Analysis of Doddington Zoo Classification for User Dependent Template Update: Application to Keystroke Dynamics Recognition

Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, Najoua Essoukri Ben Amara

Analysis of Doddington Zoo Classification for User Dependent Template Update: Application to Keystroke Dynamics Recognition

Abir Mhenni[a,b,c,*], Estelle Cherrier[c], Christophe Rosenberger[c], Najoua Essoukri Ben Amara[b]

[a] *ENIT, University of Tunis El Manar, BP 94, Rommana 1068 Tunis, Tunisia*

[b] *LATIS- Laboratory of Advanced Technology and Intelligent Systems, ENISo, University of Sousse, BP 526, 4002 Sousse, Tunisia*

[c] *Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France*

Abstract

Biometric authentication systems are increasingly considered in different access control applications. Regarding that users have completely different interactions with these authentication systems, several techniques have been developed in the literature to model distinctive users categories. Doddington zoo is a biometric menagerie that defines and labels user groups with animal species to reflect their behavior with the biometric systems. This menagerie was developed for different biometric modalities including keystroke dynamics. The present study proposes a user dependent adaptive strategy based on the Doddinghton zoo, for the recognition of the user's keystroke dynamics. The novelty of the proposed approach lies in applying an adaptive strategy specific to the characteristics of each user of the Doddinghton zoo menagerie aiming to solve the intra-class variation problems. The obtained results demonstrate competitive performances on significant keystroke dynamics datasets WEBGREYC and CMU.

Keywords: Keystroke dynamics authentication, Password security, Adaptive

[☆] Fully documented templates are available in the elsarticle package on CTAN.

[*] Corresponding author

Email addresses: abirmhenni@gmail.com (Abir Mhenni), estelle.cherrier@ensicaen.fr (Estelle Cherrier), christophe.rosenberger@ensicaen.fr (Christophe Rosenberger), najoua.benamara@eniso.rnu.tn (Najoua Essoukri Ben Amara)

strategy, Doddington zoo classification, KNN-GA

1. Introduction

Nowadays, biometric systems are widely used for numerous applications [1] like physical access control, electronic payment, etc given the increasing need to solve their security issues [2, 3, 4]. In this paper, we propose a more reliable authentication approach based on the investigation of the user's keystroke dynamics. Indeed, keystroke dynamics analyses the rhythm a person exhibits while typing on a keyboard [5, 6]. Hence, keystroke dynamics is considered as a behavioral biometric modality, like signature dynamics, gait and voice.

The use of this behavioral biometric modality remains a major challenge, since an efficient description of the user's keystroke dynamics is needed to overcome the problem of intra-class variation over time [7, 8]. In fact, the typing rhythm of the user changes according to several factors like the user's emotional state, their activeness, the password mastery; etc. Adaptive strategies, are one of the most interesting solutions to remedy to the intra-class variations [9, 10] for behavioral biometric systems. They consist in updating the biometric reference template describing the typing rhythm of the user at each access verification. These strategies depend generally on five parameters [11]:

- Reference modeling : defines the representation of the user's model. It can be represented by a single sample, a gallery or a cluster;
- Adaptation criterion : decides to launch the adaptation process;
- Adaptation mode : can be supervised or semi-supervised;
- Adaptation periodicity : can be online (applied immediately after the query acceptance) or offline (applied after a specific period or after the collection of a particular number of accepted queries);
- Adaptation mechanism : determines how to apply the modifications to the reference. It can be an additional, replacement or a combined mechanism.

These adaptation strategies are efficient solutions to intra-class variations for behavioral biometric modality among them the keystroke dynamics one, which we consider in this paper. But, it has been proven that applying the same adaptation strategy to all users is not the best solution, as the users behaviors are generally different. Doddington zoo is a menageries that characterizes users into multiple animal categories [12]. It consists in grouping users according to their behavioral specificities when dealing with the authentication process. For that purpose, we propose a novel adaptation method that is appropriate to the user's typing rhythm. The main contribution of this paper is to propose a user dependent template update strategy based on the Doddington zoo classification. To the best of our knowledge, the use of the Doddington zoo menagerie for template update purpose based on keystroke dynamics data has never been reported in the literature. Our experiments are carried out on real data coming from two well known datasets in the literature WEBGREYC and CMU.

The reminder of this paper is organized as follows. In the next section, we present some related work concerning Doddington zoo categorization. In section 3, the proposed adaptive strategy specific to the keystroke dynamics of each user's category is described. Section 4 details the experiments and the obtained results. Finally, conclusion and perspectives are drawn in section 5. This invited article supports and improves the results of the original "User Dependent Template Update for Keystroke Dynamics Recognition" [13].

2. Related work

Keystroke dynamics is a behavioral modality that presents the favor of being non intrusive, inexpensive and weakly constrained for the user [5, 14, 15]. The major drawback of this modality is that it suffers from large intra-class variation [7, 8] due to aging problems. In fact, the keystroke dynamics of the user varies as time elapses according to different situations. This variability may be due to the familiarity with the password after a time span, the user's humor and activeness and the changing of the keyboard layout. In fact a recent work [16],

demonstrated that the performances of the keystroke dynamics system can be deteriorated when changing some circumstances like changing the used keyboard type (AZERTY or QWERTY).

Adaptive strategies [9, 10] also known as template update strategies, are an interesting solution to overcome the intra-class variability.

2.1. Adaptive strategies

The adaptive strategies have been deeply used to enhance the performances of the biometric systems for different modalities [17] like face [18, 19], voice [20] and keystroke dynamics [21]. Different types of adaptation process have been proposed in the literature:

- Adaptation of the system parameters: It generally consists in updating the parameters of the classifier depending on the user [22] or the quality of the capture [23]. Recently, in [21], the authors proposed an R^2BN adaptive model that consists in increasing the weight on the misclassified instances to provide them to the next-level classifier to perform better. The authors state that the proposed model achieved high accuracy in educational level prediction through the keystroke dynamics of the user.

- Adaptation of the decision threshold: It serves to make the considered threshold more suitable to user's characteristics overtime, thus the system selects highly-confident samples. Different threshold adaptation methods have been proposed for different modalities [24, 25]. For keystroke dynamics modality, Mhenni et al. [15] proposed an individual threshold that is adapted through the adaptation sessions and demonstrated competitive perfromances comapred to global thresholds.

- Adaptation of the biometric reference: It is employed to update the reference modeling the user's characteristics overtime. Several adaptation mechanism have been considered. The growing window mechanism [26] is one of the well known additive mechanisms. It adds each accepted query to the reference. Consequently, the size of the reference becomes extremely

large. As a replacement mechanism, we can cite the sliding window mechanism [26]. It consists in substituting the oldest sample of the reference by the newly accepted query.

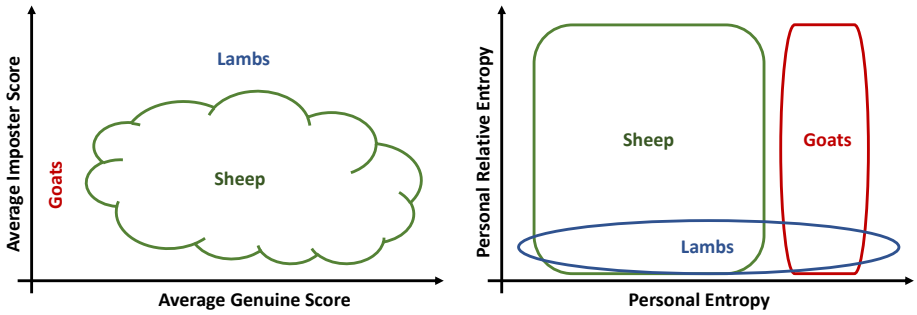
In this paper we are interested in updating both the reference and the thresholds in each adaptation session to overcome the intra-class variation problems due to the reference aging. Commonly, a unique adaptation mechanism is applied to all users of the authentication process, although it was demonstrated that biometric systems performances are subject dependent [27]. That is why, we decided to use an update strategy for each category of users in this work.

2.2. Doddington zoo menagerie

We are interested in the users classification based on the Doddington zoo [12]. It is a widely used menagerie for users behavior classification [28, 29], but, to our knowledge, it has not been associated with adaptive strategies for keystroke dynamics modality. Four categories of animals were defined, which are:

- sheep: concerns users who can easily be recognized;
- goats: represents users who are particularly difficult to recognize;
- lambs: contains users who are easy to imitate;
- wolves: consists of users who can easily imitate others.

Several approaches have been proposed to distinguish between these varieties of users as shown in Figure 1. Doddington *et al.* considered the classification based on the errors rates. Indeed, users classified as goats increase the False Non Match Rate (FNMR) of the recognition system whereas wolves and lambs increment its False Match Rate (FMR). Other research work [30] proposed to use the personal entropy and relative entropy for biometric menagerie of online signature verification. Personal entropy is computed using only genuine data. It serves to differentiate between sheep and goats class of users. Indeed, sheep



(a) FNMR and FMR based classification. (b) Entropy based classification.

Figure 1: Animals of the Doddington zoo Biometric Menagerie according to [30].

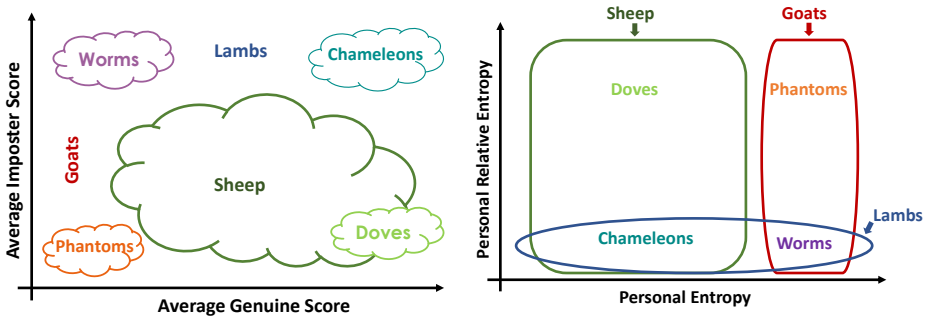
class represents users characterized by a low personal entropy and gaots class represents users marked by a high personal entropy. Relative entropy is calculated with both genuine and impostor data. It helps to distinguish lambs class, which are known by the lowest relative entropy.

Besides, sheep generally dominate the population of the zoo, goats as well as lambs constitute only a small fraction of the population. However, the wolves category constitutes a large portion of false rejection and acceptance rates.

Further, Yager and Dunstone [31] distinguished four other animal categories of users by considering simultaneously both the genuine and impostor matching scores, for each claimed identity:

- Chameleons: correspond to users who are easy to recognize and easy to attack;
- Phantoms: depict the users characterized by rejections of genuine and impostor queries;
- Doves: represent the best users because they are easy to recognize and difficult to attack;
- Worms: regroup the worst users as they are difficult to recognize and easy to attack.

The four additional sub-categories can also be distinguished thanks to the FMR and FNMR based classification or the entropy based classification as depicted in Figure 2. For the FMR and FNMR based classification, chameleons belong to the users who are known by high genuine and impostor match scores. Contrariwise, phantoms are characterized by low genuine and impostor match scores. Doves are a sub-group of sheep according to this classification methodology. They are the best users since they lead both to high genuine and low impostor match scores. Worms in the opposite, are a sub-group of goats. They represent the worst users, as they lead to low genuine and high impostor scores. This categorization was applied to different modalities like face, speech, fingerprint, iris and keystroke modalities [32], but it was not associated to an adaptive strategy specific to each category of user.



(a) FNMR and FMR based classification. (b) Entropy based classification.

Figure 2: Large animal groups distinguished by Doddington zoo according to [30].

As for the entropy based method, it exploits the personal and relative entropies to distinguish between these classes. First, chameleons are a sub-category of sheep and lambs as they are known by the lowest personal entropy and the lowest relative entropy. Second, phantoms are a sub-category of goats class regarding that they have a reference with poor data quality generated in the enrollment phase. They are characterized by a high personal entropy and a high relative entropy. Third, doves are a sub-category of sheep class. They are characterized by the lowest personal entropy and the highest relative entropy.

Finally, worms are a sub-category of goats and lambs classes. They have the highest personal entropy and the lowest relative entropy.

In this work, we are interested in the entropy based classification to distinguish between the users characteristics. For that purpose, we examined the entropy of the users of the WEBGREYC database [33] over time as illustration. We calculated the entropy of each user's set of 5 samples in chronological order of the database data. As depicted in Figure3, the characteristics of some users are stable over time such as those of user 3 and user 30. Others have an entropy that decreases over time like user 34. This means that their intra-class variation decreases owing to the mastery of the password for example. However, user 4 and user 11, have an increasing entropy. Their intra-class variations increase as time elapses. Thus the need for a user specific adaptation strategy is clearly demonstrated.

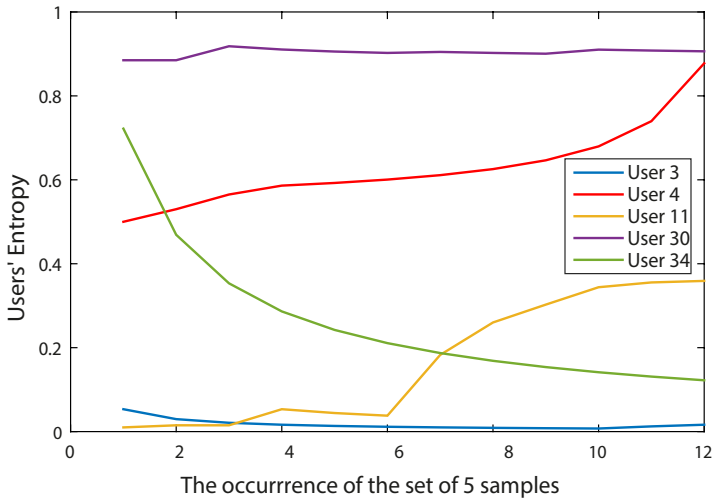


Figure 3: Personal entropy of some users of WEBGREYC database.

Hence, a user specific adaptive strategy for enhanced biometric authentication based on keystroke dynamics modality is developed. The incorporation of Doddington zoo menagerie entails the exploitation of specific parameters to each user category. Consequently, a gain in memory storage and processing time is

ensured as expanded on the next section.

3. User dependent template update

The proposed contributions investigate an authentication method based on two captures of the user's keystroke dynamics in the enrollment phase in coherence with the single enrollment adaptation strategy proposed in the previous work [34, 35]. We choose to use initially 2 samples instead of an only one sample, regarding that industrial application usually ask users to type their passwords twice. During the use of the authentication system, the reference is enriched through the user dependent adaptive strategy. Thus, based on the growing window mechanism, the size of the reference is increased until reaching the fixed maximum size. During this phase, the users are distinguished based on the evolution of their reference size over time. Indeed, the first assumption consists in considering users whose reference size increased slowly, are difficult to recognize. Hence, they are considered as goats. The second assumption considers users whose reference size increased rapidly, as they are easily recognized. Consequently, they are classified as sheep.

Once the fixed maximum size of the reference is reached, the sliding window mechanism is applied to ensure a limited size of the reference. Throughout this phase, the users categorization is ensured with the personal and the relative entropy calculation as detailed in subsection 3.4. Depending on each category of users, we define specific parameters like the reference size and the decision thresholds to overcome the users limitations. Therefore, the adaptation strategy becomes specific to each category of users. Figure 4 depicts the proposed authentication process. The steps of each phase of the process are detailed in the following:

3.1. Enrollment phase

In our previous work [34, 36], we considered an only one sample to create the user's reference. We demonstrated that the performances in the beginning

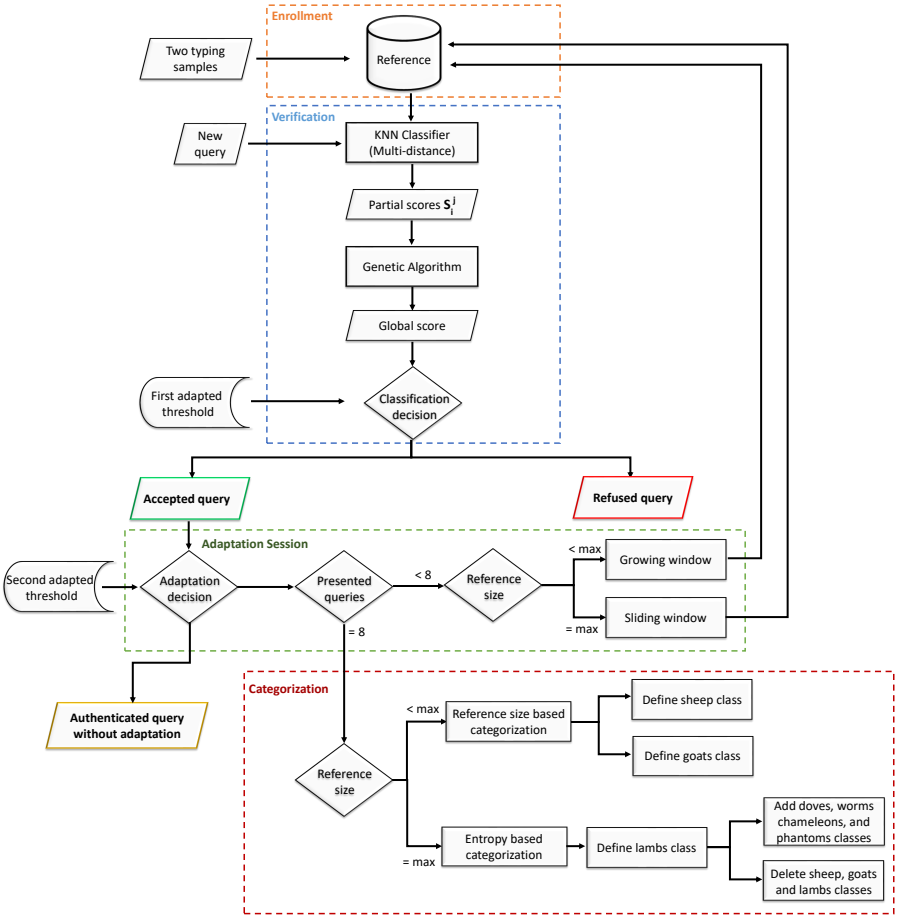


Figure 4: Description of the keystroke authentication process

of the process are low. To compensate this performance degradation, we need to enlarge the size of the reference to capture more variability. Ultimately, for nowadays password-based applications, users are usually asked to type their password and to confirm it when creating an account. Therefore, we decided to take advantage of 2 samples for creating the reference. Thus the proposed approach is suitable to industrial applications like e-commerce, e-banking, e-mailing and social media applications.

3.2. Verification phase

During the authentication process, the verification is performed by the K Nearest Neighbor (KNN) classifier based on four distances; Hamming, Euclidean, Statistical and Manhattan. These distances are chosen as long as they demonstrated the best performances when compared to other distances as detailed in [34]. Moreover, the KNN classifier has the advantage of being efficient even when considering a reduced number of features [34, 37]. Actually, the presented query is compared to each sample of the reference to retain the nearest neighbor score of them. Thus, given a user j four partial scores S_i^j are obtained through the considered distances. Afterw1ards, a vote is ensured by a Genetic Algorithm (GA) to calculate the global score as given in Equation (1).

$$GlobalScore = \sum_{i=1}^{4} w_i * S_i^j \tag{1}$$

where:

$i \in [1, 4]$ is the index of one of the considered distances ;

w_i is the weight of each partial score;

S_i^j is the similarity score obtained by the distance metric i for the user j.

In fact, the GA showed the most accurate results as a fusion function for keystroke dynamics modality [38]. The used GA-KNN classification [36] proved its efficiency within the proposed approach as we will show through the recorded results.

3.3. Adaptation phase

This phase serves to mitigate the intra-class variation problems. In the following, we detail the different components of the adaption strategy.

3.3.1. Reference modeling

At the beginning, the reference template is composed of two samples to remedy to the tedious learning phase. Therefore, the reference is relevant to the account creation process for web and mobile applications. After that, each novel

query accepted in the adaptation phase is added to the user's reference based on the growing window mechanism. We obtain a gallery of samples describing the typing rhythm of the user. Once the fixed maximum size is reached, we apply the sliding window mechanism. Indeed, we intend to limit the enrollment phase and to adapt the biometric reference to be adjusted to its intra-class variation over time.

3.3.2. Adaptation criterion

Different adaptation criteria were proposed in the literature [39, 23, 25]. We are interested in the adapted thresholds criterion that has been proposed in [15]. It has the advantage to use the double threshold verification [39] while maintaining the thresholds user dependent and adapted as time elapses. The adapted thresholds are managed by Equation (2).

$$T_j^{s+1} = T_j^s - e^{-\frac{\mu_j}{\sigma_j}} \tag{2}$$

Where T_j^s is the threshold value specific to user j during session s, μ_j and σ_j are the average of the mean vector of the reference of the user j, and the standard deviation of the standard deviation vector of the reference of the user j respectively. .

3.3.3. Adaptation mode

The adaptation is ensured in a semi-supervised mode through the KNN classifier combined with the GA. If the calculated global score is lower than the adapted thresholds, the query is used to update the reference.

3.3.4. Adaptation periodicity

The adaptation is executed online, immediately after the query acceptation. Hence, if the query satisfied the adaptation criterion, the adaptation mechanism is launched.

3.3.5. Adaptation mechanism

Concerning the adopted mechanism, we combine two existing approaches: the growing window mechanism and the sliding window one [26]. These mechanisms are frequently used for keystroke dynamics modality [40, 37]. The growing window mechanism is used to enlarge the size of the reference until the maximum size is reached. The sliding window is afterward considered to maintain a fixed reference size and to mitigate the reference aging problem. Hence, the mechanism is called "double serial mechanism".

3.4. User classification

During the two first update sessions, we start to classify users into two groups: sheep and goats. We are first interested to only these two groups because we focus on the most representative groups of the Doddington zoo.

Thereby, over the growing window phase, we assume that users, whose number of accepted queries exceeded 3 samples during the update session, are easily recognized. So, they are classified as sheep. The rest of the users, those whose number of accepted queries is less than 3, are classified as goats, as they are difficult to recognize.

Throughout the sliding window mechanism, the size of the reference is no more significant as the maximum size of the reference is reached. So, we considered the entropy measure to distinguish between the considered users groups. In fact, it was demonstrated in [30, 29] that the error rates increases when the user's entropy is higher. Thereby, both personal and relative entropies are calculated according to equations (3) and (4) respectively. For this fact, the personal entropy of the reference ref_j containing N samples of the user j is measured according to Equation (3):

$$Entropy_j = -\sum_{i=1}^{N} ref_{j_{(t)}}(i) \ \log\left(ref_{j_{(t)}}(i)\right) \tag{3}$$

The relative entropy is equally calculated according to equation (4), where $attaq_j$ is a matrix containing N samples of the keystroke dynamics of multiple users other than the user j:

$$RelativeEntropy_j = \frac{1}{2}\ (\sum_{i=1}^{N} ref_{j\,(t)}(i) \log(\frac{ref_{j\,(t)}(i)}{attaq_j(i)}) + \sum_{i=1}^{N} attaq_j(i) \log(\frac{attaq_j(i)}{ref_{j\,(t)}(i)})\ ) \tag{4}$$

Consequently, starting from session 4, we use the entropy to classify users. We initially distinguish the lambs class. Once users of this class are defined, we determine during the following sessions the remaining classes of the zoo. Once session 6 starts, classes of worms, doves, chameleons and phantoms take place and classes of sheep, goats, and lambs disappear.

4. Experiments and Results

The proposed approach was tested on two public databases of keystroke dynamics modality: WEBGREYC and CMU. WEBGREYC [33] database, contains 60 samples from 45 users. The CMU [41] database includes 400 biometric samples of 50 users.

4.1. Data stream generation

We managed user samples during the adaptation sessions as follows. Two samples of each user are considered during the enrollment phase in order to create the reference. For each adaptation session, 8 new queries are introduced to the authentication system. These queries are divided into 5 genuine samples and 3 impostor ones. Thus, we considered 12 adaptation sessions for the WEBGREYC database and 80 adaptation sessions for the CMU database.

To evaluate the proposed approach we analyzed different data stream for each adaptation session:

- Scenario 1: Presenting 5 genuine samples first, afterwards 3 imposter samples are presented to the authentication system.
- Scenario 2: Presenting alternated genuine and imposter samples.
- Scenario 3: Presenting 3 imposter samples first, afterwards 5 genuine samples are presented to the authentication system.

Generally, the first two data streams conveniently fit the actual scenarios of the password based applications. In fact, just after creating an account, the user is usually asked to enter his credentials again to gain access to his account. Consequently, at least one genuine query is guaranteed in the beginning of the process.

4.2. Biometric menagerie parameters

For each class of users, we use specific adaptation parameters. Concerning goats and worms classes, which are characterized by a high intra-class variation according to the different conducted experiments, we increased the maximum size of the reference in order to enrich the description of the keystroke dynamics of the users. The maximum size of phantoms class should be higher because this class is difficult to describe. Regarding the lambs, worms, chameleons and phantoms classes, stricter thresholds are needed to minimize the acceptance of the impostor attacks. These thresholds are generated based on Equation (5).

$$T_j^{s+1} = T_j^s - e^{-\frac{\mu_j}{2*\sigma_j}} \tag{5}$$

The fixed parameters for each user category are detailed in Table 1.

4.3. Results and Comparisons

To evaluate the performance of the proposed approach, we consider two evaluation metrics : the Error Equal Rate (EER) and the Area Under Curve (AUC).

Table 1: Specific parameters according to user's category

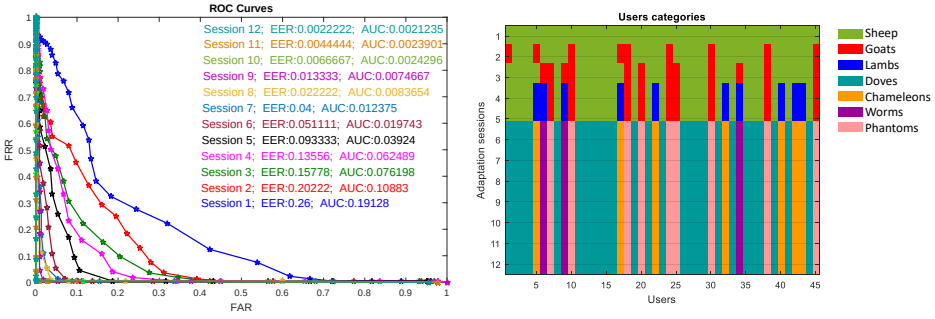
ine User category	Reference size	Thresholds
ine Sheep	10	Adapted thresholds
Goats	15	Adapted thresholds
Lambs	10	Stricter thresholds
Worms	15	Stricter thresholds
Chameleons	10	Stricter thresholds
Doves	10	Adapted thresholds
Phantoms	20	Stricter thresholds

ine

The obtained results show an interesting performance of the strategy as illustrated in Figures 5a and 6a. To illustrate the benefits of the consideration of 7 classes of the Doddington zoo in the proposed approach, we compared it to the same adaptation approach without biometric menagerie and with the consideration of only 3 classes conducted in [35] namely sheep, goats and lambs. As demonstrated in Tables 2 and 3, the proposed approach show improved performances as it proposes an adaptive strategy that is appropriate to the user's specificities. In fact, the considered users' categories encompass a wider variety of users. Hence, the adaptation method acts according to each user's particularities.

Adding doves, phantoms, chameleons and worms classes, improved the EER performances by 0.6% for the WEBGREYC database and by 0.2% for the CMU database, as demonstrated in Tables 2 and 3,. Furthermore, when compared to the same adaptation approach without biometric menagerie, the user specific adaptation approach ensures an improved EER performance of more then 2% for CMU database and 5% for WEBGREYC database.

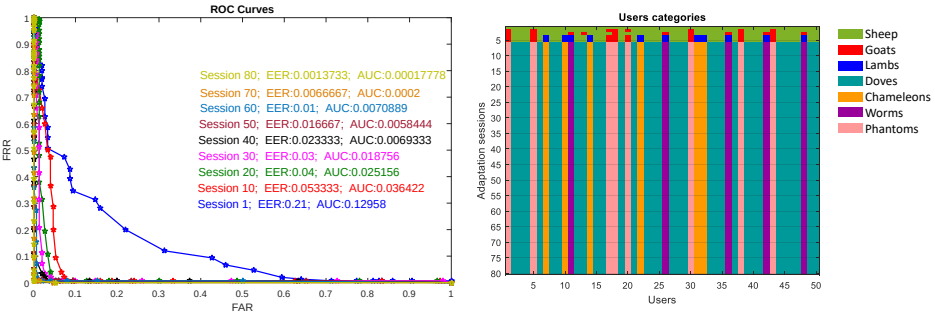
To reveal the impact of imposter attacks on the proposed approach, we tested different scenarios of the queries presentations as detailed in subsection



(a) DET curves

(b) Distribution of user classes

Figure 5: Obtained performances and the distribution of users classes for WEBGREYC database.



(a) DET curves

(b) Distribution of user classes

Figure 6: Achieved performances and the distribution of users classes for CMU database.

Table 2: Comparison of the proposed adaptation strategy for WEBGREYC database

ine Adaptation strategy	EER	AUC
ine Without Doddington menagerie [36]	5.3%	0.02
Biometric menagerie based on 3 classes [35]	0.8 %	0.003
Biometric menagerie based on 7 classes	0.2%	0.002

ine

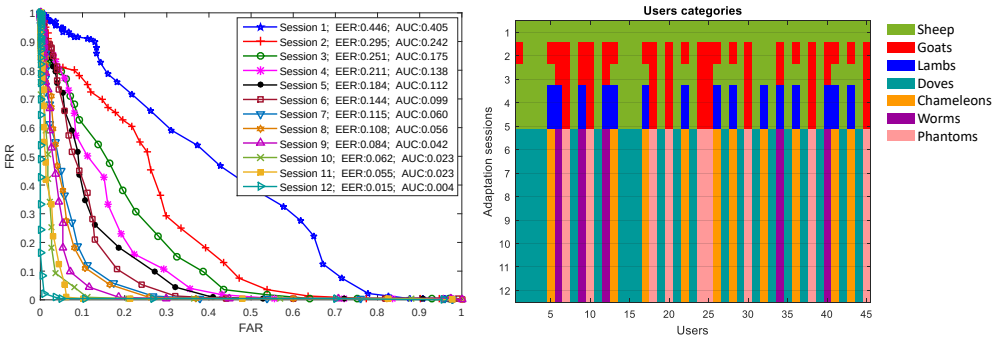
4.1. When considering the 3 imposter samples before the genuine ones (scenario

Table 3: Comparison of the proposed adaptation strategy for CMU database

ine Adaptation strategy	EER	AUC
ine Without Doddington menagerie	2.3%	0.004
Biometric menagerie based on 3 classes	0.3%	0.001
Biometric menagerie based on 7 classes	0.1%	0.0001
ine		

3), the performances are considerably decreased as demonstrated in Figure 7a. This is quite expected as the initial reference doesn't contain enough intra-class variation. Thus the recognition errors are higher in the beginning of the process. These errors decrease during the adaptation sessions through the proposed method. In addition, we illustrated the the users categorization in Figure 7. It is quite clear that the number of users belonging to goats class has increased considerably since the beginning. In fact, the percentage of goats class in adaptation session 2 raised from 20% (for scenario 1) to 53% (for scenario 2). This may be due to the inclusion of some imposter samples in the reference. Subsequently, these imposter samples will be removed as time elapses due to the proposed adaptation system. In fact, for scenario 3, the percentage of users associated to goats class in adaptation session 2 decreased to 31%. Thanks to the considered user specific parameters, the number of genuine samples included in the reference increase and the imposter samples decrease especially through the sliding window mechanism. Thus, the intra-class variation of the reference samples is reduced.

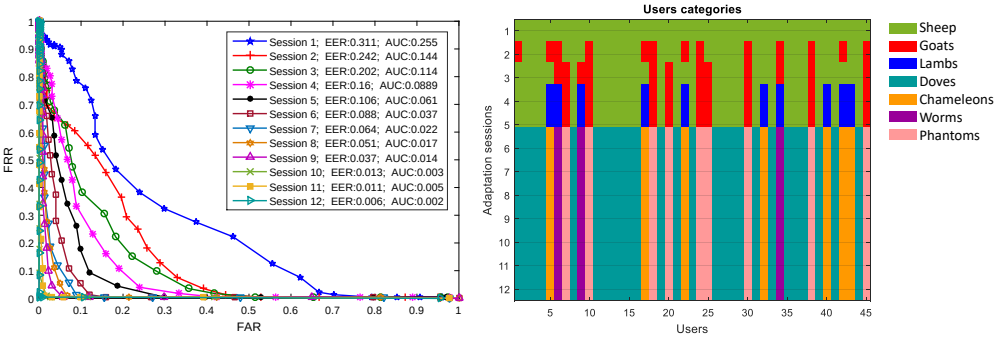
When mixing the genuine and imposter queries (scenario 2), the obtained results are better than those obtained in scenario 3 as depicted in Figure 8a and they are quite similar to those of scenario 1. The scenarios presenting better performances (1 and 2) are more realistic. In fact, adding a new account is usually transparent for any password based application. Hence, it is not evident that a hacker encounters an account since its creation.



(a) DET curves

(b) Distribution of user classes

Figure 7: Achieved performances when considering scenario 3 for WEBGREYC database.



(a) DET curves

(b) Distribution of user classes

Figure 8: Achieved performances when considering scenario 2 for WEBGREYC database.

We have furthermore performed an analysis on the variation of the reference size concerning each user category of the menagerie considered in the proposed approach. As depicted in Tables 4 and 5, the chosen parameters are optimal. Indeed, minimizing the size of the reference, guaranteed the gain in used memory space, but no improvement in the performance is recorded. Moreover, while enlarging the size of the reference, a small increase in performance is registered. Thus, the extra memory space allocated does not produce a significant influence on the obtained performance. Hence, we prove that the chosen reference sizes are the most appropriate to each user category.

Table 4: Obtained performances by varying the size of the reference for each user category for WEBGREYC database

ine Reference size	User category							Performances	
	Sheep	Goats	Lambs	Worms	Chameleons	Doves	Phantoms	EER	AUC
ine max1	5	10	5	10	5	5	15	6.5%	0.05
max2	10	15	10	15	10	10	20	2.22%	0.002
max3	15	20	15	20	15	15	25	2%	0.0017
ine									

Table 5: Obtained performances by varying the size of the reference for each user category for CMU database

ine Reference size	User category							Performances	
	Sheep	Goats	Lambs	Worms	Chameleons	Doves	Phantoms	EER	AUC
ine max1	5	10	5	10	5	5	15	5.9%	0.047
max2	10	15	10	15	10	10	20	1.37%	0.0001
max3	15	20	15	20	15	15	25	1.14%	0.0008
ine									

5. Conclusion

The contribution of this paper is to propose a user dependent adaptation strategy for keystroke dynamics authentication system. Thus the proposed strategy consists in differentiating between the users groups thanks to Doddinghton zoo theory. Afterwards, we adjust some parameters of the adaptation strategy according to the specificity of each group. Hence, we enlarge the reference size for the users suffering from large intra-class variation and we user stricter thresholds for users that are more vulnerable to hacker attacks.

The proposed approach has been validated on two significant keystroke dynamics databases, and it demonstrated enhanced EER performances equal to 2% for WEBGREYC database and 1.3% for CMU database. This performance amelioration is also relying to the KNN-GA classification method and the double serial adaptation mechanism.

As perspective, we aim to enhance the performances of the first adaptation session of the process. For that purpose, we attempt to investigate a data

augmentation approach to generate additional keystroke dynamics data providing more information about the user. Thus even scenario 3 problems will be mitigated.

References

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

[2] B. Gupta, D. P. Agrawal, S. Yamaguchi, Handbook of research on modern cryptographic solutions for computer and cyber security, IGI Global, 2016.

[3] C. Stergiou, K. E. Psannis, B.-G. Kim, B. Gupta, Secure integration of iot and cloud computing, Future Generation Computer Systems 78 (2018) 964–975.

[4] B. B. Gupta, Computer and Cyber Security: Principles, Algorithm, Applications, and Perspectives, CRC Press, 2018.

[5] M. Rybnicek, C. Lang-Muhr, D. Haslinger, A roadmap to continuous biometric authentication on mobile devices, in: Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International, IEEE, 2014, pp. 122–127.

[6] R. S. Gaines, W. Lisowski, S. J. Press, N. Shapiro, Authentication by keystroke timing: Some preliminary results, Tech. rep., DTIC Document (1980).

[7] C. Epp, M. Lippold, R. L. Mandryk, Identifying emotional states using keystroke dynamics, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM, New York, NY, USA, 2011, pp. 715–724. doi:10.1145/1978942.1979046.

[8] A. N. H. Nahin, J. M. Alam, H. Mahmud, K. Hasan, Identifying emotion by keystroke dynamics and text pattern analysis, Behaviour & Information Technology 33 (9) (2014) 987–996.

[9] L. Didaci, G. L. Marcialis, F. Roli, Analysis of unsupervised template update in biometric recognition systems, Pattern Recognition Letters 37 (2014) 151–160.

[10] N. Poh, A. Rattani, F. Roli, Critical analysis of adaptive biometric systems, IET biometrics 1 (4) (2012) 179–187.

[11] R. Giot, Contributions à la dynamique de frappe au clavier: multibiométrie, biométrie douce et mise à jour de la référence, Ph.D. thesis, Université de Caen (2012).

[12] G. Doddington, W. Liggett, A. Martin, M. Przybocki, D. Reynolds, Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation, Tech. rep., NATIONAL INST OF STANDARDS AND TECHNOLOGY GAITHERSBURG MD (1998).

[13] A. Mhenni, E. Cherrier, C. Rosenberger, N. Essoukri Ben Amara, User dependent template update for keystroke dynamics recognition, in: CyberWorlds, 2018.

[14] P. Bours, S. Mondal, Continuous authentication with keystroke dynamics, Norwegian Information Security Laboratory NISlab (2015) 41–58.

[15] A. Mhenni, E. Cherrier, C. Rosenberger, N. Essoukri Ben Amara, Keystroke template update with adapted thresholds, in: International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2016.

[16] P. Bours, J. Ellingsen, Cross keyboard keystroke dynamics, in: 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), IEEE, 2018, pp. 1–6.

[17] A. Rattani, F. Roli, E. Granger, Adaptive biometric systems, Advances in Computer Vision and Pattern Recognition. Springer International Publishing.

[18] A. Rattani, G. L. Marcialis, F. Roli, Biometric system adaptation by self-update and graph-based techniques, Journal of Visual Languages & Computing 24 (1) (2013) 1 – 9.

[19] A. Brutti, A. Cavallaro, Online cross-modal adaptation for audio–visual person identification with wearable cameras, IEEE Transactions on Human-Machine Systems 47 (1) (2017) 40–51.

[20] S. Anzar, K. Amala, R. Rajendran, A. Mohan, P. Ajeesh, M. Sabeeh, F. Aziz, Efficient online and offline template update mechanisms for speaker recognition, Computers & Electrical Engineering 50 (2016) 10–25.

[21] I. Tsimperidis, P. D. Yoo, K. Taha, A. Mylonas, V. Katos, R^2bn: An adaptive model for keystroke-dynamics-based educational level classification, IEEE Transactions on Cybernetics (2018) 1–11doi:10.1109/TCYB.2018.2869658.

[22] S. Hocquet, J.-Y. Ramel, H. Cardot, Estimation of user specific parameters in one-class problems, in: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Vol. 4, IEEE, 2006, pp. 449–452.

[23] N. Poh, J. Kittler, S. Marcel, D. Matrouf, J.-F. Bonastre, Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 1229–1232.

[24] A. Drygajlo, W. Li, K. Zhu, Q-stack aging model for face verification, in: Signal Processing Conference, 2009 17th European, IEEE, 2009, pp. 65–69.

[25] C. Pagano, E. Granger, R. Sabourin, P. Tuveri, G. Marcialis, F. Roli, Context-sensitive self-updating for adaptive face recognition, in: Adaptive Biometric Systems, Springer, 2015, pp. 9–34.

[26] P. Kang, S.-s. Hwang, S. Cho, Continual retraining of keystroke dynamics based authenticator, in: Advances in Biometrics, Springer, 2007, pp. 1203–1211.

[27] N. Poh, J. Kittler, C.-H. Chan, M. Pandit, Algorithm to estimate biometric performance change over time, IET Biometrics 4 (4) (2015) 236–245.

[28] A. Ross, A. Rattani, M. Tistarelli, Exploiting the doddington zoo effect in biometric fusion, in: IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, (BTAS), IEEE, 2009, pp. 1–7.

[29] A. Morales, J. Fierrez, J. Ortega-Garcia, Towards predicting good users for biometric recognition based on keystroke dynamics, in: European Conference on Computer Vision, Springer, 2014, pp. 711–724.

[30] N. Houmani, S. Garcia-Salicetti, On hunting animals of the biometric menagerie for online signature, PloS one 11 (4) (2016) e0151691.

[31] N. Yager, T. Dunstone, Worms, chameleons, phantoms and doves: New additions to the biometric menagerie, in: Automatic Identification Advanced Technologies, 2007 IEEE Workshop on, IEEE, 2007, pp. 1–6.

[32] N. Yager, T. Dunstone, The biometric menagerie, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2) (2010) 220–230. doi:10.1109/TPAMI.2008.291.

[33] R. Giot, M. El-Abed, C. Rosenberger, Web-based benchmark for keystroke dynamics biometric systems: A statistical analysis, in: Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on, IEEE, 2012, pp. 11–15.

[34] A. Mhenni, E. Cherrier, C. Rosenberger, N. Essoukri Ben Amara, Towards a secured authentication based on an online double serial adaptive mechanism of users' keystroke dynamics, in: International Conference on Digital Society and eGovernments (ICDS), 2018.

[35] A. Mhenni, E. Cherrier, C. Rosenberger, N. Essoukri Ben Amara, Adaptive biometric strategy using doddington zoo classification of user's keystroke dynamics, in: 2018 14th International Wireless Communications Mobile Computing Conference (IWCMC), 2018, pp. 488–493. doi:10.1109/IWCMC.2018.8450401.

[36] A. Mhenni, E. Cherrier, C. Rosenberger, N. E. B. Amara, Double serial adaptation mechanism for keystroke dynamics authentication based on a single password, Computers & Securitydoi:https://doi.org/10.1016/j.cose.2019.02.002.
URL http://www.sciencedirect.com/science/article/pii/S0167404818306059

[37] C. Ferrari, D. Marini, M. Moro, An adaptive typing biometric system with varying users model, in: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE, 2018, pp. 564–568.

[38] R. Giot, M. El-Abed, C. Rosenberger, Fast computation of the performance evaluation of biometric systems: Application to multibiometrics, Future Generation Computer Systems 29 (3) (2013) 788–799.

[39] A. Rattani, Adaptive biometric system based on template update procedures, Dept. of Elect. and Comp. Eng., University of Cagliari, PhD Thesis.

[40] R. Giot, C. Rosenberger, B. Dorizzi, Hybrid template update system for unimodal biometric systems, in: Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on, IEEE, 2012, pp. 1–7.

[41] K. S. Killourhy, R. Maxion, et al., Comparing anomaly-detection algorithms for keystroke dynamics, in: Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on, IEEE, 2009, pp. 125–134.