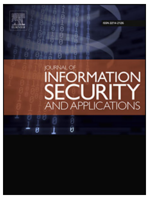
Journal of Information Security and Applications 44 (2019) 89–103

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa



AnswerAuth: A bimodal behavioral biometric-based user authentication scheme for smartphones

Attaullah Buriro[a,b,*], Bruno Crispo[a,c], Mauro Conti[d]

[a] Department of Information Engineering & Computer Science (DISI), University of Trento, Italy
[b] Department of Information Security, KFUEIT, Rahim Yar Khan, Pakistan
[c] Department of Computer Science, imec-DistriNet, KULeuven, Belgium
[d] Department of Mathematics, University of Padua, Italy

ARTICLE INFO

Article history:
Available online 7 December 2018

Keywords:
Smartphone
Sensors
Biometric authentication
Behavioral biometrics
Human-centered computing

ABSTRACT

In this paper, we present a behavioral biometric-based smartphone user authentication mechanism, namely, AnswerAuth, which relies on the very common users' behavior. Behavior, here, refers to the way a user slides the lock button on the screen, to unlock the phone, and brings the phone towards her ear. The authentication mechanism works with the biometric behavior based on the extracted features from the data recorded using the built-in smartphone sensors, i.e., accelerometer, gyroscope, gravity, magnetometer and touchscreen, while the user performed *sliding* and *phone-lifting* actions. We tested AnswerAuth on a dataset of 10,200 behavioral patterns collected from 85 users while they performed the unlocking actions, in *sitting, standing*, and *walking* postures, using *six* state-of-the-art conceptually different machine learning classifiers in two settings, i.e., with and without simultaneous feature selection and classification. Among all the chosen classifiers, Random Forest (RF) classifier proved to be the most consistent and accurate classifier on both full and reduced features and provided a True Acceptance Rate (TAR) as high as 99.35%. We prototype proof-of-the-concept Android app, based on our findings, and evaluate it in terms of security and usability. Security analysis of AnswerAuth confirms its robustness against the possible mimicry attacks. Similarly, the usability study based on Software Usability Scale (SUS)[1] questionnaire verifies the user-friendliness of the proposed scheme (SUS Score of 75.11). Experimental results prove AnswerAuth as a secure and usable authentication mechanism.

1. Introduction

Smartphones and tablets often fulfill user's desire of anytime-anywhere computing, which allows them to create their contents with ease and in an efficient manner. Furthermore, these devices also provide a portable means of accessing social networks, completing banking transactions, taking pictures and making movies along-with sharing these contents with user's family and friends. Since these devices store a growing quantity of user's private information, it becomes extremely important to keep user data secure from unauthorized access. To this end, researchers have recently designed new authentication methods specifically for smartphones and tablets. Existing user authentication techniques can be divided into four categories, i.e., passwords/PINs [2], graphical sketches [3], physical biometrics [4], and behavioral biometrics [5].

Passwords/PINs and graphical patterns are examples of classical authentication methods based on "*something user knows/remembers*". In such authentication scenarios, users must type the password or sketch they had set earlier in order to gain access to the device. These authentication methods are neither considered to be very secure [6–8], nor very convenient for the users [9]. Previous studies [10,11] reported that 70% users do not use any PIN/passwords to protect mobile phones because they consider entering their secrets more annoying compared to other telephony related problems such as lack of coverage or low voice quality [10].

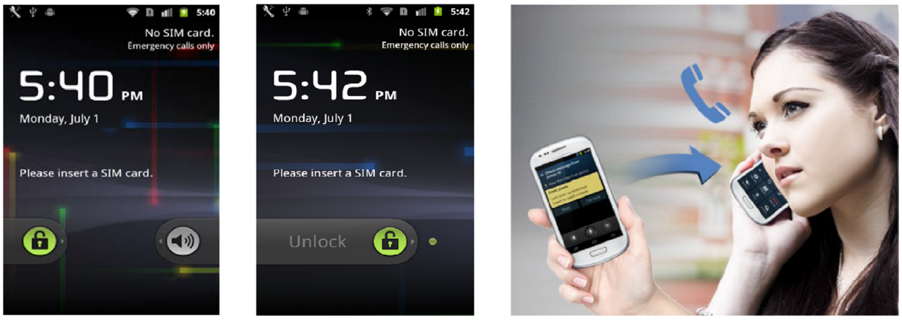
Graphical passwords use secret drawings [12], instead of secret strings of characters. User chosen graphical passwords have less entropy than traditional ones since users tend to choose symmetric figures [9], thus reducing in practice the domain space and making brute force attack feasible.

[*] Corresponding author at: Department of Information Engineering & Computer Science (DISI), University of Trento, Italy.

E-mail addresses: attaullah.buriro@unitn.it (A. Buriro), bruno.crispo@unitn.it (B. Crispo), conti@math.unipd.it (M. Conti).

[1] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

https://doi.org/10.1016/j.jisa.2018.11.008
2214-2126/© 2018 Elsevier Ltd. All rights reserved.



No SIM card.
Emergency calls only
5:40 PM
Monday, July 1
Please insert a SIM card.

No SIM card.
Emergency calls only
5:42 PM
Monday, July 1
Please insert a SIM card.
Unlock

(a) Smartphone in its default state

(b) Dragging of lock button

(c) Bringing phone towards ear [69]

Fig. 1. Different states of the our authentication mechanism.

Most of the limitations related to the use of passwords can be addressed through the use of other authentication mechanisms. Biometrics [13] refers to the establishment of the identity based on physical or behavioral modalities (sometimes called as *traits*) of an individual. Physical modalities include face, fingerprint, iris, hand geometry, etc., while behavioral biometric modalities include voice, signature, keystroke, gait, etc. Biometric authentication has multiple advantages over traditional authentication methods. They are considered more secure, as biometrics traits are hard to copy, and more reliable as they are hard to share/distribute and require user's presence at the time of authentication. However, currently deployed biometric systems have to deal with many practical problems [4], such as noise in sensed data (e.g., dirt on fingerprint sensor, and scars on finger may affect performance of fingerprint recognition), variation in luminous conditions (e.g., poor illumination of a userâs face in face recognition). Furthermore, accuracy of these biometric systems is affected by large intra-class variations, distinctiveness and non-universality (e.g., fingerprint biometric system may not extract the required features from individuals fingers, due to low quality of the ridges). Physical biometrics require expensive hardware to be reliable and robust against forgery attacks, thus increasing the cost of the device. Lastly, many of these biometric systems require active user cooperation which results in annoying the users.

We do not advocate the supremacy of behavioral biometrics over the physical biometrics, however, because of their inherent requirements for user cooperation, physical biometrics possibly result in annoying the user, whereas, since the authentication is performed unobtrusively using behavioral biometrics, they have become the preferred choice for user authentication for smartphones.

To address these issues related with smartphone usability and security, in this paper, we propose a bi-modal behavioral biometric-based authentication method based on the way a user slides the lock button and brings her phone towards her ear. Nearly all modern smartphones implement this swiping mechanism, i.e., they require sliding of the lock button on the screen to unlock the phone.

Earlier work presented in [15] used only the phone pickup movement to authenticate users. This paper enhances the initial scheme in [15] introducing a feature extraction process and evaluating the methods with multiple classifiers. Furthermore, this mechanism extends the earlier approach to a bi-modal biometric system using an additional modality - the *sliding* modality. Further, we re-run all the experiments from scratch with an higher number of users (85, compared to just 10) and higher number of samples (120) collected in the three postures, i.e., *sitting, standing*, and *walking*. Furthermore, this paper also contains the security and usability evaluation of our proposed scheme.

Our method, AnswerAuth, uses data from multiple three-dimensional physical sensors, namely, accelerometer, magnetometer, gravity and gyroscope sensors in conjunction with touchscreen data. Fig. 1a illustrates the smartphone in the default state. In order to be engaged in a call, usually a user takes her smartphone in her hand, drags the lock button to unlock (see Fig. 1b), and brings it towards her ear (see Fig. 1c). AnswerAuth leverages the combination of *sliding* behavior (action of dragging the lock button) and the *pickup* behavior (action of lifting the phone to the ear) to profile the users and use for authentication purposes.

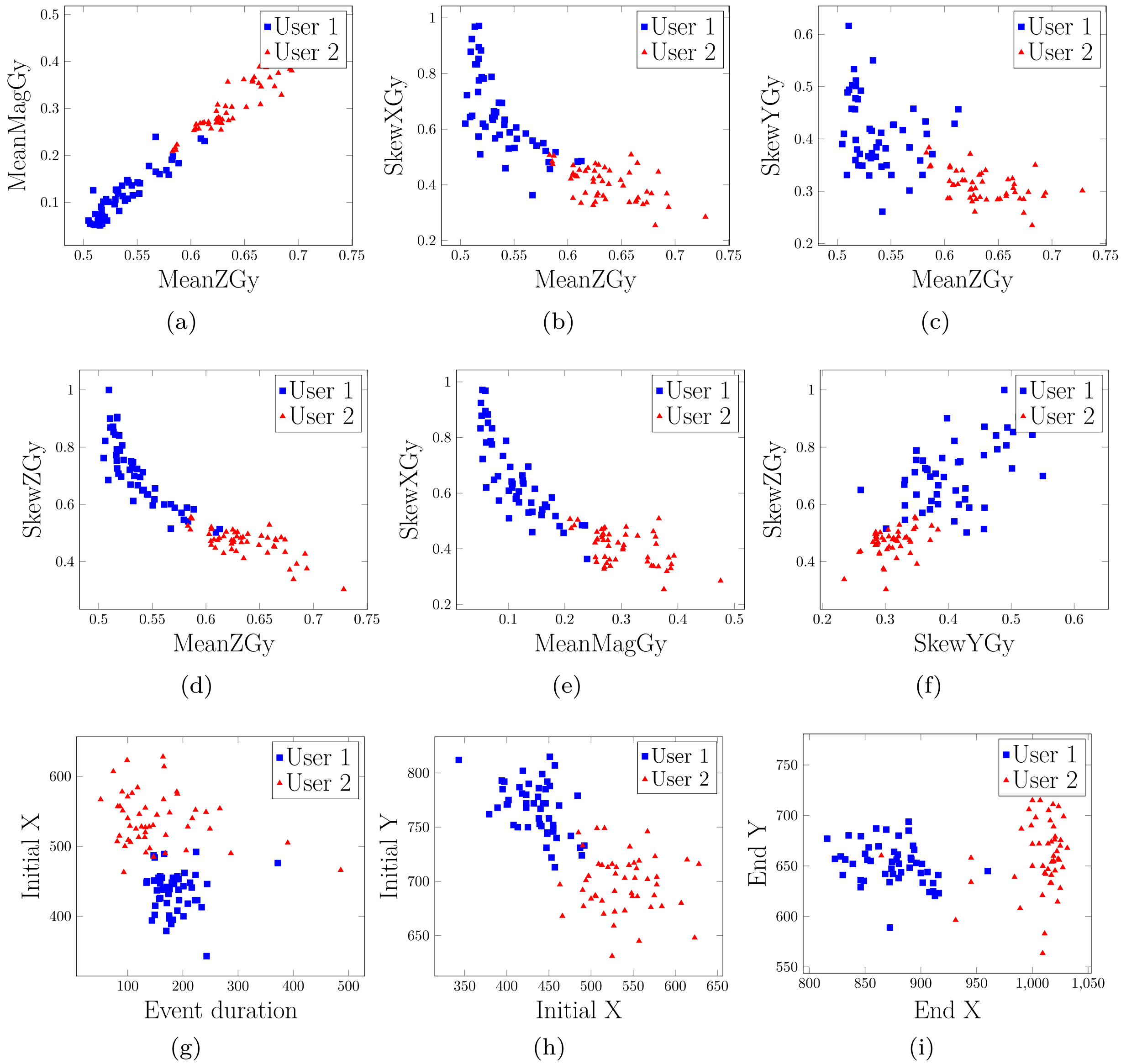
To model the *pickup* behavior, we exploited all the available physical sensors and extracted multiple statistical time domain features, i.e., Mean, Variance, Skewness and Kurtosis (see Table 3). We selected these features because they can be computed very cheaply as compared to the frequency domain features, which require computationally expensive Fourier transformations [17]. Similarly, for the *sliding* behavior, we extracted various touch-based features related with the velocity, acceleration and pressure of finger captured during the *slide-to-unlock* action (see Table 4). In this article, we consider the combination of *lock-button-drag* action and *phone-pickup* action as a *pattern*. We observed that the combination of these features is sufficiently unique from person to person (see Fig. 2) and can further be used towards designing a usable authentication method. Data from each individual sensor is preprocessed and listed features (see Tables 3 and 4) are extracted for fusion. In this way, all these features from all the sensors are concatenated to form a final feature vector for further analysis.

Since authentication is a binary-class classification problem, where data from one class is treated as a true class and other one as a potential attacker, we have used *six* state-of-the-art binary classifiers - we refer them as base classifiers for our experiments, using stratified cross validation method because of the limited number of observations (40 per user per user posture) and to provide maximum patterns for testing the classifiers. We propose four different solutions: The first leverages individual base classifiers, the second fuses these base classifiers using *Vote* classifier, the third leverages AttributeSelectedClassifier[2] (ASC) using the same set of base classifiers with *CFSEval* evaluator and bi-directional best-first search method, and the fourth combines these ASC's using *Vote* classifier. It is worth mentioning that we have used *product of probability combination rule* for this vote classifier because *average probability combination rule*[3] yielded worse results.

Interestingly, without any feature selection approach, we achieved acceptable results (the lowest accuracy from J48 classifier, i.e., 85.89% and highest accuracy from RF classifier 98.98%). In order to further improve the TAR, we classified our dataset using ASC using the same set of base classifiers. ASC is a Weka "Meta-class

[2] http://weka.sourceforge.net/doc.dev/weka/classifiers/meta/AttributeSelectedClassifier.html.

[3] https://www.programcreek.com/java-api-examples/index.php?example_code_path=weka-weka.classifiers.meta-Vote.java.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 2. Comparison of features for two users for different sensors.

classifier" capable of simultaneous feature evaluation and classification. We used same set of base classifiers with ASC using *CFSEval* along-with a bi-directional *BestFirst* search method. By doing so, we achieved a significantly improved TAR, i.e., from 85.89% to 86.94% for J48, and from 98.98% to 99.35% for RF classifier. Fusion of base classifiers gave comparatively better results than all other classifiers, except RF classifier, however, the fusion of ASC's classifiers did not yielded better accuracy, i.e., 88.4% to 92.64%. Extensive experimentation proved RF as the most accurate classifier in both settings, i.e., with full features (base) and with reduced features (ASC). Thus, our developed proof-of-the-concept prototype leverages RF as the classifier.

By embedding user authentication in the normal user action and gesture required to unlock the phone, we claim that the method is user-friendly and has the potential to gain wide user acceptability. Usability study confirms AnswerAuth as a user-friendly authentication scheme (AnswerAuth achieves a SUS score of 75.11). Additionally, our method can be implemented on almost any off-the-shelf smartphone, thus, it does not require any additional hardware.

The main contributions of the paper are listed below:

- The proposal of AnswerAuth- a user-friendly behavioral biometric-based user authentication mechanism, which is based on two very common human actions, i.e., how the user *slide-to-unlock* her smartphone (*sliding*) and and how she moves her smartphone towards her ear (*phone-pickup*).
- The evaluation of AnswerAuth, on the collected dataset of 85 users.
- Implementation of the AnswerAuth for Android phone.
- The usability evaluation of AnswerAuth based on the collected reviews from 85 users.
- The collection and sharing of data from multiple sensors in three postures, i.e., *sitting*, *standing*, and *walking* from 85 users.

Paper organization. The rest of the paper is organized as the following: In Section 2, we present the main authentication methods (based on behavioral biometrics) that have been proposed over the years, and we discuss why there is a need for an improvement. Section 3 reports our intuition and its initial assessment. Section 4 reports the detailed methodology of our conducted experiments and results are discussed in Section 5. Section 6 explains our Proof-of-the-Concept application. We evaluate AnswerAuth in terms of Security (in Section 7) and Usability (in Section 8). Finally, Section 9 concludes the paper with the summary of the work and the possible future work.

2. Related work

Researchers have been working with different user behaviors such as walking patterns (i.e., gait [16,18,28]), way of input (i.e., keystroke dynamics [19,23,33]), the way of holding [5,49] and interacting with the device [46,47,55–57] with the device and the way of bringing their phones towards their ear [15]. Interested readers are referred to this paper [20,44] for reading detailed discussion on state-of-the-art schemes available for smartphone user authentication. Additionally, the works [16,21,22] can be a good resource for understanding continuous authentication schemes. However, this Section is limited mainly to behavioral biometric-based one-shot authentication solutions proposed over the years for user authentication on smartphones.

2.1. Sensor based authentication

Smartphone sensors such as accelerometer, gyroscope, magnetometer, etc., have become main data sources for smartphone user authentication. Li et al. [29] tested specific combination of three physical sensors, i.e., accelerometer, orientation sensor and compass, in addition to the touch gestures to provide continuous user authentication. Similarly, accelerometer, touchscreen data, voice and location data were used for user authentication in [30]. Both these solutions achieved an accuracy of about 96%. *SenSec* [32] continuously collects data from accelerometer, gyroscope and magnetometer and constructs a gesture model to profile the way a user uses her phone. *SenSec* achieved 75% TAR and 71.3% TRR. The approach, presented in [18], uses acceleration signals produced when the user walks. Authors claim that their work is significantly different from classical gait recognition schemes because it does not involve any computer vision methods. By applying four matching algorithms, i.e., signal correlation method, Fast Fourier Transform, histogram and higher order moments, authors achieved an Equal Error Rate (EER) of 7%, 10%, 18% and 19%, respectively. The study [49] exploits the user's hand micro-movements (while she unlocks her smartphone using any implemented authentication scheme) to authenticate the user. More specifically, the proposed system collects 3-dimensional data from motion sensors, in the background, for a short period of time, and trains the classifier on the collected movement patterns. The data collection starts as soon as the user unlocks her smartphone. Authors reported a TAR of 96% at an EER of 4%, using MLP as the classifier, on their collected dataset of 31 qualified users. Primo et al. [51] proposed a context-aware accelerometer-based two-stage framework for user authentication. The investigated the impact of location (phone position) variations on the classification accuracy. Technically, their proposed system first infers the location of the phone (hand or pocket) and uses this information during the user authentication process. Using Logistic Regression (LR) as the classifier, they achieved as high as 82.30% accuracy while training and testing in the in-hand position.

2.2. Touch based authentication

Touch behavior has been extensively tested and used for smartphone user authentication. Recent work [25,26,34] confirms touch behavior as a potential modality for smartphone user authentication. A number of features related with time, velocity, touch-area and touch-pressure can be used. De Luca et al. [24] implemented a user password application which requires users to draw a sketch as a password. This application uses pressure, coordinates, size, speed and time to identify a valid user. The reported accuracy of the system, using Dynamic Time Warping (DTW) approach, is 77% with FRR of 19% and FAR of 21%. Angulo and Wästlund [35] proposed the use of a customized lock pattern and analyzed the touch data associated with that lock pattern. They achieved an EER of 10.39% using RF classifier. Sae-Bae et al. [36] studied specific five-finger touch gestures and reported an authentication accuracy of 90% on Apple iPad. Shahzad et al. [37] studied customized slide-based gestures for smartphone user authentication. They reported an EER of 0.5%. Authentication mechanism implemented by Sun et al. [38] requires user's arbitrary finger patterns on a specific region of the screen for unlocking the smartphone. Users were authenticated based on the geometric features extracted for the curves, drawn by finger movements. In [27], authors studied vulnerability of these touch gestures in terms of zero-effort (where attacker does not needs to make effort to spoof a gesture). More specifically, they demonstrated how a robotic device can pose a major threat to touch-based user authentication systems. Using, support vector machine and KNN as classifiers, they obtained an EER of 0.035%, and 0.13%, respectively, before robotic attacks and these EERs increased dramatically (upto 900%) after attacks. Frank et al. [57] proposed a touch-based continuous user authentication scheme for smartphones. More specifically, authors propose a classification framework that learns the touch behavior (the way a user interacts with the smartphone touchscreen) of a user during the enrolment phase and authenticates the users by monitoring the similarity of these interactions in testing phase. Applying KNN and SVM as classifiers, they achieve 0% to 4% median EER across all the scenarios; In inter-session, computed EER was $2\% - 3\%$, and below 4 in inter-week session. SVM classifier performed well compare to KNN. Another touch-based user authentication scheme [47] also exploits the touch gesture for user authentication. Their scheme selects 21 features from touch-based logs to train the chosen neural network classifier. Authors reported 7.8% and 3% error rate before and after optimizing the chosen classifier on their collected dataset of 20 users.

2.3. Sensor-enhanced touch-typing based authentication

Giuffrida et al. [31] proposed a sensor-assisted *fix-text* scheme for user authentication on Android smartphones. Authors reported 4.97% EER on passwords and 0.08% on sensory data over a dataset of 20 users. Later, Buriro et al. [33] used sensory readings to profile the users' *hold* behavior and fused it with the *free-text* password, the user enters on the touchscreen. They reported 1% EER on a dataset of 12 users. Similar research [55] leverages the way the user writes or signs on the touchscreen combined with the *hold* behavior, for user authentication on smartphones. They achieved $\approx$ 95% TAR at 3.1% FAR on the dataset of 30 users. Similarly another study [62] leverages the way the user dials any combination of 10-digit "*free-text*" from the smartphone dialpad combined with the sensory readings generated while dialing, for user authentication on smartphones. Authors reported 85.77% TAR on their collected dataset of 97 users. Another recent study [50] combines three modalities, namely, 8-digit free-text touchstroke, phone-movements while a user enters her PIN and the face, for user authentication on smartphones. They achieved as high as 99% TAR and an EER of 1% on the dataset of 95 users. Kumar et al. [45] proposed a tri-modal user authentication scheme based on swiping, typing and phone movement patterns. Authors evaluated the performance of each modality, individually, and also their fusion over their collected dataset of 28 users. Feature level fusion of the two modalities; swiping and the corresponding phone-movements modalities achieved an authentication accuracy as high as 93.33%.

AnswerAuth is different from the existing state-of-the-art in the following ways: Firstly, all the touch-based solution either leverage the timing-based features generated from keystroke / touchstroke or on the touch-points generated as a result of users' writing/signing or dialing on the touchscreen. However, our solution AnswerAuth leverages the *slide-to-unlock* action generated features, i.e., initial and final XY position, the velocity and the acceleration of the *slide-to-unlock* drag and the pressure and size of the finger. Similarly, in most of the sensor-based solutions, the data was collected continuously or in a context, for example, when the user walks [18], or when the user tries to log in to the smartphone [50]. AnswerAuth uses sensors to profile the specific movement - arm swing the user makes to lift her phone to her ear. Secondly, most of the proposed solutions utilized the collected data in the lab and under supervised conditions in one-session. This data collection scheme is cumbersome on one-hand and somehow biased because human behavior tends to vary with respect to time [58], on the other. In contrast, AnswerAuth is evaluated on a dataset of 85 users collected in the wild in three-days long experiment.

3. Our solution: ANSWERAUTH

In this section, we illustrate the main approach adopted by our solution.

3.1. Intuition assessment

AnswerAuth is based on the intuition that every user has a unique way of dragging the lock button and bringing the phone towards her ear. This phone motion is sufficiently unique and discriminating across different users. Physical sensors have the ability to measure these differences in movements. Therefore, we are considering extracted features from all the sensors and the touch related features from the user *slide-to-unlock* action.

Fig. 2 shows the scatter plots of some of the features, extracted from the accelerometer, gyroscope and the touchscreen for two users. It is evident that both the users have well separated features. We observed significant difference in the features of all the users, but due to the space limitations, we illustrate the scatter plots for two users, only.

We conducted several experiments to confirm our intuition. Specifically, we were looking for answers to two basic research questions, i.e., are the patterns (combination of *slide-to-unlock* and *phone-pickup-movement* actions) of the same user similar to each other (intra-class variations)? And the patterns of different users are different enough to be distinguished (inter-class variations)? Experiments confirmed our initial intuition. We observed a strong correlation between the patterns of the same user, and sufficient differences among the patterns of different users.

3.2. Our solution

AnswerAuth leverages *touch-based* and *phone-pickup* movement features generated as a result of *slide-to-unlock* and *phone-pickup* actions to identify a legitimate user. We observed sufficient variations in these gestures across multiple users, hence, these gestures could potentially be used for authentication purposes. We solve the problem of user authentication with four different models. In the first model, we select 6 significantly different classifiers and use them to authenticate a real user. It is worth mentioning that all of these classifiers were applied in their default settings. Fig. 3a illustrates this scenario, i.e., data from each sensors is preprocessed and the features are extracted. These features are fused together to make a feature vector of 128 features and stored as the template in the main database. Our selected features for this experiment are listed in Tables 3 and 4. For user authentication, the query pattern is matched with the pre-stored patterns and the user is authenticated based on the decision of these classifiers. In the second model, illustrated in Fig. 3b, the outcomes of all the classifiers are combined using *vote* classifier to authenticate/reject the user. In the third model, illustrated in Fig. 3c, the set of classifiers used in the first model was used after applying the feature selection process using ASC Meta-Class classifier. In the final model, illustrated in Fig. 3d, the outcomes of all the ASC classifiers are combined using *vote* classifier.

3.3. Considered sensors

Current smartphones are fully equipped with a wide range of sensors, i.e., motion, position and environmental sensors. In our scheme, we used all the 3-dimensional physical sensors, i.e., accelerometer, gravity, gyroscope, magnetometer along-with the touchscreen. Additionally, we applied High Pass Filter (HPF) and Low Pass Filter (LPF) to obtain HPF and LPF acceleration readings. By applying HPF, we obtained exact acceleration applied on the device by the user, and by LPF, we obtained the apparent transient forces acting on the device due to the users' activity. Thus, we used 3 variants of accelerometer sensors, i.e., Raw, LPF and HPF [5]. We further explain the working of our selected sensors as follow:

3.3.1. Accelerometer sensor

This sensor measures the acceleration of phone in three dimensions, namely, X, Y and Z directions. In this way, it provides the movement of the smartphone in a 3-dimensional space. Furthermore, acceleration of this phone movement varies from person to person and it can be computed through accelerometer sensor.

3.3.2. Gravity sensor

This sensor measures the applied force of gravity (m/s^2) on the smartphone in three dimensions. In simple words, it provides magnitude and direction of the force of gravity applied on the phone. The coordinate system and the unit of measurement of gravity sensor are the same as of the accelerometer sensor.

3.3.3. Gyroscope sensor

This sensor measures the smartphone rate of rotation (rad/s) in three dimensions. The sensor's coordinate system is the same as the one used for the acceleration sensor. The counter-clock-wise rotation is positive, i.e., an observer if looking from some positive location on the three axes at a device positioned on the origin world, is considered positive.

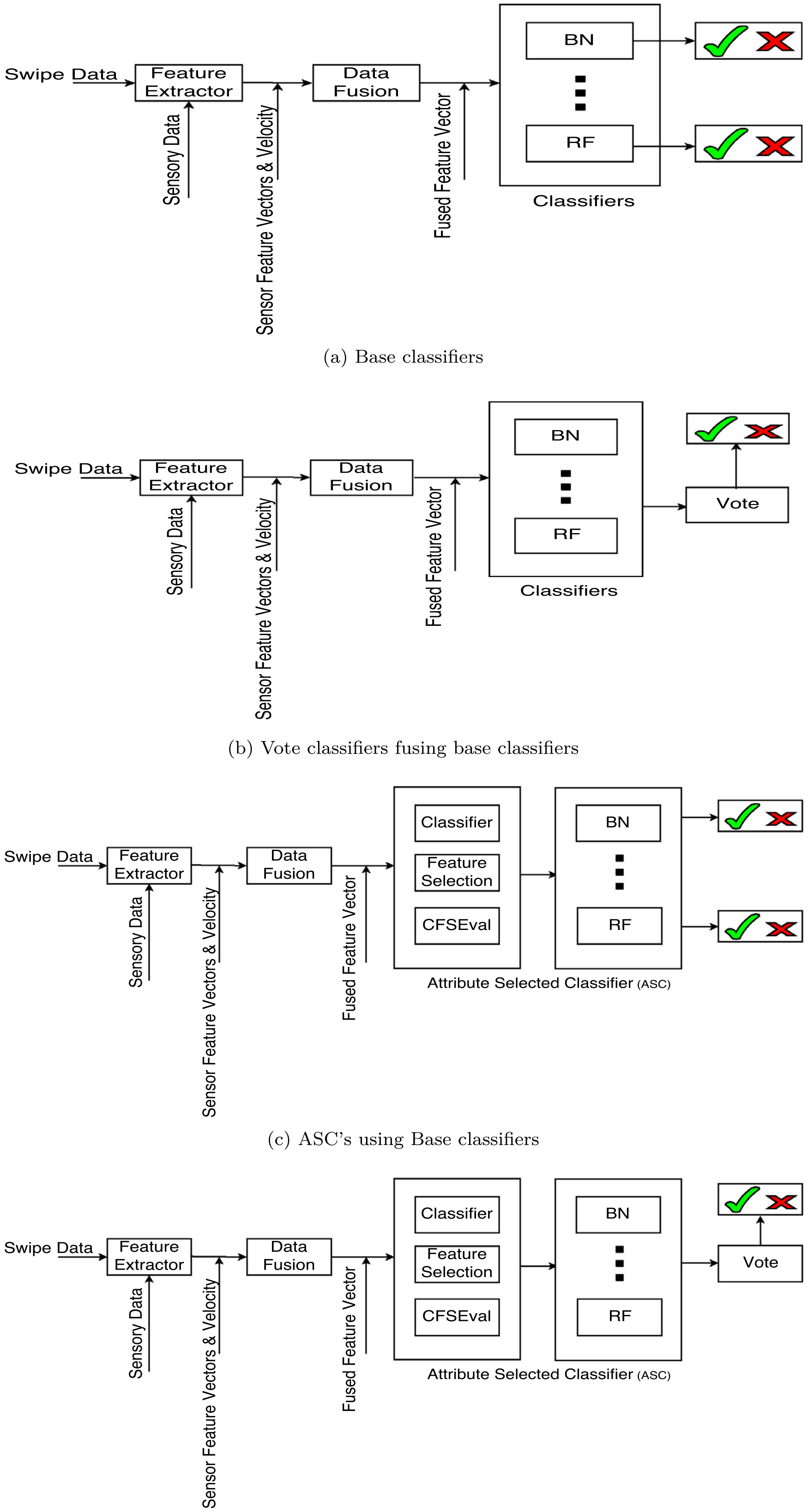
3.3.4. Magnetometer sensor

The magnetometer sensor measures the strength and/or direction of the magnetic field (μT) in three dimensions. It differs from the compass as it does not provide point north. The magnetometer measures the Earth's magnetic field if the device is placed in an environment absolutely free of magnetic interference.

All the above sensors generate continuous streams in X, Y and Z directions. We have added a fourth dimension to all of these sensors and name it *magnitude*. Magnitude has been tested in the context of smartphone user authentication [33,48,54], and has proved to be very effective in classification accuracy. The magnitude is mathematically represented as:

$$S_a = \sqrt{(a_x^2 + a_y^2 + a_z^2)}, \tag{1}$$

where S_M is the resultant dimension and a_x, a_y and a_z are the accelerations along the X, Y and Z directions.



(a) Base classifiers

(b) Vote classifiers fusing base classifiers

(c) ASC's using Base classifiers

(d) Vote classifier fusing ASC's

Fig. 3. Models of our authentication mechanism.

Table 1
Classifiers summary.

Types of classifier	Weka version	Notation
BayesNET	BayesNET	BN
NaiveBayes	NaiveBayes	NB
Support Vector Machine	SMO	SVM
K Nearest Neighbor	IB1 (KNN with K=1)	IB1
Decision Tree	J48	J48
Decision Tree	Random Forest	RF

3.3.5. Touchscreen

Touchscreens are the input devices designed for providing users an interface to interact with the device. Some touchscreens are single touch and some of them are multi-touch. Finger acts as an input tool to interact with them. We chose Android device for this implementation because of its market supremacy [1] and its tremendous popularity in smartphone users. Android supports a variety of touchscreens and touchpad.[4] We can determine the starting and ending point of a pointer, the direction of the finger movement (in `x` and `y` coordinates), and its velocity as it moves across the touchscreen, by using android MotionEvent [63] class. An object of this class is used to report the performed event. Whenever some action is performed on the screen, a touch-event is reported with a specific action code along-with touched area (`xy` coordinates), pressure and size and orientation of the touched area, etc. The action code represents the state of the touch action, e.g., `Action_Down` represents the start of a touch action while `Action_Up` represents the end of a touch action. As the name suggests, Android `VelocityTracker` class tracks down the motion of the pointer on the touchscreen. User usually drags or slides the lock button for unlocking their smartphones. The sliding or dragging of the lock button is done in horizontal direction. We use `getXVelocity()` method to compute the magnitude of finger movement on touchscreen of the smartphone, and later this measured velocity magnitude is taken into account. By definition, we can say that sliding velocity is the ratio of the total distance covered by this dragging and the time taken to do so.

3.4. Considered classifiers

3.4.1. Base classifiers

In order to check the quality of extracted features and to suggest the best classifying algorithm for user authentication in smartphones, we have tested our dataset with *six* conceptually different classification techniques (see Table 1). Interested readers are referred to this book [65] for better understanding of their working. Some of these classifiers are reported to be among the top 10 machine learning algorithms [39–41], as such they have been used extensively for smartphone user authentication. Random Forest classifier also proved itself as the best classifier due to its simplicity, robustness to overfitting and quicker learning [41], hence it has been widely tested for smartphone user authentication in recent studies [33,42,43,49]. All these classifiers have been applied for testing our dataset, using an open source, portable, GUI-based Weka workbench. We have used all the classifiers in their default settings, because we are more interested to investigate the role of feature space and efficacy of classifiers without even applying any optimization technique. It is reasonable to assume that our initial results can be further improved by fine tuning the classifiers' parameters.

3.4.2. Combining base and ASC's - vote classifier

Multiple techniques of combining classifiers have been widely tested and evaluated in pattern recognition over the years. Such combination methods may lead to a significant reduction in the error rates. Additional advantage of combining classifiers is the robustness of the system against the possible problems that each individual classifiers may have observed on the dataset. We have used Weka *Vote* classifier to fuse the outcomes of our chosen classifiers, in two different settings, i.e., we have combined our base classifiers using *average probability rule* and ASC's using *product of probability rule*.

3.5. Success metric

In this work, we use the following measures to compute our error rates.

- *True Acceptance Rate (TAR)*: The fraction describing the ratio of successful login attempts to all the attempts made by the legitimate user.
- *False Acceptance Rate (FAR)*: The fraction describing the ratio of successful login attempts to all the attempts by an adversary.
- *False Rejection Rate (FRR)*: The fraction describing the ratio of unsuccessful login attempts to all the attempts by the legitimate user.
- *True Rejection Rate (TRR)*: The fraction describing the ratio of unsuccessful login attempts to all the attempts by an adversary.
- *Accuracy*: It is the ratio of correct decision to all the decisions.

$$Accuracy = \frac{TAR + TRR}{TAR + FAR + FRR + TRR}. \tag{2}$$

- *Receiver Operating Characteristics (ROC)*: Recognition results can be elaborated through ROC curve; plotting TAR against FAR. Usually the values of FAR are plotted on the horizontal axis with TAR on vertical axis.

4. Experimental analysis

This section explains how we evaluated AnswerAuth.

4.1. Data collection

We developed a customized Android application, namely, *AuthCollector*, which can be installed on any Android smartphone starting from Android version 4.0.4. As *SENSOR_DELAY_GAME* (50 samples/sec) was found more accurate for authentication in recent studies [33,55], we decided to use this delay for data collection purposes.

We outsourced the experiment and shared the *AuthCollector* implementation with a crowd-sourcing platform - Ubertersters[5] to test the application. Ubertesters through their "Hire Testers" service allows access to their crowd of professional testers globally to test the application in real life conditions and on real devices. We paid € 25 per hour, as compensation, for each participant. Ubertersters recruited 100 users, in total, but some of them were disqualified (15) because of (i) the non-availability of the required sensor(s), (ii) their patterns have more Not A Number (NaNs) values and (iii) the users had less than 30 observations in an activity, etc. We setup a web page with the complete explanation of the experiment and its potential outcome and details of *AuthCollector* application, i.e., the user consent, the questionnaire to collect demographics data, and the procedure to install/uninstall the application. Participants were requested to answer to the demographic questions and install the application, provide *slide-to-unlock* and *phone-pickup-movement* samples in different activities (3 activities, i.e., *sitting, standing,and walking*) and keep the application running for at least 3 days. *AuthCollector* required 3 sessions

[4] http://source.android.com/devices/tech/input/touch-devices.html.

[5] https://ubertesters.com/.

Table 2
User demographics (M = Male, F = Female, U = Undisclosed, R = Right, L = Left, B = Both).

Information	Description
No. of Users	85
Sample Size	11,200
Devices	Android Smartphones with `4.4.x` version
No. of Sessions	3
Unsupervised Conditions	Yes
Gender	55(m), 30(f)
Handedness	70(R), 8(L) & - 7(B)
Age Groups	80 (20 – 35), 5 (36 – 60)

Table 3
List of selected features from sensory readings.

No.	Lift Features			
1–4	MeanX	MeanY	MeanZ	MeanM
5–8	STDX	STDY	STDZ	STDM
9–12	SkewX	SkewY	SkewZ	SkewM
13–16	KurtX	KurtY	KurtZ	KurtM

Table 4
List of selected features from sensory readings.

No.	Touch Features				
1–5	Event_Duration	InitialX	InitialY	EndX	EndY
6–10	VX_min	VX_max	VY_min	VY_max	VX_avg
11–15	VY_avg	VX_std	VY_std	VX_var	VY_var
16–20	AX_min	AX_max	AX_avg	AX_std	AX_var
21–25	AY_min	AY_max	AY_avg	AY_std	AY_var
26–30	P_min	P_max	P_avg	P_std	P_var

in 3 days, from the users to complete the experiment. It required 30 min users' interaction on the first day and 15 min on each of the following two days. In this way, we collected 20 samples from each user in each activity on the first day and 10 samples in each activity on the two subsequent days. In this way, each participant had to test the application for 1 hour. Since the experiment had to be performed in the wild and within an hour, we had to fix the number of behavioral samples to be obtained. Behavioral biometric matchers require fair number of training samples to provide higher accuracy, however, asking for two many samples could have annoyed the users and also they might not have completed the experiment within due course of time. As such, we decided to collect 40 behavioral patterns per activity per user. In total, we collected 120 samples from each user (in total 10,200), in our 3-days long experiment. We embedded the demographic questionnaire (see Appendix A) with the application and in order to participate in the experiment, the participants had to answer those demographic related questions and install the application. Table 2 summaries the collected demographic information.

4.2. Feature extraction

The most critical part of designing any authentication mechanism is the selection of appropriate features - that provide most relevant information to model user's behavior, but not at the computational cost. Statistical time domain features have shown to be very productive for modeling the users' behaviors in previous studies [33,55]. Additionally, these chosen time domain features are computationally cheaper as compared to their frequency domain counterparts (due to the expensive Fourier transformation).

We collected 4 data streams from every 3-dimensional sensor. We extracted 4 statistical features, namely, Mean, Standard Deviation (STD), Skewness, and Kurtosis, from each of the data stream. We define below our extracted features:

- *Mean:* Average value of the sensor dimension.
- *Standard Deviation:* Standard Deviation of the sensor dimension.
- *Skewness:* Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point[6].
- *Kurtosis:* Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers.[7]

We extracted 16 features from each sensor to form a feature vector (see Table 3). AnswerAuth leverages 6 sensory readings and 16 features per sensor makes 96 feature long vector. We also added time offset as a feature to this feature vector and formed a final feature vector of 97 features, to model the *phone-pickup-movement* behavior. Similarly, we extracted 31 touch-based *slide-to-unlock* features to form a final feature vector for this behavior (see Table 4). So, the final feature vector is the horizontal concatenation of the two behaviors and is 128 features long.

4.3. Feature concatenation

The fusion of data as early as possible may increase the recognition accuracy of the system [60]. However, the fusion of data at sensor level may not yield better results (as compared to the fusion at other levels) because of the presence of noise during data acquisition. As such, the fusion at feature level is expected to provide better results, because the feature representation communicates much more relevant information. The extracted feature set from the data through multiple sources can be combined together to form a new feature set. We have fused the extracted features from our data sources (the accelerometer, the gyroscope sensor, etc., and the touchscreen) at feature level in order to provide maximum relevant information to the authentication system.

4.4. Feature subset selection

Feature subset selection is the method of selecting a subset of relevant features to be used in the model construction. Basic applications of these techniques are in domains where there are many features and comparatively less observations. These methods are useful because they point out the important features that could possibly lead towards improvement in classification accuracy.

A feature subset selection method also involves a search technique for proposing new feature subset, and the evaluation method which provides scores of different subsets. The subset with highest score is likely to be picked and further used for classification. The details of feature selection techniques and their performance comparison is discussed in [59].

In order to further improve the performance of our base classifiers and ensemble *Vote* classifiers, we later evaluated our dataset with correlation-based feature selection (*CFS*) feature selection scheme (see Fig. 3). We have used a meta customized Attribute Selected Classifier (ASC) to perform in parallel both features selection and automatic classification.

4.5. Analysis

We have used Weka experimenter workbench for the classification of our dataset. Every class has to be compared with every other class present in the dataset. Due to the limited number of observations (40 per user), we have performed a 10-fold

[6] https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm.

[7] https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm.

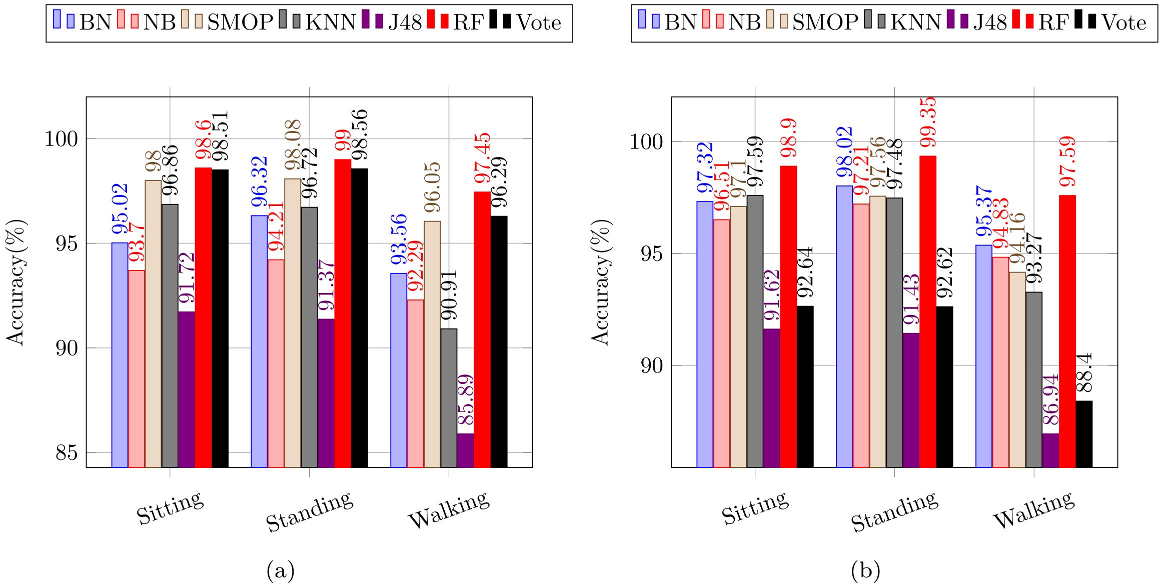


Fig. 4. Accuracy of all the base (Fig. 4a) and ASC (Fig. 4b) classifiers, in all the activities, averaged over 85 users.

cross-validation with 10 runs for training and testing the classifiers. Cross-validation looks justified as each available sample is tested. Alternately, the train/test method could test only the testing samples. Since we modeled this problem as the binary class classification problem, the training was performed on some samples of an owner (+ve) and impostor (-ve) and tested with all the remaining owner samples and impostor samples from all the users (one by one). More technically, one user was taken as the genuine user and her samples were compared with all the sample of all the users. Reported values are the average results obtained for all the users.

5. Results

Our classifiers have solved the problem of verifying a user, that is 1:1 matching between a query sample of an unknown person and the person's pre-stored biometric template. The results for the base and ASC's and vote in terms of accuracy and TAR can be seen in Fig. 4 and Fig. 5, respectively. The performance of all the classifiers clearly indicate that all the subjects are being recognized with a high probability confirmed by their TAR.

5.1. Authentication using base classifiers

We have tested our dataset with our chosen base classifiers (see Table 1), as per our proposed model (see Fig. 3a). Surprisingly, every base classifier performed well and provided acceptable authentication results possibly because of the productive feature space. RF classifiers is the best with providing 98.87%, 98.98%, and 96.8% TAR in *sitting, standing*, and *walking* activities, respectively. These initial results confirm that the features we selected were meaningful and there is a room for further improvement in accuracy. Initial authentication results of these base classifiers in terms of accuracy and TAR are illustrated in Figs. 4 and 5, respectively. From both the figures, it is evident that RF classifier outperformed all the other chosen classifiers on both full (using base classifiers) and reduced features (ASC settings).

5.2. Authentication: vote classifier combining base classifiers

We have used the *average probability combination rule* for classifier fusion. Each of the base classifier used under vote classifier had to predict the average probability for each class label for every test sample. This rule returns the mean of the probability distribution for each of the base classifiers learned within vote classifier. The class with highest probability is chosen as the decided class. Our model for classifier fusion can be seen in Fig. 3b.

The results of vote classifier (fusion of base classifier) are depicted in Figs. 4 and 5. We achieved 98.47%, 98.61%, and 95.25%, TAR in *sitting, standing*, and *walking* activities, respectively. Similarly the achieved accuracy (using Eq. 4) in these states is 98.51%, 98.56%, and 96.25%, respectively.

5.3. Authentication: AttributeSelectedClassifier with cfsEval and bi-directional search method

Feature selection algorithms are meant to automatically search for the best subset of features in a provided dataset. The notion of "best" means the subset that may provide the highest classification accuracy. The main idea behind the feature selection is to identify the best or good enough feature combination that may lead to the improvement in classification performance.

Machine learning experts, Weka designers [64] in particular, do not recommend to apply attribute selection (especially supervised attribute selection) on all datasets and then run an evaluation (such as cross-validation) on the dimensionally reduced data. This approach will yield overly optimistic error rates because the attribute selection process has seen data from the test folds as well. However, the same can be done through the ASC. The justified way of applying feature selection with a classifier is to wrap the attribute selection process with the classifier itself. This is achieved with Weka meta classifier, the AttributeSelectedClassifier. This technique requires an attribute selection method and a base classifier for its operation. It is worth noting that both attribute selection method and the base classifiers have access only to the training data or folds during cross validation.

Since we have used six base classifiers we have one ASC for each of these classifiers. For example our first ASC uses BayesNET as its classifier and *CFSEval* as attribute selection method and so on. Our attribute evaluation -*CfsEval*, is the same for all the ASC's along-with *BestFirst (bi-directional)* search method. The reason behind the use of this configuration is to speed up the classification process through ASC's and its performance of this configuration is comparable with other evaluation methods.

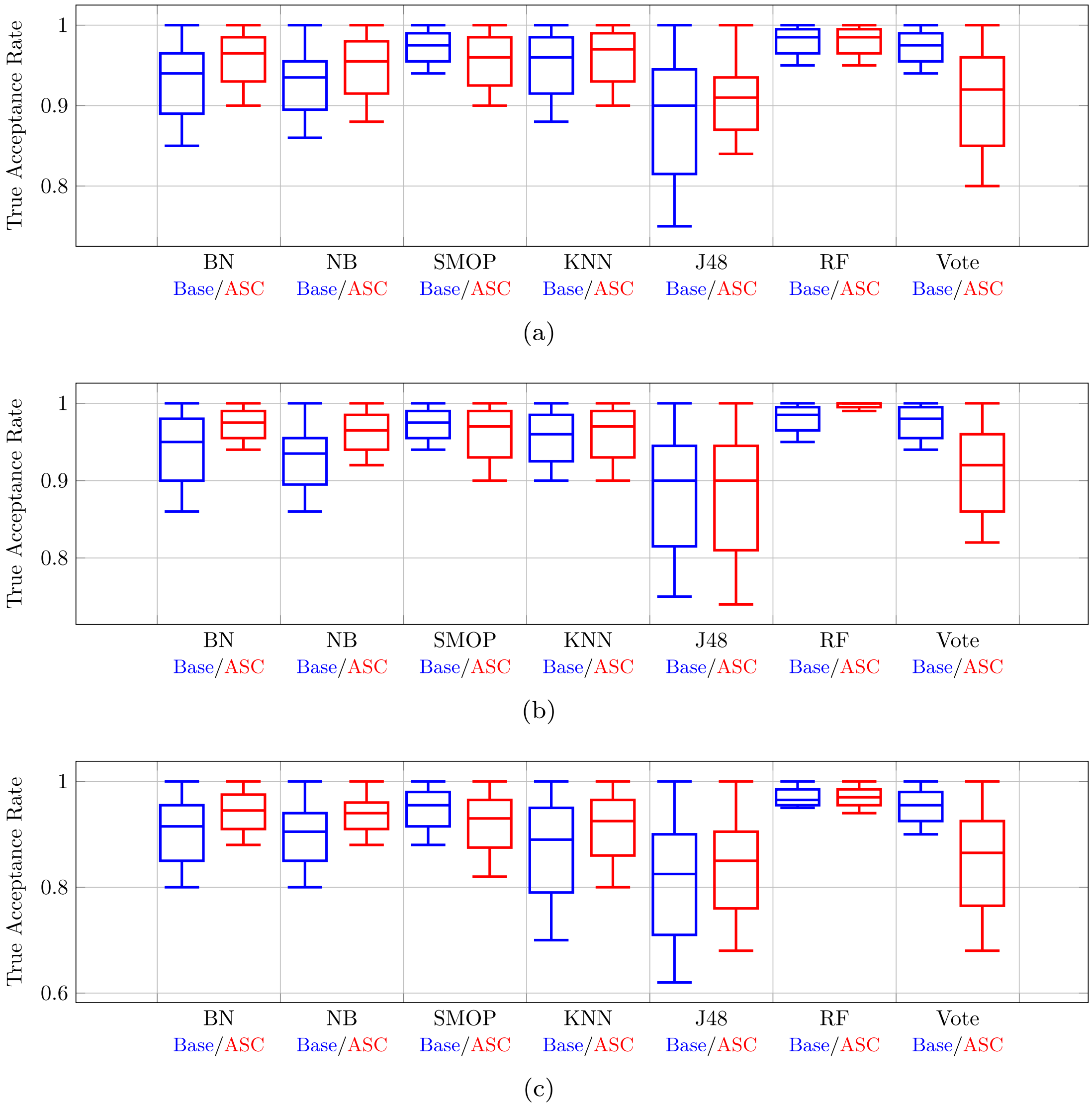


Fig. 5. Comparison of TAR for different Base and ASC classifiers averaged over 85 users in *sitting* (Fig. 5a), *standing* (Fig. 5b), and *walking* (Fig. 5c) activity.

Authentication is performed as shown in Fig. 3c and the results of these ASC's in terms of overall accuracy and TAR are summarized in Figs. 4 and 5, respectively. We show these results in the same figures with other classifiers in order to decrease the space and increase the readability. It is evident that the performance of each base classifier improved after their use as ASC classifier over reduced features. RF ASC classifier also achieved slight improvement, i.e., 99.03%, 99.35%, and 97.0% TAR as compared to Base RF of 98.87%, 98.98%, and 96.8%, in *sitting, standing*, and *walking* activities, respectively.

5.4. *Authentication:* vote *classifier combining* AttributeSelectedClassifier with *cfsEval*

As already stated, the reason behind the fusion of outputs of different classifiers was to increase the TAR as compared to the best individual classifier (RF is best under attribute selection method). We report a TAR of our vote classifier (with *product of probability combination rule*) of 88.4% as shown in Fig. 4.

Our classifiers have solved the problem of verifying a user i.e., 1:1 matching between a query sample of an unknown person and the person's pre-stored biometric template. The results in terms of their accuracy, averaged TAR and individual user's TAR are shown in Figs. 4, 5 and B.8, respectively. We have also fused the outputs of the base classifiers using vote classifiers and their predicted authentication results are also shown in Figs. 4, 5 and B.8.

In most of the biometric authentication systems, the performance is measured through recognition rate. This recognition rate can further be specified by two values namely FRR and FAR. There is always a trade off between these two parameters and people chose their acceptable values based on their applications. For an example 0.001% FAR is selected for fingerprint in military and border crossing applications (which often results in higher FRR).

The classification results can be optimized by selecting certain FAR, which might be acceptable for the scenario, in which this classification is being performed. For an example, in any authentication mechanism if there are more false rejections, the classifier performance can be optimized by fine tuning, by accepting a higher FAR. The trade off between FAR and FRR for our experiments is presented in Fig. 6. It is evident that in both Base and ASC settings, RF classifier is found extremely accurate.

We also show the results of the best performing classifier, i.e., RF classifier in terms of ROC curves (see Fig. 6). We show an average ROC of all the users obtained through Vertical Averaging(VA)

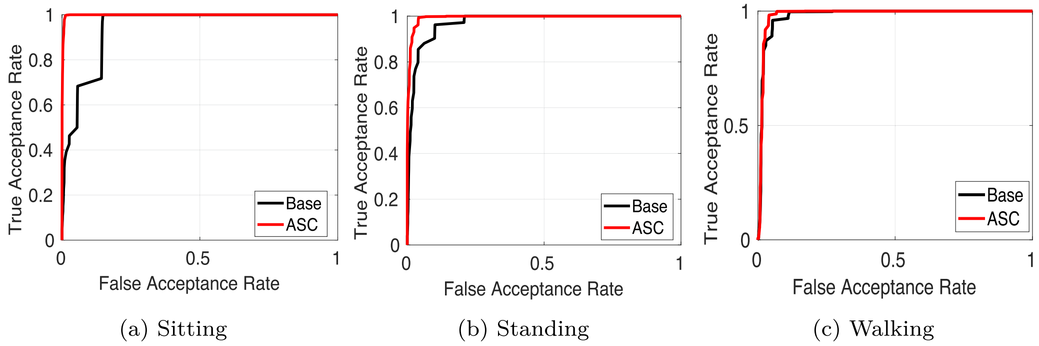


Fig. 6. Comparison of ROC curves for RF classifier, i.e, in Base and ASC settings.

[14]. In this scheme, the averages of the TAR rates are plotted against the researcher-defined fixed FAR. RF classifier is found consistent in all the activities and in each setting, i.e., Base and ASC settings and outperformed all the classifiers because it can reduce the variance and its ability to counter overfitting. Other classifiers were less accurate for several reasons, e.g., because the number of training samples was less or they required Gaussian distributed data which might not be true for the dataset.

All the classifiers performed well on our collected dataset, especially, RF classifier is the top ranked classifier on both base and ASC settings (see Figure. B.8 attached as Appendix B). Since, this classifier worked well, we show the distribution of its decision for every user. It is evident that its ASC version achieved higher accuracy than its base counterparts. We observed that the RF classifier in either case may be the best choice, considering its accuracy, the time required for decision making, and countering the effect of overfitting.[8]. We will prototype our proof-of-the-concept app using this classifier and with the selected ASC features.

6. Proof-of-the-concept application

We developed the proof-of-the-concept final prototype of AnswerAuth based upon our findings. Final prototype of AnswerAuth leverages RF as the classifier. AnswerAuth can be installed on any Android phone running Android `4.4.4` version or higher. AnswerAuth requires minimal configuration, i.e., the user may select both the modalities or any one of them. AnswerAuth also allows the users to decide by themselves the number of training samples, i.e., how many times users would perform swiping and picking up gestures to train the RF classifier. In any case, the user is assisted by AnswerAuth by displaying the suggestion. The same process needs to be performed later for authentication purposes.

7. Security analysis

We claim AnswerAuth as extremely secure because it depends on multiple hidden features generated from person-specific *swiping* and *phone-pickup* movements. The same claim is proved from our conducted experiments. We explain below the evaluation and the obtained results:

7.1. Evaluation

We recruited 6 more testers to assess the robustness of AnswerAuth and performed the additional experiments. We explained our testers the complete experiment, the modalities, and the purpose of the experiment (to impersonate the user behavior). We installed our proof-of-the-concept prototype application on HUWAEI GRA-L09 smartphone. For the random attacks, we relied on our previously connected dataset, however, for the mimic attack, we took one tester as the legitimate user and trained the RF classifier on his 30 training samples, and the remaining testers act as the would-be impostors for the time being. The training process is performed in front of the would-be impostors with the intention that they would learn the *swiping* and *phone-pickup* gesture and effectively attempt to mimicking the trained behavior. The process is repeated till the behavior of each tester is attempted to be mimicked. RF classifier checks for the similarity between the training samples and each of the incoming adversarial attempt and shows, as a toast, the binary outcome: authenticated or rejected and we saved the outcome as FAR and TRR.

7.2. Random attacks

AnswerAuth is not so easy to be successfully attacked by a random attacker, as it relies on very private person-specific behavior; velocity and acceleration related features from *swiping* and motion-based features from *phone-pickup* actions. These movements have shown be secure and extremely difficult [5], if not impossible, hence it would be extremely difficult to spoof this behavior, especially by a random attacker. The same is actually evident from our conducted experiments. Our conducted experiments involved a Zero-effort or random attack where we compared the samples of each of the valid user with all the remaining random attackers. It is worth recalling that RF classifier performed well in this scenario; as we obtained as high as $\geq 99\%$ TAR, on reduced features.

7.3. Mimic attacks

In the mimic attack scenario, we picked a valid user for the training. Each valid user trained the smartphone by performing *swiping* and *phone-pickup* actions 10 times in *sitting, standing*, and *walking* positions each, respectively, in front of the attackers and later they had to spoof the legit user's behavior. The would-be attackers were later asked to spoof the legit user's behavior and try to break the login process in 30 attempts in any of their preferred activity. In this way, each trained behavior was attempted to be spoofed in 150 attempts. It is worth mentioning that some of the would-be attackers also required more demonstration and practice and hence were given as much time for practice, as desired. In total, we collected 180 valid user training samples and 900 spoof attempts in this scenario. Just one (1) out of 900 attempts went successful and the behavior got spoofed. Obtained result prove AnswerAuth as a robust authentication mechanism.

[8] https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.

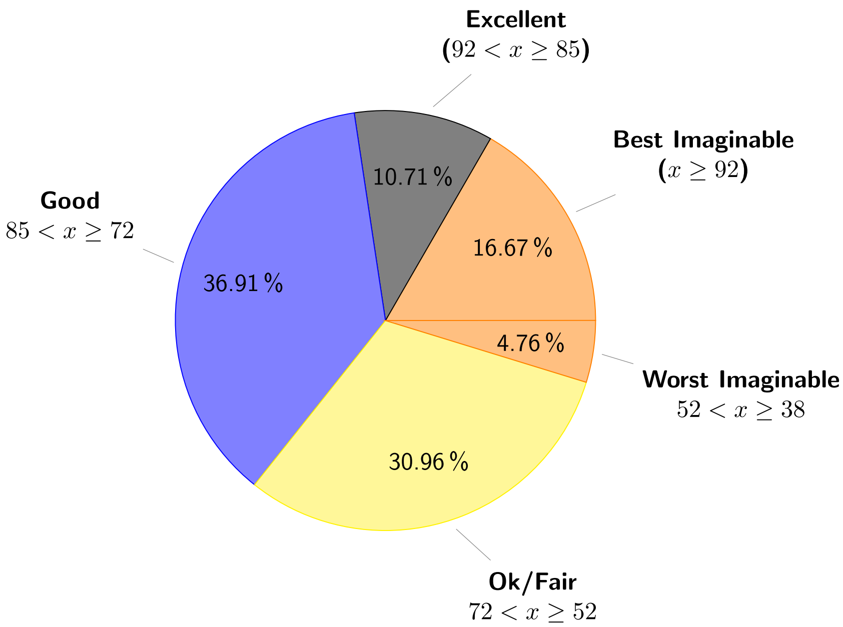


Fig. 7. Break-up of the obtained SUS score.

8. Usability analysis

We are of the opinion that AnswerAuth would be widely accepted because it leverages the two very common user behaviors and every Android smartphone user is assumed to be familiar with the swiping and pick up gestures. We applied the Software Usability Scale[9] (SUS) scale to gather the participants' reviews and assess the usability of AnswerAuth. SUS is a 10-questions based questionnaire and is widely used to evaluate the usability of authentication scheme [52,53,55]. It is worth-mentioning that we replaced the word "System" with "AnswerAuth".

SUS requires the users to record their response on a given 5-point scale ranging from "Strongly Disagree" to "Strongly Agree". The users' impression is converted into a SUS score between 0 and 100. The obtained SUS score, x, is categorized as follows: (i) Best Imaginable ($x \geq 92$), (ii) Excellent ($92 < x \geq 85$), (iii) Good ($85 < x \geq 72$), (iv) OK ($72 < x \geq 52$), (v) Poor ($52 < x \geq 38$), and (vi) Worst Imaginable ($38 < x \geq 25$).

We show the breakup of our obtained SUS score in Fig. 7. More than 64% testers considered AnswerAuth as either "Best Imaginable ($\approx 17\%$)", "Excellent ($\approx 11\%$)" or "Good ($\approx 37\%$)". AnswerAuth achieves an overall mean score of 75.11(standard average score is 68 [61]). From the SUS grading key, the score ≥ 72 reveals high probability of wide user acceptance.

9. Conclusions & future work

We have proposed a fully transparent bi-modal behavioral biometric-based solution - AnswerAuth for smartphone user authentication. Our approach can independently be used for user authentication as well as in conjunction with existing authentication mechanisms, on smartphones. We implemented and tested the system with a dataset comprising of 10,200 patterns (120 from each sensor) from 85 users in three common user activities, i.e., *sitting*, *standing* and *walking*. We have classified these patterns of different users by *six* conceptually different classifiers. Experiments show that the classifiers we chose and the features we extracted are good enough to accurately identify a valid user. We have obtained an accuracy as high as 98.98% using RF as classifier (without any feature subset selection) and as high as 99.35% over reduced features.

In order to further improve the authentication results, we fused the outputs of listed base and ASC's through *Vote* classifier and achieved an accuracy as high as 98.56% on full features in *standing* activity using the *average probability* and 92.64%, using the *product of probability*, combination rules, respectively.

The authentication results for all the classifiers are acceptable, which indicates the effectiveness of our features set. However, we utilized RF as the classifier in our final proof-of-the-concept Application.

We have developed a proof-of-the-concept Android app based on our findings in this work. We evaluate our prototype in terms of security (to explore the robustness against the potential attacks) and usability (to check its usefulness from the users' perspective). Initial obtained results indicate positive security and usability evaluation.

As a future work, we are going to address the problem of fast and seamless detection of the users' current activity [55] as some papers show that the performance of the system varies in different activities. In this way, we would be able to better determine the environmental context. Additionally, we are also going to solve the AnswerAuth training problem by collecting transparently these actions while the user swipes and brings the smartphone to her ear while getting engaged in a call. This way after reaching the best number (e.g., 40) samples, AnswerAuth would notify the user about the availability of the modality for onward authentication. Additionally, we will investigate the impact of combining this mechanism with other biometric modalities like voice, or gait, for example, on its security and usability.

Acknowledgment

Authors would like to thank all the volunteers for their participation in the experiments and groupmates for valuable and insightful thoughts. A special thanks goes to Mr. Filippo Del Frari, Sandeep Gupta, and Getachew Demissie Abebe, for their support in implementation and data collection process.

[9] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

The work was partially supported by the EIT Digital project: Android App Reputation Service (ARTS), by the European Training Network for CyberSecurity (NeCS) grant number 675320 and by the project XProbes funded by the Provincia Autonoma di Trento.

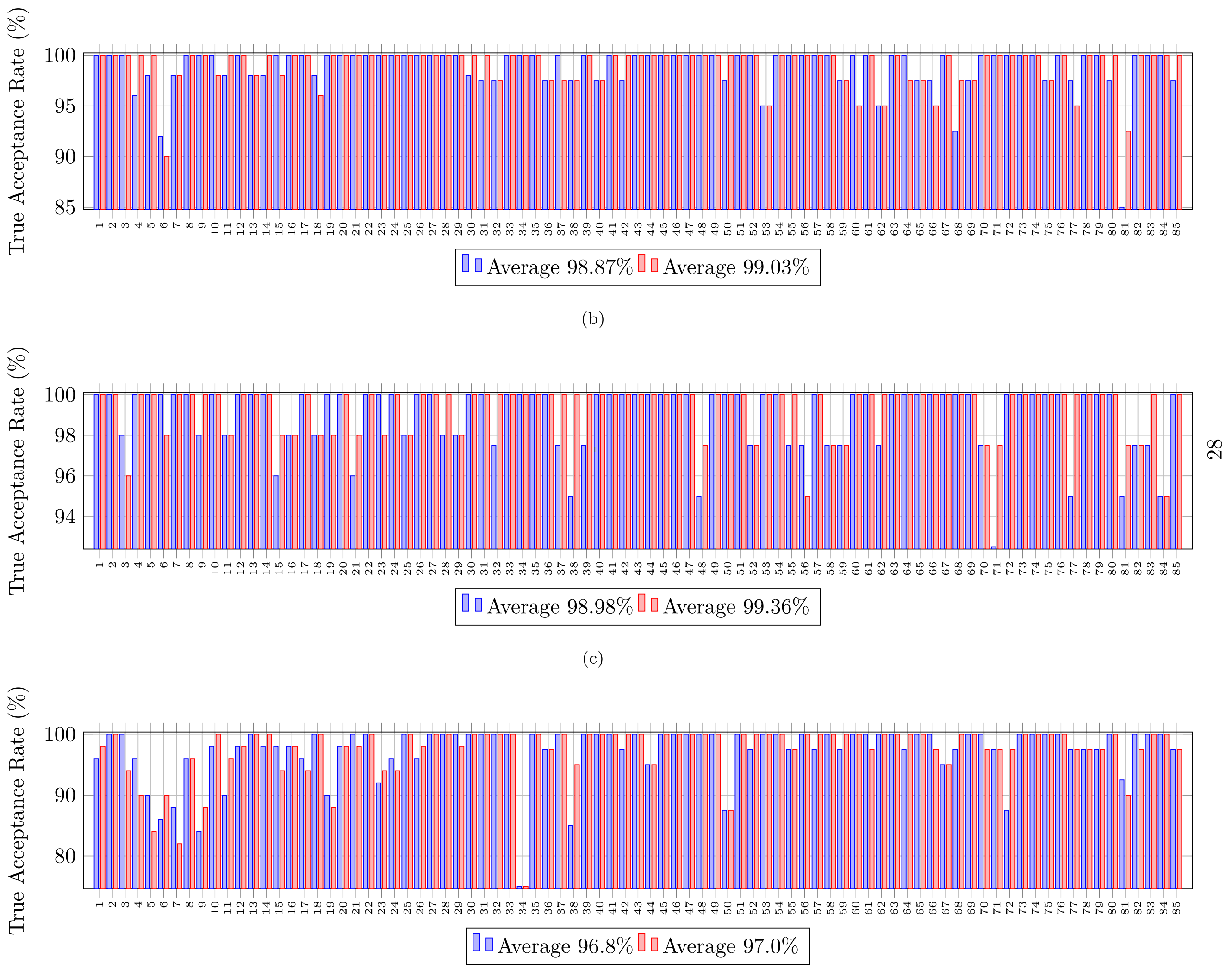
Mauro Conti is supported by a Marie Curie Fellowship funded by the European Commission (agreement PCIG11-GA-2012-321980). This work is also partially supported by the EU TagItSmart! Project (agreement H2020-ICT30-2015-688061), the EU-India REACH Project (agreement ICI+/2014/342-896), by the project CNR-MOST/Taiwan 2016-17 "Verifiable Data Structure Streaming", the grant n. 2017–166478 (3696) from Cisco University Research Program Fund and Silicon Valley Community Foundation, and by the grant "Scalable IoT Management and Key security aspects in 5G systems" from Intel.

Appendix A. demographic questionnaire

- What is your gender?
  1. Male
  2. Female
  3. I don't want to disclose
- How old you are?
  1. ≤ than 20 years.
  2. > 20 years and ≤ 35 years.
  3. > 35 years and ≤ 60 years.
  4. > than 60 years.
  5. I don't want to disclose
- Which hand(s) do you use for interacting with your smartphone?
  1. Right
  2. Left
  3. Both
  4. I don't want to disclose

Appendix B. TAR comparison for base and ASC RF classifier for individual users

(a)



(b)

(c)

Fig. B.8. TAR comparison of base RF and asc RF classifier in *sitting* (Fig. B.8a), *standing* (Fig. B.8b) and *walking* (Fig. B.8c) position for individual user.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jisa.2018.11.008.

References

[1] Smartphone, Smartphone OS market share, Q1 2015. 2015. Available at http://www.idc.com/prodserv/smartphone-os-market-share.jsp.

[2] Morris R, Thompson k. Password security: a case history. Commun ACM 1979;22(11):594–7.

[3] Chiang HY, Chiasson S. Improving user authentication on mobile devices: a touchscreen graphical password. In: Proceedings of the 15th ACM international conference on Human-computer interaction with mobile devices and services; 2013. p. 251–60.

[4] Jain AK, Ross A, Pankanti S. Biometrics: a tool for information security. IEEE Trans Inf Forensics Secur 2006a;1(1):125–43.

[5] Buriro A. Behavioral biometrics for smartphone user authentication. University of Trento; 2017. Ph.d. thesis.

[6] Akkermans AHM, Kevenaar TAM, Schobben DWE. Acoustic ear recognition for person identification. In: Proceedings of the fourth IEEE workshop on automatic identification advanced technologies; 2005. p. 219–223205.

[7] Schaub F, Deyhle R, Weber M. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In: Proceedings of the 11th international conference on mobile and ubiquitous multimedia; 2012. p. 56–66.

[8] Tari F, Ozok A, Holden SH. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In: Proceedings of the second symposium on Usable privacy and security; 2012.

[9] Chiasson S, van Oorschot PC, Biddle R. A usability study and critique of two password managers.. In: Proceedings of the Usenix security; 2012.

[10] Jakobsson M, Shi E, Golle P, Chow R. Implicit authentication for mobile devices.in: Proceedings of the 4th USENIX conference on Hot topics in security, 2009.

[11] Survey, Survey says 70% password don't protect mobiles: download free mobile toolkit. 2014. Available at http://nakedsecurity.sophos.com/2011/08/09/free-sophos-mobile-security-toolkit.

[12] Aviv AJ, Gibson K, Mossop E, Blaze M, Smith JM. Smudge attacks on smartphone touchscreens. In: Proceedings of the WOOT; 2009.

[13] Jain AK, Bolle R, Pankanti S. Biometrics: personal identification in networked society. Springer; 1999.

[14] Fawcett T. ROC Graphs: notes and practical considerations for researchers. Mach Learning 311 2004:1–38.

[15] Conti M, Zachia-Zlatea I, Crispo B. Mind how you answer me!: transparently authenticating the user of a smartphone when answering or placing a call. In: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security; 2011. p. 249–59.

[16] Wu G, Wang J, Zhang Y, Jiang S. A continuous identity authentication scheme based on physiological and behavioral characteristics. Sensors 2018;18(1):179.

[17] Anjum A., Ilyas M.U. Activity recognition using smartphone sensors. in: Proceedings of the IEEE, consumer communications and networking conference (CCNC)2013b;:914–919.

[18] Mantyjarvi J, Lindholm M, Vildjiounaite E, Makela S-M, Ailisto HA. Identifying users of portable devices from gait pattern with accelerometers. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP'05). Ii–973

[19] Bergadano F, Gunetti D, Picardi C. User authentication through keystroke dynamics. ACM Trans Inf Syst Secur (TISSEC) 2002b;5(4):367–97.

[20] Alzubaidi A, Kalita J. Authentication of smartphone users using behavioral biometrics. IEEE Commun Surv Tutor 2016;18(3):1998–2026.

[21] Patel VM, Chellappa R, Chandra D, Barbello B. Continuous user authentication on mobile devices: recent progress and remaining challenges. IEEE Signal Process Mag 2016;33(4):49–61.

[22] Eberz S, Rasmussen KB, Lenders V, Martinovic I. Evaluating behavioral biometrics for continuous authentication: challenges and metrics. In: Proceedings of the ACM on asia conference on computer and communications security; 2017. p. 386–99.

[23] Clarke NL, Furnell SM. Authenticating mobile phone users using keystroke analysis. Int J Inf Secur 2007b;6(1):1–14.

[24] De Luca A, Hang A, Brudy F, Lindner C, Hussmann H. Touch me once and i know it's you!: implicit authentication based on touchscreen patterns. In: Proceedings of the SIGCHI conference on human factors in computing systems; 2012. p. 987–96.

[25] Zheng N., Bai K., Huang H., Wang H. You are how you touch: User verification on smartphones via tapping behaviors. 2012. Technical Report, WM-CS-2012-06.

[26] De Luca A, Hang A, Brudy F, Lindner C, Hussmann H. A novel non-intrusive user authentication method based on touchscreen of smartphones. In: Proceedings of the International symposium on biometrics and security technologies (ISBAST); 2013. p. 212–16.

[27] Serwadda A, Phoha VV. When kids' toys breach mobile phone security. In: Proceedings of the ACM SIGSAC conference on computer & communications security; 2013. p. 599–610.

[28] De Luca A, Hang A, Brudy F, Lindner C, Hussmann H. Gait-based recognition of humans using continuous HMMs. In: Proceedings on Fifth IEEE international conference on automatic face and gesture recognition; 2002. p. 336–41.

[29] Li L, Zhao X, Xue G. Unobservable re-authentication for smartphones. In: Proceedings of the NDSS; 2013.

[30] Shi W, Yang J, Jiang Y, Yang F, Xiong Y. Senguard: passive user identification on smartphones using multiple sensors. In: Proceedings of the IEEE 7th international conference on, wireless and mobile computing, networking and communications (WiMob); 2011. p. 141–8.

[31] Giuffrida C, Majdanik K, Conti M, Bos H. I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics. In: Proceedings of the International conference on detection of intrusions and Malware, and vulnerability assessment; 2014. p. 92–111.

[32] Zhu J, Wu P, Wang X, Zhang J. Sensec: mobile security through passive sensing. In: Proceedings of the international conference on computing, networking and communications (ICNC); 2013. p. 1128–33.

[33] 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 Workshops; 2015. p. 27–34.

[34] Lai K, Konrad J, Ishwar P. Towards gesture-based user authentication. In: Proceedings of the IEEE ninth international conference on advanced video and signal-based surveillance (AVSS); 2012. p. 282–7.

[35] Angulo J, Wästlund E. Exploring touch-screen biometrics for user identification on smart phones. Priv Ident Manag Life 2012:130–43.

[36] Sae-Bae N, Memon N, Isbister K, Ahmed K. Multitouch gesture-based authentication. IEEE Trans Inf Forens Secur 2014:568–82.

[37] Shahzad M, Liu AX, Samuel A. Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it. In: Proceedings of the 19th annual international conference on Mobile computing & networking; 2013. p. 39–50.

[38] Sun J, Zhang R, Zhang J, Zhang Y. Touchin: Sightless two-factor authentication on multi-touch mobile devices. In: Communications and Network Security (CNS), 2014 IEEE Conference on. IEEE; 2014. p. 436–44.

[39] A tour of the top 10 algorithms for machine learning newbies. Available at https://towardsdatascience.com/a-tour-of-the-top-10-algorithms-for-machine-learning-newbiesdde4edffae11.

[40] The 10 algorithms machine learning engineers need to know. b. Available at https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html.

[41] Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res 2014;15(1):3133–81.

[42] Lee W-H, Lee RB. Implicit smartphone user authentication with sensors and contextual machine learning. In: Proceedings of the 47th Annual IEEE/IFIP international conference on dependable systems and networks (DSN); 2017. p. 297–308.

[43] Singha T.B., Nath R.K., Narsimhadhan A.V. Person recognition using smartphones' accelerometer data. 2017. arXiv:1711.04689.

[44] Meng W, Wong DS, Furnell S, Zhou J. Surveying the development of biometric user authentication on mobile phones. IEEE Commun Surv Tutor 2015;17(3):1268–93.

[45] Kumar R, Phoha VV, Serwadda A. Continuous authentication of smartphone users by fusing typing, swiping, and phone movement patterns. In: Proceedings of the IEEE 8th international conference on biometrics theory, applications and systems (BTAS), 17(3); 2016. p. 1–8.

[46] Xu H, Zhou Y, Lyu MR. Towards continuous and passive authentication via touch biometrics: an experimental study on smartphones. In: Proceedings of the symposium on usable privacy and security(SOUPS), 14; 2014. p. 187–98.

[47] Meng Y, Wong DS, Schlegel R. Touch gestures based biometric authentication scheme for touchscreen mobile phones. In: Proceeding of the international conference on information security and cryptology; 2014. p. 331–50. Others

[48] Zheng N, Bai K, Huang H, Wang H. You are how you touch: user verification on smartphones via tapping behaviors. In: Proceedings of the international conference on network protocols (ICNP); 2014. p. 221–32.

[49] Buriro A, Crispo B, Zhauniarovich Y. Please hold on: unobtrusive user authentication using smartphone's built-in sensors. In: Proceedings of the IEEE international conference on identity, security and behavior analysis (ISBA-2017); 2017a. p. 1–8.

[50] Akhtar Z, Buriro A, Crispo B, Falk TH. Multimodal smartphone user authentication using touchstroke, phone-movement and face patterns. In: Proceeding of the 5th IEEE global conference on signal and information processing (GlobalSIP-2017); 2017. p. 1368–72.

[51] Primo A, Phoha VV, Kumar R, Serwadda A. Context-aware active authentication using smartphone accelerometer measurements. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2014. p. 98–105.

[52] Van Nguyen T, Sae-Bae N, Memon N. DRAW-A-PIN: authentication using finger-drawn PIN on touch devices. In: Proceedings of the computers & security, vol 66. Elsevier; 2017. p. 115–28.

[53] Trewin S, Swart C, Koved L, Martino J, Singh K, Ben-David S. Biometric authentication on a mobile device: a study of user effort, error and task disruption. In: Proceedings of the 28th ACM annual computer security applications conference (ACSAC 2012); 2017. p. 159–68.

[54] 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. 2015. arXiv:1501.01199.

[55] Buriro A, Crispo B, Delfrari F, Wrona K. Hold and sign: a novel behavioral biometrics for smartphone user authentication. In: Proceedings of the Security and privacy workshops (SPW) co-located with IEEE S&P; 2016. p. 276–85.

[56] Song Y, Cai Z, Zhang Z-L. Multi-touch authentication using hand geometry and behavioral information. In: Proceedings of the IEEE symposium on security and privacy (SP); 2017. p. 357–72.

[57] Frank M, Biedert R, Ma E, Martinovic I, Song D. Touchalytics: on the applicability of touchscreen input as a behavioral biometric for continuous authentication. IEEE Trans Inf Forensics Secur 2013;8(1):136–48.

[58] Buriro A, Akhtar Z, Crispo B, Gupta S. Mobile biometrics: towards a comprehensive evaluation methodology. In: Proceedings of the IEEE international conference on security technology (ICCST-2017); 2017b. p. 1–6.

[59] Hall MA. Correlation-based feature selection for machine learning. The University of Waikato; New Zealand; 1999. Ph.d dissertation.

[60] Jain AK, Ross AA, Nandakumar K. Introduction to biometrics. Springer; 2011.

[61] Sauro J. Measuring usability with the system usability scale (SUS). 2011. Available at http://www.measuringu.com/sus.php.

[62] Buriro A, Crispo B, Gupta S, Del Frari F. Dialerauth: a motion-assisted touch-based smartphone user authentication scheme. In: Proceedings of the ACM SIGMM workshop on Biometrics methods and applications; 2018.

[63] Tracking movement Tracking movement. 2014. Available at http://developer.android.com/training/gestures/movement.html.

[64] Attribute selected classifier Attribute selected classifier. More Data Mining with Weka MOOC course2015; Retrieved from https://www.youtube.com/watch?v=-BKP0gpDlbI.

[65] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning, 112. Springer; 2013.