BioTouchPass: Handwritten Passwords for Touchscreen Biometrics

Ruben Tolosana, Ruben Vera-Rodriguez, Member, IEEE, and Julian Fierrez, Member, IEEE

Abstract—This work enhances traditional authentication systems based on Personal Identification Numbers (PIN) and One-Time Passwords (OTP) through the incorporation of biometric information as a second level of user authentication. In our proposed approach, users draw each digit of the password on the touchscreen of the device instead of typing them as usual. A complete analysis of our proposed biometric system is carried out regarding the discriminative power of each handwritten digit and the robustness when increasing the length of the password and the number of enrolment samples. The new e-BioDigit database, which comprises on-line handwritten digits from 0 to 9, has been acquired using the finger as input on a mobile device. This database is used in the experiments reported in this work and it is available together with benchmark results in GitHub. Finally, we discuss specific details for the deployment of our proposed approach on current PIN and OTP systems, achieving results with Equal Error Rates (EERs) ca. 4.0% when the attacker knows the password. These results encourage the deployment of our proposed approach in comparison to traditional PIN and OTP systems where the attack would have 100% success rate under the same impostor scenario.

Index Terms—Biometrics, passwords, PIN, OTP, handwriting, touch biometrics, mobile, deep learning, RNN, LSTM, DTW, e-BioDigit database

I. INTRODUCTION

MOBILE devices have become an indispensable tool for most people nowadays [1]. The rapid and continuous deployment of mobile devices around the world has been motivated not only by the high technological evolution and new features incorporated but also to the new internet infrastructures like 5G that allows the communication and use of social media in real time, among many other factors. In this way, both public and private sectors are aware of the importance of mobile devices for the society and are trying to deploy their services through user-friendly mobile applications ensuring data protection and high security.

Traditionally, the two most prevalent user authentication approaches have been Personal Identification Numbers (PIN) and One-Time Passwords (OTP). While PIN-based authentication systems require users to memorize their personal passwords, OTP-based systems avoid users to memorize them as the security system is in charge of selecting and providing to the user a different password each time is required, e.g., sending messages to personal mobile devices or special tokens. Despite the high popularity and deployment of PIN- and OTP-based authentication systems in real scenarios, many studies have highlighted the weaknesses of these approaches [2], [3]. First, it is common to use passwords based on sequential digits, personal information such as birth dates, or simply words such as “password” or “qwerty” that are very easy to guess. Second, passwords that are typed on mobile devices such as tablets or smartphones are susceptible to “smudge attacks”, i.e., the deposition of finger grease traces on the touchscreen can be used for the impostors to guess the password [4]. Finally, password-based authentication is also vulnerable to “shoulder surfing”. This type of attack is produced when the impostor can observe directly or use external recording devices to collect the user information. This attack has attracted the attention of many researchers in recent years due to the increased deployment of handheld recording devices and public surveillance infrastructures [5], [6]. Biometric recognition schemes are able to cope with these challenges by combining both a high level of security and convenience [7].

This study evaluates the advantages and potential of incorporating biometrics to password-based mobile authentication systems, asking the users to draw each digit of the password on the touchscreen instead of typing them as usual. This way, the traditional authentication systems are enhanced by incorporating dynamic handwritten biometric information. One example of use that motivates our proposed approach is on internet payments with credit cards. Banks usually send a numerical password (typically between 6 and 8 digits) to the user’s mobile device. This numerical password must be inserted by the user in the security platform in order to complete the payment. Our proposed approach enhances such scenario by including a second authentication factor based on the user biometric information while drawing the digits. Fig. 1 shows a general architecture of our proposed password-based mobile authentication approach. The three following main modules are analyzed in this study: i) enrolment set, ii) password generation, and iii) touch biometric system. Depending on the final application (i.e., PIN or OTP), the handwritten digits can be first recognized using for example an Optical Character Recognition (OCR) system in order to verify the authenticity of the password. After this first authentication stage, the biometric information of the handwritten digits is compared in a second authentication stage to the enrolment data of the claimed user, comparing each digit one by one. In this study we focus on the second authentication stage based on the behavioral information of the user while performing the handwritten digits as the recognition of numerical digits has already shown to be an almost solved problem with errors close to 0% [8], [9]. Therefore, in this study we make the
assumption that impostors pass the first stage of the security system (i.e., they know the password of the user to attack) and thus, the attack would have 100% success rate if our proposed approach was not present.

The main contributions of this study are related to our proposed architecture, the competitive results obtained with respect to related research, and our experimental findings:

- We survey and compare advantages and limitations of recent research on touch biometrics for mobile authentication.
- We incorporate biometrics to password-based mobile authentication. Two different state-of-the-art approaches are studied for the similarity computation: i) Dynamic Time Warping (DTW), which is widely used in many different fields such as handwriting and speech; and ii) Recurrent Neural Networks (RNNs), which are specific deep learning architectures considered for modelling sequential data with arbitrary length.
- We perform a complete analysis of the touch biometric system regarding the discriminative power of each handwritten digit. In addition, we analyze the robustness of our proposed approach when increasing the length of the handwritten password and the number of available enrolment samples per user.
- We discuss specific details for the deployment of our proposed approach on current PIN- and OTP-based authentication systems, including various strategies for password generation.
- We achieve better results than other verification schemes such as the handwritten signature and graphical passwords, as well as other recent works on touch biometrics.
- We introduce the new e-BioDigit database, which comprises on-line handwritten numerical digits from 0 to 9 for a total of 93 users, captured on a mobile device using finger touch interactions. Handwritten digits were acquired in two different sessions in order to capture the intra-user variability. This database is publicly available to the research community.

The remainder of the paper is organised as follows. Sec. II summarizes related works in touch biometrics for mobile scenarios. Sec. III describes our proposed touch biometric system. Sec. IV describes the new e-BioDigit database, which comprises on-line handwritten numerical digits from 0 to 9. Sec. V and VI describe the experimental protocol and results achieved using our proposed approach, respectively. Sec. VII discusses specific details for the deployment of our proposed approach on current PIN- and OTP-based authentication systems, including password generation strategies. Finally, Sec. VIII draws the final conclusions and points out some future work lines.

II. RELATED WORKS

A. Handwriting Biometrics and Beyond

Touch biometrics are becoming a very attractive way to verify users on mobile devices [18], [20]. Table I summarizes relevant approaches in this area. For each study, we include information related to the verification method, features, classifiers and datasets considered. We also report in Table I the verification performance for the two impostor scenarios commonly considered in this area: i) imiation attack, the case in which impostors have some level of information about the user being attacked [26]; and 2) random attack, the case in which no information about the user being attacked is known. Note that most algorithms and experimental conditions vary between the listed works, e.g., the amount and type of training and testing data. Therefore, Table I should be mainly interpreted in general terms to compare different scenarios of use based on touch biometrics, but not individual algorithms.
In [10], Angulo et al. evaluated the use of lock pattern dynamic systems for user authentication. Users were asked to draw three different lock patterns a certain number of times (50 trials for each pattern), with each pattern consisting of six dots. Authors considered a total of 11 timing-related features extracted from the finger-in-dot time (i.e., the time in milliseconds from the moment the participant finger touches a dot to the moment the finger is dragged outside the dot area), and the finger-in-between-dots time (i.e., representing the speed at which the finger moves from one dot to the next) achieving results above 10.0% EER for imitation attacks.

In [11], Lacharme et al. incorporated biometric dynamic features related to the position of the finger, pressure, finger size and accelerometer sensor to the traditional Android unlock patterns, achieving a final 15.0% EER for imitation attacks using a matching algorithm based on Hamming Distance.

Zezschwitz et al. presented in [12] a similarity metric for Android unlock patterns to quantify the effective password space of user-defined gestures. The proposed metric was evaluated using 506 user-defined patterns revealing very similar shapes that only differ by simple geometric transformations such as rotation. Consequently, they presented an approach to increase the pattern diversity in order to strengthen user lock patterns.

Other studies have focused on the potential of keystroke biometrics for user authentication on mobile scenarios. In [13], Buschek et al. introduced qualitative aspects like personal expressiveness in order to enhance traditional keystroke biometric systems based on quantitative factors such as error rates and speed. They introduced a dynamic font personalisation framework, TapScript, which adapted a finger-drawn looking font according to user behavior and context, such as finger placement, device orientation, and position of the user while typing (i.e., walking or sitting) - resulting in a handwritten-looking font. Following their new approach, users were able to distinguish pairs of typists with 84.5% accuracy and walking/sitting scenarios with 94.8%. The same authors compared touching/sitting scenarios with 94.8%. The same authors compared touch-specific features between three different hand postures (i.e., one-thumb, two-thumb and index finger typing) and evaluation schemes: Gaussian Model without covariance (GM), k-Nearest-Neighbours (kNN) and Least Squares Anomaly Detection (LSAD). Authors concluded that spatial touch features reduces the Equal Error Rates (EER) by 26.4 - 36.8% compared to the traditional temporal features.

Biometric verification systems based on touchscreen gestures (i.e., scrolling, zooming and clicking) while using mobile devices in scenarios such as document reading, web surfing or free tasks are gaining a lot of impact nowadays [13], [16], [17], [18]. These approaches enable active or continuous authentication schemes, in which the user is transparently authenticated [27], [28]. Different features and algorithms

### TABLE I

COMPARISON OF DIFFERENT TOUCH BIOMETRIC APPROACHES FOR MOBILE SCENARIOS. ACC = ACCURACY.

| Study | Method | Features | Classifiers | Verification | Performance | Imitation | # Participants (Dataset) |
|-------|--------|----------|-------------|--------------|-------------|----------|-------------------------|
| Angulo et al. (2011) | Lock Pattern Dynamics | Timing-related Features | Random Forest | - | EER = 10.39% | 32 |
| Lacharme et al. (2016) | Lock Pattern Dynamics | Dynamic Features | Hamming Distance | - | EER = 15.0% | 34 |
| Zezschwitz et al. (2016) | Lock Pattern Dynamics | Shape Features | Greedy Clustering | - | - | 506 |
| Buschek et al. (2015) | Keystroke | Font Adaptation Features | Manual | Acc = 94.8% | - | 91 |
| Buschek et al. (2015) | Keystroke | Touch-specific Features | GM, kNN, LSAD | EER = 13.74% | - | 28 |
| Li et al. (2013) | Touchscreen Gestures | Static Features | SVM | EER = 3.0% | - | 75 |
| Sae-Bae et al. (2014) | Touchscreen Gestures | Distance between Points | DTW | EER = 1.58% | - | 34 |
| Shen et al. (2016) | Touchscreen Gestures | Static Features | SVM, Random Forest, kNN, Neural Networks | EER ~ 3.0% | - | 71 |
| Fierrez et al. (2018) | Touchscreen Gestures | Static Features | SVM, GMM | EER = 10.7% | - | 190 |
| Sae-Bae et al. (2014) | Handwritten Signatures | Histogram Static Features | Manhattan Distance | EER = 5.04% | - | 180 |
| Tolosana et al. (2017) | Handwritten Signatures | Dynamic Features | DTW | EER = 0.5% | EER = 17.9% | 65 |
| Khan et al. (2011) | Graphical Passwords | Predefined Symbols | Exact Match | - | - | 100 |
| Martinez-Diaz et al. (2016) | Graphical Passwords | Dynamic Features | DTW, GMM | EER = 3.4% | EER = 22.1% | 100 |
| Kutzner et al. (2015) | Handwritten Password | Static and Dynamic Features | Bayes-Nets, kStar, kNN | FAR = 10.42% | FRR = unknown | 32 |
| Nguyen et al. (2017) | Handwritten Digits | Dynamic Features | DTW | - | EER = 4.84% | 20 |
| Tolosana et al. (2018) | Handwritten Digits | Dynamic Features | DTW | - | EER = 5.5% | 93 |
| Proposed Approach | Handwritten Digits | Dynamic Features | DTW, RNNs | - | EER = 3.8% | 93 |
have been proposed in this field achieving very good results against random attacks. In [16], the authors proposed a set of 22 multitouch gestures using characteristics of hand and finger movements with an algorithm robust to orientation and translation achieving a final result of 1.58% EER. In [18], a set of 100 static features extracted from swipe gestures and systems based on Support Vector Machines (SVM) and Gaussian Mixture Models (GMM) were considered obtaining performances up to 10.7% EER. Very good results have been also achieved in [15], [17] using verification algorithms such as SVM, kNN, Random Forest and Neural Networks.

Handwritten signature is one of the most socially accepted biometrics as it has been used in financial and legal agreements for many years [29], [30], [31], [32], and it also finds applications in mobile scenarios. In [33], the authors explored the use of new algorithms based on RNNs on traditional desktop scenarios for pen-based signature recognition achieving results below 5.0% EER for imitation attacks. However, a considerable degradation of the system performance with results around 20.0% EER is obtained for imitation attacks when testing on mobile scenarios using finger touch as input [19], [20]. The main reason for such degradation of the system performance when using finger touch compared to pen-based desktop scenarios is the fact that users tend to modify the way they sign between both approaches, e.g., users who perform their signatures using closed letters with a pen input tend to perform much larger writing executions when using the finger. Besides, users whose signatures are composed of a long name and surname (or two surnames) tend to simplify some parts of their signatures due to the small surface of the screen to sign on. Graphical passwords were studied in [21], [22]. In [23], the authors proposed an approach based on graphical passwords (doodles) achieving final results above 20.0% EER for imitation attacks. The main reason for such degradation of the system performance lays down on the specific task that the user needs to perform to be authenticated, e.g., doodles were difficult to memorize for most of the users as they didn’t use them on a daily basis.

Finally, strongly related to the present work, in [23], [24] the authors proposed the use of handwritten passwords to be authenticated. In [23], Kutzner et al. asked the users to perform an 8-digit password on the screen of a tablet device. For each handwritten password, a total of 25 static and dynamic features were extracted and tested using many different authentication algorithms. However, the authentication scenario considered in that approach restricts the deployment of the technology in real mobile applications as: i) the authors considered a large number of training samples (12), and ii) it seems to be only applicable to devices with large screens (such as tablets) as it would be very difficult for the users to perform such a long password (8 digits) on a screen of much smaller size. In [24], Nguyen et al. evaluated the use of handwritten touch biometrics for PIN-based authentication systems. Their proposed authentication approach overcame some of the drawbacks previously cited as they asked users to draw each digit of the PIN one by one. A final 4.84% EER was achieved using a biometric system composed of 5 dynamic features and a matcher algorithm based on DTW. Finally, a preliminary study of the work presented here was published in [25]. In that work we performed an initial analysis of the touch biometric system only for OTP authentication schemes. In addition, DTW was the only approach considered for the similarity computation. The study presented here extends the preliminary analysis carried out in [25]. The main improvements over [25] are:

- Our preliminary touch biometric system in [25] based on DTW has been extended by incorporating RNN deep learning architectures. To the best of our knowledge, this is the first work to date that studies recurrent Siamese networks to model handwritten password authentication systems.
- Our analysis in [25] studied only OTP schemes. Here we also study PIN authentication, as depicted in Fig. 1.
- The system architecture includes 2 new blocks with respect to [25]: i) enrolment set, and ii) password generation; which are discussed in the text and evaluated experimentally.
- Sec. IV has been included to survey and compare advantages and limitations of recent research on touch biometrics for mobile authentication.
- The results achieved in the present study outperform our initial results presented in [25] with a final 3.8% EER, a relative improvement of 30.9% EER compared to [25]. This result outperforms other touch biometric approaches and considers fewer enrolment samples.
- Sec. VII has been included to discuss specific details for the deployment of our proposed approach on practical PIN and OTP authentication systems, including various strategies for password generation.

B. Two-Factor Password Authentication

The incorporation of biometric information on traditional password-based systems can improve the security through a second level of user authentication. Two-factor authentication approaches have been very successful in the last years. These approaches are based on the combination of two authentication stages. For example: i) the security system checks that the claimed user introduces its unique password correctly, and ii) its behavioral biometric information is used for an enhanced final verification [24], [34]. This way the robustness of the security system increases as impostors need more than the traditional password to get access to the system. This approach has been studied in previous works. In [10], the authors proposed a two-factor verification system based on timing-related features for dynamic lock patterns, achieving a final average EER of 10.39% for imitation attacks. A similar two-factor authentication approach was proposed in [11] for traditional Android unlock patterns but considering biometric dynamic features related to the position of the finger, pressure, finger size and accelerometer sensor achieving a final 15.0% EER for imitation attacks. Two-factor authentication approaches have also been expanded to physiological biometric traits. In [35], Jenkins et al. proposed a system based on features extracted for periocular images acquired using an iPhone 5, achieving very good results for the task of identification.
In this work we evaluate the potential of touch biometric verification systems based on time functions [36]. Signals captured by the digitizer (i.e., X and Y spatial coordinates) are used to extract a set of 21 time functions for each numerical digit sample (see Table II). Information related to pressure, pen angular orientations or pen up/downs broadly used in other biometric traits such as handwriting and handwritten signature is not considered here as this information is not available in all mobile devices when using the finger touch as input.

Sequential Forward Floating Search (SFFS) algorithm is used for the DTW algorithm in some of the experiments in order to select the best subsets of time functions for each handwritten digit and improve the system performance in terms of EER (%).

### III. Touch Biometric System

##### A. Digit-based Feature Extraction

In this work we evaluate the potential of touch biometric verification systems based on time functions [36]. Signals captured by the digitizer (i.e., X and Y spatial coordinates) are used to extract a set of 21 time functions for each numerical digit sample (see Table II). Information related to pressure, pen angular orientations or pen up/downs broadly used in other biometric traits such as handwriting and handwritten signature is not considered here as this information is not available in all mobile devices when using the finger touch as input.

Sequential Forward Floating Search (SFFS) algorithm is used for the DTW algorithm in some of the experiments in order to select the best subsets of time functions for each handwritten digit and improve the system performance in terms of EER (%).

### B. Similarity Computation

1) **Dynamic Time Warping**: DTW is used to compare the similarity between genuine and query input samples, finding the optimal elastic match among time sequences that minimizes a given distance measure. Scores are obtained as 

\[ \text{score} = e^{-D/K} \]

where \( D \) and \( K \) represent respectively the minimal accumulated distance and the length of the warping path [37].

2) **Recurrent Neural Networks**: some of the fields in which RNNs have caused more impact in the last years is in handwriting recognition and writer identification [38], [39]. This study explores the potential of RNNs for the task of handwritten passwords. In particular, we consider an adaptation of our original RNN systems proposed in [33] for the task of on-line handwritten signature verification. In that work we proposed RNN systems based on a Siamese architecture. The main goal was to learn a dissimilarity metric from data minimizing a discriminative cost function that drives the dissimilarity metric to be small for pairs of genuine samples from the same subject, and higher for pairs of samples coming from different subjects. Both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) systems were studied in [33], outperforming a state-of-the-art DTW system in challenging scenarios where skilled forgeries were considered. In addition, in [33] we also studied bidirectional schemes (i.e., BLSTM and BGRU), which allow access to future context, achieving much better results compared to the original schemes that only had access to past and present contexts.

In this study we adapt the original BLSTM system proposed in [33] to handwritten passwords for touchscreen biometrics. To the best of our knowledge, this is the first work to date that

---

**TABLE II**

| # | Feature |
|---|---------|
| 1 | X-coordinate: \( x_n \) |
| 2 | Y-coordinate: \( y_n \) |
| 3 | Path-tangent angle: \( \theta_n \) |
| 4 | Path velocity magnitude: \( v_n \) |
| 5 | Log curvature radius: \( \rho_n \) |
| 6 | Total acceleration magnitude: \( a_n \) |
| 7-12 | First-order derivative of features 1-6: \( \dot{x}_n, \dot{y}_n, \theta_n, v_n, \rho_n, \dot{\theta}_n \) |
| 13-14 | Second-order derivative of features 1-2: \( \ddot{x}_n, \ddot{y}_n \) |
| 15 | Ratio of the minimum over the maximum speed over a 5-samples window: \( v'_n \) |
| 16-17 | Angle of consecutive samples and first-order derivative: \( \alpha_n, \alpha'_n \) |
| 18 | Sine: \( s_n \) |
| 19 | Cosine: \( c_n \) |
| 20 | Stroke length to width ratio over a 5-samples window: \( r^5_n \) |
| 21 | Stroke length to width ratio over a 7-samples window: \( r^7_n \) |
The software for capturing handwritten numerical digits was a Samsung Galaxy Note 10.1 general purpose tablet. This device has a 10.1-inch LCD display with a resolution of 1280×800 pixels.

Regarding the acquisition protocol, subjects had to perform handwritten numerical digits from 0 to 9, one at a time. The acquisition setup and some examples of the handwritten numerical digits of the e-BioDigit database are depicted in Fig. 5. Additionally, samples were collected in two sessions with a time gap of at least three weeks between them in order to consider inter-session variability, very important for behavioral biometric traits [40]. For each session, users had to perform a total of 4 numerical sequences from 0 to 9 using the finger as input. Therefore, there are a total of 8 samples per numerical digit and user.

The software for capturing handwritten numerical digits was developed in order to minimize the variability of the user during the acquisition process. A rectangular area with a writing surface size similar to a 5-inch screen smartphone was considered, see Fig. 3(a). A horizontal line was represented on top of the drawing rectangular area, including two buttons “OK” and “Cancel” to press after writing if the sample was good or bad respectively.

The database comprises a total of 93 users. Regarding the age distribution, the majority of the subjects (85.0%) are between 17 and 27 years old, as the database was collected in a university environment (36.6% between 17 and 21). Regarding the gender, 66.7% of the subjects were males and 33.3% females whereas for the handedness distribution, 89.2% of the population was right-handed.

V. EXPERIMENTAL PROTOCOL
The experimental protocol designed in this study intends to cover all details of the two following main modules of our proposed password-based touch biometric system (see Fig. 1):

- **Enrolment Set:** When designing biometric authentication systems for real applications, there are usually two conflicting factors: i) the amount of data requested to the user during the enrolment, and ii) the security level provided by the biometric system. From the point of view of the security system, it seems clear that the ideal case would be to have as much information of the user as possible. However, in most real scenarios, the feasibility and success depend on the development of user-friendly applications.

This aspect has shown to be crucial for different tasks such as the handwritten signature. In [41], we evaluated this effect using statistical systems based on HMM and GMM, achieving an absolute improvement of 11.7% EER when training the user models with 41 genuine signatures instead of just 4. In this work, we analyze the intra-user variability on this new authentication scenario and perform a complete analysis of how the biometric system performance changes with the number of enrolment samples acquired per digit.

- **Password Generation:** The selection of a password that is robust enough for a specific application is a key factor. The number of digits that comprise the password depends on the scenario and level of security considered in the

Fig. 3. (a) Acquisition setup. (b-d) examples of different handwritten numerical digits of the e-BioDigit database. X and Y denote horizontal and vertical position versus the time samples.
Table III

| #    | Time-function description                        |
|------|-------------------------------------------------|
| 1    | X-coordinate: \( x_n \)                         |
| 2    | Y-coordinate: \( y_n \)                         |
| 7-8  | First-order derivate of features 1-2: \( x_n, y_n \) |
| 13-14| Second-order derivate of features 1-2: \( x_n, y_n \) |

Fig. 4. Examples of the digit 7 performed by two different users.

For the development of our proposed handwritten touch biometric systems, \( N \) genuine signatures (up to 4) from the first session can be used as enrolment samples, whereas the 4 remaining genuine samples from the second session are used for testing. Impostor scores are obtained by comparing the \( N \) enrolment samples with one genuine sample of each of the remaining users (simulating this way the imitation attack in which the impostor knows the password).

For the evaluation of our proposed touch biometric system, different scenarios are generally considered regarding the number of available enrolment samples per user (i.e., \( N_{vs1} \)), in which the final score is performed as the average score of \( N \) one-to-one comparisons. In addition, in case of using passwords composed of several digits, the final score is produced after averaging the different one by one digit score comparisons.

It is important to highlight that the inter-session variability problem is also considered in the experimental protocol carried out in this work as genuine digit samples from different sessions are used for enrolment and testing samples respectively. This effect has proven to be very important for many behavioral biometric traits such as the case of the handwritten signature [40].

VI. EXPERIMENTAL RESULTS

A. One-Digit Analysis.

This section analyzes the potential of each numerical digit (i.e., from 0 to 9) for the task of user authentication. We consider three different systems: i) a baseline DTW system, ii) an adapted DTW considering feature selection, and iii) a system based on RNNs.

Experimental results on the evaluation dataset for these three systems are shown in Table IV and V in terms of EER (%) for the cases of 1vs1 and 4vs1 comparisons, respectively.

1) DTW Baseline System: In order to provide an easily reproducible framework, we first consider a baseline system based on DTW with the same fixed time functions for all numerical digits. Table III shows the time functions selected, which are commonly used as baseline in other biometric traits such as the handwritten signature [20], [32].

Analyzing the first rows of Tables IV and V we can see how very good authentication results are obtained by the DTW Baseline System taking into account that we only consider one digit and the same time functions for all numerical digits.

Analyzing in Table IV the extreme scenario of having just one available digit sample during the enrolment (1vs1), the numerical digit 7 achieves the best result with 22.5% EER. In addition, other numerical digits such as 4 or 5 achieve similar results with EERs below 25.0%. This first experiment puts in evidence the discriminative power of each handwritten digit. Fig. 4 shows examples of the digit 7 performed by two
different users in order to observe the low intra- and high inter-user variability of this number. This effect is produced as different users tend to perform a specific digit in a different way, i.e., starting from a different stroke of the digit or even removing some of them such as the crossed horizontal stroke of the number 7.

Analyzing in Table IV the scenario of using four enrolment samples (4vs1), an average absolute improvement of 3.2% EER is achieved compared to the 1vs1 scenario showing the importance of acquiring more than one sample during the enrolment stage, if possible. For this scenario, the digit 4 achieves the best result with 18.0% EER.

2) **DTW Adapted System:** We now apply SFFS over the development dataset in order to enhance the DTW touch biometric system through the selection of specific time functions for each handwritten digit. Fig. 5 shows the number of times each time function is selected in our DTW Adapted System from the 21 total time functions described in Table II. In general, we can highlight the importance of $x_n$, $y_n$, time functions as they are selected for 70% of the numerical digits. In addition, time functions $x'_n$, $y'_n$ related to $X$ and $Y$ time derivatives seem to be very important as they are selected for near half of the digits. Other time functions such as $\rho_n$, $\rho'_n$, $\alpha_n$, and $s_n$ related to geometrical aspects of the numerical digits are proven not to be very useful to discriminate between genuine and impostor users.

The second rows of Tables IV and V show the results achieved for each digit using our DTW Adapted System over the evaluation dataset for both 1vs1 and 4vs1 cases, respectively. In general, better results are achieved compared to the DTW Baseline System. Analyzing the 1vs1 scenario, our DTW Adapted System achieves an average absolute improvement of 2.0% EER, being the numerical digit 5 the one that provides the best result with a 21.7% EER. Analyzing the 4vs1 scenario, our DTW Adapted System achieves an average absolute improvement of 1.6% EER, being again the numerical digit 5 the one that achieves the best result with a 16.9% EER. These results put in evidence the importance of considering different time functions for each digit in order to develop more robust biometric authentication systems against attacks.

3) **BLSTM System:** We now explore the potential of state-of-the-art deep learning technology applied to our touch biometric data. Our proposed end-to-end writer-independent BLSTM system is trained using only the 50 users of the development dataset. Samples from all numerical digits (i.e., from 0 to 9) and development users are considered together during training as we intend to distinguish between genuine and impostor handwritten digit samples regardless of the user and the numerical digit. This approach resulted in better generalisation results compared to the case of training one system per numerical digit. Therefore, our BLSTM system is trained considering two different cases: i) pairs of genuine
digit samples drawn by the same user, and ii) pairs of genuine and impostor digit samples, one performed by the claimed user and the other one by an impostor. For each case there are a total of 4 train samples × 4 test samples × 10 numerical digits × 50 users ≈ 8,000 comparisons, having the same number of genuine and impostor comparisons. Our BLSTM System has been implemented under Keras using Tensorflow as back-end, with a NVIDIA GeForce RTX 2080 Ti GPU. Adam optimizer is considered with a learning rate of 0.001 and a loss function based on binary cross-entropy.

The third rows of Tables [V] and [V] show the results achieved for each digit using our BLSTM System over the evaluation dataset for both 1 vs 1 and 4 vs 1 cases, respectively. In general, better results are achieved compared to the DTW Baseline System. Analyzing the 1 vs 1 scenario, our BLSTM System achieves an average absolute improvement of 1.4% EER, being the numerical digit 5 the one that provides the best result with a 19.6% EER. Analyzing the 4 vs 1 scenario, our BLSTM System achieves an average absolute improvement of 0.9% EER, being again the numerical digit 5 the one that achieves the best result with a 17.4% EER.

Finally, we compare our BLSTM System to the DTW Adapted System. In general, very similar results have been achieved for both authentication systems. For example, analyzing the 1 vs 1 case in Table [V] the BLSTM System has outperformed the DTW Adapted System for the 50% of the numerical digits (i.e., 0, 1, 5, 8, and 9), proving the potential of deep learning technologies even when just a single enrolment sample is considered. Despite these improvements, the DTW Adapted System outperforms slightly the BLSTM System in general, achieving an average absolute improvement of 0.5% and 0.7% EER for the 1 vs 1 and 4 vs 1 cases, respectively.

**B. Digit Combinations**

This section explores the robustness of our proposed approach when increasing the length of the password and also the number of available enrolment samples. The DTW Adapted System is considered in this analysis as it has outperformed the other systems studied. Regarding the type of digits that comprises the password, repetitions of the same numerical digits are allowed. However, the number of repetitions is restricted to 4, e.g., “2 5 8 8 8 8”. The reason for this limitation is motivated due to only 4 samples were acquired per digit during the second session of the e-BioDigit database. Table [VI] shows the evolution of the system performance in terms of EER (%) on the evaluation dataset when increasing the length of the handwritten password (from 1 to 8 digits) and also the number of available enrolment samples (from 1 to 4).

First, we analyze how the length of the handwritten password affects the system performance. In general, a considerable system performance improvement is achieved when adding more handwritten digits to the password. For example, for the case of having just one enrolment sample per user (1 vs 1), a password that is composed of just two handwritten digits achieves a 14.0% EER, an absolute improvement of 7.7% EER compared to the case of using a password with just one digit. This result is further improved when increasing the number of handwritten digits of the password with a final 8.5% EER for the case of considering a 6-digit password. However, there seems to exist a limit in the system performance improvement with the number of digits that comprise the password. In our experiments, the best results are obtained for passwords with a length of 6 and 7 digits.

Now, we analyze the effect of the number of available enrolment samples on the system performance. In general, the system performance improves with the number of enrolment samples. For example, for the case of having just one enrolment sample and a password composed of just one digit, the biometric system achieves a 21.7% EER. This result is further improved when increasing the number of enrolment samples to 4, achieving a final value of 16.9% EER, an absolute improvement of 4.8% EER. However, there seems to exist a limit in the system performance improvement with the number of enrolment samples. In our experiment, very similar results are obtained when considering 3 or 4 enrolment samples, achieving a final value of 3.8% EER when considering 3 enrolment samples and a handwritten password of 7 digits. This interesting finding is different compared to other behavioral biometric traits such as the handwritten signature as the system performance keeps improving even with large number of enrolment samples [41]. This effect may be due to the lower intra-user variability of our proposed touch biometric approach compared to other behavioral biometrics as well as the DTW similarity computation algorithm considered.

Finally, we pay attention to the content and the number of possible combinations of the best handwritten passwords using

**TABLE VI**

**EVOLUTION OF THE SYSTEM PERFORMANCE IN TERMS OF EER (%) ON THE EVALUATION DATASET. THE BEST SYSTEM PERFORMANCE ACHIEVED AND THE CORRESPONDING HANDWRITTEN DIGITS SELECTED ARE SHOWN THE CORRESPONDING HANDWRITTEN DIGITS SELECTED ARE SHOWN ON THE TOP AND BOTTOM OF EACH CELL RESPECTIVELY.**

| # Enrolment samples | # Digits that comprise the password |
|---------------------|-----------------------------------|
| 1                   | 1 [5], 2 [5, 8], 3 [5, 7, 9], 4 [1, 5, 7, 9], 5 [2, 5, 6, 7, 8], 6 [2, 3, 5, 6, 7, 8], 7 [1, 2, 3, 5, 6, 7, 8], 8 [2, 3, 4, 5, 6, 7, 8] |
| 2                   | 1 [5], 2 [5, 8], 3 [5, 7, 9], 4 [1, 5, 7, 9], 5 [2, 5, 6, 7, 8], 6 [2, 3, 5, 6, 7, 8], 7 [1, 2, 3, 5, 6, 7, 8], 8 [2, 3, 4, 5, 6, 7, 8] |
| 3                   | 1 [5], 2 [5, 8], 3 [5, 7, 9], 4 [1, 5, 7, 9], 5 [2, 5, 6, 7, 8], 6 [2, 3, 5, 6, 7, 8], 7 [1, 2, 3, 5, 6, 7, 8], 8 [2, 3, 4, 5, 6, 7, 8] |
| 4                   | 1 [5], 2 [5, 8], 3 [5, 7, 9], 4 [1, 5, 7, 9], 5 [2, 5, 6, 7, 8], 6 [2, 3, 5, 6, 7, 8], 7 [1, 2, 3, 5, 6, 7, 8], 8 [2, 3, 4, 5, 6, 7, 8] |
our proposed touch biometric system so as to achieve the best system performance. Table VI indicates in the bottom of each cell the best handwritten digits selected but not their order, as the final score of our proposed touch biometric system is produced after averaging the different one by one digit score comparisons. Therefore, for the case of having a password comprised of \( n \) digits, there are a total of \( n! \) possible password combinations (note that in our experiments we did not have any case of repetitions of digits achieving the best results).

C. Comparison to the State of the Art

Our proposed approach is now compared to other state-of-the-art biometric authentication approaches described in Table I. In order to perform a fair analysis, we compare our proposed approach to all studies that consider the same type of impostors, i.e., imitation attacks.

In general, our proposed approach achieves better results than other touch biometric approaches. For the case of lock pattern dynamic systems [10], [11], the best system performance reported was an average 10.39\% EER. Our proposed approach also outperforms other biometric methods such as the handwritten signature or graphical passwords [20], [22]. In [20], the authors proposed handwritten signature verification systems adapted to mobile scenarios, i.e., using mobile devices such as smartphones and tablets with the finger as input, achieving EERs around 20.0\%. In [22], the authors proposed the use of graphical doodles and pseudosignatures (i.e., simplified versions of the signatures drawn with the finger). EERs above 20.0\% were obtained in both cases for imitation attacks.

Finally, our proposed approach has been compared to other state-of-the-art authentication systems based on handwritten passwords. In [23], the authors proposed the use of handwritten passwords with a fixed length of 8 characters, achieving a final False Acceptance Rate (FAR) of 10.42\% when using a total of 12 training samples per user (the False Rejection Rate FRR was not provided by the authors). In [24], Nguyen et al. evaluated the potential of drawing each digit of a 4-digit PIN one by one, achieving a final result of 4.84\% EER when considering a total of 5 enrolment samples. Our proposed approach achieves a final value of 3.8\% EER and it is able to mitigate the limitations of [23] about the size of the touchscreen, as users perform numerical digits one at a time. Additionally, we only consider 3 enrolment samples and not 5 as in [24] in order to improve the usability of our approach.

VII. PASSWORD GENERATION AND SYSTEM SETUP

In this section we discuss specific details for the deployment of our proposed approach in real scenarios considering the same experimental protocol described in Sec. V. The DTW Adaped System has been considered for this analysis.

First, we focus on PIN-based systems. For this scenario, we propose to use passwords based on 4 digits as users have to memorize them and it is not feasible from the point of view of the user to consider longer passwords. Regarding the enrolment stage, we propose to request 3 enrolment samples per digit to each user. We consider this as something feasible for real applications as users would have to perform a total of \( 4 \times 3 \) samples/digit = 12 samples, i.e., 12 samples \( \times \) 2 seconds/sample \( \simeq \) 25 seconds.

Once we have fixed the number of enrolment samples and digits parameters, we design what type of passwords we let users to use (i.e., we design the Password Generation module in Fig. 1). The following cases are considered regarding both the system performance and number of possible combinations: i) ALL password combinations are allowed, and ii) only combinations using the BEST 4 digits selected in Table VI and with no repetitions (recall in Sect. VI-B we obtained that the most discriminative password combinations in terms of touch biometric information didn’t include repeated digits). Fig. 6 shows the EER distribution values obtained for all possible password combinations. On the box, the central mark indicates the median, and the left and right edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually. In general, we can see that the 75\% of password combinations provide results below 16.2\% EER. Analyzing the case ALL, the system performance results achieved are between 5.9\% and 35.7\% EER with a total of \( 10^4 \) combinations. The performance is improved in the case BEST with a 5.9\% EER for all considered combinations. However, users would be able to choose only among \( 4! \) combinations (i.e., 24). Besides, the security level of the first authentication stage would decrease as fewer password combinations would be possible. Therefore, a good choice could be to select all possible passwords that provide results in a range of EERs. For example, permitting between 5.9\% and 10.0\% EER. This approach would allow users to choose among 2,956 different 4-digit passwords.

Now, we analyze the OTP-based system. For this scenario, we propose to use passwords composed of 7 digits, similar to current OTP-based applications, as users do not have to memorize the password, i.e., the system is in charge of selecting and providing different passwords to the user each time is required. Regarding the enrolment stage, we also
propose to request 3 enrolment samples per digit so users would have to perform a total of 10 digits \( \times 3 \) samples/digit = 30 samples, i.e., 30 samples \( \times 2 \) seconds/sample \( \approx 1 \) minute.

Once we have fixed both the number of enrolment samples and the length of the password, we analyze the content of the passwords. For this scenario, the following cases are considered: i) ALL digit combinations are allowed, and ii) only combinations using the BEST 7 digits selected in Table VII with no repetitions. Table VII depicts the number of possible combinations as well as the EER (%) for both cases. Analyzing the case in which users can choose any possible combination, the system performance results achieved are between 3.8% and 14.0% EER. However, it is important to remark that for this case (longer passwords) results were obtained due to experimental restrictions using the SFFS algorithm and limiting the maximum number of digit repetitions to 4, so the final 14.0% EER might get a bit worse in practice when considering all possible digit combinations. This approach is further improved in the case BEST with a final 3.8% EER. For this scenario we propose to use this second case as there would be a total of 7! (i.e., 5,040) combinations that provide the best system performance for our proposed touch biometric approach.

VIII. CONCLUSIONS

This work evaluates the advantages and potential of incorporating handwritten touch biometrics to password-based mobile authentication systems. The new e-BioDigit database that comprises handwritten numerical digits from 0 to 9 were collected in two sessions for a total of 93 subjects. Handwritten numerical digits were acquired using the finger touch as the input on a Samsung Galaxy Note 10.1 general purpose tablet device.

For the new e-BioDigit database, we report a benchmark evaluation using two different state-of-the-art approaches: i) DTW in combination with the SFFS function selection algorithm, and ii) RNN deep learning technology. Both approaches have been compared, achieving very good results even for the case of using just a single enrolment sample. In addition, we perform a complete analysis of the touch biometric system regarding the discriminative power of each handwritten digit, and the robustness of our proposed approach when increasing the length of the password and the number of enrolment samples per user.

Our proposed approach achieves good results with EERs ca. 4.0% when considering imitation attacks, outperforming other traditional biometric verification traits such as the handwritten signature or graphical passwords on similar mobile scenarios. Additionally, we discuss specific details for the deployment of our proposed approach on current PIN- and OTP-based authentication systems.

Future work will be oriented to enlarge the current e-BioDigit database in order to consider lower- and uppercase letters and also to train more complex deep learning architectures.

ACKNOWLEDGMENTS

This work has been supported by projects: BIBECA (MINECO), Bio-Guard (Ayudas Fundación BBVA a Equipos de Investigación Científica 2017) and by UAM-CecaBank. Ruben Tolosa is supported by a FPU Fellowship from Spanish MECID.

REFERENCES

[1] M. Salehan and A. Negahban, “Social Networking on Smartphones: When Mobile Phones Become Addictive,” Computers in Human Behavior, vol. 29, no. 6, pp. 2632–2639, 2013.
[2] J. Bonneau, C. Herley, P. Oorschot, and F. Stajano, “The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes,” in Proc. IEEE Symposium on Security and Privacy, 2012, pp. 553–567.
[3] J. Galbally, I. Coisel, and I. Sanchez, “A New Multimodal Approach for Password Strength Estimation — Part I: Theory and Algorithms,” IEEE Transactions on Information Forensics and Security, vol. 12, pp. 2829–2844, 2017.
[4] A. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. Smith, “Smudge Attacks on Smartphone Touch Screens,” in Proc. of the 4th USENIX Conference on Offensive Technologies, 2010, pp. 1–7.
[5] D. Shakila, R. Kumar, A. Serwadda, and V. Ploha, “Beware, Your Hands Reveal Your Secrets!” in Proc. of the 2014 ACM SIGSAC Conference on Computer and Communications Security, 2014.
[6] Q. Yue, Z. Ling, X. Fu, B. Liu, W. Yu, and W. Zhao, “My Google Glass Sees Your Passwords!” in Black Hat USA, 2014.
[7] W. Meng, D. Wong, S. Furnell, and J. Zhou, “Surveying the Development of Biometric User Authentication on Mobile Phones,” IEEE Communications Surveys Tutorials, vol. 17, no. 3, pp. 1268–1293, 2015.
[8] L. Wen, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus, “Regularization of Neural Networks using DropConnect,” in Proc. of the 30th International Conference on Machine Learning, 2013, pp. 1058–1066.
[9] M. Liang and X. Hu, “Recurrent Convolutional Neural Network for Object Recognition,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3367–3375.
[10] J. Angulo and E. Westlund, “Exploring Touch-Screen Biometrics for User Identification on Smart Phones,” J. Camenisch, B. Crispo, S. Fischer-Hübner, R. Leenes, G. Russello (Eds.), Privacy and Identity Management for Life, Springer, pp. 130–143, 2011.
[11] P. Lacharme and C. Rosenberger, “Synchronous One Time Biometrics With Pattern Based Authentication,” in Proc. 11th Int. Conf. on Availability, Reliability and Security, ARES, 2016.
[12] E. von Zerschwitz, M. Eiband, D. Buschek, S. Oberhuber, A. D. Luca, F. Alt, and H. Hussmann, “On Quantifying the Effective Password Space of Grid-based Unlock Gestures,” in Proc. of the International Conference on Mobile and Ubiquitous Multimedia, 2016, pp. 201–212.
[13] D. Buschek, A. D. Luca, and F. Alt, “There is more to Typing than Speed: Expressive Mobile Touch Keyboards via Dynamic Font Personalisation,” in Proc. of the International Conference on Human-Computer Interaction with Mobile Devices and Services, 2015, pp. 125–130.
[14] ——, “Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices,” in Proc. of the CHI Conference on Human Factors in Computing Systems, 2015, pp. 1393–1402.
[15] L. Li, X. Zhao, and G. Xue, “Unobservable Reauthentication for Smartphones,” in Proc. 26th Network and Distributed System Security Symposium, NDSS, 2013.
[16] N. Sae-Bae, N. Memon, K. Ishibster, and K. Ahmed, “Multitouch Gesture-Based Authentication,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 4, pp. 568–582, 2014.
[17] C. Shen, Y. Zhang, X. Guan, and R. Maxion, “Performance Analysis of Touch-Interaction Behavior for Active Smartphone Authentication,” IEEE Transactions on Information Forensics and Security, vol. 11, no. 3, pp. 498–513, 2016.
[18] J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales, “Benchmarking Touchscreen Biometrics for Mobile Authentication,” IEEE Trans. on Information Forensics and Security, vol. 13, 2018.

2https://github.com/BiDAlab/eBioDigitDB
[19] N. Sae-Bae and N. Memon, “Online Signature Verification on Mobile Devices,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 6, pp. 933–947, 2014.

[20] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-García, “Benchmarking Desktop and Mobile Handwriting across COTS Devices: the e-BioSign Biometric Database,” PLOS ONE, 2017.

[21] W. Zhou, M. Aalsalem, and Y. Xiang, “A Graphical Password Based System for Small Mobile Devices,” International Journal of Computer Science, vol. 5, no. 2, pp. 145–154, 2011.

[22] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, “Benchmarking the Accuracy of BioSign and Other Online Signature Verification Systems,” in Proc. 5th International Conference on Information Security, ICIS, 2013.

[23] T. Nguyen, N. Sae-Bae, and N. Memon, “DRAW-A-PIN: Authentication using finger-drawn PIN on touch devices,” Computers and Security, vol. 66, pp. 115–128, 2017.

[24] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, “Incorporating Touch Biometrics to Mobile One-Time Passwords: Exploration of Digit Strokes,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.

[25] A. K. Jain, A. Ross and P. Flynn (Eds.), Handbook of Biometrics, Springer, pp. 189–209, 2008.

[26] R. Plamondon and S. Srihari, “On-Line and Off-Line Handwriting Recognition: a Comprehensive Survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 63–84, 2000.

[27] R. Plamondon, G. Pirlo and D. Impedovo, “Online Signature Verification,” D. Doermann and K. Tomber (Eds.), Handbook of Document Image Processing and Recognition, Springer, pp. 917–947, 2014.

[28] M. Díaz, M. Ferrer, D. Impedovo, M. Malik, G. Pirlo, and R. Plamondon, “A Perspective Analysis of Handwritten Signature Technology,” ACM Computing Surveys, pp. 103–112, 2018.

[29] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-García, “Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics,” IEEE Access, vol. 6, pp. 5128–5138, 2018.

[30] A. Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, “Touch Me Once and I Know It’s You! Implicit Authenticiation based on Touch Screen Patterns,” in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2012, pp. 987–996.

[31] J. Jenkins, J. Shelton, and K. Roy, “One-Time Password for Biometric Systems: Disposable Feature Templates,” in Proc. SoutheastCon, 2017.

[32] R. Tolosana, R. Vera-Rodriguez, J. Ortega-García, and J. Fierrez, “Pre-processing and Feature Selection for Improved Sensor Interoperability in Online Biometric Signature Verification,” IEEE Access, vol. 3, pp. 478–489, 2015.

[33] M. Martinez-Diaz, J. Fierrez and S. Hangai, “Signature Matching,” S.Z. Li and A. Jain (Eds.), Encyclopedia of Biometrics, Springer, pp. 1382–1387, 2015.

[34] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, “A Novel Connectionist System for Unconstrained Handwriting Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 855–868, 2009.

[35] J. Galbally, D. Impedovo, and Y. Bengio, “End-To-End Online Writer Identification With Recurrent Neural Network,” IEEE Transactions on Human-Machine Systems, vol. 47, pp. 103–112, 2017.

[36] J. Galbally, M. Martinez-Diaz, and J. Fierrez, “Aging in Biometrics: An Experimental Analysis on On-Line Signature,” PLOS ONE, vol. 8, no. 7, p. e69897, 2013.

[37] R. Tolosana, R. Vera-Rodriguez, J. Ortega-García, and J. Fierrez, “Update Strategies for HMM-Based Dynamic Signature Biometric Systems,” in Proc. 7th IEEE Int. Workshop on Information Forensics and Security, 2015.

[38] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez, “Performance Evaluation of Handwritten Signature Recognition in Mobile Environments,” IET Biometrics, vol. 3, pp. 139–146, 2014.