Continuous Implicit Authentication through Touch Traces Modelling

Thomas Karanikiotis[1], Michail D. Papamichail[1], Kyriakos C. Chatzidimitriou[1], Napoleon-Christos I. Oikonomou[1], Andreas L. Symeonidis[1], and Sashi K. Saripalle[2]
[1]Electrical and Computer Engineering Dept., Aristotle University of Thessaloniki, Thessaloniki, Greece
[2]ZOLOZ, Kansas City, MO 64108, USA
{thomas.karanikiotis,mpapamic,kyrcha,napoleon.oikonomou}@issel.ee.auth.gr, asymeon@eng.auth.gr, sashi@zoloz.com

***Abstract*—Nowadays, the continuously increasing use of smartphones as the primary way of dealing with day-to-day tasks raises several concerns mainly focusing on privacy and security. In this context and given the known limitations and deficiencies of traditional authentication mechanisms, a lot of research efforts are targeted towards continuous implicit authentication on the basis of behavioral biometrics. In this work, we propose a methodology towards continuous implicit authentication that refrains from the limitations imposed by small-scale and/or controlled environment experiments by employing a real-world application used widely by a large number of individuals. Upon constructing our models using Support Vector Machines, we introduce a confidence-based methodology, in order to strengthen the effectiveness and the efficiency of our approach. The evaluation of our methodology on a set of diverse scenarios indicates that our approach achieves good results both in terms of efficiency and usability.**

***Index Terms*—Implicit Authentication, Smartphone Security, Touch Traces Modelling, Support Vector Machines**

I. Introduction

According to recent statistics [1], the number of mobile phone users is continuously increasing. In the year 2015, the number of mobile phone purchases around the world exceeded 7 billion devices [2]. This turn towards mobile phones, especially smartphones, is more than evident considering that over a third of the population worldwide uses them on a daily basis [3]. This massive penetration of smartphones into everyday life originates from the fact that current devices are powerful enough to support all kinds of day-to-day activities (work, entertainment, socializing etc.) by providing numerous applications, which are just a tap away. This fact is reflected in the number of mobile applications that is expected to reach 258 billions in 2022 [4].

In that context and given the fact that the day-to-day transactions performed via the various applications contain a lot of personal information (i.e. passwords, photos, messages, bank accounts etc.), security and privacy are of utmost importance. As a result, given that attacks are constantly becoming more sophisticated and thus question the traditional authentication approaches (i.e. passwords) in terms of effectiveness, several research efforts are targeted towards the construction of more advanced methodologies. Towards this direction, a lot of attention has been drawn in methodologies that employ continuous implicit authentication on the basis of behavioral biometrics [5], [6], which aim at identifying invariant features of the human behavior during different activities such as speaking, typing, or walking [7].

The main idea behind continuous implicit authentication is to take advantage of data that originate from the interaction of the user with the mobile device (e.g. taps and swipes) in order to calculate a series of features that can provide an effective way of describing and distinguishing an individual among others. This approach exhibits several advantages against the widely used traditional authentication schemes that mainly involve the use of passwords or other active authentication methodologies such as face or speech recognition. At first, traditional approaches have been designed to provide an entry point authentication and thus fail to detect and recognize a challenger after this point, which is not the case in continuous authentication where each interaction is always evaluated regrading its compliance with the constructed behavioral profiles. Furthermore, continuous authentication approaches outperform against traditional authentication techniques in terms of usability. Given the results of several studies [8], [9], usability is of major importance for end-users as in an effort to refrain from remembering long and complex passwords, they tend to select easy-to-guess passwords. Simple passwords like “pass” are subject to statistical guessing or pattern identification attacks [10]. There are even cases where users disable all authentication methods in the sake of usability [11].

In the context of continuous authentication, many state-of-the-practice approaches are directed towards modeling the behavior of users based on the way they interact with their mobile devices by employing taps and swipes [12], [13]. A representative example is the one of [14], where the authors propose a non-cooperative and non-intrusive method for on-device authentication taking advantage of a number of features that originate from modelling natural human kinematics. However, while the existing approaches provide satisfactory results, they have two major limitations. The first is that they use a controlled laboratory-environment application that involves a set of certain functionalities and thus is limited to a small number of predefined actions [15]. This comes in contrast to many real-world usage scenarios where the users navigate through an application in their own way and without specific

978-1-7281-8913-0/20/$31.00 ©2020 IEEE
DOI 10.1109/QRS51102.2020.00026

rules to follow. The second limitation originates from the fact that these approaches are usually performed with a limited number of users (subjects) and thus the produced models that are not easily generalizable [16].

In this work, we design and implement a continuous implicit authentication methodology based on gestures' data. In an effort to overcome the aforementioned limitations, we made use of a real-world application, publicly available, through which gestures data from many different users and devices were collected. This application enabled capturing the behavior of end-users under various conditions and usage environments. Upon using the provided dataset of the gestures coming from the application, we compute a series of features that quantify how each user interacts with the mobile device. These features are then used as the information basis upon which we apply artificial intelligence techniques in order to train various continuous implicit authentication models. The results obtained by evaluating our continuous authentication approach under various real-world usage scenarios showed that our methodology can provide efficient and effective continuous authentication capabilities.

The rest of this paper is organized as follows. Section II reviews the related work in the area of continuous implicit authentication and discusses any limitations of existing approaches. Our experimental setup along with a detailed discussion on our benchmark dataset is given in Section III, while Section IV presents our modelling approach towards continuous implicit authentication. Finally, Section V evaluates our approach against different baselines, while Section VI concludes this work and provides interesting insight for future research.

II. Related Work

There are several studies that aspire to model the behavior of end-users (usually referred to as subjects) based on their interaction with the mobile device. A common practice involves using taps and swipes measurements as ground truth information upon which certain features are calculated. These features are then used in order to train models that enable continuous implicit authentication. Alzubaidi and Kalita [5] tried to summarize and analyze the approaches made towards continuous authentication, highlighting the methodology applied to each of them, the datasets used and the evaluation approaches.

De Luca et al. [12] used data from participants using touchscreens, using both only simple unlocking patterns (horizontal and vertical swipes) and all unlocking patterns. The results showed that modelling using complex unlocking patterns achieves satisfying results. Xu et al. [17] recruited 32 different users to use a device with a data acquisition tool and specific instructions. The authors used the Support Vector Machine (SVM) classifier. The distinctiveness performance of the model was pretty high in cases in which the number of users involved was quite small. Additionally, Feng et al. [18] calculated features regarding taps, swipes and zoom gestures made by 23 original users and 100 guest users. Touch-based Identity Protection Service (TIPS) was tested both off- and on-device, achieving almost 90% accuracy.

Using a similar approach, Zhao et al. [19] proposed Graphic Touch Gesture Feature (GTGF), which converts touch sequences to images using all gestures made by a user. For the experiments, 30 users were recruited and an Equal Error Rate of 10 to 20% was achieved. Saravanan et al. [20] collected gestures from 20 participants and trained a multiclass classifier, achieving accuracy of almost 100%, and a unary classifier, achieving an average accuracy of around 97%. Feng et al. [21] used all kind of touch gestures made by users with three different classifiers: Decision Trees, Random Forest and Bayes Net Classifier. Gestures were collected from 40 different users, achieving False Acceptance Rate and False Rejection Rate below 2% when additional sensors data are used.

On the other hand, Mahbub et al. [22] captured only touch sequences (swipes) varying in length between 1 and 3637 touch data points for 48 different subjects. The authors trained one binary classifier for each user following the one-vs-all approach, using k-nearest neighbors (kNN), Gaussian kernel Support Vector Machines (RBF-SVM), Naive Bayes (NB), Linear Regression (LR), Random Tree estimation followed by Linear Regression (RT+LR), Random Forest estimator (RF), and Gradient Boosting Model (GBM). According to the obtained results, the Random Forest classifier appears to outperform all others. Moreover, Gong et al. [23] used also only horizontal and vertical swipes to discriminate the original user from attackers. The authors achieved Equal Error Rate of 1-2% for random attacks and 16-23% for targeted attacks for the horizontal strokes. In a more general approach, Angulo and Wästlund [24] used data from 32 different participants, using a Random Forest classifier, which achieved a mean Equal Error Rate of about 10%.

Wang et al. [13] extended the above methodology by making use of a real-world application that applies no restrictions to the end-users. The authors collected taps and swipes data from 20 different users (subjects) and used the gathered data for computing 59 different features, training two different classifiers: SVM and RF. Upon evaluating their models using Area Under Curve metric (AUC) for detecting unauthorized access, the authors were able to achieve an AUC score of 80% to 96%. In a similar approach, Frank et al. [25] developed a mobile application in scenarios that simulate the daily actions made by mobile users. The gestures made by the users were used to train and test the two selected classifiers, kNN and SVM. The Equal Error Rate achieved when only one stroke is used in classification is about 13% and when multiple strokes are used to provide a classification output the Equal Error Rate converges to 2-3%.

Finally, trying to extend the number of users participating in the Continuous Implicit Authentication experiments, Zheng et al. [15] used over 80 subjects, who had to type certain Personal Identification Numbers (PINs). The Equal Error Rate achieved for each password the users had to type is below 10%.

The previous approaches, despite achieving some good results, include either one or both of the limitations noted before, that is the small number of individuals (subjects) participating in the experiments and the low-scale controlled environment in which the subjects interacted with the experiments' devices, along with the predefined actions they ought to take.

Taking into account the continuous authentication methodologies described in all the studies, we followed the state-of-the-practice paradigm according to which taps and swipes data can provide valuable information towards the successful behavioral profiling of end-users. However, considering the limitations that occur while trying to perform behavioral profiling based on certain actions and in an effort to provide a thorough continuous authentication methodology applicable in a wide range of scenarios, we built our approach using a real-world mobile and tablet application that not only can be attractive to a wide audience and enable the long-term active involvement of end-users through gamification, but also offers numerous different modelling and design capabilities and enable capturing the end-user's behavior under various conditions (focus on speed or accuracy, focus on taps or swipes, perform free or directed movements etc.).

III. Experimental Setup

As already noted, the gestures data coming from the interaction of the user with the screen of the mobile phone could be the main object towards Continuous Implicit Authentication methodology. The gestures, after first coming through a features extraction level, can be used to train the selected Continuous Implicit Authentication model, which then will be responsible to recognize the legitimate owner of the device, leaving him uninterrupted, while the external non-authorized users will be denied the use of the device. In order to avoid the limitations of many previous approaches, the application that accomplishes the data collection procedure needs to simulate real-world daily scenarios, in which the user interacts with the device in his/her own way, without any guided predefined moves and actions. At the same time, the application should be open to public, available to many individuals of all ages, genders and level of expertise to download and use, running on many different devices with various specifications. Towards this goal, we made use of a dataset provided online [26] that was collected using a mobile phone application, called "BrainRun", which includes various games and through which the data collection mechanism is applied.

A. BrainRun Overview

In an attempt to capture the different characteristics of the users interacting with a mobile phone, Papamichail et al. [26] developed "BrainRun"[1], which is a brain training game aiming at boosting cognitive skills of individuals through a series of interactive gaming types. "BrainRun" is open to public and enables the collection of gestures and sensors data from many different users/devices. It includes a set of 5 different game types (”Focus", ”Mathisis", ”Memoria", ”Reacton" and ”Speedy"). Through navigation and game-playing, gestures are collected, forming the dataset in which our Continuous Implicit Authentication modelling approach will be applied. Each game-type is specifically designed to collect different kind of hand gestures, such as taps, horizontal swipes, vertical swipes, swipes and taps combined etc.

B. Benchmark Dataset

The "BrainRun" dataset contains the users playing the game, the games they played, the devices used and the gestures made by them in order to navigate through or play the game. Table I depicts some statistics about the provided dataset.

TABLE I
BrainRun Statistics

Metric	Value
Number of users	2,221
Number of devices	2,418
Number of games played	106,805
Total number of gestures	3,110,101
Number of taps	2,463,115
Number of swipes	646,986

Figure 1 illustrates the number of games played by the users in each one of the game-types, while Table II depicts the number of taps and swipes, as well as the total number of gestures, collected by every game-type.

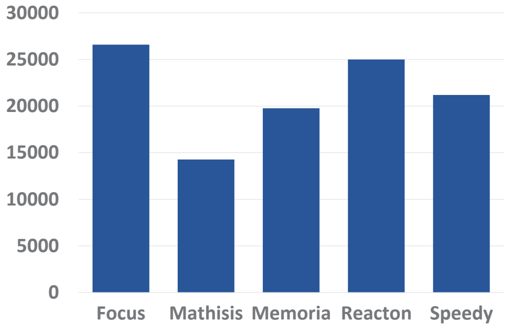


Fig. 1. Number of games per game-type

TABLE II
Number of gestures per game-type

Game-type	Taps	Swipes	Total gestures
Focus	56,836	304,975	361,811
Mathisis	60,483	200,098	260,581
Memoria	631,052	8,888	639,940
Reacton	350,345	86,654	436,999
Speedy	1,111,801	18,802	1,130,603

[1] http://brainrun.issel.ee.auth.gr/

TABLE III
CALCULATED FEATURES

Feature Name	**Feature Description**
Horizontal Trace Length	Calculated distance between first and last point of swipe in the horizontal axis
Vertical Trace Length	Calculated distance between first and last point of swipe in the vertical axis
Slope	The slope of the straight line that best approaches the swipe's trace
Mean Squared Error	The mean squared error between the swipe's points and the straight line
Mean Absolute Error	The mean absolute error between the swipe's points and the straight line
Median Absolute Error	The median absolute error between the swipe's points and the straight line
Coefficient of Determination	The coefficient of determination between the swipe's points and the straight line
Horizontal Acceleration	Mean acceleration of user's movement in the horizontal axis
Vertical Acceleration	Mean acceleration of user's movement in the vertical axis
Horizontal Mean Position	Mean position of user's gesture in the horizontal axis
Vertical Mean Position	Mean position of user's gesture in the vertical axis

IV. CONTINUOUS IMPLICIT AUTHENTICATION MODELLING

A. Research Objectives

Our approach towards Continuous Implicit Authentication lies on the classification of one individual against everyone else. On this basis, only one specific individual's gestures are known in advance and the selected classifier is trained on them. This is going to be also the original user, who is eligible to use the device. Everyone else is a part of the "attackers" (e.g. the "universe"), who should be detected by the model and whose actions will subsequently lock the device.

The classifier used for this purpose is going to be an 1-class classifier, trained on the gestures of the original user and tested both on the gestures of the original user and on the gestures of everyone else. The primary target of the classification problem is to minimize the False Acceptance Rate (FAR), the percentage of the unknown users or attackers who are badly classified as the original user and thus gain access to the device with unknown consequences. This is also the main goal of the Continuous Implicit Authentication methodology. At the same time, it is also very important to minimize as much as possible the False Rejection Rate (FRR). The FRR counts the percentage of the occasions in which the original user is not recognized by the model and thus is prohibited from using the device, until he/she activates it again. Equations 1 and 2 depict the calculations of FAR and FRR respectively.

$$FAR = \frac{attacker\ accepted\ swipes}{attacker\ total\ swipes} \tag{1}$$

$$FRR = \frac{original\ user\ rejected\ swipes}{original\ user\ total\ swipes} \tag{2}$$

However, as it is obvious from the experiments, the FAR is a metric that cannot be used in all occasions, so the number of gestures coming from an attacker and being accepted by the model will occasionally be used as the appropriate metric.

B. Model Construction

For the Machine Learning model, towards the "One against the universe" scenario described above, we selected the One-Class Support Vector Machine (SVM) as the main classifier, because of its successful application in a wide range of applications [27]. In our case, SVM is going to model the feature space of the original user and detect the features, and thus the corresponding gestures, that do not belong into it.

From the gestures collection, we calculate the features needed for the construction of the classification model. For this purpose, we divide the whole dataset of gestures into three parts. The first part consists of all the tap-gestures of the users, the second part includes the swipe-gestures and the third part contains the swipe-gestures with duration less than a predefined value (for example 70ms), which we will call "Fake-swipes". These gestures have been categorized as "Swipes" from the device, due to the sampling of many points, but, in fact, they are "Taps", in which the finger of the user activated more than one point in the device. From the whole group of gestures, we discard "Taps" and "Fake-swipes" and use only "Swipes", because they contain much more information, produce better features and have been shown to give better classification results.

For every swipe collected and stored in the gestures collection, multiple points on the screen have been activated by the hand movement and sampled by the device. The raw data of every swipe are coming through the features extraction layer for the features calculation process to be applied. Table III depicts the features we calculate for each swipe.

It should be noted that, for the features "Horizontal Trace Length", "Vertical Trace Length", "Horizontal Mean Position" and "Vertical Mean Position", we have experimented both with normalized feature values according to the dimensions of the device, in order to discard their effect on the classification, and without.

From the game-types of BrainRun, swipes are collected through "Mathisis", "Focus" and "Reacton", as "Memoria" and "Speedy" contain only taps. At the same time, "Reacton" forces users to tap or swipe in a specific area, which may not be helpful for the classification. Because of these facts, in our experiments there will be used only swipes from "Mathisis" and "Focus" games, training and testing models separately for horizontal swipes ("Mathisis") and vertical swipes ("Focus").

Figure 2 illustrates the system process for predicting new

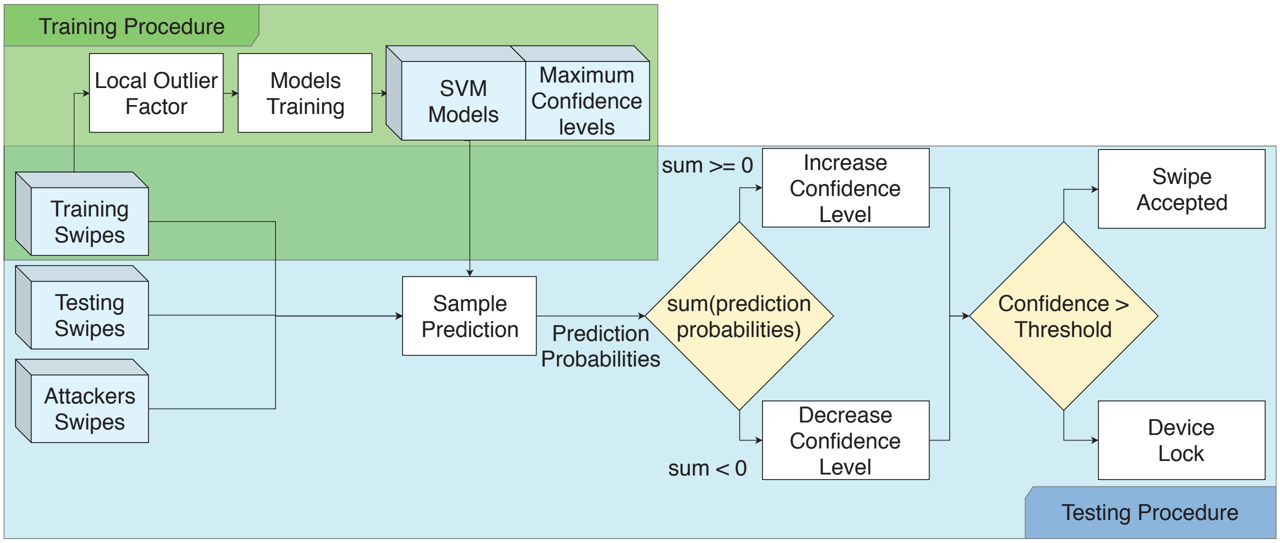


Fig. 2. CIA prediction procedure.

gestures and recognizing the original user of the device. In the classification experiments, the classification accuracy of the SVM One-Class is further assisted by the use of system's **confidence level**. The confidence level is a classification boosting technique introduced to our system, which refers to the certainty of the system that the current user of the device is the original one. The confidence level is initialized at a predefined value (e.g. 60%) and there is also a predefined threshold (e.g. 35%) under which the device is locked (confidence level immediately reaches 0). Every gesture coming through the system, changes the current confidence level according to the classification result. In the case in which the gesture is classified as belonging to the original user, the confidence level for the user is increased by a value. On the opposite side, if a gesture is classified as an outlier (not a known user), the confidence level is decreased by a value and may force the system to lock. In that occasion, the device is unlocked only by using a traditional authentication mechanism, such as password, and then the confidence level of the system is restored to its predefined starting value. Additionally, the confidence level of the system may be decreased by a step over time, in order to force the device to lock, should the device be left unused for a time interval. Equation 3 depicts the changes of the confidence level (e.g. CL) over time and upon a classification decision. For our experiments, the parameter a was set to 0 (no need to drop confidence level over time in the experiments), b was set to 5% and c was set to 9% for the experiments in "Focus" game-screen and 15% for the experiments in "Mathisis" game-screen.

$$CL = \begin{cases} CL - a * t & \text{over time} \\ CL + b & \text{if } \textit{original user} \\ CL - c & \text{if } \textit{attacker} \end{cases} \quad (3)$$

Furthermore, before the training of the model, one additional layer is inserted, the preprocessing step, in which the samples of the original user that will be used for the appropriate training of the model are processed to eliminate some outliers, e.g. swipes that diverge from the typical behavior of the user and that are made randomly. This layer is accomplished with the use of the *Local Outlier Factor* model, which doesn't need training and labels all the samples as inliers or outliers. The training samples are passed through this model and the swipes that are labelled as inliers are the only ones that will be used in the further steps. We are not going to use parameter optimization at a large scale for this preprocessing model, as we only need to cut off some extreme outliers.

Towards the implementation and training of the SVM model described so far, one limitation arises. The parameters of the SVM One-Class model that can be tuned to achieve better accuracy (e.g. ν and γ parameters) need to be optimally selected according both to original user's and to attackers' swipes. However, in practice and in a real-world scenario this is not possible, as, during the training procedure, only the swipes of the original user will be available (the training set in our approach). This limitation could be of high importance, due to the fact that we expect from the individuals to have much different behaviors and thus the features extracted from each one will have quite high variance, leading to bad assessments from the model.

In an attempt to overcome this limitation, we applied the following approach. Instead of using one single SVM One-Class model with a pair of values for the ν and γ parameters, we used a big number of models with various ν-γ pairs, covering a big area of the fine-tuning procedure. The training procedure is exactly the same for each model, while, in the testing phase, every new gesture is classified by every model, which returns its classification decision. Then, the decisions made from all models are summarized and the final prediction is made with majority voting, e.g. by selecting the most supported decision.

Another technique used to boost the accuracy of the classification is the **certainty of prediction**. This term is used to describe a value, which determines how confident a specific model is for its prediction on the current gesture, or, in other words, the percentage of certainty for the label given to the current sample. Obviously, this term is similar (and is used instead of) the prediction probabilities of Machine Learning models. Because of the certainty of prediction, the system will especially focus on the samples, for which the classification model is quite sure about their labels, partially ignoring the confusing samples. For the calculation of the certainty of prediction, each SVM One-class model calculates the distance of the sample from the hyperplane designed by the model. This distance is used in equation 4 to calculate the certainty of prediction (e.g. COP) of each sample, based on the maximum distance from the hyperplane of the training samples.

$$COP(s) = \begin{cases} -1 & \text{if } (d(s) < -|max\_dist|) \\ 1 & \text{if } (d(s) > |max\_dist|) \\ \frac{d(s)}{max\_dist} & \text{otherwise} \end{cases} \quad (4)$$

where s is the sample that is going to be classified, $d(s)$ is the distance of the sample from the hyperplane calculated by the model and max_dist is the maximum distance from the hyperplane for all the training samples.

The certainty of prediction is used along with the application of many SVM models with various ν-γ parameters. Each model returns its prediction probability on the classification of the current gesture as a percentage which represents the model's confidence for the label assigned to the sample. Then, all prediction probabilities are summed, giving the final decision of the system, focusing on the models that are more confident for their result, as shown in equation 5.

$$dec(s) = \begin{cases} 1 & \text{if } (\sum_{\nu,\gamma} COP_{SVM_{\nu,\gamma}}(s) \geqslant 0) \\ -1 & \text{if } (\sum_{\nu,\gamma} COP_{SVM_{\nu,\gamma}}(s) < 0) \end{cases} \quad (5)$$

where $dec(s)$ is the classification decision made by the system on the sample s.

V. Evaluation

The performance of our proposed methodology was evaluated along three axes. First of all, we tested our system in real-world scenarios, e.g. testing the "one-vs-the-universe" approach. In this approach, an original user was selected from the pool in the available dataset, while the rest of the users consisted the set of the attackers, trying to trick the system and make use of the device. In a second level of evaluation, we tested the performance of each individual SVM model (with standard ν-γ parameters) introduced to our system. Finally, we tested the proposed methodology in practise, employing a real-world application that includes the classification model and that is tested upon several subjects.

A. System Evaluation

In order to test the overall system's performance, an original user is selected from the dataset, who trains the proposed models using a part of his/her available swipes (e.g. the training set of the original user), using a 0.75 training-testing split. Subsequently, the models are tested using this training set first, then the original user's testing set and finally the testing set containing the swipes of all the available attackers.

We initially tested the performance of our models without the use of the aforementioned classification boosting techniques, e.g. the confidence level, the certainty of prediction and the preprocessing of the original user's training set with the use of the Local Outlier Factor classifier. Certainly, the metrics used in this approach are the FAR, measuring the number of one attacker's swipes that have been accepted by the system as a percentage of the total number of the available swipes of the selected attacker, and the FRR, which counts the number of the original user's swipes that have been rejected by the system as a percentage of the total number of available original user's swipes. The FAR can then easily be transformed into the number of the attacker's swipes that have been accepted by the system. Attacker accepted swipes is a metric which indicates the amount of swipes an attacker can participate in before the system locks down.

As a second test towards the building of the whole model described above, we introduced the system's confidence level into the model used in the previous attempt. It is thus obvious that the system and its evaluation metrics need to be changed. For the evaluation of our models and in an attempt to simulate a real-world scenario, during the feed-forward phase of the training and the testing set of the original user and once the system locks (multiple misclassified samples that lead the confidence level to drop below the selected threshold), the original user can unlock the device and continue its use, thus continue the testing of the modelling with the rest of his/her swipes. As a consequence, for the evaluation of our approach, the FRR can still be used, as the whole set of the original user's swipes are going to pass through the system and the number of the swipes that lead the system to lock are going to be counted.

On the other hand, as would happen in a real-world scenario, every attacker who is used to evaluate the system, cannot unlock the device once it locks. The available swipes of every attacker are being shuffled and passed through the system, until it recognizes the unauthorized use and locks. As a result, the FAR metric cannot be used any more, since the whole set of the attacker's swipes are not passing through the system and the number of left-unused swipes is not important. Subsequently, for the evaluation of the system during the attacking phase and for every attacker used, we are going to count the number of swipes that are accepted by the system,

TABLE IV
EXPERIMENTS RESULTS

Experiment	Mathisis - FRR	Focus - FRR
Simple models	29.51%	51.15%
Addition of Confidence Level	4.50%	8.29%
Addition of Prediction Probabilities	4.73%	5.66%
Addition of Preprocessing and Limit on Swipe Points	5.91%	6.77%

TABLE V
COMPARISON TO OTHER APPROACHES

Approach	FRR	# of Subjects	Environment
Yang *et al*. [28]	10%	200	Specifically designed
Feng *et al*. [18]	9%	23	Specifically designed
Gong *et al*. [23]	4-8%	25	Specifically designed
Lee *et al*. [29]	0.9%	35	Specifically designed
Ours	4.7-5.7%	2,221	Real-world, Public

which is a valuable metric, showing the number of actions an attacker would manage to make until he/she is recognized.

In the next experiments, following again the real-world scenario presented above, we introduced the certainty of prediction and the preprocessing stage successively into the system, as well as an additional preprocessing step, which removes the swipes containing less than 3 data points (points sampled in the screen during the gesture) or more than 10 data points, as these gestures do not represent the normal swiping behavior and the modelling of the user becomes more complicated.

TABLE VI
THE PERFORMANCE OF EACH SVM MODEL

	Model Parameters			
ν	**Majority**	**0.01**	**0.29**	**0.13**
γ	**Voting**	**0.00005**	**0.0009**	**0.0005**
FRR	2.33%	0.00%	6.09%	3.76%
Acc. Swipes	1.48	4.53	1.26	1.50

Table IV depicts the FRR achieved in all the above experiments, both on Mathisis (e.g. horizontal) and on Focus (e.g. vertical) swipes.

From the results presented above, we conclude that the whole system presented in the previous chapter can model quite well the behavior of the mobile user. The introduction of the confidence level drops the FRR of the original user significantly, both in the Mathisis and the Focus screen. At the same time, the number of accepted swipes from the attackers is below 1 for the case of "Simple model", as the models are quite strict, while in the rest of the experiments, the number of attacker accepted swipes is slightly over 1, due to the fact that the addition of the system's confidence needs at least one swipe to drop the initial level. Nonetheless, in a real-world scenario the one swipe that an attacker could use, is not able to cause a significant threat to the device and the user's data.

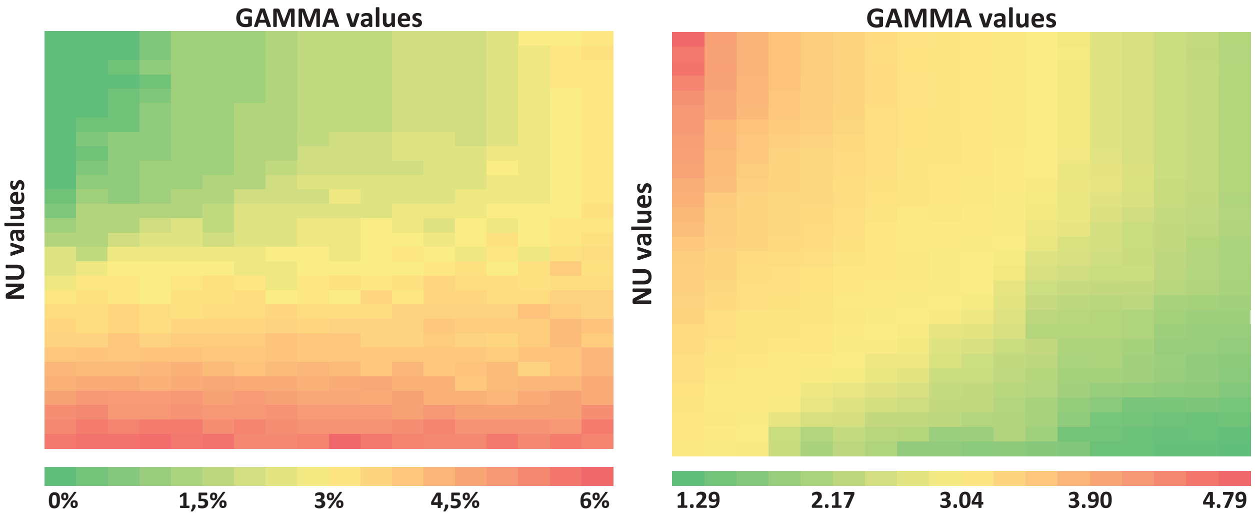
Additionally to the previous evaluation, we created a histogram of the attackers that successfully completed only one swipe until the device locked, subsequently causing no harm to the device, the attackers that managed to submit 2 to 5 swipes, taking some limited actions on the device and, finally, the attackers that managed to deceive the system, make more than 5 swipes and use the device uninterrupted. Table VII depicts the number of occasions in which the system needed only one swipe, two to five or more than five swipes to recognize an attacker.

TABLE VII
FREQUENCY OF ATTACKER ACCEPTED SWIPES

Number of accepted swipes	Frequency
1	737,488
2 - 5	42,585
>5	9,407

The number of swipes that were needed by the system to recognize an attacker presented above, prove that the system could provide satisfactory security, as in the huge majority of the occasions, the attacker was only able to take a limited number of actions, before the device was locked, being unable to harm the device and its owner.

In order to fully evaluate our approach against state of the practice strategies, we compared our results to 4 similar approaches to the Continuous Implicit Authentication problem. The system's confidence level included in our approach does not allow the use of the False Acceptance Rate metric, as already noted, which makes the comparison regarding the attackers not possible. Table V depicts the full comparison between the False Rejection Rate of the original user of a device achieved in these approaches, the number of subjects used for the evaluation and the specific environment (*e.g.* the application) in which the experiments were conducted. Our approach achieves one of the best results, while, contrary to the huge majority of the literature approaches, it combines a very big number of subjects participating in the experiments with an open-to-the-public real-world application for data gathering.



(a)

(b)

Fig. 3. Performance results of each SVM model regarding (a) the False Rejection Rate (FRR) and (b) the number of attacker accepted swipes

B. Models Evaluation

In an attempt to test our decision of introducing many SVM models into the whole system, with different ν and γ parameters, we evaluated the performance of each model inside the whole system. Using only one model every time, we counted the FRR and the number of attackers' accepted swipes achieved by the specific model.

Figure 3 illustrates the performance of each SVM model with its own ν-γ parameters values upon the FRR on the swipes of the original user and upon the number of the swipes belonging to attackers that are being accepted by the system. Table VI depicts the FRR and the number of accepted swipes made by attackers for randomly selected models and for the whole system containing all the SVM models. It is obvious that a unique selection of ν and γ values would be a difficult task, as the values that achieve a lower FRR seem to maximize the number of attacker accepted swipes and vice versa. Our approach of combining all the SVM models through majority voting smooths out the models' performances.

C. Case Study

Our final evaluation axis lies on the real-world applicability of the modelling, e.g. the use of our model in real-world scenarios, counting the same metrics for the system's performance on individuals using the device. The experiment consisted of two different stages. In the first part ("Random Attacks"), the individuals taking part in this task are asked to use the device and play a game (similar to the ones of "BrainRun") without any more information, while we count the FRR of the original user and the number of the attackers' swipes that have been accepted by the system. In the second stage ("Targeted Attacks"), the attackers are asked to watch the original user's behavior and try to mimic it. By doing so, the attackers are trying to deceive the system, just like in a real-world attacking scenario. For the experiment, 10 different users (attackers) were used.

Table VIII depicts the metrics measured in this case study, both in the first stage of the experiment and in the second (attacking) task of the attackers.

TABLE VIII
CASE STUDY PEFORMANCE RESULTS

	FRR	Accepted Swipes
Random Attacks	2.56%	2.52
Targeted Attacks		3.67

From the results depicted in Table VIII, it is obvious that the system works quite well in the normal mode, discriminating the original user from the unknown individuals. In the second task, in which the attackers are aware of the behavior of the original user, the proposed methodology cannot achieve its best performance, as the attackers are able to mimic the selected features of the original user for a small number of gestures. However, in both cases the results are satisfactory and the main goal is achieved.

VI. CONCLUSIONS AND FUTURE WORK

The results obtained from Table IV prove that the addition of the confidence level granted a significant boost to the overall system's performance, as some misclassifications cannot lock the original user out of the device, but they simply drop the confidence level, which is then increased again by the right recognition of the original user. The only drawback of the confidence level is that the system may allow an attacker to make more swipes, until the confidence level reaches and drops below the threshold. It should be noted that the number of attacker accepted swipes in the no-confidence experiments cannot be directly compared with the rest of the experiments, as in this specific experiment the whole set of attackers' swipes is coming through the system, while in the rest of the experiments, an attacker tests the system only until it locks.

At the same time, the use of many different SVM models has smoothed out the peculiarities of each individual model, avoiding the drawback of selecting a unique set of ν-γ parameters that can lead either on high FRR values or on a big number of attackers' accepted swipes. This justifies our decision of selecting multiple SVM models instead of fine-tuning a single model. In this context, while the number of different SVM models is big, it cannot be perceived as a drawback, as these models are going to be trained once when the application starts and the device will only perform predictions thereafter. In case a retraining procedure is needed, it can take place in a dedicated server, thus eliminating any time- or resource-consuming tasks from running in the device.

The addition of the prediction probabilities and the preprocessing stage, as well as the limit of the data points included in each swipe, has some ambiguous results. In the Mathisis experiments with horizontal swipes the more complicated the model becomes, the more misclassifications of the original user are happening, while less swipes from the attackers are being accepted. On the other hand, the classification boosting techniques seem not to influence the accepted swipes of the attackers on Focus experiments, while they introduce a drop in the FRR.

Finally, the complete system introduced in this paper achieves pretty good results, recognizing the legitimate owner of a mobile device, who uses the device uninterrupted, and identifying any attacker trying to use the device, not allowing its use. The above results are even more important, as the experiments were conducted with gestures coming from a real-world application, simulating daily scenarios and allowing its use at the user's will, while the evaluation phase was conducted with data coming from more than 2,000 users. Since our approach is independent from the environment the gestures come from, our methodology could easily be generalized and used in various usage scenarios and applications with different scope and context, such as communications apps, just by changing the training dataset.

In a future attempt, different features could be tested upon the above described system, containing more information on the behavior of the user. Moreover, the use of the user's taps could provide additional information and improve the classification. Additionally, the system's confidence level could be further improved by changing its current value according to the models' prediction probabilities. Last but not least, in an effort to further explore the effectiveness of the constructed methodology, the attacking scenarios could be automated, trying to mimic the behavior of the original user.

Acknowledgements

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code: T1EDK-04673).

References

[1] "Number of mobile phone users," https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/, 2016, [Online; accessed April-2020].

[2] "List of countries by number of mobile phones in use," https://en.wikipedia.org/wiki/List_of_countries_by_number_of_mobile_phones_in_use, 2019, [Online; accessed April-2020].

[3] "Smartphone user penetration as percentage of total global population from 2014 to 2021," https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/, 2017, [Online; accessed April-2020].

[4] "Number of mobile app downloads worldwide," 2018, [Available at https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/; retrieved April-2020]. [Online]. Available: https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/

[5] A. Alzubaidi and J. Kalita, "Authentication of smartphone users using behavioral biometrics," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1998–2026, thirdquarter 2016.

[6] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, "Continuous user authentication using multi-modal biometrics," *Computers and Security*, vol. 53, pp. 234 – 246, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404815000875

[7] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Oct 2010, pp. 306–311.

[8] E. Chin, A. P. Felt, V. Sekar, and D. Wagner, "Measuring user confidence in smartphone security and privacy," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: ACM, 2012, pp. 1:1–1:16. [Online]. Available: http://doi.acm.org/10.1145/2335356.2335358

[9] W. Melicher, D. Kurilova, S. M. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek, "Usability and security of text passwords on mobile devices," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 527–539. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858384

[10] S. Schechter, C. Herley, and M. Mitzenmacher, "Popularity is everything: A new approach to protecting passwords from statistical-guessing attacks," in *The 5th USENIX Workshop on Hot Topics in Security (HotSec '10)*. USENIX, August 2010, this paper scheduled to be made public 12:01AM, July 19. [Online]. Available: https://www.microsoft.com/en-us/research/publication/popularity-is-everything-a-new-approach-to-protecting-passwords-from-statistical-guessing-attacks/

[11] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner, "Are you ready to lock?" in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '14. New York, NY, USA: ACM, 2014, pp. 750–761. [Online]. Available: http://doi.acm.org/10.1145/2660267.2660273

[12] 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," *Conference on Human Factors in Computing Systems - Proceedings*, 05 2012.

[13] X. Wang, T. Yu, O. Mengshoel, and P. Tague, "Towards continuous and passive authentication across mobile devices: An empirical study," in *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WiSec '17. New York, NY, USA: ACM, 2017, pp. 35–45. [Online]. Available: http://doi.acm.org/10.1145/3098243.3098244

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

[15] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," in *2014 IEEE 22nd International Conference on Network Protocols*, Oct 2014, pp. 221–232.

[16] G. Xue, "Unobservable re-authentication for smartphones," 5 2013.

[17] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. Menlo Park, CA: USENIX Association, 2014, pp. 187–198. [Online]. Available: https://www.usenix.org/conference/soups2014/proceedings/presentation/xu

[18] T. Feng, J. Yang, Z. Yan, E. Munguia Tapia, and W. Shi, "Tips: context-aware implicit user identification using touch screen in uncontrolled environments," 02 2014.

[19] X. Zhao, T. Feng, and W. Shi, "Continuous mobile authentication using a novel graphic touch gesture feature," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sep. 2013, pp. 1–6.

[20] P. Saravanan, S. Clarke, D. H. Chau, and H. Zha, "Latentgesture: Active user authentication through background touch analysis," pp. 110–113, 04 2014.

[21] T. Feng, Z. Liu, K. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen, "Continuous mobile authentication using touchscreen gestures," in *2012 IEEE Conference on Technologies for Homeland Security (HST)*, Nov 2012, pp. 451–456.

[22] U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results," *CoRR*, vol. abs/1610.07930, 2016. [Online]. Available: http://arxiv.org/abs/1610.07930

[23] N. Z. Gong, M. Payer, R. Moazzezi, and M. Frank, "Forgery-resistant touch-based authentication on mobile devices," in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS '16. New York, NY, USA: ACM, 2016, pp. 499–510. [Online]. Available: http://doi.acm.org/10.1145/2897845.2897908

[24] J. Angulo and E. Wästlund, "Exploring touch-screen biometrics for user identification on smart phones," in *Privacy and Identity Management for Life*, J. Camenisch, B. Crispo, S. Fischer-Hübner, R. Leenes, and G. Russello, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 130–143.

[25] 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," *CoRR*, vol. abs/1207.6231, 2012. [Online]. Available: http://arxiv.org/abs/1207.6231

[26] M. D. Papamichail, K. C. Chatzidimitriou, T. Karanikiotis, N.-C. I. Oikonomou, A. L. Symeonidis, and S. K. Saripalle, "Brainrun: A behavioral biometrics dataset towards continuous implicit authentication," *Data*, vol. 4, no. 2, p. 60, May 2019. [Online]. Available: http://dx.doi.org/10.3390/data4020060

[27] K.-K. Seo, "An application of one-class support vector machines in content-based image retrieval," *Expert Systems with Applications*, vol. 33, no. 2, pp. 491 – 498, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417406001655

[28] L. Yang, Y. Guo, X. Ding, J. Han, Y. Liu, C. Wang, and C. Hu, "Unlocking smart phone through handwaving biometrics," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 1044–1055, 2015.

[29] W. Lee and R. B. Lee, "Sensor-based implicit authentication of smartphone users," in *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2017, pp. 309–320.