani, K.: HMOG: A New Biometric Modality for Continuous Authentication of Smartphone Users. arXiv preprint arXiv:1501.01199, 2015.

[Zh14] Zheng, N.; Bai, K.; Huang, H.; Wang, H.: You are how you touch: User verification on smartphones via tapping behaviors. In: IEEE International Conference on Network Protocols (ICNP). pp. 221–232, 2014.