

A Secure Biometric Authentication Scheme Based on Robust Hashing

Yagiz Sutcu
Polytechnic University
Six Metrotech Center
Brooklyn, NY 11201
ysutcu01@utopia.poly.edu

Husrev Taha Sencar
Polytechnic University
Six Metrotech Center
Brooklyn, NY 11201
taha@isis.poly.edu

Nasir Memon
Polytechnic University
Six Metrotech Center
Brooklyn, NY 11201
memon@poly.edu

ABSTRACT

In this paper, we propose a secure biometric based authentication scheme which fundamentally relies on the use of a robust hash function. The robust hash function is a one-way transformation tailored specifically for each user based on their biometrics. The function is designed as a sum of properly weighted and shifted Gaussian functions to ensure the security and privacy of biometric data. We discuss various design issues such as scalability, collision-freeness and security. We also provide test results obtained by applying the proposed scheme to ORL face database by designating the biometrics as singular values of face images.

Categories and Subject Descriptors

E.m [Data]: Miscellaneous – *biometrics, security, robust hashing.*

General Terms

Security, Design, Human Factors.

Keywords

Authentication, Biometrics, Robust Hashing, Security, Privacy.

1. INTRODUCTION

Today, as a member of technology driven society, we are faced with many security and privacy related issues and one of them is reliable user authentication. Although for most of the cases, traditional password based authentication systems may be considered secure enough, the level of security is limited to relatively weak human memory and therefore, it is not a preferred method for systems which require high level of security. An alternative approach is to use biometrics (fingerprints, iris data, face and voice characteristics) instead of passwords for authentication. Higher entropy and uniqueness of biometrics make them favorable in so many applications which require high level of security, and recent developments of biometrics technology enable widespread use of biometrics-based authentication systems.

Despite the qualities of biometrics, they have also some privacy and security related shortcomings. In the privacy point of view,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM-SEC '05, August 1–2, 2005, New York, New York, USA.

Copyright 2005 ACM 1-59593-032-9/05/0008...\$5.00.

most of the biometrics-based authentication systems have common weakest link which is the need for a template database. Typically, during the enrollment stage, every user presents some number of samples of their biometric data and using this information, some descriptive features of that type of biometric (i.e., singular values, DCT coefficients, etc.) are extracted. Analyzing these extracted features, templates for each and every user are constructed. During authentication, a matching algorithm tries to match the biometric data acquired by a sensor with the templates stored in the template database. According to the result of the matching algorithm, authentication succeeds or fails. This enrollment and authentication process is illustrated in Figure 1.

Table 1. Properties of different authentication techniques [6]

Method	Examples	Properties
What you know	User ID Password PIN	Shared Easy to guess Forgotten
What you have	Cards Badges Keys	Shared Duplication Lost or stolen
What you know + What you have	ATM card + PIN	Shared PIN is weakest link
Something unique about user	Fingerprint Face, Iris, Voice, ...	Not possible to share Forging difficult Cannot be lost or stolen

Main weakness of the biometrics is the fact that, if biometrics compromised, there is no way to assign a new template, and therefore, storing biometric templates should be avoided. However, unlike passwords, the dramatic variability of biometric data and the imperfect data acquisition process prevents the use of secure cryptographic hashing algorithms for securing the biometrics data. Secure cryptographic hashing algorithms such as MD-5 and SHA-1 give completely different outputs even if the inputs are very close to each other. This problem made researchers to ask the following question: Is it possible to design a robust hashing algorithm such that, the hashes of two close inputs are same (or close) whereas inputs which are not that close will give completely different outputs?

In recent years, researchers have proposed many different ideas to overcome this problem. Juels and Wattenberg [1] proposed a fuzzy commitment scheme which simply uses quantization idea to define closeness in the input space. Depending on the

quantization level, if noisy biometric data is close enough to its nominal value determined at the time of enrollment, user will be successfully authenticated. Later, Juels and Sudan [4] proposed “fuzzy vault” scheme, which combines the polynomial reconstruction problem with error correcting codes, in order to be able to handle unordered feature representations. Tuyls et al. [2], [3] also used error-correction techniques with quantization to handle the variability of biometric data. Ratha et al. [6] and Davida et al. [5] were among the first to introduce the concept of cancelable biometrics. In [6], the main idea is to use a noninvertible transform to map biometric data to another space and store that mapped template instead of the original one. This approach will give the opportunity to cancel that template and corresponding transformation when the biometric data is compromised. Vielhauer et al. [7] also proposed a simple method to calculate biometric hash values using statistical features of online signatures. The idea behind their approach can be summarized as follows: After the determination of the range of feature vector components, the length of extended intervals and corresponding offset values of each interval are calculated. At the time of authentication, extracted feature values are first normalized using the length and offset values determined previously and then rounded accordingly to get the hash value. Although this approach is simple and fast, hash values cannot be assigned freely due to nature of the scheme and this makes the collision resistance performance of the proposed method questionable. Furthermore, need for storing the offset and interval length values for each individual is another weakness from the security point of view. More recently, Connie et al. [10], Teoh et al. [11] and Jin et al. [12] proposed similar bio-hashing methods for cancelable biometrics problem. A detailed survey of all these approaches can be found in [7].

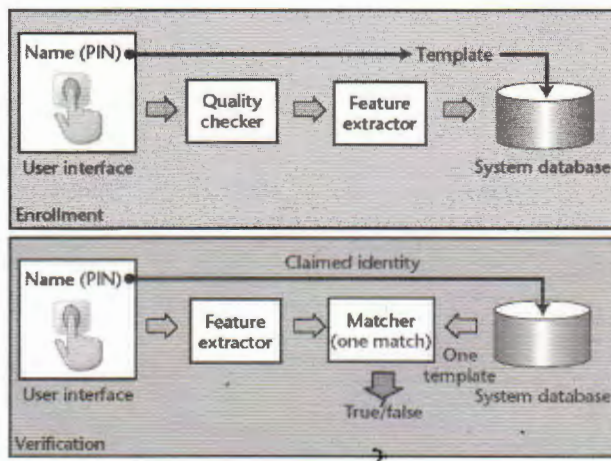


Figure 1. Enrollment and authentication process of a biometric authentication system [13].

In this paper, we analyze the performance and feasibility of a biometric based authentication system which relies on the sequential use of a robust hash function and a cryptographic hash function (i.e., MD-5, SHA-1). The robust hash function is a one-way function designed as a sum of many Gaussian functions. In section 2, we give the details of our approach and discuss related design issues and challenges. In section 3, we elaborate on the

setup and present simulation results. Our conclusions and the scope of future work are provided in Section 4.

2. PROPOSED SCHEME

In [6], Ratha et al. proposed the use of a noninvertible distortion transform, in either the signal domain or the feature domain to secure the biometric data of the user. This will not only eliminate the need for storing biometric template in the database but also provide flexibility to change the transformation from one application to another to ensure the security and privacy of biometric data. Figure 2 simply illustrates that noninvertible transformation idea such that, the value of a feature x is mapped to another space (y) meaning that, given y , it is not possible to find the value of x since the inverse transform is one-to-many.

However, in this setup matching process needs to be performed in transformed space, and it is not a trivial task to design such a transform because of the characteristics of the feature vector. Typically, depending on the type of biometric used and feature extraction process, the components of feature vectors take different values changing in some range, rather than taking precise values, and therefore candidate transform has to satisfy some smoothness criteria. While providing robustness against to variability of same user's biometric data, that transformation also has to distinguish different users successfully.

Apart from the difficulty in design of such transformations, the smoothness properties of that transformation might reveal the range information of the feature vector components to some extent. Furthermore, overlapping or even close ranges may pose another problem for this design and especially it becomes more difficult to satisfy the required robustness.

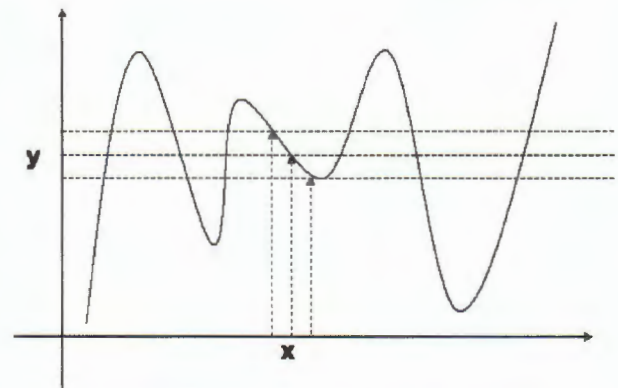


Figure 2. An one-way transformation example.

In this context, other than the one-way transform and error tolerance requirements, there are other important design issues that need to be addressed. One concern is the scalability of the overall system. Since the number of users may vary over the time, the design has to be flexible enough to accommodate new user addition and deletion. That is, it should be possible to create new accounts at minimum cost as well as providing collision free operation. Another design issue is the user-dependence of these transformations. If not impossible, it is extremely difficult to design such a single non-invertible transformation for each user that satisfies all design specifications. Finally, output space of the candidate transformation needs to be quantized in order to make it

possible to combine this transformation with a secure hashing algorithm.

Considering these issues, we propose an alternate form of one-way transformation which is combined with a secure cryptographic hash function. The one-way transformation is designed as a combination of various Gaussian functions to function as robust hash. The cryptographic hash is used to secure the biometric templates stored in the database.

In this approach, we simply assume that every component of the n -dimensional feature vector is taking some value in some range without imposing any constraint on the values and ranges as follows:

$V_i = [v_{i1}, v_{i2}, \dots, v_{in}]^T$ is the n -dimensional feature vector of i^{th} user of the system and

$$\bar{v}_{ij} - \delta_{ij} \leq v_{ij} \leq \bar{v}_{ij} + \delta_{ij} \quad i = 1, \dots, N; j = 1, \dots, n$$

where $2\delta_{ij}$ determine the range of the j^{th} component of the feature vector of the i^{th} user and N is the total number of the users.

In the enrollment stage, enough number of samples of biometric data is acquired from users. Using these data, range information of each user's feature vector (δ_{ij}) is obtained. Once this information is determined, every component of the feature vectors are considered separately and a single Gaussian (red Gaussian in Figure 3) is fitted to corresponding range considering the output value assigned to that component of the feature vector. Let us explain this fitting operation with the help of an example.

Consider j^{th} component, v_{ij} , of the feature vector of user i . Assume that v_{ij} takes values between $(v_{ij} - \delta_{ij})$ and $(v_{ij} + \delta_{ij})$ and also assume that o_{ij} is the assigned output value for that component of the feature vector. Set of points to be used for Gaussian fitting will be:

$\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ where

$$(x_1, y_1) = (v_{ij} - \delta_{ij}, o_{ij}); (x_2, y_2) = (v_{ij}, o_{ij} + r) \text{ and}$$

$$(x_3, y_3) = (v_{ij} + \delta_{ij}, o_{ij}) \text{ with } r \text{ is a uniformly selected random number between } 0 \text{ and } 1.$$

After that stage, some number of fake Gaussian functions are generated and combined with the first one in order to cover the whole range and hide the real range information and this process will be repeated n times for every user. This process is illustrated in Figure 3.

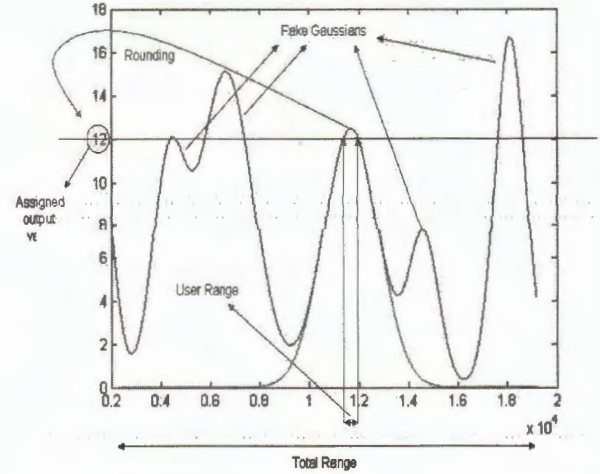


Figure 3. Design process of proposed one-way transformation.

Certainly the parameters of these transformations are determined and given to the users by an authorized, trusted third party and furthermore this information is stored in a smartcard or a token which needs to be used at the time of authentication.

Authentication process will be performed in the following manner: Firstly, user's biometric data will be acquired with a sensor and his/her feature vector will be extracted. Secondly, one-way transformation, stored in the smart-card, will be generated, and it will be evaluated at the extracted feature vector component values. Lastly, values obtained after quantization will be concatenated together to form a string and then hashed. The hashed value will be compared to user's entry for authentication, as illustrated in Figure 3.

Assuming the fact that hashing algorithm used in this scheme is secure, for an attacker who has access to the database, determining the real values of the feature vector by looking at hashed values stored in the database will not be possible. Furthermore, even though the information on the smartcard is compromised, it still remains difficult for an attacker to guess the real values of the biometric data of the user by only analyzing the shape of one-way transformation of that user.

This approach is also scalable not only because of the fact that generating gaussians is relatively a simple task, but also it is possible to generate and assign different output values for each and every component of a feature vector while satisfying collision-free operation. Considering a number of potential users, one can generate m -by- n matrix (where m is the total number of users and n is the dimensionality of the feature vector) ensuring that any two rows of this matrix are not identical. By the time of a new user account needed, one row from that matrix will be assigned to that user and his/her one-way transformation will be designed using these values.

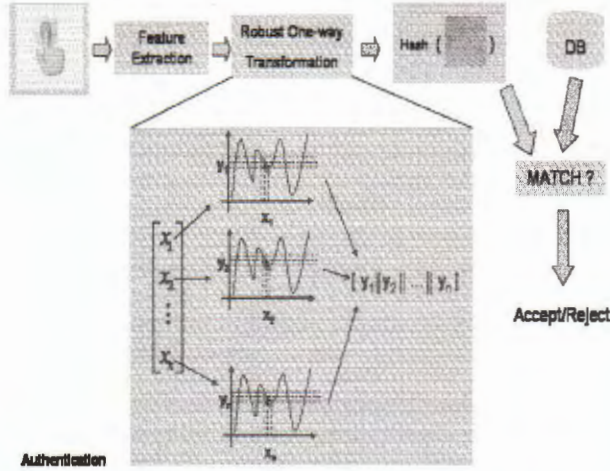


Figure 4. Authentication process of proposed scheme.

However, since the range information is hidden by the peaks of the gaussians, these transformations are not used in an efficient manner. This weakness of the proposed approach may be observed by an intelligent attacker and help him/her to reduce brute force guessing space for biometric data. To be able to reduce this leakage of information, number of fake gaussians should be as high as possible but also these fake gaussians need to have variance and magnitude parameter values close to real gaussian fitted to the real range. But in this case, especially if the length of user range is relatively high with compared to the length of overall range for a specific component of his/her feature vector, it will not be possible to generate so many number of fake gaussians. The reason for that constraint is the consequence of the fact that, summation of overlapping tails of gaussians will have a relatively high value and this will make the design difficult and resulting transformation will have a poor hiding quality.

Finally, since the proposed approach is generic, type of biometric data may be changed regularly to assure the privacy and security of the system. The proposed approach is tested on the ORL face database using simple singular value based feature vectors and performance of the scheme will be presented in the following section.

3. EXPERIMENTAL RESULTS

In recent years, singular values have been introduced as the feature vector for face recognition and other applications. In this study, we also used singular values as feature vector for testing our scheme and in the following sub-sections, we will give a brief explanations about singular value decomposition and its properties and then explain our experimental setup.

3.1 Singular Value Decomposition

Let us first introduce the singular value decomposition of a matrix.

Theorem 1 (Singular Value Decomposition)

If $A \in R^{m \times n}$, then there exist orthogonal matrices $U \in R^{m \times m}$ and $V \in R^{n \times n}$ such that

$$A = U \Sigma V^T \text{ where } \Sigma = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p),$$

with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ and $p = \min(m, n)$

Following theorem provides the necessary information about the sensitivity of singular values of a matrix.

Theorem 2 (Perturbation)

Let $\tilde{A} = A + E \in R^{m \times n}$, be a perturbation of A and let $\tilde{A} = \tilde{U} \tilde{\Sigma} \tilde{V}^T$ be SVD of $\tilde{A}$, then $|\lambda_i - \tilde{\lambda}_i| \leq \|E\|_2$ for $i = 1, \dots, p$ where $\|E\|_2$ is induced -2 norm of E

Since SVD is one of the well-known topics of linear algebra, we omitted to give detailed analysis of this subject and interested reader may find more details in [9].

3.2 Experiments and Results

The ORL face database [8] is created for face recognition related research studies and as a result, differences of facial expressions of the subjects are more than acceptable limits for a biometric authentication system. However, since creating a new set of face images for our study is not trivial, we decided to make our preliminary tests using this database.

ORL face database consists of 10 different images of 40 distinct subjects and the size of each image is 92×112 , 8-bit grey levels. In our simulation, we randomly divide each 10 samples of subjects into two parts namely, training and test sets while training set has 6 of the images, test set has the remaining 4 samples. In our simulations, only first 20 singular values of the images are considered and none of the data pre-processing techniques (such as principal component analysis (PCA), linear discriminant analysis (LDA), etc) are used.

The performance of the proposed scheme is determined in terms of basic performance measures of biometric systems, namely, False Acceptance Rate (FAR) and False Rejection Rate (FRR). However, another type of performance measure that needs to be considered is due to the possibility that a one-way transformation designed for a particular user can be used in authentication of another user. (This is the likelihood of user X authenticating himself as user Y while using user Y's smartcard.) This type of error can be interpreted as a factor contributing to FAR. For the sake of clarity, we will denote such errors by FAR-II.

In our analysis, we first extract a feature vector from the set of training images, and then determine the range of variation for each feature vector component. The range for each component is calculated by measuring the maximum and minimum values observed in the training set and expanding this interval by some tolerance factor (e.g., 5% or 10%) in order to account for the possible variation in a feature value that is not represented within the available training images. Our results obtained for 5% and

10% tolerance factors are given in Tables 2 and 3. It should be remembered that in our experiments, we used 6 out of 10 images (available for each person) to estimate the range and tested the scheme on the rest of the images

Table 2. FRR results

Correct Authentication Ratio	# of correctly authenticated subjects (5% tolerance)	# of correctly authenticated subjects (10% tolerance)
4/4	2	15
3/4	8	10
2/4	13	10
1/4	13	4
0/4	4	1
Total	40	40

Table 3. FAR-II results

Incorrect Authentication Ratio	# of incorrectly authenticated subjects (5% tolerance)	# of incorrectly authenticated subjects (10% tolerance)
0/39	12	1
1/39	12	7
2/39	9	3
3/39	6	4
$\geq 4/39$	1	25
Total	40	40

Table 2 summarizes the *FRR* performance of the proposed scheme in the following manner: First column stands for the correct authentication ratio, which is the ratio of correctly authenticated number of unseen test images to the total number of unseen images — 4 images. On the other hand, each row shows the number of persons that were successfully authenticated at a given authentication ratio. For example, the number 2, which stands in the second column of first row indicates that; there are 2 subjects (out of 40), who are authenticated successfully for all of the test images. Similarly, the number 4 (second column and fifth row) denotes that there are 4 subjects (out of 40) that were not authenticated at all, indicating that the assumed tolerance factor is not satisfactory.

In Table 3, *FAR-II* performance of our scheme is presented in a similar manner. For a given user, all remaining (39) users are tried to be authenticated using that user's smart-card (one-way transform function) and presenting their own biometric data and results obtained are summarized in Table 3. First column of Table 3 represents the ratio of incorrectly authenticated users to the number of remaining users — 39 users. For example, there are 12 (out of 40) users who were never confused by any other user, meaning that, none of the remaining 39 users were authenticated as one of them. On the other hand, with a tolerance factor of 10% there are 25 users whose authentication data were collided with at least 4 of the remaining 39 users.

In our scheme, any of the users who uses his/her own smart-card, is authenticated as another user, which means, *FAR* is zero. However, false acceptance results (*FAR-II*), presented in Table 3, which actually indicate the rate of being authenticated as another user using other user's smart-card. One of the reasons to observe such a relatively high false acceptance rate (especially with a

tolerance factor of 10%) is due to nature of ORL face database which contains images captured under extensively varying conditions. As a result, actual range information of the singular values could not be estimated efficiently due to the high variations depending on the differences of facial expressions of the subjects. It should be noted that, to further improve the performance one can employ data pre-processing techniques such as PCA or LDA. It is reasonable to expect that, when appropriate pre-processing techniques are employed along with higher dimensional feature vectors (e.g., more than 20 singular values), performance of the proposed scheme will be better. These considerations will be the parts of our future work.

4. CONCLUSION AND FUTURE WORK

We proposed a secure biometric based authentication scheme which employs a user-dependant one-way transformation combined with a secure hashing algorithm. Furthermore, we discussed its design issues such as scalability, collision-freeness and security. We tested our scheme using ORL face database and presented simulation results. Preliminary results show that, proposed scheme offers a simple and practical solution to one of the privacy and security weakness of biometrics-based authentication systems namely, template security.

In order to improve the results, our future focus is three-fold: (1) To find a more flexible and efficient way to design one-way transformations with less parameters; (2) To find a metric for measuring and comparing data hiding quality of these one-way transformations. (3) To test our approach on larger databases also with different types of biometric data.

5. REFERENCES

- [1] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in Proc. 6th ACM Conf. Computer and Communications Security, G. Tsudik, Ed., 1999, pp. 28–36.
- [2] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in Proc. 4th Int. Conf. Audio and Video-Based Biometric Person Authentication, 2003, pp. 393–402.
- [3] E. Verbitskiy, P. Tuyls, D. Denteneer, and J. P. Linnartz, "Reliable biometric authentication with privacy protection," presented at the SPIE Biometric Technology for Human Identification Conf., Orlando, FL, 2004.
- [4] A. Juels and M. Sudan, "A fuzzy vault scheme," in Proc. IEEE Int. Symp. Information Theory, A. Lapidot and E. Teletar, Eds., 2002, p. 408.
- [5] G. I. Davida, Y. Frankel, and B. J. Matt, "On enabling secure applications through off-line biometric identification," in Proc. 1998 IEEE Symp. Privacy and Security, pp. 148–157.
- [6] N. Ratha, J. Connell, and R. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," IBM Syst. J., vol. 40, no. 3, pp. 614–634, 2001.
- [7] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric Cryptosystems: Issues and Challenges," Proceedings of the IEEE, Vol. 92, No. 6, June 2004.
- [8] The ORL Database of Faces, available at <http://www.uk.research.att.com/facedatabase.html>
- [9] Strang, G., "Introduction to linear algebra", 1998, Wellesley, MA, Wellesley- Cambridge Press.

- [10] T. Connie, A. Teoh, M. Goh, and D. Ngo, "Palmhashing: a novel approach for cancelable biometrics", Elsevier Information Processing Letters, Vol. 93, (2005) 1-5.
- [11] A. B. J. Teoh, D.C.L. Ngo, and A. Goh, "Personalised cryptographic key generation based on facehashing", Elsevier Computers & Security, Vol. 23, (2004), 606-614.
- [12] A. T. B. Jin, D.N.C Ling, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenized random number", Elsevier Pattern Recognition, Vol. 37, (2004) 2245-2255.
- [13] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric Recognition: Security and Privacy Concerns", IEEE SECURITY & PRIVACY, March/April 2003.