A probabilistic analysis of the biometrics menagerie existence: a case study in fingerprint data

Márjory Da Costa Abreu[1], **Rayron Medeiros de Araújo**[1]

[1]Departamento de Informática e Matemática Aplicada
Universidade Federal do Rio Grande do Norte (UFRN) – Natal, RN – Brazil

marjory@dimap.ufrn.br

Abstract. *The use of biometrics, until very recently, has been restricted to high-security environments and criminal identification applications, for economic and technological reasons. However, with its popularisation, it has been noted that users within the system may have different degrees of accuracy. Some people may have trouble authenticating, while others may be particularly vulnerable to imitation. Recent studies have investigated and identified these types of users, giving them the names of animals: Sheep, Goats, Lambs, Wolves, Doves, Chameleons, Worms and Phantoms. The aim of this study is to evaluate the existence of these animals in a database of fingerprints by proposing a new way of their identification, based on the performance of verification between subjects samples.*

1. Introduction

In biometric authentication systems, it is common to have some inconsistencies in the accuracy of different users. Thus, biometric systems designers are interested in identifying groups of users who have poor performance because they can be the cause of a disproportionate number of verification errors. An analysis of these users, and common features between them, can expose a fundamental vulnerability in a biometric system, and, by addressing this vulnerability it is possible to design more reliable systems [Yager and Dunstone 2009].

Doddington [Doddington et al. 1998] introduced the concept of menagerie biometrics (biometrics zoo) and identified four groups of users, which gave the following animal names: sheep, goats, lambs and wolves. More recently, Yager and Dunstone [Yager and Dunstone 2007] identified four other groups of users: doves, chameleons, worms and phantoms. Each of these animals has a different behaviour in the system.

Although it is accepted that these animals exist, their current identification protocol is very much dependent on the performance of the system (features used and classification algorithm, for example). Moreover, there is not, to our knowledge, a detailed investigation of how these animals are related with each other as well as how the exploration of this information can be used in real life applications.

This paper presents an investigation of the presence of animals in a fingerprint identification biometric system and evaluate the causes of their appearances. The main aim is to propose a new method of investigation and analysis of biometric systems, which differs from traditional methods which focus on statistical evaluation of global errors, such as ROC (Receiver Operating Characteristic) curves and EER (Equal Error Rate).

These statistics are useful for evaluating a biometric system as a whole, but they might ignore idiosyncrasies associated with users characteristics.

2. The biometric menagerie

A biometric-based authentication system can vary significantly, in terms of recognition performance, from one user to another. The vast majority of users have a satisfactory performance, while others are falsely rejected by the system or easily impersonated by impostors. As already mentioned, several groups of problematic users were identified, and each one was given the name of an animal which, similarly, reflects its behaviour. The concept of biometric menagerie was formalised and the first four animals can be described as follows:

a. **Sheep** represent the majority of the population and are usually easy to identify by the system;

b. **Goats** are users usually difficult to identify by the system. These individuals tend to have a low match scores when compared with themselves. They represent a disproportionate increase in the FRR (False Rejection Raet);

c. **Lambs** are individuals easy to impersonate. Other users tend to have a relatively high match scores when compared with lamb users. They represent a disproportionate increase in the FAR (False Acceptance Rate);

d. **Wolves** are good at impersonating users. When compared against other users, they tend to have a high match scores. Like lambs, they represent a disproportionate increase in the FAR.

In this classification, only goats, lambs and wolves can have a negative impact on the system error rate, therefore users in these categories are called weak users. Different investigations [Doddington et al. 1998, Hicklin et al. 2005, Poh and Kittler 2008, Poh and Kittler 2009] have confirmed that the weak users constitute only in a small fraction of the population of a biometric system, however, their contribution in the error rate can be disproportionately high.

In a recent investigation by Yager and Dunstone [Yager and Dunstone 2007], four new animals were proposed, as follows:

e. **Worms** are individuals who, like goats, tend to have a low match scores when compared against themselves. In addition, they also tend to have a high match score when compared against other individuals;

f. **Chameleons** are individuals who receive high match scores when compared against any user, either himself, or another user;

g. **Phantoms** are individuals who receive low match scores when compared against any user, including themselves;

h. **Doves** represent the most “perfect” users of the biometric system. Such as sheeps, they receive high match scores when compared against themselves, but they also receive low match score when compared against other users.

It is important to note that the animals do not necessarily represent a distinct and mutually exclusive subgroup of users [Doddington et al. 1998, Wayman 2004]. In fact, it is possible that they do not even exist in a real system. Animals can be better understood as a tendency of behaviour and, thus, an individual may be more susceptible to attack than another. Doddington conducted their study based on data from speech recognition, but the concept of animals can be applied to several areas of biometric identification. Subsequently, other studies have shown the existence of animals in other biometric modalities. Wayman [Wayman 2004] demonstrated the existence of lambs and wolves in fingerprint data with a high degree of significance; Wittman [Wittman et al. 2006] examined the existence of animals in face recognition; Poh and Kittler [Poh and Kittler 2008] showed, individually, the phenomenon in different biometrics; and others, such as [Yager and Dunstone 2007].

These studies mentioned previously presented methods for dealing with the animals existence, but, when regarding the relationship between user groups, there are important questions to be considered. Firstly, the fact that a user is, notoriously, a lamb might mean that he is more likely to be goat, for example. Doddington reported a positive correlation between lambs and wolves. Indeed, lambs and wolves reflect a symmetry of matching algorithms. Wittman showed that goats, on average, present also bad recognition rate when compared against other users. In other words, goats are terrible wolves.

3. Statistical analysis

In order to analysis the existence of the goats, lambs and wolves, we have used the well-known Kruskal-Wallis method to test the null hypotheses with a level of significance at 0.01. The weak users belonging to these animals can be identified, if they exist, with a statistical framework based on the concept of percentiles of match score, as proposed by [Doddington et al. 1998]. We obtain the p-th percentile of the N ordered values calculating the rank $r = \frac{p}{100} \times N + \frac{1}{2}$. The percentile value of each group of animals is very dependent on the database as well as the nature of the application. The percentile value used for identifying goats, lambs and wolves was 2.5, the same used by Doddington.

One limitations of the Kruskal-Wallis test is its necessity of having at least five samples in each distribution. Thus, in our investigation, each sample have its own animal score. From the user’s template, we have calculated the performance of the animals as follow:

- Goats:
  - for each user, we have compared each template against all the templates of himself, storing the worst result.
  - we have tested the null hypothesis from the previous five results.
  - the average of the five results is the goat score (genuine score).
  - the users with goat score less than 2.5 percentile are considered goats.
- Lambs:
  - for each user, we have compared all the samples of the others users with each sample of this user, storing the best (higher) result. As each sample is attacked by all others users, we have five results by user.

   - the average of these five results is the lamb score (impostor score).
   - the users with lamb score greater than 97.5 percentile are considered lamb.
- Wolves:
   - the test for the wolves is symmetric to the lambs test, thus, we only need to reverse the comparison.

As the definition of worms, chameleons, phantoms and doves are based on the relationship between genuine and impostor score of the users, the test of existence of these animals groups use the values already calculated for the first animals.

The null hypothesis for worms, chameleons, phantoms and doves is that the genuine and impostor score performance for each user are independent and, therefore, there are about $1/16$ users in each animal group. Thus, we use the concept of ranks and quartiles to calculate the potential animals in each animal group and we test the null hypotheses using the Equation (1) [Yager and Dunstone 2010].

$$f(c; n, p) = \sum_{i=c}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{1}$$

As an example, we can calculate the amount of chameleons by analysing individuals who have both high genuine and impostor score ($\mathcal{G}_H \cap \mathcal{I}_H$). The Figure 1 shows the distribution region where we find the potential chameleons.

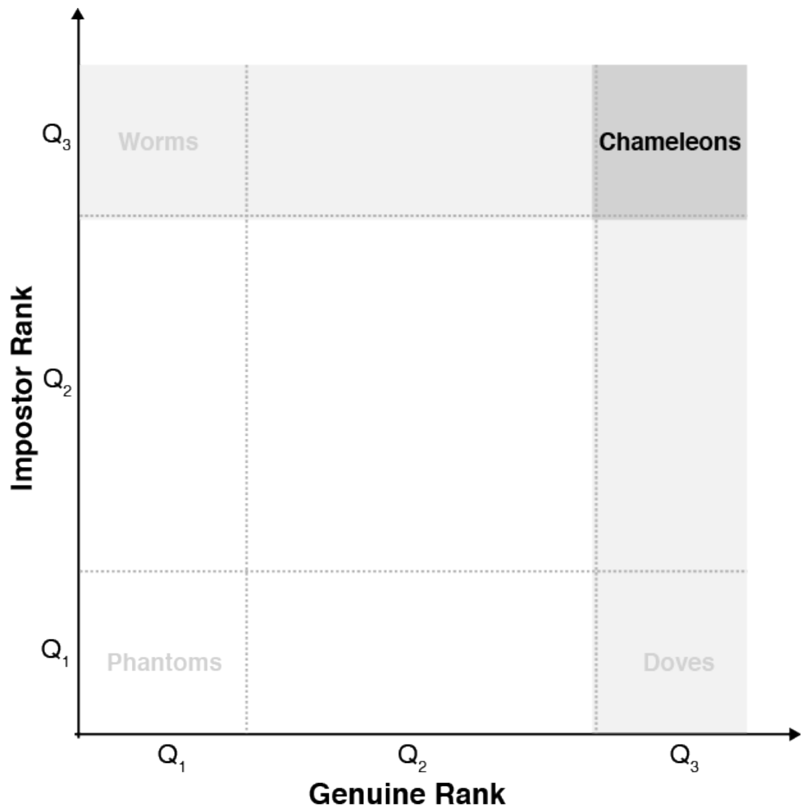

Figure 1. Distribution region where we find potential chameleons

Once the quantity of potential chameleons is calculated, we define a confidence level α for the existence of the animal in the system. For our experiments $\alpha = 0.05$. The null hypotheses is rejected if $f(c; n, p) < \alpha$, where, in this case, c is the quantity of potential chameleons, n is the size of the users population and p the probability of an individual be chameleon. If the null hypotheses is rejected, the users in that region will be considered chameleons.

4. Results and discussion

In our study we have analysed the presence of animals in the fingerprint-based biometric system, have identified the weak users and discussed the reason why these users are in that weak group. We have also analysed the performance behaviour of each finger separately and together.

4.1. Experimental Data

In order to perform our investigation, we have used the CASIA-FingerprintV5 [ref 2013] database, that contains 20.000 fingerprint images of 500 individuals, where each individual contributed with 40 impressions (5 for each finger, except the little finger). The individuals were instructed to rotate the fingers with various levels of pressure to generate significant intra-class variations. Figure 2 shows two images from the CASIA-FingerprintV5 database.

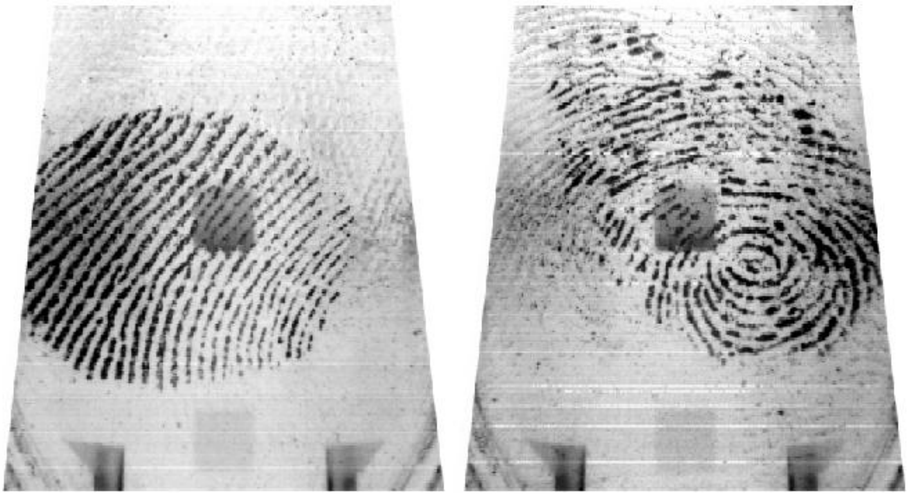

Figure 2. Example of images in the CASIA-FingerprintV5

In our experiments, we have only used the thumb finger of each individual for simplicity, therefore, we ended up with 2500 samples.

We have extracted the fingerprint features of the images using the MINDTCT [Watson et al. 2007b] feature extractor and have used the BOZORTH3 [Watson et al. 2007a] matcher to calculate the similarity between fingerprints. Both, MINDTCT and BOZORTH3 matcher, are modules of the NBIS distribution.

4.2. Results for the analysis of existence: Goats

Figure 3 shows the mean distribution of worst genuine match score with an interval of 97.5% of confidence level. As a high genuine score represents a better matching, any user

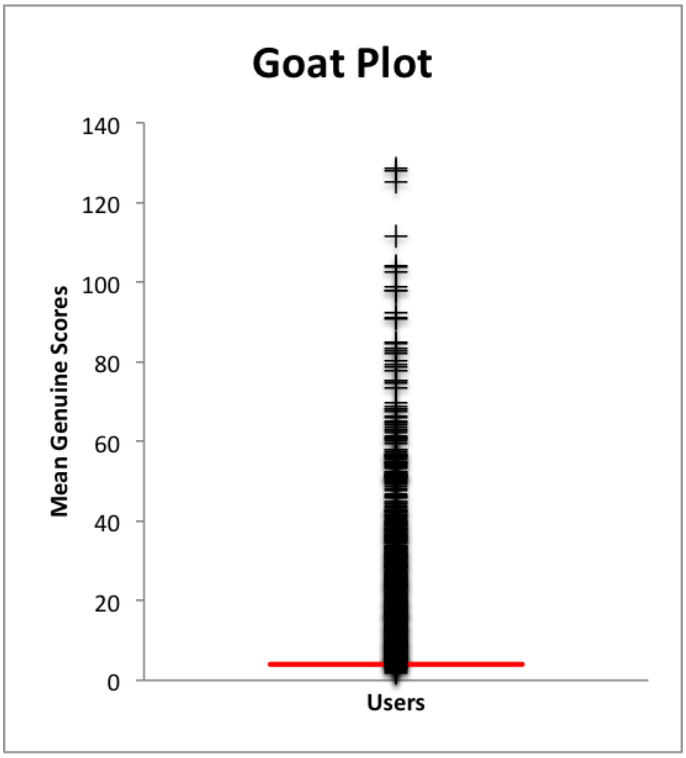

Figure 3. Distribution of the mean worst genuine match score.

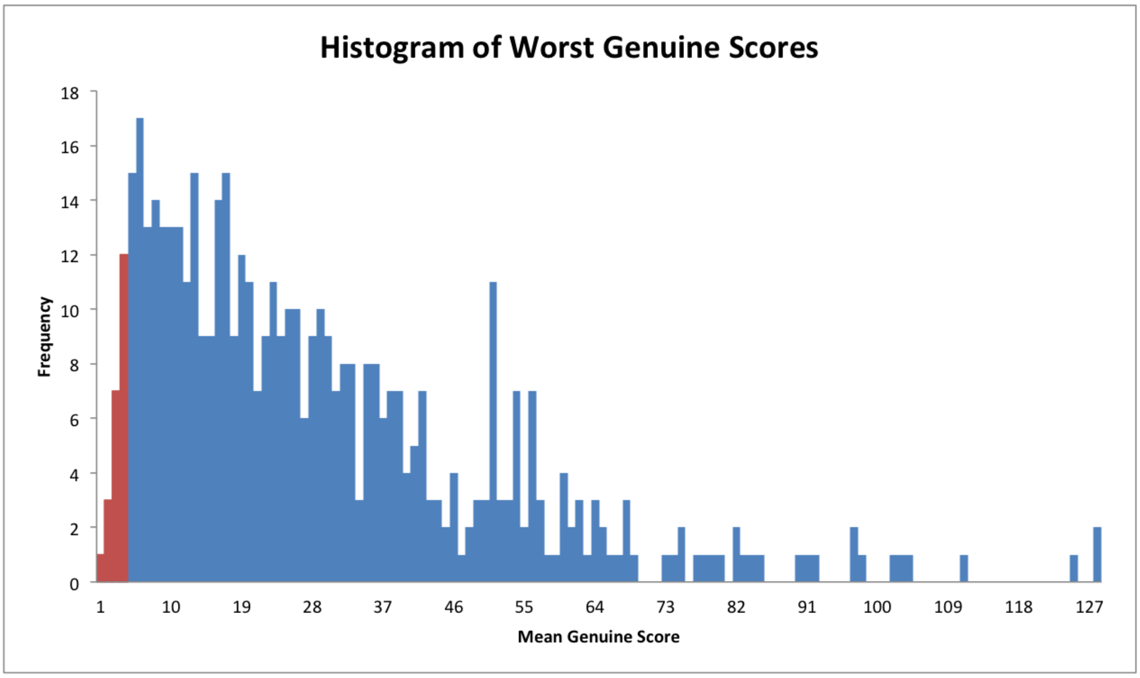

Figure 4. Histogram of the mean worst genuine match score.

with a goat statistic value which is below the line can be considered as a goat relative to other users. However, the presence of goats is not clear. If there were no user dependencies, as proposed by the null hypotheses, only one user in forty would be below the line.

The histogram of the mean for the worst genuine match scores is shown in Figure 4. The values corresponding to the users below the 2.5 percentile are shown in red. Clearly, the mean of the worst genuine scores are concentrated in the lower values. In fact, 45.2% of users have the mean of the worst genuine score less than 20. On the other hand, 24.4% of users have the genuine match score greater than 40 (recommended value for the BOZORTH3 commercial use).

The Kruskal-Wallis test was applied to 500 users on the CASIA-FingerprintV5 and we have concluded that there are no significant evidence of the goat presence. Therefore, the null hypothesis was not rejected with 0.05 of significance and the existence of goats, at least for the system used, cannot be confirmed.

Figure 5 shows 3 samples of each of the two individuals with worst goat score. It can be noted that there is a significant intra-class variation among samples of each user and the samples of different regions are very noisy.

Figure 5. Above: 3 samples of the user with worst goat score; Below: samples of the user with the second worst goat score.

4.3. Results for the analysis of existence: Lambs and Wolves

Figure 6 shows the impostor score distribution for lambs (right) and for wolves (left) with a 97.5% confidence interval. As a low impostor score represents a worst matching, any user below the line can be considered, respectively, a lamb and wolf relative to the other users. The histogram of the lambs and the wolves is shown on Figure 7.

Once again, the Kruskal-Wallis test was applied to the users and we can conclude that there are no significant evidences either of the lambs or the wolves presence. There-

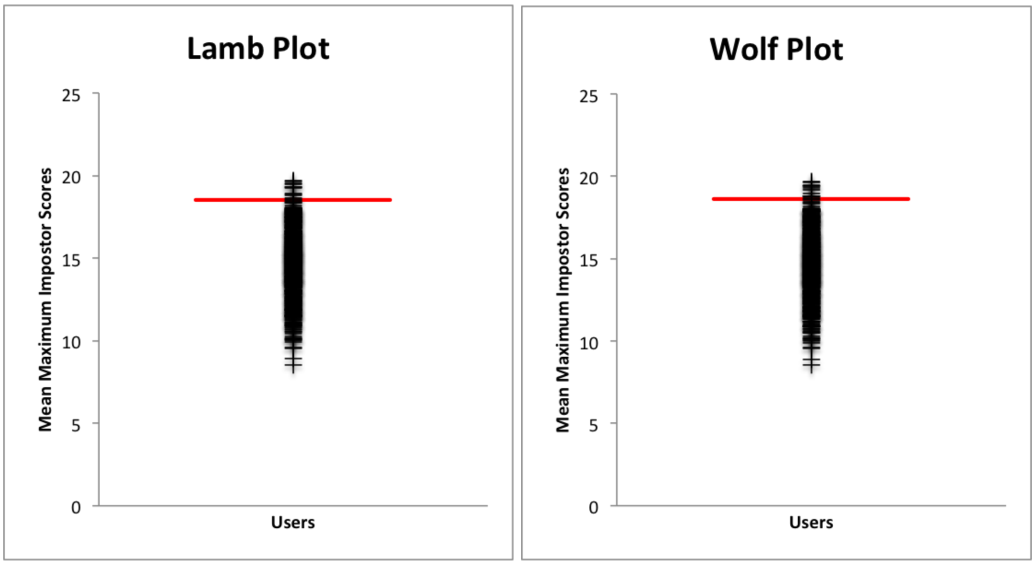

Figure 6. Distribution of the mean higher impostor match score.

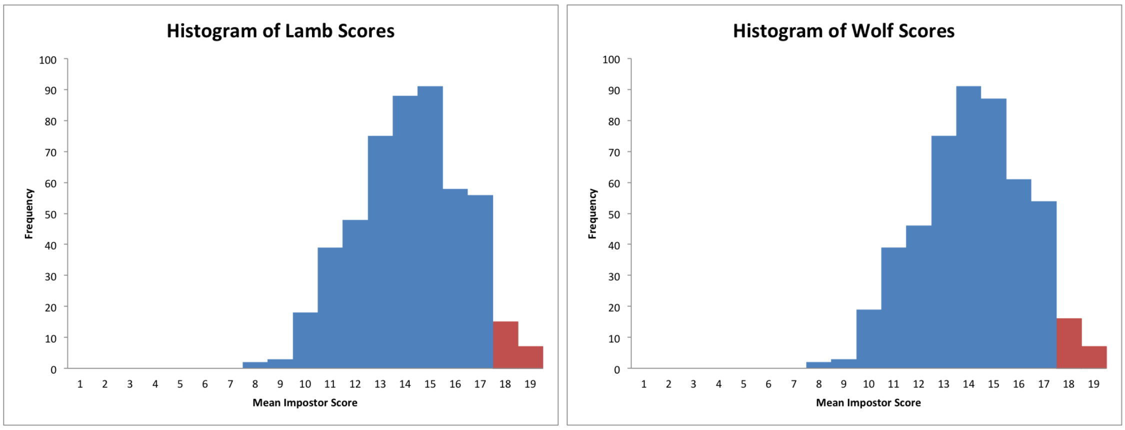

Figure 7. Histogram of the mean higher impostor match score.

fore, the null hypothesis was not rejected with 0.05 of significance and the lambs and wolves existence cannot be confirmed.

It is possible to note that among the 13 users with higher lamb score, 12 are within the 13 users with higher wolf score. This can be explained as a reflection of the symmetry of the matching algorithm BOZORTH3. We can also note that among the 13 users with worst wolf scores, 2 users are among the worst goat scores, demonstrating that the relationship between wolves and goats, reported by [Wittman et al. 2006], although less expressive, also exists in this system and for this database.

Figure 8 shows samples from each of the two individuals with a higher tendency to lamb and wolf. It can be said that these users are attacking each other. Due to the matching algorithm nature, which takes into account the distance and the orientation between minutiae pairs, it is difficult to identify, by the images, common features.

Figure 8. Above: 3 samples of the user with higher tendency to lamb and wolf; Below: samples of the user with the second higher tendency.

4.4. Results for the analysis of existence: Chameleons, Phantoms, Worms and Doves

The Figure 9 shows the relationship between genuine and impostor match score. Table 1 summarises the experimental outcomes and the associated probability values for chameleons, phantoms, worms and doves using the new approach proposed in this paper. It is possible to note a significant presence or absence of each animal. A possible interpretation for a system like this (containing, or not, each animal) is that, in some cases, the probability of a user be falsely rejected is not independent of the probability of being falsely accepted. More importantly, these values show that the probability $< 0.01\%$ of existence (or absence) of the animals are the result of chance.

Table 1. Results and probabilities for presence or absence of the animals.

	Worms	**Doves**	**Chmlns**	**Phntms**
Test	absence	absence	present	present
Probab.	<0.01%	<0.01%	<0.01%	<0.01%
qty. users	10	17	55	50

Due to the significant absence of worms and doves, Figure 9 illustrates a strong positive correlation between the average genuine and the average impostor match score. This result shows a significant population of phantoms in the lower left corner. Further analysis shows that the major of phantoms are users which the fingerprints are quite damaged.

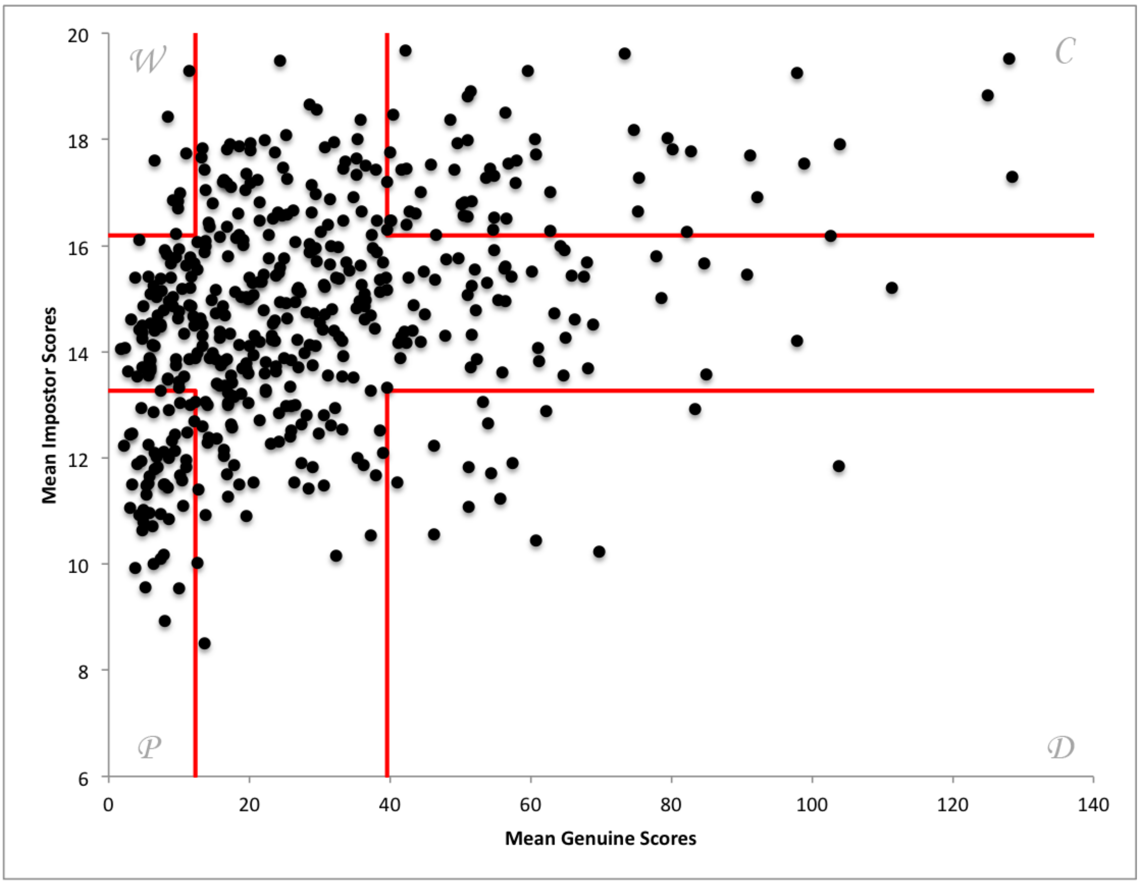

Figure 9. Relationship between genuine and impostor match scores.

4.5. Finger analysis

It is well accepted that the presence of goats, lambs and wolves can harm the user authentication, however, if a user is a goat in one of its finger will also be for remaining others? In order to verify how is the behaviour of the performance finger by finger, we have identified users for each animal. In this experiment, we have only considered the average of the scores within the user value of 2.5 percentile. Therefore, we have considered only the users who have a higher animal behaviour tendency.

The distribution of fingers considered as belonging to the user group that have higher tendency for each animal is shown in Figure 10, while Table 2 shows the percentage of occurrences per finger. Initially, we can note that the number of fingers belonging to wolves and lambs are the same. This, again, reflects the symmetry in the matching algorithm, although 16% of the users belong to only one group of these two animals.

The number of users who have only a finger belonging to one of the animals is represented by the majority. In this case, the mere use of a biometric authentication system by using more than one finger can solve this specific problem.

Analysing the goat samples, we can note that approximately 10% of the users have three or more fingers considered as a goat, and a user has five fingers.

As in the analysis of the existence of animals, we want know if the score distribution is finger independent. If this hypothesis is true, this means that any finger can be

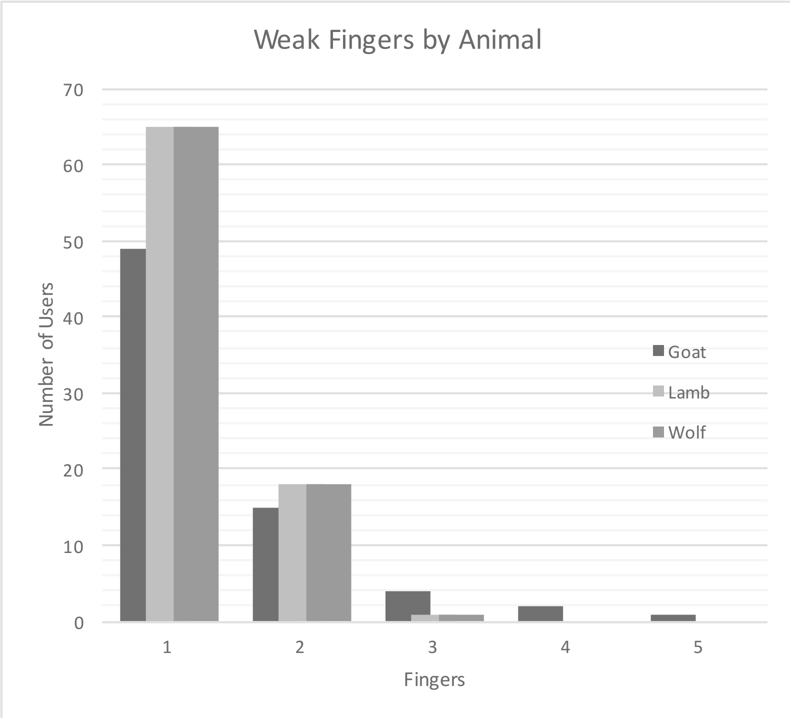

Figure 10. Number of fingers by user belonging to each animal.

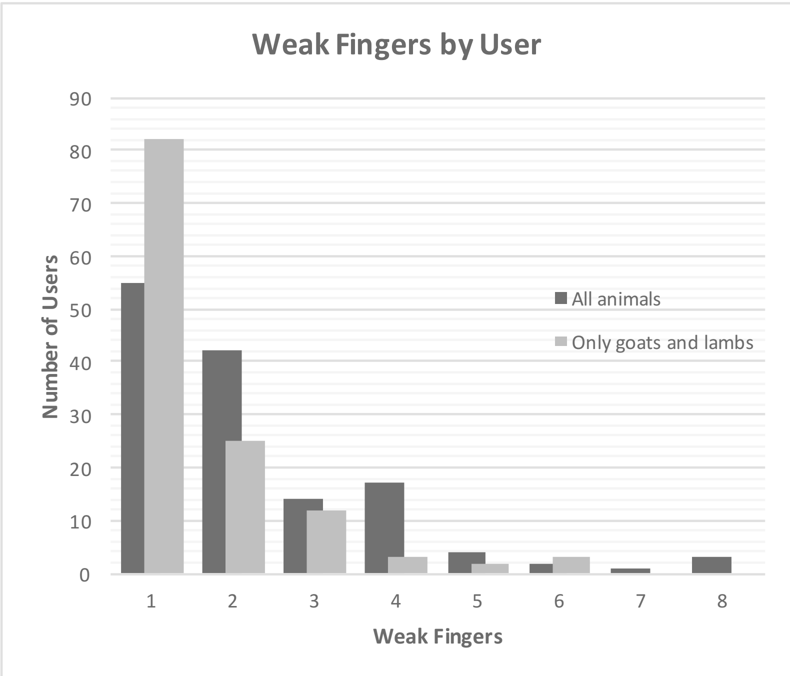

Figure 11. Number of weak fingers by user.

chosen for authentication resulting in the system performance being the same. We have once again used the Kruskal-Wallis test to verify this independence, and thus, the null hypothesis is that this score distribution is finger independent.

In this experiment, the distribution for each finger is represented by the mean

Table 2. Percentage distribution of users by fingers occurrence for each animal.

	1 finger	2 fgrs	3 fgrs	4 fgrs	5 fgrs
Goat	69.01%	21.13%	5.63%	2.82%	1.41%
Lamb	77.38%	21.43%	1.19%	0%	0%
Wolf	77.38%	21.43%	1.19%	0%	0%

score of all users. Therefore, we have, for each finger, the mean score of all user in the 2.5 percentile. We perform this analysis for each animal and, for all of them, the null hypothesis was rejected with 0.01 of significance level.

5. Final remarks

This paper has introduced an investigation of the presence of the biometric menagerie in a fingerprint-based biometric authentication system. It was shown that, despite what is claimed in the literature, there is no statistical significance evidence for the existence of goats, lambs and wolves. This is mainly due to the fact that, although the users tend to have individual match score distributions of genuine and impostor match score, some factors can reduce this effect, such as a good algorithm for the minutiae extraction.

On the other hand, we have found evidence of the existence of the other four animals (worms, doves, chameleons, and phantoms). These animals are defined in terms of the relationship between genuine and impostor match scores, and it was shown that they are present or absent significantly in real biometric fingerprints. This presence or absence of these animals do reflect the properties of the matching algorithm as well as the population of users.

We can, therefore, conclude that the reasons for the existence of a particular group of animals are varied and complex. They depend on a number of factors, including the processes used in the registration, feature extraction and matching algorithm, the quality of the captured fingerprint by the sensor and the intrinsic properties of the user population.

References

(2013). Casia-fingerprintv5.

Doddington, G., Liggett, W., Martin, A., Przybocki, M., and Reynolds, D. (1998). Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In *ICSLP'98*.

Hicklin, A., Watson, C., and Ulery, B. (2005). The myth of goats: How many people have fingerprints that are hard to match? Technical report, NIST.

Poh, N. and Kittler, J. (2008). A methodology for separating sheep from goats for controlled enrollment and multimodal fusion. In *Biometrics Symposium, 2008. BSYM '08*.

Poh, N. and Kittler, J. (2009). A biometric menagerie index for characterising template/model-specific variation. In *Advances in Biometrics*. Springer Berlin Heidelberg.

Watson, C. I., Garris, M. D., Tabassi, E., Wilson, C. L., Mccabe, R. M., Janet, S., and Ko, K. (2007a). User's guide to export controlled distribution of nist biometric image software (nbis-ec).

Watson, C. I., Garris, M. D., Tabassi, E., Wilson, C. L., Mccabe, R. M., Janet, S., and Ko, K. (2007b). User's guide to nist biometric image software (nbis).

Wayman, J. L. (2004). Multifinger penetration rate and roc variability for automatic fingerprint identification systems. In *Automatic Fingerprint Recognition Systems*. Springer New York.

Wittman, M., Davis, P., and Flynn, P. (2006). Empirical studies of the existence of the biometric menagerie in the frgc 2.0 color image corpus. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*.

Yager, N. and Dunstone, T. (2007). Worms, chameleons, phantoms and doves: New additions to the biometric menagerie. In *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*.

Yager, N. and Dunstone, T. (2009). *Biometric System and Data Analysis: Design, Evaluation, and Data Mining*. Springer Publishing Company, Incorporated.

Yager, N. and Dunstone, T. (2010). The biometric menagerie. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*.

