Behavioural Authentication Based on Smartphone Protected Personal Communication Data

Takoua Guiga, Christophe Rosenberger, Jean-Jacques Schwartzmann

Behavioural Authentication Based on Smartphone Protected Personal Communication Data

Takoua Guiga[1,2], Christophe Rosenberger[2], Jean-Jacques Schwartzmann[1]
[1] Orange Labs, France
[2] Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

***Abstract*—Smartphones have become ubiquitous in everyday life, storing and generating a huge amount of sensitive personal data which make them vulnerable to increasing security and privacy threats. While protecting smartphones has become a necessity, existing traditional authentication methods, which are mainly PINs and passwords, are facing remarkable drawbacks and behavioural biometrics-based authentication was adopted as the best alternative to ensure better protection. This paper presents a comparative study of many behavioural authentication solutions using smartphone personal communication data. Different approaches are compared such as using Distance Minimization, K-means and Support Vector Machine (SVM) as classification method. The data privacy protection by using the BioHashing algorithm is also considered in the paper. The authentication approaches were tested on a dataset of 93 users with more than 16.000 samples and show promising results with an EER of 10% without any data protection with the One Class SVM method and an EER remarkably lower than 1% for the 3 adopted methods with data privacy protection.**

***Index Terms*—Behavioural Authentication, Smartphone, Privacy protection, BioHashing, Classification methods.**

I. Introduction

For the time being, what is not irritating and won't let you down? That true friend: your smartphone, which has become an essential companion in our daily life, on which we rely to do practically everything. The one that knows you more than anyone else by holding a huge amount of your personal information such as account credential and credit card details, business information, social media accounts, emails, images, voices,etc. Yet, that will remains true until the battery runs out, being lost or stolen, infected by malware or attacked by social engineering.

Hereby, the boundless use of smartphones for extended range of activities, and the increasing amount of sensitive personal data that are stored and generated, gives rise to security and privacy threats. Actually, any unauthorized access to this device, could have serious consequences and may turn it form a friend to a nightmare. According to Malware Statistics, trends and facts in 2019, presented by the security detective website [1], malware attacks have been increased for the last ten years across all mobile platforms, from 12.40 malware in 2009 to 812.67 malware in 2018, additionally to the augmented number of stolen or lost smartphones over the last few years. Thus, protecting this device is becoming a necessity.

Existing traditional authentication methods, which are mainly PINs codes and passwords, are facing remarkable drawbacks, while remaining vulnerable to different types of attacks as mentioned in the recent literature [2]. So, they are neither notably appropriate to use, as they are regularly forgotten, nor perfectly secure, since they are vulnerable enough to be guessed or stolen. This fact induces a usability problem as well, since 70% of smartphone users consider PINs and passwords really annoying and prefer not using them, reported by a recent study [3].

Hence, in order to strengthen the security of these devices, researchers have expanded their interest in developing more performing authentication mechanisms based on biometric modalities, considering them unique to a single person and nearly impossible to compromise. It is about exploiting different data based on measurements and characteristics of user's body parts, called physiological modalities such as, face recognition, fingerprint, hand geometry, iris recognition, etc. However, none method is without limitation, it has been proved that these modalities have several drawbacks. Basically, physiological biometric data can be spoofed [4] and are susceptible to potential privacy pitfalls. As discussed in [5], it is true that biometric data are unique identifiers but they are not secret: fingerprint is leaved on everything we touch, faces can be easily acquired and voice can be simply recorded. Thus, the potential collection and use of biometric data without the knowledge of its owner, without his/her consent or personal control makes this information very sensitive. Consequently, always in order to improve authentication methods, behavioural biometric modalities have been revealed as the analysis of a behavioural trait, learned and acquired over time, considering the manner in which people react and how they perform something such as walking (Gait Recognition) [6], signing (Signature Recognition) [7], and typing on the keyboard (Keystroke dynamics) [8].

One of the main advantages of behavioral biometrics is being dependent on the user actions and habits, which makes them perfect candidates for transparent user authentication

[9], as samples are recorded seamlessly. According to Nathan Clarke [10], transparent authentication can be achieved by any authentication approach that is able to obtain the sample required for verification non-intrusively.

This paper presents, in one hand, a comparative study of behavioural authentication using smartphone communication data by using different methods such as distance Minimization, K-means and Support Vector Machine (SVM) as classification techniques. On the other hand, this paper analyses the efficiency of data privacy protection by adopting the biohashing algorithm. In this study, we use real data from 93 users with more than 16.000 samples.

Section II describes the related works on behavioral authentication solutions on mobile devices. Section III presents the used classification methods (distance minimization, Kmeans, SVM) and the BioHashing algorithm for the privacy data protection. Section IV describes the proposed protocol and details the used dataset. Section V reports obtained results and section VI concludes the paper and present some perspectives and future works.

II. Related works

This section is dedicated to a state of the art both for behavioral authentication on mobile devices and privacy issues of biometrics.

Behavioral authentication solutions that provide transparent authentication are a fast growing area. This is especially due to the Active Authentication project [11]. The Defense Advanced Research Projects Agency offers to move beyond password by using transparent authentication mechanism. This means most users will authenticate themselves using biometric sensors.

Thus, Google announced in May 2016 the Abacus project [12], a multi-modal, seamless and continuous authentication system designed to replace the login/password pair. The authors of [13] proved that combining the location with a standard authentication increases the global trust in that authentication. In addition, this article shows that the two main locations arising for a user are home and workplace. This implies to continuously know where the user is and therefore compromises users privacy. The location property and especially the one offered by the Global Positioning System (GPS) sensors embedded in modern smartphones represents relevant features.

The authors in [14] offer a solution to authenticate users using the geolocation and the phone calls. They obtain an EER of 5.4% with the 6 last phone calls. However, the privacy aspect is not taken into account. In [15], the authors combine different authentication modalities and also include the text message content. To proceed with the text message information, the messages must be read. This implies a privacy leakage.

Less sensitive data can be exploited to perform behavioral authentication. This is the case of gait recognition [16]. However, the authors in [17] have shown that combining location information with gait recognition increases the global performance of the system. By combining those data, they obtained an ERR of 10% on a dataset of 13 users. However, privacy protection is not taken into account.

To the best of the authors knowledge, there are few proposed solutions in the literature dealing with privacy concerns. The authors in [18] use an homomorphic encryption scheme. In [19], the authors address the problem of online authentication using implicit information and store the data directly on the mobile phone, thus delegating the authorization server role to the mobile phone. This permits to mitigate the privacy problem but does not solve the cancellability issue.

In [20], the authors considered privacy issues in the designing of an authentication scheme based biometric features. This solution permits to solve both the privacy problem and cancellability issue. This paper aims at improving this solution by using less sensitive information (phone calls statistics). In the next section, different classification methods are compared to reach this objective.

III. Proposed methods

This section first presents different classification techniques used in the authentication system in order to evaluate their relative performance. We start with a simple distance based minimization method (Distance minimization), then applying an unsupervised classification algorithm (K-means), to finish with a supervised classification algorithm (SVM). Second, we focus on privacy protection.

A. Classification methods

1) Distance minimization: It represents a simple technique consisting in affecting the unknown individual to the class having the most similar individuals. Considering a dataset (X_i,U_i), i=1:n with U_i having K different values (K classes). An unknown individual is affected to the class J if the distance between its parameters and an individual belong to the class J is minimal among all individuals.

2) K-Means: It is one of the most popular unsupervised algorithm. Generally, unsupervised algorithms make decisions from datasets using only input vectors without implying known or labelled classes. K-means stores k centroids that uses to define clusters. A data point is considered to be in a particular cluster if it is closer to that cluster's centroid than any other centroid. Based on [21], in the clustering problem,a training set x^1,...,x^m is given, and the goal is to group the data into a few cohesive clusters. Principally, For given feature vectors for each data point $x^i \in \mathrm{R}_n$, the intention is to predict k centroids

and a label C^i for each data point. The k-means clustering algorithm is as follows:

- Initialize cluster centroids u_1,u_2,...,$u_k \in \mathrm{R}^n$ randomly
- Repeat until convergence:
  For every i, set

$$C^i = \arg\min_j(||x^i - u_j||^2) \tag{1}$$

  For each j, set

$$u_j = \frac{\sum_{i=1}^{m} 1\{C^i = j\}x^i}{\sum_{i=1}^{m} 1\{C^i = j\}} \tag{2}$$

3) SVM: The Support Vector Machine (SVM) classification algorithm, is among the best supervised learning algorithms and is widely used in many types of applications [22]. This technique is a two-class classification method that aims generally to separate negative data from positive ones. The algorithm then searches for the hyperplane (in the linear case) that ensures this separation by maximizing the margin distance between the two classes. Given a set of learning data (x_i, y_i) for i=1..n (n = size of data), with $x_i \in \mathrm{R}^d$ and $y_i \in \{-1, 1\}$, train a classifier to find:

$$\mathbf{f}(x_i) = \begin{cases} < 0 & \text{if } y_i\text{= -1} \\ \geq 0 & \text{if } y_i\text{= 1} \end{cases} \tag{3}$$

The decision function f(x_i) depends on whether the data are linearly separable or not. When data are linearly separable, the decision function is as follows:

$$\mathbf{f}(x_i) = \sum_{i=1}^{n} w_i x_i + b \tag{4}$$

where $w_i \in \mathrm{R}^n$ is the weight vector, b the bias, and x_i the data variable. In most classification problems, the data are non-linearly separable. So the solution to classify this data is to project them into a larger space where the data becomes linearly separable using a function called Kernel K, and the decision function becomes:

$$\mathbf{f}(x_i) = \sum_{i=1}^{n} w_i K(x_i, x) + b \tag{5}$$

The kernel depends on the number of data and the complexity of the presented problem and it can be Linear, Sigmoid, Polynomial or RBF as shown in the following table I:

TABLE I
SVM KERNEL TYPES

Kernel's Type	Function formula
Linear	K(x,y)= x.y
Sigmoid	K(x,y)= tanh(ax.y+b)
Polynomial	K(x,y)= $(ax.y + b)^d$
RBF	K(x,y)= exp $(-\|\|x - y\|\|^2 / \sigma^2$

B. Privacy protection

In this paper, the use of smartphone communication data for the user authentication is investigated. As the verification process could be done by a server considered as honest but curious, a privacy protection of collected data is required.

The concept of privacy protection of biometric data has been defined in 2001 in a seminal paper [23]. Since then, many methods have been proposed among random projections approaches [24], BioHashing methods [25], Bloom filters [26], to cite just a few. A complete review of cancelable biometric systems can be found in [27]. Very recently, Teoh *et al.* [28] proposed a new two-factor scheme to protect the biometric template by transformation. Compared with previous works, this method is based on localized random projection and on the rank correlation. Moreover, the obtained results show that this system is strongly resistant against the main attacks. These good results are the consequences of their technical called *Index-Of-Max* which can be viewed as a machine learning on the plain database. For this previous constraint, the comparaison to this method is not adopted where the BioSystem is tuned for a particular basis. More generally, a security analysis of the biometric system protecting the biometric template based on transformations [29] are considered. In the following, the BioHashing algorithm is particularly detailed [25], one of the popular template protection schemes.

The BioHashing algorithm is applied on biometric templates that are represented by real-valued vectors of fixed length (so the metric used to evaluate the similarity between two biometric features is the Euclidean distance). It generates binary templates of length lower than or equal to the original length (here, the metric D_T used to evaluate the similarity between two transformed templates is the Hamming distance). This algorithm has been originally proposed for face and fingerprints by Teoh *et al.* in [25]. Then, the BioHashing algorithm transforms the biometric template $T = (T_1, \ldots T_n)$ into a binary template $B = (B_1, \ldots B_m)$, with $m \leq n$, as following:

The specificity of the BioHashing algorithm is that it uses a one way function and a random seed of m bits. It is important to note that every enrolled biometric feature uses a different seed in order to create a specific BioCode. The performance of this algorithm is ensured by the scalar products with the orthonormal vectors. The quantization process of the last step ensures the non-invertibility of the data (even if n = m, because each coordinate of the input T is a real value, whereas the coordinates of the output B is a single bit). Finally, the random seed guarantees both the diversity and revocability properties.

Algorithm 1 BioHashing

Inputs
$T = (T_1, \ldots, T_n)$: biometric template,
K_z: secret seed
Output $B = (B_1, \ldots, B_m)$: BioCode
Generation with the seed K_z of m pseudorandom vectors $V_1, \ldots, V_m$ of length n,
Orthogonalize vectors with the Gram-Schmidt algorithm,
for $i = 1, \ldots, m$ **do** compute $x_i =< T, V_i >$.
end for
Compute BioCode:

$$B_i = \begin{cases} 0 & \text{if} \quad x_i < \tau \\ 1 & \text{if} \quad x_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

IV. Proposed protocol

A. The smartphone personal dataset

The idea is to use behavioral data from a smartphone communication to authenticate a user. Thus, a real private dataset is collected, including the communication behavior of 93 users recorded during one month. Each user has 8 collected information as follows:

- The start date and time of a call;
- The phone number of the caller;
- The phone number of the callee;
- The type of communication (Text message or phone call);
- The number of consumed units in a communication (seconds for a phone call, or number of SMS);
- The type of a call (outgoing call, incoming call);
- The latitude of a cell;
- The longitude of a cell;

A set of the 8 values defines each user profile representing his/her behavior based on a phone communication as shown in figure 1

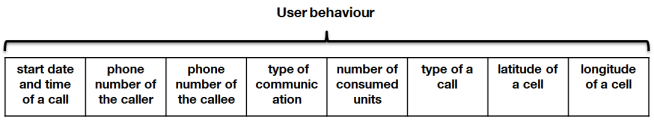

Fig. 1. User behaviour based on smartphone communication data

The number of samples is different from a user to another, since data is collected in a real and unique way for each user (different numbers of calls, different callee, etc.). The total number of samples is 16143. Samples are then divided equally and randomly swapped into two sets: the reference dataset, dedicated to train the model and the test dataset is used to provide an evaluation of the model.

B. Protocol description

In this section, we describe the experimental protocol used to evaluate the proposed solution both in terms of performance and in terms of privacy protection.

In order to evaluate the capacity of the authentication system for a user based on his/her behaviour dataset described in section IV-A, three classification techniques have been used: The distance minimization technique, the K-means algorithm and the support vector machine (SVM) algorithm, as mentioned in section III-A.

It is about calculating a similarity score between samples of the test dataset and the model created by the reference dataset. The lower the score is, the higher is the similarity between them. It is checked whether the score higher than a certain threshold to correctly authenticate a user. For each user, two different scores are computed: the legitimate scores, comparing test samples of one user to the model of the same user and the impostor scores, comparing samples of all users with the model of one user.

The performance of the authentication system is evaluated by three main measures which are the False Acceptance Rate (FAR) defined as the proportion of times a system grants access to an unauthorized person, the False Rejection Rate (FRR) which is the proportion of times a biometric system fails to grant access to an authorized person and the Equal Error Rate (EER) defining the common value where the FAR and the FRR values intersect. The lower the EER is, the better is the system performance.

Two scenarios are adopted:

- Scenario 1: evaluating the system performance without data protection
- Scenario 2: evaluating the system performance with data protection

First, the three classification methods are performed separately in order to compare their performances using the behavioural dataset in terms of EER value (which has to be minnimized). Then, the Biohashing algorithm is applied in the interest of protecting biometric data privacy and to analyze its impact on classification algorithms in terms of authentication performance.

Furthermore, as the reference dataset and the test dataset are constructed with a random draw for a set ratio. It is highly recommended to average the resulting EER value for each random draw. The ratio is the percent of dividing the training data and testing data. So, for a fixed value of a ratio, the EER value is computed n times, and then the average of the n obtained values is considered as the final EER value assigned to a fixed ratio.

V. Results

This section presents the experimental results obtained in order to evaluate the system performance for both proposed

scenarios.

1) Scenario 1: evaluating the system performance without any data protection: This first scenario basically evaluates the performance of the different classification algorithms presented in section III without any data protection, to figure out the best approach for a better authentication.

Distance minimization is implemented with a simple computation of the euclidean distance between the test dataset and the model described in the proposed protocol. A one Class SVM is used with RBF kernel which gives better results than other kernel types. K-means algorithm is computed with K=3 centroids. Results are presented in terms of EER values for a ratio = 80% (thus, 20% of data are used for testing) and are summarized in table III.

TABLE II
COMPARING EER VALUES OF DISTANCE MINIMIZATION, K-MEANS AND ONE CLASS SVM , WITHOUT DATA PROTECTION

Classification methods	EER (%)
Distance Minimization	13.9
K-means	37
One class SVM	10.4

The ROC curves presenting FRR values against FAR values for the 3 evaluated methods are exposed in figure 2.

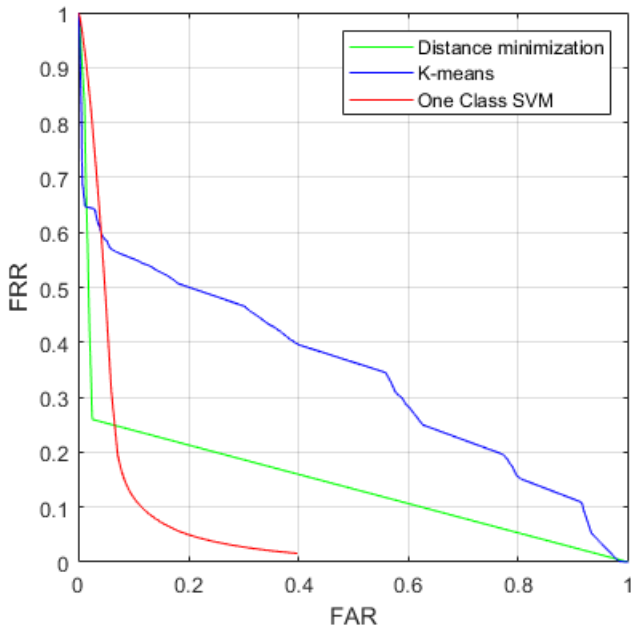

Fig. 2. ROC curves for Distance Minimization, K-means and One class SVM methods (ratio=0.8)

It is clear that better results are obtained with the One class SVM classifier with an EER= 10.4%. Furthermore, the EER values depend on the used ratio value. For this reason, the following figures 3, 4, 5 show respectively the EER variation depending on the ratio variation for the distance minimization, K-means and the one class SVM methods, in order to determine the best ratio for every method. Clearly for the 3 methods, the higher is the ratio, the lower is the EER.

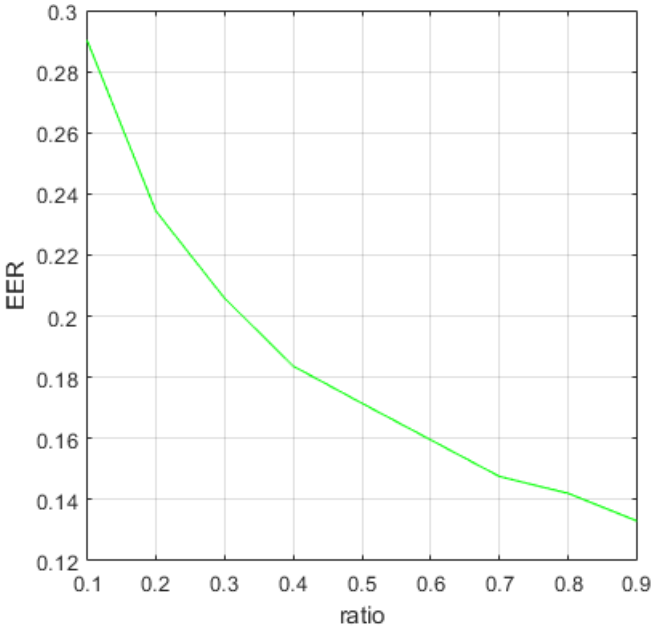

Fig. 3. EER variation for Distance minimization method

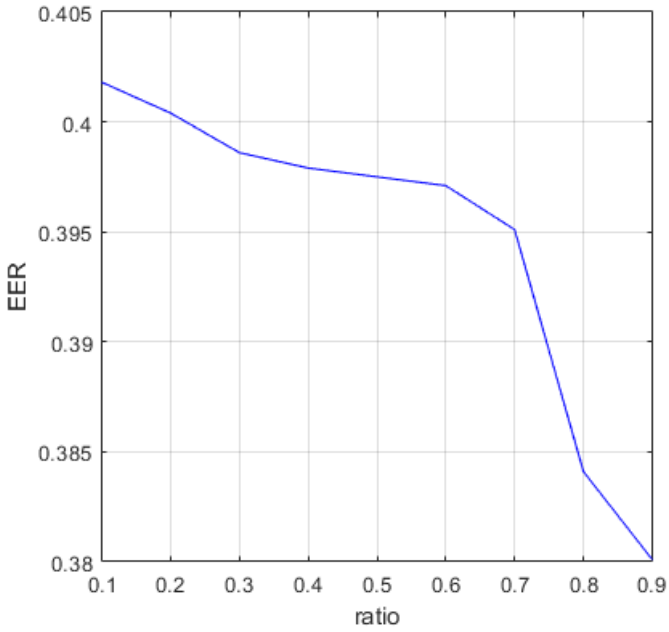

Fig. 4. EER variation for K-means method

2) Scenario 2: evaluating the system performance with data protection: In this second scenario, the data privacy is considered. Collected data are protected with the Biohashing algorithm. The Biohashing algorithm is implemented with the

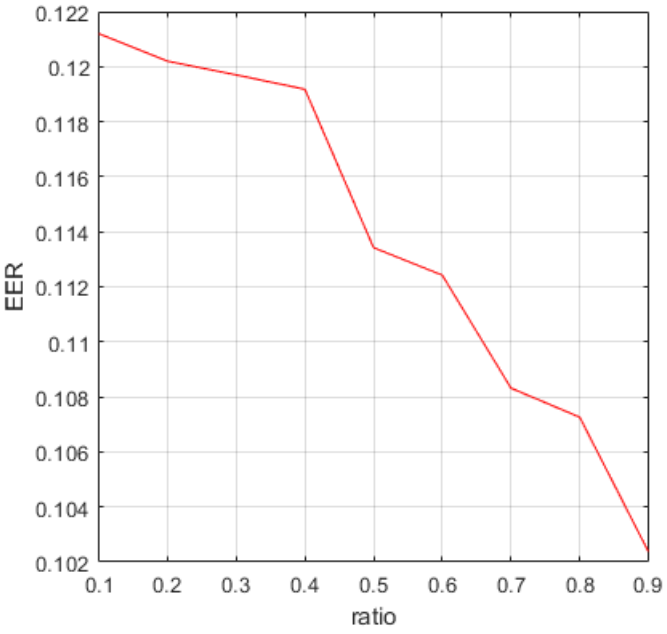

Fig. 5. EER variation for One class SVM method

3 classification methods, considering the user identifier as a unique seed for every user (for example, for user 1, seed = user identifier = 1 and for user 93, seed = user identifier = 93) and the BioCode is of length n, with [m,n]= size(data) as mentioned in section III-B. BioCodes of different users are presented in the form of a barcode as shown in figure 6 . In real operations, we can use a PIN code as seed to protect the collected data to generate the BioCode.

Fig. 6. Barcodes of different users BioCodes

Experimental results with the 3 classification methods are presented in terms of EER values for a ratio = 0.8 and are summarized in table III.

TABLE III
COMPARING EER VALUES OF DISTANCE MINIMIZATION, K-MEANS AND ONE CLASS SVM , WITH DATA PRIVACY PROTECTION

Classification methods	EER (%)
Distance Minimization	0.21
K-means	0.74
One class SVM	0.19

The ROC curves presenting FRR values against FAR values for the 3 evaluated methods with the BioHashing algorithm are exposed in figure 7.

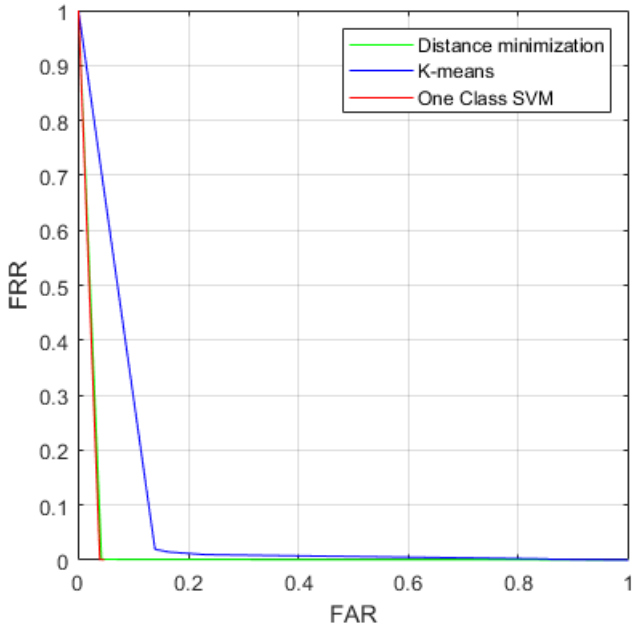

Fig. 7. ROC curves for Distance Minimization, Kmeans and one class SVM with Biohashing (ratio=0.8)

It is clearly remarkable that the performance of the 3 classification methods with the BioHashing algorithm is improved compared to the unprotected case. One class SVM remains the best with an EER= 0.10 %, but we can also notice that the performance of the K-means algorithm with privacy protection has improved significantly and the EER goes from a value of 37% to 0.7% , which is not the best result but is still interesting.

To better compare the difference, figures 8, 9 and 10 show respectively the performance with and without privacy protection of Distance Minimization method, K-means and One class SVM method.

Also, as studied in the first scenario, the EER values vary according to the ratio values. For that, figure 11, 12,13 display the variation of EER depending on the ratio values with data privacy protection respectively for Distance Minimization method, K-means and One class SVM method.

We can see clearly that the proposed solutions with protection with the BioHashing algorithm provide very good results.

VI. CONCLUSION AND FUTURE WORK

This study presented a behavioural biometric authentication approach based on smartphone personal communication data in order to improve smartphone security and privacy protection of sensitive data by providing a robust user authentication. For this purpose, three classification methods

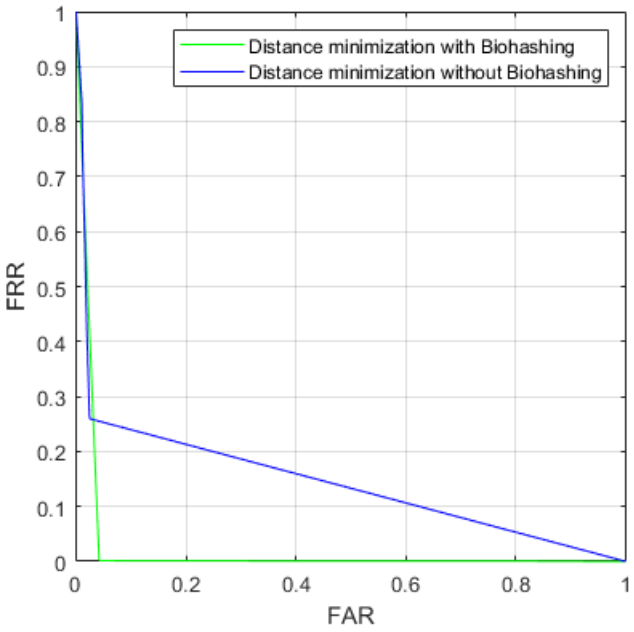

Fig. 8. ROC curves for Distance Minimization with and without Biohashing (ratio=0.8)

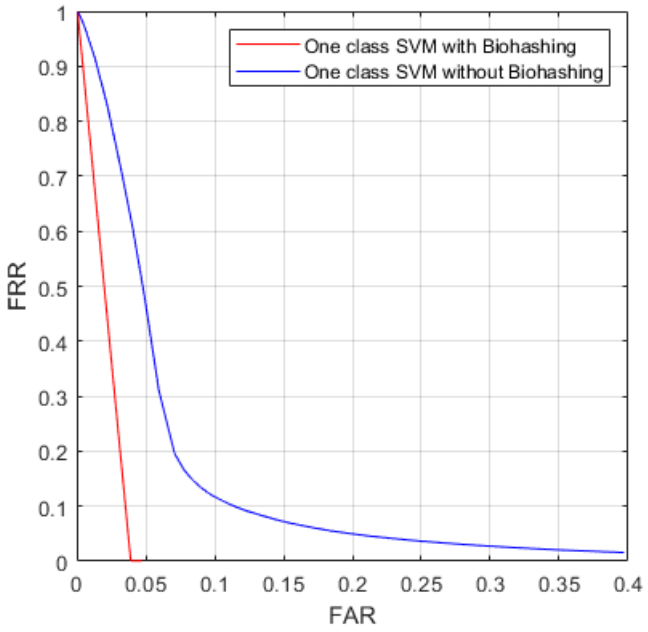

Fig. 10. ROC curves for One Class SVM with and without Biohashing (ratio=0.8)

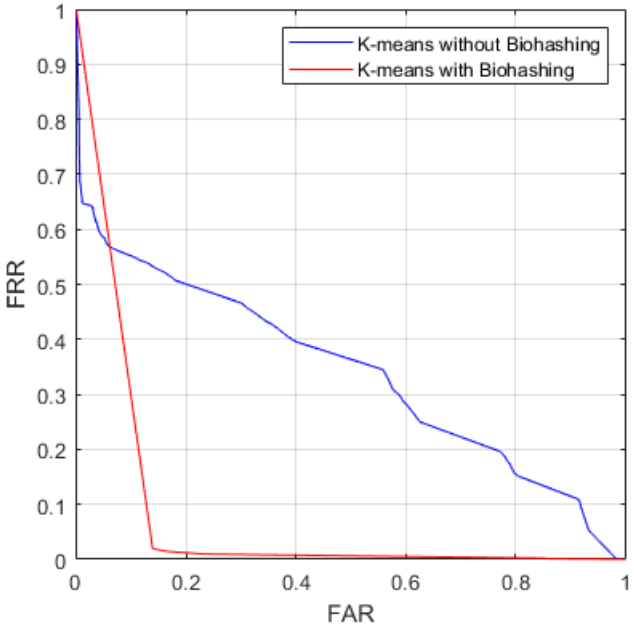

Fig. 9. ROC curves for K-means with and without Biohashing (ratio=0.8)

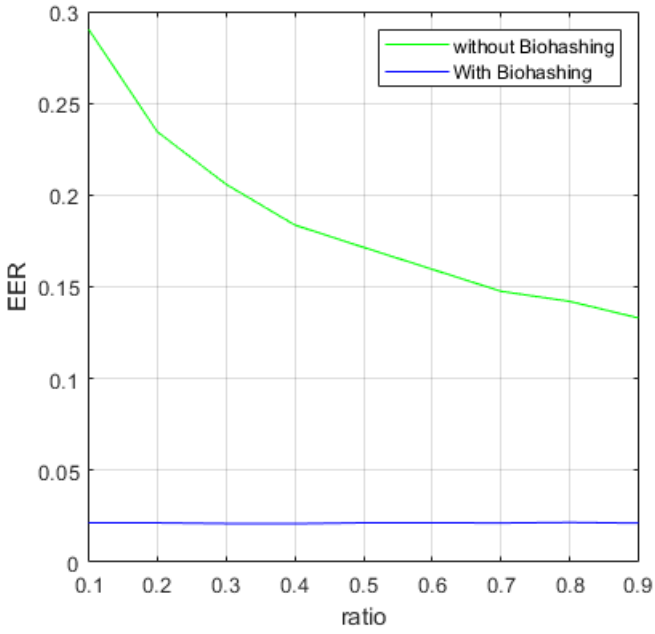

Fig. 11. EER variation for Distance Minimization method with and without Biohashing

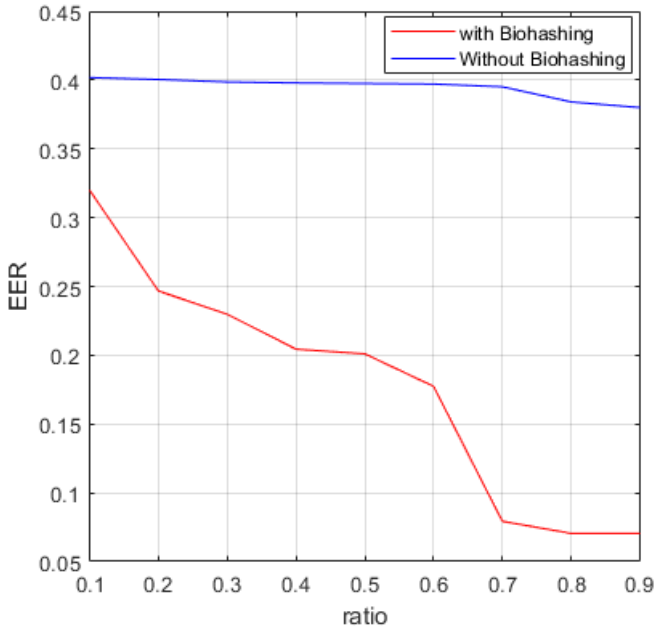

Fig. 12. EER variation for K-means method

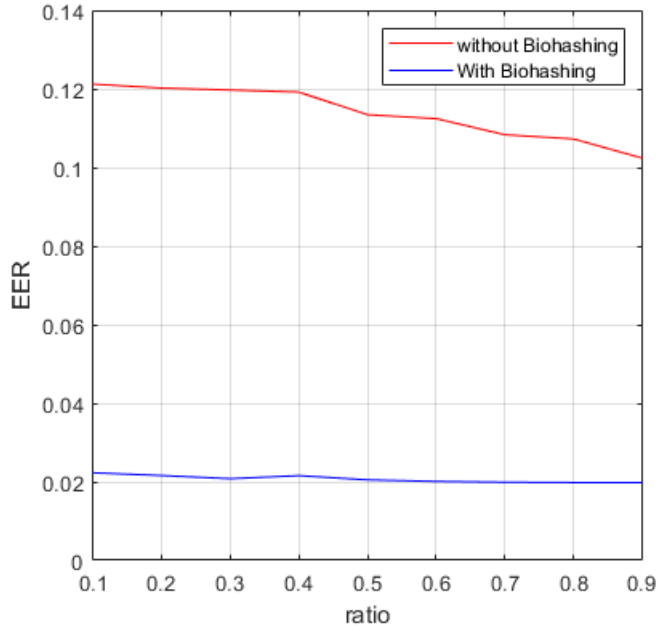

Fig. 13. EER variation for One class SVM method with and without Biohashing

were used to evaluate the proposed approach performances in terms of deciding on the best classifier for a better authentication based on EER value. In a second scenario, the data privacy is considered and protected with the BioHashing algorithm to evaluate the system performances with data protection.

The obtained experimental results are promising with an EER value of 10% for One Class SVM without data protection which make it the best classifier for a better authentication comparing to the other adopting classifiers. K-means method shows an EER value of 37% which is considered high in comparison with the Distance Minimization and One Class SVM techniques. However, when implementing the BioHashing algorithm for data privacy protection, EER values for the three methods are lower than 1% and One Class SVM classifier remains the best with an EER value of 0.1%. Great improvements are observed when using the BioHashing algorithm with the smartphone personal communication data which proves an interesting approach for a robust user authentication.

Concerning future works, combining this approach with other biometrics modalities is highly conceivable in order to improve user authentication. Furthermore, in a ubiquitous digital environment, multidevice authentication is becoming the solution for more secure and protected user authentication, where devices of the same user can interact mutually in order to ensure a high confidence score, sufficient enough to be able to communicate between them the authentication result found on a single device. For this reason, results with smartphone communication personal dataset found in this paper are considered to be associated to other types of devices (laptop for example),with the aim of ensuring more robust authentication in a ubiquitous digital environment.

REFERENCES

[1] A.Zacks, “Malware statistics, trends and facts in 2019,” 2018. [Online] Available:https://www.safetydetective.com/blog/malware-statistics/. [Accessed: 15- Apr- 2019].

[2] O. Berkman and O. M. Ostrovsky, “The unbearable lightness of pin cracking,” in *Financial Cryptography and Data Security*, S. Dietrich and R. Dhamija, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 224–238.

[3] H. M. Wood, *The use of passwords for controlled access to computer resources*. US Department of Commerce, National Bureau of Standards, 1977, vol. 500, no. 9.

[4] Z. Akhtar, C. Micheloni, and G. Foresti, “Biometric liveness detection: Challenges and research opportunities,” *IEEE Security Privacy*, vol. 13, pp. 63–72, 09 2015.

[5] R. Belguechi, E. Cherrier, V. Alimi, P. Lacharme, and C. Rosenberger, “An overview on privacy preserving biometrics,” in *Recent Application in Biometrics*. IntechOpen, 2011.

[6] L. Lee and W. E. L. Grimson, “Gait analysis for recognition and classification,” in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, May 2002, pp. 155–162.

[7] R. Plamondon and S. N. Srihari, “Online and off-line handwriting recognition: a comprehensive survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, Jan 2000.

[8] N. L. Clarke and S. M. Furnell, "Authenticating mobile phone users using keystroke analysis," *International Journal of Information Security*, vol. 6, no. 1, pp. 1–14, Jan 2007. [Online]. Available: https://doi.org/10.1007/s10207-006-0006-6

[9] S. Alotaibi, S. Furnell, and N. Clarke, "Transparent authentication systems for mobile device security: A review," in *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, Dec 2015, pp. 406–413.

[10] N. Clarke, *Transparent user authentication: biometrics, RFID and behavioural profiling*. Springer Science & Business Media, 2011.

[11] R. Guidorizzi, "Security: Active authentication," *IT Professional*, vol. 15, pp. 4–7, 07 2013.

[12] Google, "Google Abacus project," http://www.androidcentral.com/project-abacus-atap-project-aimed-killing-password, [Online; accessed 29-April-2019].

[13] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley, "Casa: Context-aware scalable authentication," in *SOUPS '13 Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013.

[14] F. Li, N. Clarke, M. Papadaki, and P. Dowland, "Active authentication for mobile devices utilising behaviour profiling," *International Journal of Information Security*, 2013.

[15] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, "Text-based active authentication for mobile devices," in *ICT Systems Security and Privacy Protection*. Springer, 2014, pp. 99–112.

[16] M. Derawi and P. Bours, "Gait and activity recognition using commercial phones," *Computers & Security*, 2013.

[17] M. Tanviruzzaman and S. I. Ahamed, "Your phone knows you: Almost transparent authentication for smartphones," in *Computer Software and Applications Conference (COMPSAC), 2014 IEEE 38th Annual*. IEEE, 2014, pp. 374–383.

[18] N. A. Safa, R. Safavi-Naini, and S. F. Shahandashti, "Privacy-preserving implicit authentication," in *IFIP International Information Security Conference*. Springer, 2014, pp. 471–484.

[19] M. Nauman, T. Ali, and A. Rauf, "Using trusted computing for privacy preserving keystroke-based authentication in smartphones," *Telecommunication Systems*, vol. 52, no. 4, pp. 2149–2161, 2013. [Online]. Available: http://dx.doi.org/10.1007/s11235-011-9538-9

[20] J. Hatin, E. Cherrier, J.-J. Schwartzmann, and C. Rosenberger, "Privacy Preserving Transparent Mobile Authentication," in *International Conference on Information Systems Security and Privacy (ICISSP)*, Porto, Portugal, Feb. 2017. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01659952

[21] C.Piech, "Kmeans," 2013. [Online] Available:http://stanford.edu/cpiech/cs221/handouts/kmeans.html/. [Accessed: 22- Apr- 2019].

[22] A. Ng, "Cs229 lecture notes," *Intelligent Systems and their Applications IEEE*, 01 2000.

[23] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[24] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, "Sectored random projections for cancelable iris biometrics," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1838–1841.

[25] A. Teoh, D. Ngo, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern recognition*, vol. 40, 2004.

[26] C. Rathgeb, F. Breitinger, C. Busch, and H. Baier, "On application of bloom filters to iris biometrics," *IET Biometrics*, vol. 3, no. 4, pp. 207–218, 2014.

[27] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 54–65, 2015.

[28] Z. Jin, J. Y. Hwang, Y. L. Lai, S. Kim, and A. B. J. Teoh, "Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 393–407, Feb 2018.

[29] C. Rosenberger, "Evaluation of biometric template protection schemes based on a transformation," in *International Conference on Information Systems Security and Privacy (ICISSP)*, 2018.