To appear in International Conference On Biometrics (ICB) 2019

Actions Speak Louder Than (Pass)words: Passive Authentication of Smartphone* Users via Deep Temporal Features

Debayan Deb, Arun Ross, Anil K. Jain
Michigan State University
East Lansing, MI, USA
{debdebay, rossarun, jain}@cse.msu.edu

Kwaku Prakah-Asante, K. Venkatesh Prasad
Ford Motor Company
Dearborn, MI, USA
{kprakaha, kprasad}@ford.com

Abstract

Prevailing user authentication schemes on smartphones rely on explicit user interaction, where a user types in a passcode or presents a biometric cue such as face, fingerprint, or iris. In addition to being cumbersome and obtrusive to the users, such authentication mechanisms pose security and privacy concerns. Passive authentication systems can tackle these challenges by unobtrusively monitoring the user's interaction with the device. We propose a Siamese Long Short-Term Memory (LSTM) network architecture for passive authentication, where users can be verified without requiring any explicit authentication step. On a dataset comprising of measurements from 30 smartphone sensor modalities for 37 users, we evaluate our approach on 8 dominant modalities, namely, keystroke dynamics, GPS location, accelerometer, gyroscope, magnetometer, linear accelerometer, gravity, and rotation sensors. Experimental results find that a genuine user can be correctly verified 96.47% *a false accept rate of 0.1% within 3 seconds.*

1. Introduction

The Digital Age has ushered in a large number of devices that store and generate information. Among these devices, smartphones are the most widely used [36]. An interesting phenomenon has been observed in the smartphone age, where users appreciate the convenience of services at their fingertips and implicitly store more and more valuable and private data, *e.g.*, banking and payment details, health records, etc. This has inadvertently sparked a community of hackers that dedicate their time and effort to gain access to smartphones in order to steal sensitive data [39]. Therefore, securing access to mobile devices by authenticating users is of utmost importance.

*Smartphones are multi-purpose mobile computing devices with strong hardware capabilities and extensive mobile operating systems, facilitating wide internet and multimedia functionalities [47].

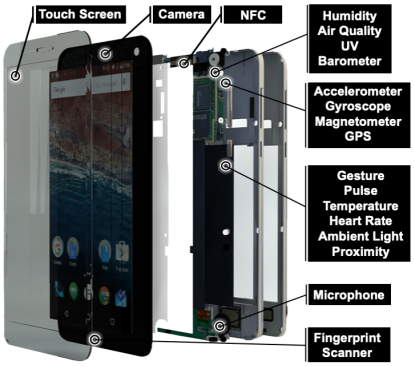

Figure 1: Authentication on smartphones by exploiting sensorial data has become an active field of research due to the growing number of available sensors in smartphones. We show 8 dominant modalities used in our passive authentication.

Current authentication schemes on mobile platforms require explicit user interaction with the device, referred to as *explicit authentication*, in order to gain access to it. The entry point for device access is typically a passcode or a biometric cue such as face, fingerprints or iris [5]. Passwords and PINs have long been viewed as the pinnacle of securing information and controlling access to mobile devices. However, these knowledge-based authentication schemes are prone to social engineering hacks, guessing and over-the-shoulder attacks [16]. With recent advances in technology, smartphones are getting better at authenticating users by learning their biological traits, such as face, fingerprint, or iris, which are believed to be unique to individuals [27]. Since these traits are innate to an individual, they are regarded as more reliable than knowledge-based authentication schemes. On the downside, biometric authentication raises privacy concerns related to how one's biometric data will be stored and protected. In addition, spoof attacks at the biometric sensor level [40], and possible theft of biometric templates stored inside the device, are among the growing

concerns related to biometric-based authentication.

Although the use of explicit authentication schemes is widespread, they are both cumbersome and obtrusive as the user needs to actively focus on the authentication step before utilizing the device. For instance, an average user unlocks their phone around 80 times a day [9], which can be a source of frustration even for the most avid of users. It is also estimated that, on average, a smartphone user spends over 4 hours per day on their device [23]. Unsurprisingly, more and more users prefer to set simple and weak passwords, increase the inactive period for lock-out time, or disable the authentication step completely [35], [44]. In addition, PIN codes, passwords, and biometric scans are well-suited for one-time authentication but are not effective in detecting intrusion after successful authentication by the genuine user when unlocking the phone. *Passive* authentication systems tackle these challenges by providing an additional layer of security by frequently and unobtrusively monitoring the user's interaction with the device. In this paper, we propose a passive user authentication scheme for smartphones where users are not required to participate in any explicit authentication step.

The first and foremost difficulty in designing user authentication schemes for smartphones lies in gathering data from the wide array of sensors available in a smartphone, as well as from a variety of users. Smartphone users today are becoming more sensitive to their privacy and more aware of spywares, which may avert potential users from providing their data. In this paper, we acquired a dataset comprising of measurements from 30 sensor modalities in Android phones they normally use for 37 users.

Besides data collection, a major challenge in designing a robust passive authentication system for smartphones involves extracting robust features from noisy[1] data. In addition, the robustness and accuracy of the authentication scheme needs to be thoroughly evaluated and inference should be performed in real-time. On average, a smartphone user's session lasts for just 72 seconds [33] and, therefore, the time required to authenticate the user should be as small as possible. It merely takes around 1.2 and 0.91 seconds to unlock an iPhone using FaceID and TouchID, respectively [1].

We propose a Siamese Long Short-Term Memory (LSTM) architecture for extracting deep temporal features from the data corresponding to a number of passive sensors in smartphones for user authentication.

Concisely, the contributions of the paper are as follows:

- Proposed a passive user authentication method based on keystroke dynamics, GPS location, accelerometer, gyroscope, magnetometer, linear accelerometer, gravity, and rotation modalities that can unobtrusively verify a genuine user with 96.47% TAR at 0.1% FAR within 3 seconds.
- Acquired a dataset comprising of measurements from 30 different smartphone sensors for 37 users around the world. An Android application was designed to log data unobtrusively from the users' smartphones.
- Analyzed changes in accuracy when (1) multiple modalities are fused, and (2) authentication time is varied. Increasing the number of fused modalities boosts the accuracy, whereas the TAR at 0.1% FAR drops from 99.87% to 96.47% for authentication times of 10 and 3 seconds, respectively.

[1] Smartphone sensors are prone to provide time-variant sources of noise leading to inaccurate measurements [24].

2. Related Work

2.1. Passive Smartphone Authentication

Currently, there are around 2.5 billion active smartphone users in the world [45]. With this increasing number, accurate, fast and robust authentication on smartphones has become an active area of research. Early work on passive smartphone authentication was based on touchscreen analysis [19], [30]. Frank *et al.* proposed a classification framework, namely Touchalytics, and achieved an EER of 4% on a dataset comprising of 41 users using touchscreen input data [21]. An obvious limitation of touchscreen recognition for passive authentication is the requirement of substantial explicit input from the user.

Smartphones today are shipped with an array of sensors (see Figure 1). A topic of increasing number of studies has focused on passive smartphone authentication via motion sensors. For instance, Derawi *et al.* investigated authenticating users based on their gait patterns [18] and achieved an equal error rate of 20.1%. In [32], the authors proposed a continuous motion-based authentication system using data from accelerometer and gyroscope sensors and obtained an EER of 18.2%. These high error rates do not meet smartphone security requirements.

2.2. Multimodal Biometric Systems

Most of the passive authentication studies have focused on a single sensing modality for authentication. Authenticating a user on their smartphone based on a *single* biometric modality becomes very challenging when the authentication time window is short. In addition, given the task the user is engaged in, the amount of data and the availability of different sensor modalities fluctuates. A robust passive authentication scheme must be able to adapt to the high intra-user variability observed in human-smartphone interaction.

Sitova *et al.* introduced a multimodal approach to passive smartphone authentication via accelerometer, gyroscope, and touch-screen observations [43]. Using a one-class SVM classifier, an EER of 7.16% was achieved;

Table 1: A few related work on multimodal passive smartphone authentication.

Study	Modality	Dataset Statistics	Classifier	Accuracy	Auth. Time*
HMOG [43]	Movement, tap, keys	100 users, 24 sessions†	Scaled Euclidean	EER 7.16%	60-120 seconds
Hold and Sign [8]	Movement, signature	30 users†	Multilayer Perceptron	95% TAR @ FAR = 3.1%	235 seconds
Touchalytics [21]	Touch gestures	41 users†	kNN, SVM	EER 3.0%	11-43 seconds
Mahbub *et al.* [31]	Movement and others	48 users for 2 months	Hidden Markov Model	Accuracy 96.6%	N/A
Fridman *et al.* [22]	Stylometry, app & web usage, GPS	200 users for 5 months	SVM, n-gram	EER 5.0%	60 seconds
This study	8 modalities in Table 2	37 users for 15 days	Siamese LSTM	96.47% TAR @ FAR=0.1%	3 seconds

TAR = true accept rate; FAR = false accept rate; EER = equal error rate
† No time span available for this study
* Time required before authentication

however, the authors deferred real-world scenarios, such as investigating authentication accuracy when the user is not engaged in typing, to future work. Authentication in a continuous setting has been studied in the past. Specifically, for smartphones, continuous authentication performance for touch gestures have been widely studied [48], [17], [21], [2], [20], [38], [11]. Keystroke dynamics is also another popular modality for studying continuous authentication on mobile phones [13], [12], [7], however, the devices used in these studies have a hardware keyboard for interfacing with the device and not a touch-based keyboard. Frank *et al.* and Serwadda *et al.* studied gait for exploring continuous authentication on smartphones [21], [41]. Niinuma *et al.* explored a continuous authentication scheme using the user's face and color of clothing for verification using a webcam [34]. In [15], the authors explored a passive authentication system for smartphones using face recognition. However, unobtrusively acquiring face images may be invasive to the user's privacy. Using a Siamese convolution neural network, Centono *et al.* showed a 97.8% accuracy in verifying the genuine user [10] using accelerometer, gyroscope, and magnetometer sensor modalities. However, the study does not consider the temporal dependence between samples. DeepAuth, on the other hand, used a LSTM architecture to classify a genuine/impostor user via accelerometer, gyroscope, and magnetometer samples [3]. However, they considered a small dataset where each user's session lasts for only 10-13 minutes.

Fusing decisions from multiple modalities to authenticate the user has been demonstrated to be very useful [42]. The majority of multimodal biometric systems fuse classifiers at the score level based on min, max, or sum rules [29]. In the proposed approach, we adopt the sum of scores fusion technique, which has been shown to perform well in multimodal biometric systems compared to other fusion schemes [26].

Limited studies on passive smartphone authentication have utilized multimodal biometric systems but, to the best of our knowledge, they have (1) all considered a small pool of modalities, (2) not evaluated the temporal performance of intrusion detection, and (3) not considered the temporal dependence of features across modalities. In this study, we propose a Siamese LSTM network to address temporal dependencies. A brief list of related work on passive smartphone authentication is given in Table 1.

3. System Overview

In the off-line phase, an authentication model is trained via the proposed methodology for each modality. During deployment, the incoming data from the smartphone sensor modalities are continuously monitored. If the incoming data successfully passes the authentication criteria, a decision is made that the current user is indeed the legitimate owner of the devices. Otherwise, the system locks out the user from the device and expects an explicit authentication method such as a password, or biometrics such as fingerprint scans.

4. Dataset

The dataset used in this work consists of measurements from 30 sensors, currently present in most commonly used smartphones, for 37 users. Data for each user was collected over a period of 15 days. Users that participated in the data collection process were primarily students from universities, across different countries, who are also regular smartphone users[2]. Dataset statistics are given in Table 3.

An Android application[3] was built that passively acquired data from the sensors. This application automati-

[2]The users were contacted by the authors and a consent form along with the link to the data collection application was sent. Users were remunerated with USD 50 for their efforts.

[3]https://play.google.com/store/apps/details?id=com.debayan.continuousdatacollect

Table 2: The eight dominant sensor modalities considered in our study.

Keystroke Dynamics	Key hold time, finger area and finger pressure
GPS Location	User's GPS location (latitude, longitude)
Accelerometer	Smartphone's acceleration in X, Y, Z plane
Gyroscope Gesture	Rate of rotation of the device in X, Y, and Z planes
Magnetometer	Earth's magnetic field in X, Y, and Z planes
Gravity Sensor	Direction and magnitude of gravity
Linear Acceleration	Linear acceleration in X, Y, and Z planes
Rotation Sensor	Device's rotation in X, Y, and Z planes

Table 3: Dataset Statistics. A record is a measurement for a sensor modality. Users that participated in the study are primarily students located in different regions of the world, including USA, India, Turkey, Brazil, and Dominican Republic.

No. of users	37
Data collection duration	15 days
No. of sensor modalities	30
Total number of records	6.7M
Average number of records per user	180K
Male to Female Ratio (%)	57/43
Age Range (years)	18 - 56

cally turns on whenever the smartphone boots up and continuously runs in the background while passively recording sensor data. In order to collect keystroke dynamics, we also built a custom soft-keyboard, installable from the data collection application.

To the best our knowledge, our dataset is unique due to (i) its rich sensor space (30 different sensors), and (ii) the manner in which data was acquired keeping the real-world scenario in mind. First, no user interaction with the data collection application is required, thereby enabling the users to use their smartphones as they generally would in their everyday lives. In order to simulate real world performance, data from users were collected from their own personal devices with no restrictions placed on the usage patterns, the Android device, or the Android OS version. Data from the modalities were acquired continuously, even when the user is not actively interacting with their smartphone.

In this study, we evaluate authentication performance of our proposed method on eight modalities (see Table 2), (i) keystroke dynamics, (ii) GPS location, and (iii) accelerometer, (iv) gyroscope, (v) magnetometer, (vi) linear acceleration, (vii) gravity, and (viii) rotation. We chose these modalities due to their popularity in literature [4], [43], [22], and because these sensors are available in all smartphones. In addition, among all the 30 modalities, measurements from these eight sensors are most abundant and distinctive.

Keystroke Dynamics By modeling user's typing rhythms and mannerisms, it is possible to authenticate smartphone users. We record the finger pressure, finger area, and hold time whenever a user types a character on their smartphone (similar to [4]). The exact characters typed are not logged and therefore, the keystroke patterns collected are non-invasive in nature.

GPS Location For every user in the dataset, a pair of latitude and longitude coordinates is recorded whenever the device was moved. Location is considered as a measure of an individual's characteristic and, therefore, we hypothesize that distinguishable patterns can be found in a user's location pattern.

Movement We evaluate the authentication performance for all the six movement sensors in our dataset: accelerometer, gyroscope, magnetometer, linear accelerometer, gravity, and rotation. Measurements are recording in three axes, X, Y, and Z, for all six sensors. The sampling rate for all six sensor data stream is 1 Hz, *i.e.* one measurement per second.

In our dataset, intra-user chronological gaps in measurements exists due to the high-variability in user's behavior. For instance, they may switch off their smartphone, or the phone may shut off due to battery drain. Even though these intra-user variations in the dataset pose additional challenges, it also better simulates real-world scenarios where such situations are common.

5. Methodology

As shown in Figure 1, smartphones today are shipped with an array of sensors, including global positioning system (GPS), accelerometer, gyroscope, magnetometer, and others. In this paper, our goal is to utilize data from these modalities in order to verify whether the genuine owner of the device is logged in. There are two main passive smartphone authentication model training strategies, namely, *on-line* and *off-line*. On-line approach trains an authentication model using samples pertaining to the smartphone user for a certain period of time before deploying the model. A major limitation to on-line approach involves training an individual model for each user. As a consequence, it is challenging to accurately evaluate the overall authentication performance across all the users due to high variance. In addition, the required amount of data, the duration of data collection before model deployment, and privacy concerns of storing the training data are ongoing challenges.

Off-line approaches, on the other hand, train a common authentication model that learns salient representations for individual modalities. In this approach, the same trained model is deployed when users install the application. Moreover, the users can avail of the authentication mechanism immediately after the installation of the application. For these reasons, we propose an off-line learning strategy for passive smartphone authentication. In particular, for each of the eight modalities, we train a Siamese LSTM network to learn deep temporal features. Samples from users are transformed into an embedding space learned by the Siamese

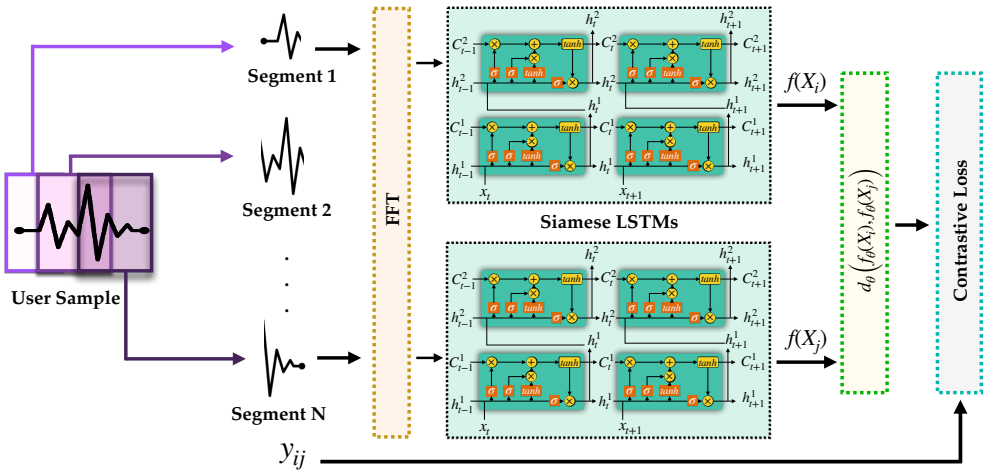

Figure 2: Architecture of the proposed model. Here, $\{X_i, X_j\}$ are input segment pairs, y_{ij} is the label, and $f_\theta(\cdot)$ and d_θ denote the embedding and distance function, respectively. The parameters of the Siamese LSTM network are denoted by θ.

network. Therefore, during deployment, only the features extracted from the incoming data are required for authentication, thereby, eliminating the need to store private data on the device. The proposed method is comprised of three modules: (1) data sampling, (2) preprocessing, and (3) Siamese LSTM. Figure 2 outlines the overall architecture of the proposed method.

Sampling Module Suppose that we extract a D-dimensional data sample for a given sensor modality (for instance, accelerometer has data in 3 axes, namely X, Y, and Z). The number of samples for each user can vary, and our dataset is comprised of chronologically irregular measurements for the same user due to various reasons such as their phones being switched off. Therefore, we segment the data by moving a window of fixed size T (authentication time window) over the sequential data with a pre-defined shift of T_{shift} and build overlapping fixed-sized segments. Hence, for each user, we have a set of $D \times T$ segments.

Preprocessing Module The outputs of the sampling module contain measurements from a modality in their original domain, namely the time domain. The frequency domain can handle and remove noise, while also retaining the discriminating patterns in the data within sequential data [37]. We map measurements from the time domain to frequency domain only for the movement sensors, *i.e.* accelerometer, gyroscope, magnetometer, linear accelerometer, gravity, and rotation. Fast Fourier Transform (FFT) [14] is utilized to convert time domain signals on each feature dimension to frequency domain signals. The output of the FFT vectors are concatenated with samples in the time domain so that we can utilize information from both the domains.

Siamese LSTM Our goal is to obtain highly discriminative features for each modality that can distinguish samples from genuine and impostor users. In other words, we would like to learn information-rich transformation of the data from modalities into an embedding space that can preserve distance relation between training samples. Suppose we are given a pair of input samples, $\{X_i, X_j\}$. Let y_{ij} be a label, such that, $y_{ij} = 0$, if X_i and X_j belong to the same user, and $y_{ij} = 1$, otherwise. Our objective is to map input samples to an embedding space where two samples from the same user are closer together and two samples from different users are far apart. A Siamese network architecture, which is a neural network architecture comprising of two identical sub-networks, is well-suited for such verification tasks [6], [46]. In this manner, relationship between two input samples can be learned. In a Siamese network, weights between the two sub-networks are shared and the weights are updated based on the label, y_{ij}.

A Siamese Convolutional Neural Network (CNN) was previously proposed for passive smartphone authentication [10]; however, CNNs are not well-suited to capture the temporal dependence within samples. We leverage Long Short-Term Memory (LSTM) [25] as our Siamese architecture to model patterns in users' data. LSTM, a variant of Recurrent Neural Networks (RNN), is designed for classifying, processing and making predictions on time series data. In our approach, we stack two LSTMs in order to learn hierarchical representation of the time series data. The first LSTM outputs a sequence of vectors, $h_1^1, \ldots, h_T^1$ which are then fed as input to the second LSTM. The last hidden state, h_T^2, of the second LSTM represents the final non-linear embedding, denoted by $f_\theta(\cdot)$, where θ represents the parameters of the Siamese LSTM network. This hierar-

Table 4: Authentication performance of the 8 modalities. Across 5 folds, the mean and std. dev. of the TARs at 1.0% and 0.1% FAR are given. The best performing model (highlighted in light gray) is the proposed Siamese LSTM (first column).

Modality	Proposed Siamese LSTM		Siamese CNN [10]		LSTM [3]		Euclidean Distance	
	1.0% FAR	0.1% FAR	1.0% FAR	0.1% FAR	1.0% FAR	0.1% FAR	1.0% FAR	0.1% FAR
Keystroke Dynamics	81.61 ± 13.65	58.71 ± 14.61	71.12 ± 15.67	43.87 ± 11.39	59.87 ± 18.82	26.73 ± 16.75	12.11 ± 7.60	8.20 ± 7.60
GPS	78.34 ± 4.76	52.21 ± 5.76	63.23 ± 10.82	39.23 ± 8.26	51.87 ± 9.97	21.42 ± 7.76	21.32 ± 1.32	12.76 ± 1.65
Accelerometer	74.56 ± 5.64	37.74 ± 6.67	67.28 ± 6.61	35.98 ± 7.33	64.83 ± 3.73	23.11 ± 3.82	13.06 ± 1.87	8.01 ± 0.81
Gyroscope	44.15 ± 7.53	15.18 ± 3.50	28.14 ± 7.69	11.33 ± 2.21	36.68 ± 7.43	8.89 ± 2.39	8.15 ± 1.23	6.54 ± 0.81
Magnetometer	74.15 ± 7.52	46.19 ± 8.83	60.91 ± 9.07	32.21 ± 6.87	26.33 ± 9.26	10.26 ± 8.33	18.85 ± 5.15	16.51 ± 5.88
Linear Accelerometer	50.19 ± 14.86	28.39 ± 16.20	46.35 ± 17.62	27.97 ± 18.74	29.89 ± 9.54	11.53 ± 6.78	8.91 ± 1.38	7.66 ± 0.95
Gravity	69.95 ± 4.35	32.24 ± 2.49	61.87 ± 7.95	31.92 ± 4.32	52.43 ± 5.64	32.12 ± 4.14	18.07 ± 5.41	10.98 ± 3.36
Rotation	74.85 ± 4.78	41.52 ± 3.02	61.91 ± 4.14	30.08 ± 1.29	56.75 ± 4.86	35.21 ± 3.98	17.96 ± 5.6	13.33 ± 3.72

chy of hidden layers allows for more salient representation of the time-series data. We denote the embedding size of both LSTMs as C.

In order to train the Siamese LSTM network, we define a pairwise contrastive loss function. For a given pair of input samples, the Euclidean distance between the two output feature vectors from the two sub-networks are fed to the contrastive loss function. This loss function regulates large or small distances depending on the label associated with the pair of samples, y_{ij}. In this manner, we ensure that the Euclidean distance between the pairs, $d_\theta(X_i, X_j)$, where

$$d_\theta(X_i, X_j) = ||f_\theta(X_i) - f_\theta(X_j)||_2$$

is small for genuine pairs and large for impostor pairs. The contrastive loss function is defined as [46]:

$$\ell_\theta = \Sigma_{i,j=1}^{N} L_\theta(X_i, X_j, y_{ij}), \qquad \text{where}$$
$$L_\theta = (1 - y_{ij})\frac{1}{2}(d_\theta)^2 + (y_{ij})\frac{1}{2}\{max(0, \alpha - d_\theta)\}^2$$

where, $\alpha > 0$ is called the margin.

6. Experimental Results

The contrastive loss function is optimized using Adam [28] optimizer. The embedding size (the number of hidden units) is fixed at 16. Segment shift is set to $T_{shift} = 1$ for all the experiments.

6.1. Individual Modality Performance

Since a separate Siamese LSTM model is trained for each modality, we first evaluate the authentication performance of each individual modality. We perform 5-fold cross-validation such that each fold comprises of 29 users for training and 8 users for testing. The proposed method is trained and tested on 20-second segments ($T = 20$). In order to train and evaluate our model, pairs of segment samples are generated. Genuine pairs are all possible pairs of segments from the same user. A sample from a user is paired with another user's segment randomly to constitute an impostor pair. On average, each user has around 180,000 genuine pairs and the same number of impostor pairs across all the modalities.

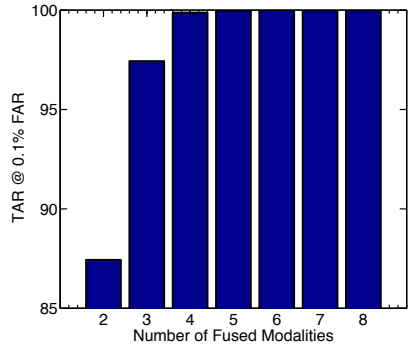

Figure 3: Best performing subsets of modalities. Increasing the number of fused modalities boosts the overall accuracy which saturates at a subset of 4 modalities.

In Table 4, authentication performance of the proposed method is compared with three baselines: (1) Siamese CNN proposed by Centeno *et al.* [10], (2) a single LSTM network proposed by Amini *et al.* [3], and (3) Euclidean distance classifier. The proposed Siamese LSTM network outperforms Siamese CNN due to LSTM's capability of capturing temporal dependencies. In addition, Siamese architectures are better suited for preserving distances between pairs of input samples. Thus, the proposed architecture has a better authentication performance than a single LSTM. Since Euclidean distance is used in the proposed method, we also investigate the authentication performance without learning a non-linear transformation of the temporal data. For this purpose, we obtain thresholds at 1.0% and 0.1% FARs by computing the Euclidean distances between the raw segment pairs in the training set. These thresholds are used to compute the TARs. We observe that using deep temporal features significantly improves the authentication performance.

6.2. Fusion of Modalities

The performance of the individual modalities is far from satisfactory. Given a short authentication time window, relying on a single modality for authentication is, therefore,

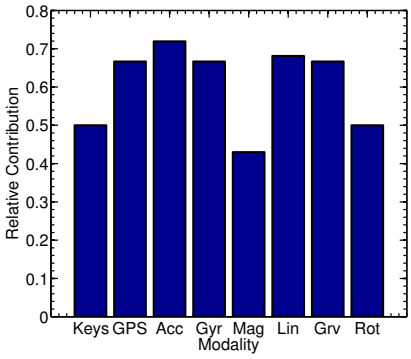

Figure 4: Relative contribution of each modality.

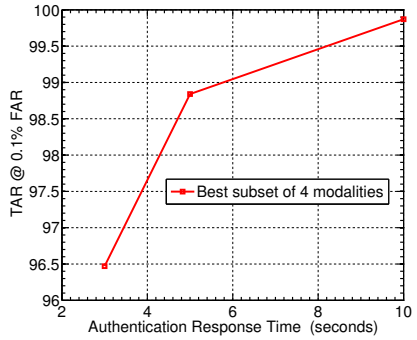

Figure 5: TAR at 0.1% FAR for authentication time windows varying from 3 to 20 seconds.

not practical for robust and accurate authentication. In addition to poor performance of individual modalities, all the modalities may not even be available at any given time of the day. For instance, if the user is not using the keyboard, we will not have access to keystroke dynamics. Therefore, for a reliable passive authentication system, it is imperative that we rely on a combination of modalities. We evaluate the authentication performance for all possible subsets of modalities on fusing 2, 3, 4, 5, 6, 7, and 8 modalities together (totaling 247 different subsets).

For all the 247 different combinations, the best performing subsets are shown in Figure 3. It is observed that: (1) increasing the number of fused modalities boosts the overall authentication performance; (2) the performance saturates after fusing 4 different modalities. A subset of accelerometer, linear accelerometer, magnetometer, and rotation achieves 99.87% TAR at 0.1% FAR.

6.3. Modality Contribution

The next natural question to ask is, *'Which modality contributes the most when all the 8 modalities are fused?'* Let $M = \{m_1, \ldots, m_8\}$ be the set of all modalities. When a modality, say, m_i, is not considered in the fusion, the drop in overall authentication performance is computed by $(TAR_M - TAR_{M'})$, where M' contains all the modalities except m_i and TAR_M denotes the True Accept Rate (%) at 0.1% FAR on fusing all modalities in M. The contribution for modality, m_i, is defined as $(TAR_M - TAR_{M'})/(100 - TAR_{M'})$. We plot the relative contributions of the eight modalities in Figure 4. Note that a high performing individual modality may not necessarily have a high relative contribution. For instance, magnetometer has the lowest contribution to the overall authentication performance out of the eight modalities. Linear accelerometer, on the other hand, has a relatively high contribution even though it performs poorly on its own. This is likely because the comparison scores obtained from the linear accelerometer model is complementary to scores from other modalities.

6.4. Temporal Information

A common phenomenon in passive authentication systems involves a trade-off between the authentication time and recognition performance. Figure 5 shows the authentication performance for segment sizes of $T = 3$, 5, and 10 seconds. We find that accuracy slightly drops with decreasing authentication time windows, likely due to lack of information content required to successfully authenticate the user. Therefore, the window size can be chosen depending on the application at hand. When shorter authentication times are desired, then more number of modalities should be considered for a secure application.

7. Conclusions

We have proposed a Siamese LSTM architecture for passive authentication of smartphone users. We collected a dataset comprised of measurements from 30 sensor modalities, for 37 smartphone users, over a time period of 15 days, while they are engaged in their daily activities on their own smartphones. We evaluated the authentication performance under various scenarios for 8 dominant modalities, namely keystroke dynamics, GPS location, accelerometer, gyroscope, magnetometer, linear accelerometer, gravity, and rotation. We summarize our findings as follows:

- The proposed method can passively authenticate a smartphone user with TAR of 99.87% and 96.47% at 0.1% FAR within 10 and 3 seconds, respectively.
- Relying on a single modality for authentication is not reliable. Fusing a subset of 4 modalities achieves 99.87% TAR @ 0.1% FAR.

With the growing number of sensors found in smartphones, it is important to explore robust and unobtrusive passive authentication approaches. In the future, we plan on acquiring data from new smartphone sensors and fusing additional smartphone modalities.

References

[1] 9To5Mac. iPhone X Face ID versus Touch ID which is faster? https://bit.ly/2D4bxc3, 2017. 2

[2] A. Acien, A. Morales, R. Vera-Rodriguez, and J. Fierrez. Multilock: Mobile active authentication based on multiple biometric and behavioral patterns. *arXiv preprint arXiv:1901.10312*, 2019. 3

[3] S. Amini, V. Noroozi, A. Pande, S. Gupte, P. S. Yu, and C. Kanich. Deepauth: A framework for continuous user re-authentication in mobile apps. In *ACM International Conference on Information and Knowledge Management*, pages 2027–2035. ACM, 2018. 3, 6

[4] M. Antal and L. Z. Szabó. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices. In *CSCS, 2015*, pages 343–350. IEEE, 2015. 4

[5] Biometric Update. Fujitsus latest smartphone using Delta ID iris recognition technology. https://bit.ly/2CMzDII, 2016. 1

[6] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah. Signature verification using a" siamese" time delay neural network. In *NIPS*, pages 737–744, 1994. 5

[7] A. Buchoux and N. L. Clarke. Deployment of keystroke analysis on a smartphone. In *Australian Information Security Management Conference*, page 48, 2008. 3

[8] A. Buriro, B. Crispo, F. Delfrari, and K. Wrona. Hold and sign: a novel behavioral biometrics for smartphone user authentication. In *SPW, 2016 IEEE*, pages 276–285. IEEE, 2016. 3

[9] Business Insider. The average iPhone is unlocked 80 times per day. https://read.bi/2EcOzkb, 2016. 2

[10] M. P. Centeno, Y. Guan, and A. van Moorsel. Mobile based continuous authentication using deep features. In *MobiSys*, pages 19–24. ACM, 2018. 3, 5, 6

[11] A. Chan, T. Halevi, and N. Memon. Touchpad input for continuous biometric authentication. In *IFIP CMS*, pages 86–91. Springer, 2014. 3

[12] N. L. Clarke and S. Furnell. Advanced user authentication for mobile devices. *Computers & Security*, 26(2):109–119, 2007. 3

[13] N. L. Clarke and S. M. Furnell. Authenticating mobile phone users using keystroke analysis. *International Journal of Information Security*, 6(1):1–14, 2007. 3

[14] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965. 5

[15] D. Crouse, H. Han, D. Chandra, B. Barbello, and A. K. Jain. Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data. In *ICB*, pages 135–142. IEEE, 2015. 3

[16] Data Genetics. PIN analysis. https://bit.ly/1cZJIvI, 2012. 1

[17] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012. 3

[18] M. O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *IIH-MSP, 2010*, pages 306–311. IEEE, 2010. 2

[19] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen. Continuous mobile authentication using touchscreen gestures. In *2012 IEEE HST*, pages 451–456. Citeseer, 2012. 2

[20] J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales. Benchmarking touchscreen biometrics for mobile authentication. *IEEE TIFS*, 13(11):2720–2733, 2018. 3

[21] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE TIFS*, 8(1):136–148, 2013. 2, 3

[22] L. Fridman, S. Weber, R. Greenstadt, and M. Kam. Active authentication on mobile devices via stylometry, application usage, web browsing, and gps location. *IEEE Systems Journal*, 11(2):513–521, 2017. 3, 4

[23] Hackernoon. How Much Time Do People Spend on Their Mobile Phones in 2017? https://bit.ly/2qXgA98, 2017. 2

[24] I. Hazan and A. Shabtai. Noise reduction of mobile sensors data in the prediction of demographic attributes. In *IEEE/ACM MobileSoft*, pages 117–120. IEEE Press, 2015. 2

[25] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. 5

[26] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005. 3

[27] A. K. Jain, K. Nandakumar, and A. Ross. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79:80–105, 2016. 1

[28] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[29] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE PAMI*, 20(3):226–239, 1998. 3

[30] C.-C. Lin, C.-C. Chang, D. Liang, and C.-H. Yang. A new non-intrusive authentication method based on the orientation sensor for smartphone users. In *SERE, 2012 IEEE*, pages 245–252. IEEE, 2012. 2

[31] U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa. Active user authentication for smartphones: A challenge data set and benchmark results. In *BTAS*, pages 1–8. IEEE, 2016. 3

[32] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, and G. Taylor. Learning human identity from motion patterns. *IEEE Access*, 4:1810–1820, 2016. 2

[33] Nielsen Norman Group. Mobile User Experience: Limitations and Strengths. https://bit.ly/1REinpX. 2

[34] K. Niinuma, U. Park, and A. K. Jain. Soft biometric traits for continuous user authentication. *IEEE TIFS*, 5(4):771–780, 2010. 3

[35] Norton. Norton Survey Reveals One in Three Experience Cell Phone Loss, Theft. https://bit.ly/2UbVX3J, 2011. 2

[36] Ofcom. Communications Market Reports. https://bit.ly/2IgfUVM, 2018. 1

[37] L. R. Rabiner and B. Gold. Theory and Application of Digital Signal Processing. *Englewood Cliffs, NJ, Prentice-Hall, Inc.*, 1975. 5

[38] A. Roy, T. Halevi, and N. Memon. An hmm-based behavior modeling approach for continuous mobile authentication. In *2014 IEEE ICASSP*, pages 3789–3793. IEEE, 2014. 3

[39] Scott Wright. The Symantec Smartphone Honey Stick Project. https://symc.ly/2AsSDdz, 2012. 1

[40] Security Research Labs. Fingerprints are not fit for secure device unlocking. https://bit.ly/2X5cQ1M. [Online; accessed 24-November-2018]. 1

[41] A. Serwadda, V. V. Phoha, and Z. Wang. Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. In *BTAS*, pages 1–8. IEEE, 2013. 3

[42] T. Sim, S. Zhang, R. Janakiraman, and S. Kumar. Continuous verification using multimodal biometrics. *IEEE PAMI*, 29(4):687–700, 2007. 3

[43] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani. Hmog: New behavioral biometric features for continuous authentication of smartphone users. *IEEE TIFS*, 11(5):877–892, 2016. 2, 3, 4

[44] Sophos. Survey says 70% don't password-protect mobiles. https://bit.ly/2DbPUXd, 2016. 2

[45] Statista. Number of smartphone users worldwide from 2014 to 2020 (in billions). https://read.bi/2EcOzkb, 2018. 2

[46] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Closing the gap to human-level performance in face verification. deepface. In *IEEE CVPR*, 2014. 5, 6

[47] Wikipedia. Smartphone. https://bit.ly/1Je5FpM, 2018. 1

[48] H. Xu, Y. Zhou, and M. R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *SOUPS*, volume 14, pages 187–198, 2014. 3