CORRELATION-BASED PASSWORD GENERATION FROM FINGERPRINTS

Gurgen Khachatrian, Hovik Khasikyan

Abstract: *In this paper, a methodology for reliable password generation from fingerprints is developed. In contrast to traditional biometric systems, proposed algorithm does not authenticate user by matching his or her biometrics. Reference data gives no information about the password and fingerprint. In hand with cryptography, this method can provide highly secure protection for cryptographic keys used in Digital Signatures and Digital Rights Management systems.*

Keywords: *Password Generation, Confidentiality, Authentication, Privacy, Security, Fingerprints, Image Processing, Pattern Recognition, Template Matching.*

ACM Classification Keywords: *D.4.6 Security and Protection (K.6.5)*

Introduction

In most cryptographic applications, secret keys such as private keys in public key systems or symmetric keys are protected by passwords. Secret keys are stored on the computer and encrypted by hashed password. They are released for user authentication by entering a proper password by the user.

However, conventional passwords are relatively simple and sometimes easy to guess or to break. People remember only short passwords. What is more, they tend to choose passwords, which are easily cracked by dictionary attacks [1, 2, 3].

Biometric technologies provide alternative solution to this problem. In traditional biometric systems user is authenticated by matching his or her biometrics. In these systems, there is no secret key, which can be guessed or lost. Instead of decrypting the secret key with a password, system performs matching of input biometrics with corresponding biometrics stored in the database. If matching is successful, the secret key is released.

Nonetheless, biometric systems have their vulnerabilities [4]. The main weakness of these systems is in their design. Since biometric data is stored locally (on smart card or another computer), possession of the smart card or access to the database gives access to the biometric template of the users. This gives attacker an opportunity to break current and all the other authentication systems using the same biometric identifier. What is more, authentication process in this design is completely decoupled from the key release and outputs only one bit accept/reject decision. This makes the system vulnerable against Trojan horse attacks (which can simply overwrite the decision bit).

Based on the above observations, in this paper a password generation method is proposed from fingerprints. The generated passwords are not stored on any device. Only a small reference data should be kept in database or on smart card with all the other credentials of the user. Reference data gives no information about the password. Thus, the passwords cannot be guessed, lost, shared or stolen.

The accuracy of the system is evaluated on a large database of fingerprints and gives promising results for the use in commercial authentication, where very low False Rejection Rate (FRR) is required for a given False Acceptance Rate (FAR).

Previous Works

The proposed approach is inspired in particular by the “Generation of Secret Key from Fingerprint Image” of Maslennikov [5]. In order to generate keys from fingerprint images in [5] a correlation based method was used, which is described in detail in “A Correlation-Based Fingerprint Verification System” [6].

In correlation based verification, system selects appropriate template images from the primary fingerprint and then uses template-matching techniques to locate them on the secondary image. If their locations are the same as they were on the primary image then the owner is recognized as genuine (Figure 1).

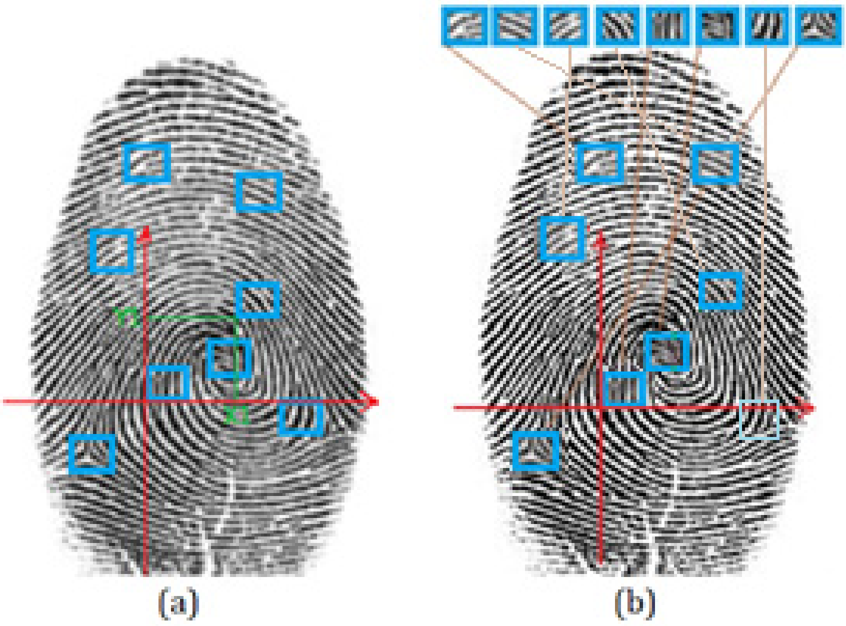

Figure 1: Correlation-based template matching. (a) primary fingerprint and chosen templates; (b) secondary fingerprint and located templates.

In [5], these locations were used to generate a cryptographic key. Having all the templates located on the secondary image, system determined coordinates of the aligned templates, considering the first template as a starting point in the coordinate system. Coordinates are in the form: X1, Y1, X2, Y2, X3, Y3 …, where X1 shows the horizontal distance between the first template and the second one, Y1 is the vertical distance between the first template and the second one. X2 is the horizontal distance between the first and third templates and so on. Using these coordinates and some additional data (e.g. passwords), in [5] system generated cryptographic key.

The weaknesses of the system were high FRR and long execution time. To be authenticated user had to scan his/her finger 5 to 7 times [5]. These factors made the system uncomfortable and not practical.

Fingerprint Processing

Fingerprint is one of the noisiest image types. In order to reduce the noise and enhance useful information, first Gaussian Smoothing [7] algorithm is applied to mitigate noisy points on fingerprint image, and then the image is binarized using adaptive image binarization techniques [8].

Using traditional threshold based methods to convert gray scale fingerprint image into black white shows low accuracy. The reason is that different points of finger are pressed with different strength on the screen of scanner. Results of binarization with different threshold values are illustrated in Figure 2.

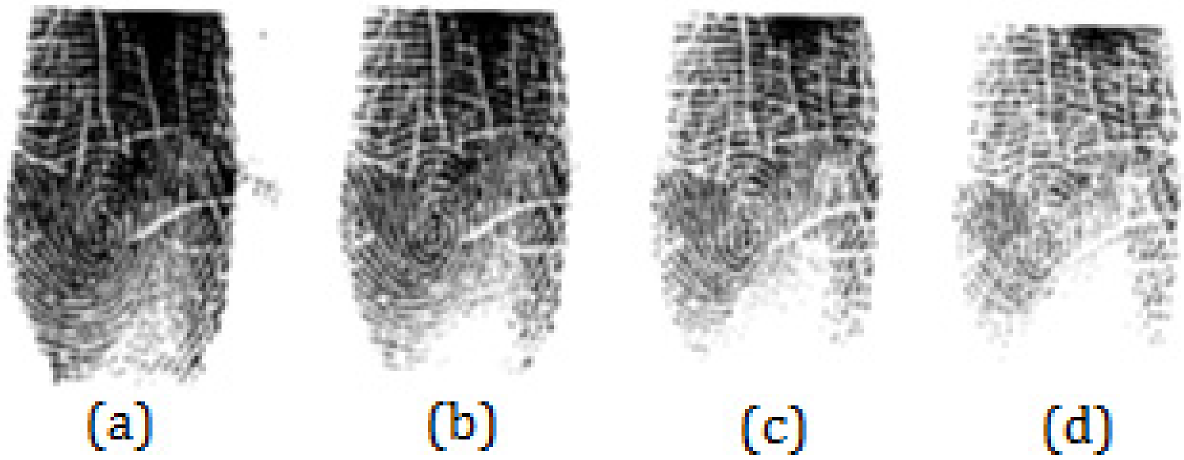

Figure 2: Results of binarization with different threshold values (threshold value increases from left to right).

From Figure 2 it can be seen, that for a small threshold values valleys at the upper part become too dark. For bigger threshold values, ridges at the bottom disappear. This problem can be solved using adaptive binarization algorithm. In this case there is no universal threshold value for whole image, but for each pixel its own threshold value is calculated separately. For each pixel a square with specific sizes is chosen, where the total sum of pixel intensity is calculated. If the gray value of the pixel is greater than threshold, then it is set to white, otherwise it is black. This method is much more accurate, if the size of the square for threshold calculation is selected appropriately. Analyses have shown that the best size of the square is different for different fingers. The best results are obtained when block size is a little bigger than twice of the thickness of the ridge. This can be explained with the fact, that the square of this size is highly probable to contain equal quantities of black and white pixels. In such case, it works better than histogram correction algorithms. The result of adaptive binarization with correct block size is illustrated in Figure 3.

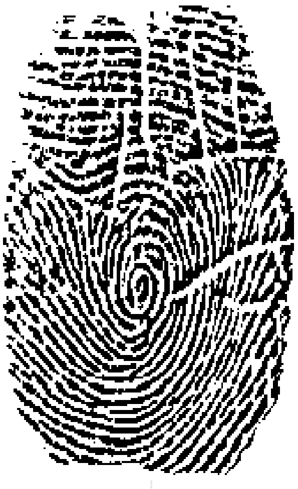

Figure 3: Results of adaptive

Shapes of the Reference Images

The shape of reference images is very important for the accuracy of the system. In the previous works square templates were used in template matching, but because of physical features of fingers these templates do not show very accurate results. Analyses on the local database have shown, that skin of the fingers is deformed mainly in horizontal axis, whereas on the vertical axis the distortion is minimal. Because of these concerns, the local database was analyses for all possible rectangular patterns.

The most optimal shape of the templates was found to be a wide rectangle with its length equal two to four times height. This ratio is justified with the previously stated explanation of shape distortions. All the further calculations in this work are implemented for the best rectangular shape for the whole database. However, the best shape of templates is individual for each finger and can be found by distinct analysis.

Template Selection

The template selection process is the least attractive part of correlation-based verification. The number of operations is very high and makes the system very slow.

The quality of a template is considered high, if it is located on the secondary image easily and precisely. In other words the template should fit as well as possible at the same location, but as badly as possible at the other locations. This feature is the same as the uniqueness of the template. To count the uniqueness of the template, template is compared with all the other templates on image pixel wise.

In this work a novel template selection methodology is proposed, based on specific features of the templates with high uniqueness. Analyses show that for the unique templates there is specific distribution of similar patterns. The most similar templates, which are deterministic for counting the final uniqueness of the template, are located close to the original. Figure 4 illustrates this distribution.

The yellow square shows the location of the original template. Red circles show locations of the first four closest templates (by Hamming distance) to the original one. The green triangles are next 5 to 10 templates, blue points 11 to 20. This was the general way of distribution for the majority of fingerprints of the local database. Nevertheless, this is not a rule. For some fingerprints very unique templates are found, for which very similar patterns were located far from it (Figure 5).

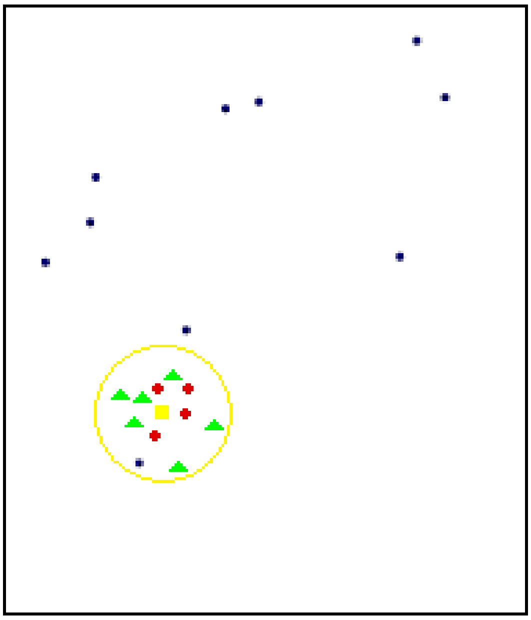

Figure 4: Example of similar pattern distribution.

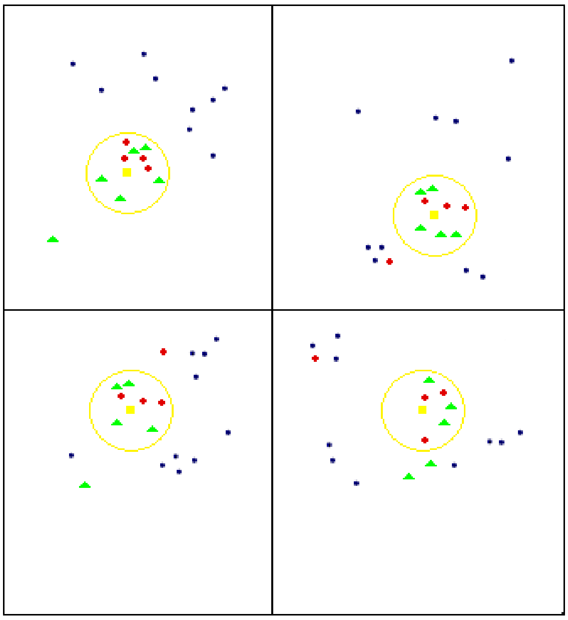

Figure 5: Distribution of the similar patterns for the most unique templates.

To use this novel characteristic some special measures are required in our algorithm. First, for each template its uniqueness is calculated by comparing it with its neighbors. Then these templates are sorted by uniqueness value. After sorting, the most unique templates are chosen and verified for validity. The decision is reached by aligning new templates on the image and locating the closest templates. If algorithm finds very close template far from the original, the template is treated as invalid. It is removed to another place in the queue, according to its new uniqueness value. The next template by uniqueness is considered as candidate and the same procedures are implemented until all the required templates are found.

One more observation was carried out, that when counting the uniqueness of a template, the neighbor templates (in two pixel distance) should not be considered. These templates are very similar to the original and have misleading effect, when counting the uniqueness of the template. Furthermore, since approximations should be made with the coordinates of the templates, there is no difference between the original template and the closest neighbors.

Template Matching

For faster template matching it is important to exclude comparisons with obviously dissimilar templates. In this work it is proposed to use the first lines of the reference image and the candidate template to make a decision. The first lines are used to find out whether two templates have anything in common or the system should omit this candidate and continue pixel wise search. The further comparisons are permeated or cancelled based on the correspondence of these lines and on the predefined threshold value. The threshold value should be about 70%; not more, otherwise it affects accuracy. For higher confidence, the threshold value can be decreased down to 55%. For the threshold value less than 50% there is no acceleration at all, because practically all candidate templates satisfy this requirement.

Password Generation

In this paper an approximation method is proposed for reliable password generation from the extracted coordinates. Analyses show that for the fingerprint image with sizes 240x280 these coordinates waive 0 to 5 pixels from their original locations (Table 1).

	X1	Y1	X2	Y2	X3	Y3	X4	Y4	X5	Y5	X6	Y6	X7	Y7
Original	72	-27	-25	47	-36	110	-73	46	-58	-113	-29	77	-12	-70
Imprint1:	72	-27	-24	48	-36	111	-72	50	-58	-112	-28	79	-12	-71
Imprint2:	73	-27	-27	46	-37	109	-75	45	-59	-114	-32	76	-12	-70
Imprint3:	72	-27	-26	47	-37	110	-75	47	-60	-113	-32	77	-13	-70
Imprint4:	72	-27	-25	47	-36	110	-74	46	-59	-113	-30	77	-12	-70
Imprint5:	72	-27	-25	47	-36	110	-73	46	-59	-113	-29	76	-13	-70
Imprint6:	72	-27	-23	47	-35	110	-71	46	-58	-113	-26	78	-12	-70
Imprint7:	72	-27	-26	48	-36	110	-73	49	-59	-112	-30	78	-12	-69

Table 1: Coordinates of the localized templates

Thus, with precision of five pixels, the final result will be exact. Therefore, ([X0- Xn] mod 10) is kept in the database for each of the reference images. In the example at the table 1

ΔX1 = ([X0- X1] mod 10) = 2, ΔY1 = ([Y0- Y1] mod 10) = 3, ΔX2 = ([X0- X2] mod 10) = 5 etcetera.

Testing and Analysis

All the tastings and analyzes are performed on the local database of fingerprints. The database consists totally of flat (dab) impressions. There are 320 fingerprints of 40 different persons (8 imprints of index fingers from each person). The images have been captured using U.are.U 4500 optical fingerprint reader [9]. All fingerprint images are recorded at 512 dpi and as 256 gray tone images (8-bit grayscale). After scanning images are resized to 240x280 pixels and stored in the database. The fingers have been pre-scanned to insure a representative mix of varying quality impressions, ranging from those of poor quality to those of excellent quality (Figure 7).

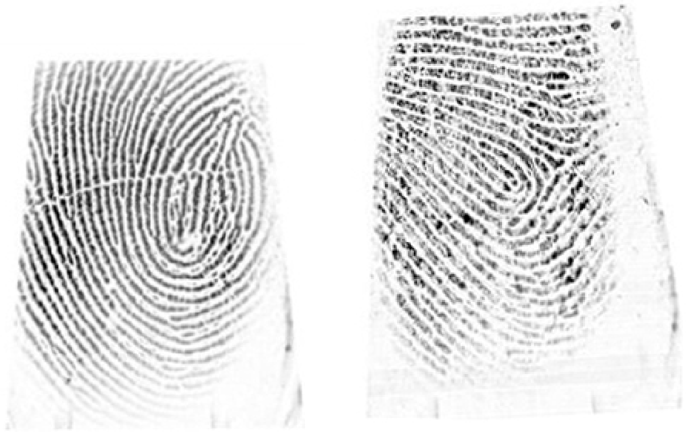

Figure 7: Examples of fingerprints from the local database

The experimental demonstrate that the proposed algorithm has acceptable accuracy for the use in commercial authentication (where very low False Rejection Rate is required for a given False Acceptance Rate). The FRR of the algorithm equals 3.35%, while FAR is less than 0.1%.

Entropy

Entropy of the system is important to avoid generating the same passwords for different users. To demonstrate the distribution of the coordinate values, for each coordinate X1, Y1, X2… the quantity of accepted values is counted. Figure 8 illustrates this distribution for the first position.

Analyses have shown that there is a dangerous peak between -3 to 1. The reason for accumulations is that the second template by its uniqueness, was mostly located very close to the first one. The third and fourth templates are also presupposed to be close, so their values were between -5 to 5.

These accumulations made the system vulnerable, because attacker could use the most favorable values and break the system after some iteration of the most favorable values.

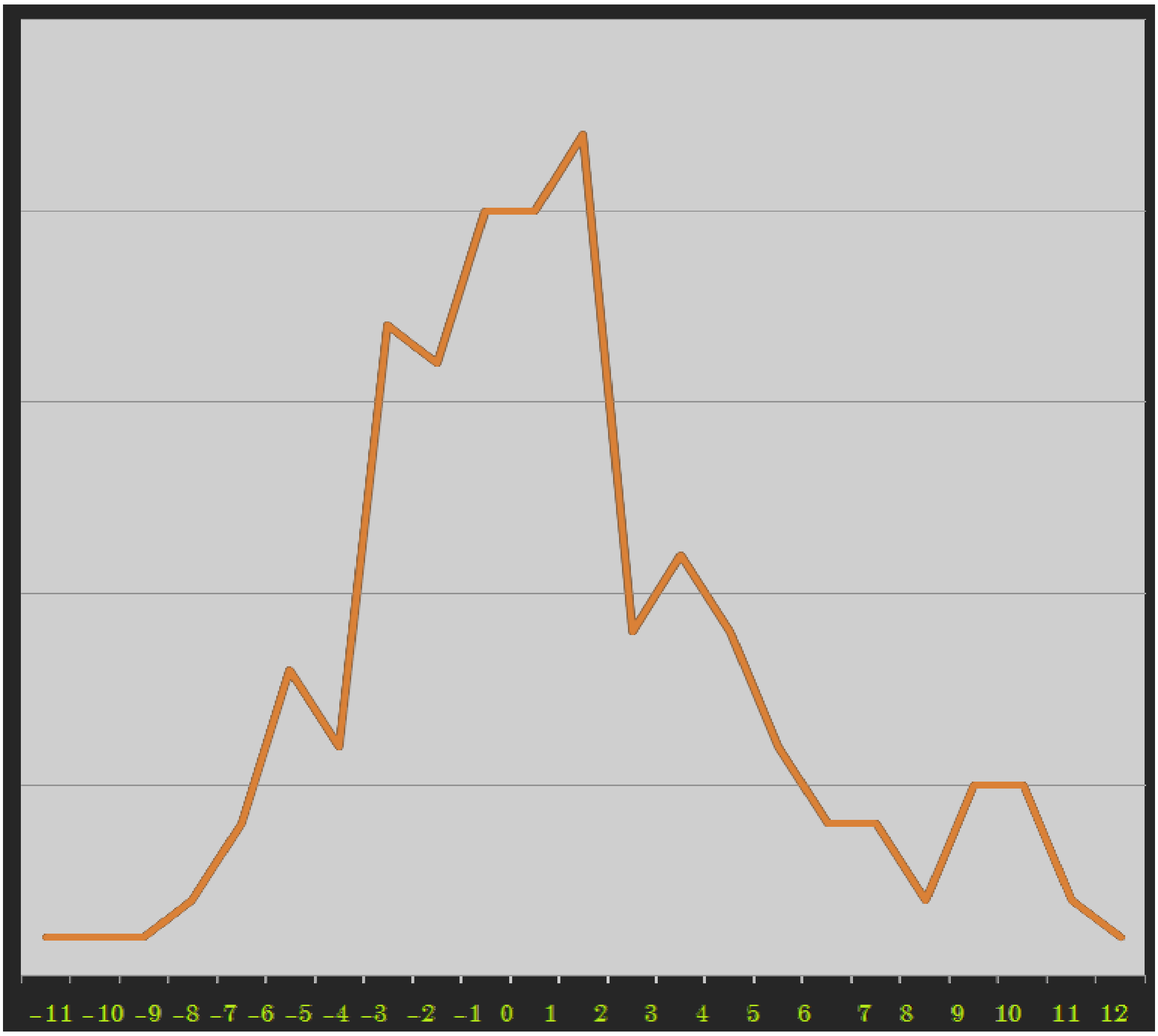

Figure 8: Accumulations of the coordinate values.

In order to handle this problem permutations of the reference templates have been introduced at the enrollment phase. After locating all required templates, their indexes are randomly shuffled, so the most unique template is stored from the first to the last positions randomly. The result of this shuffles on the distribution are depicted in the Figure 9.

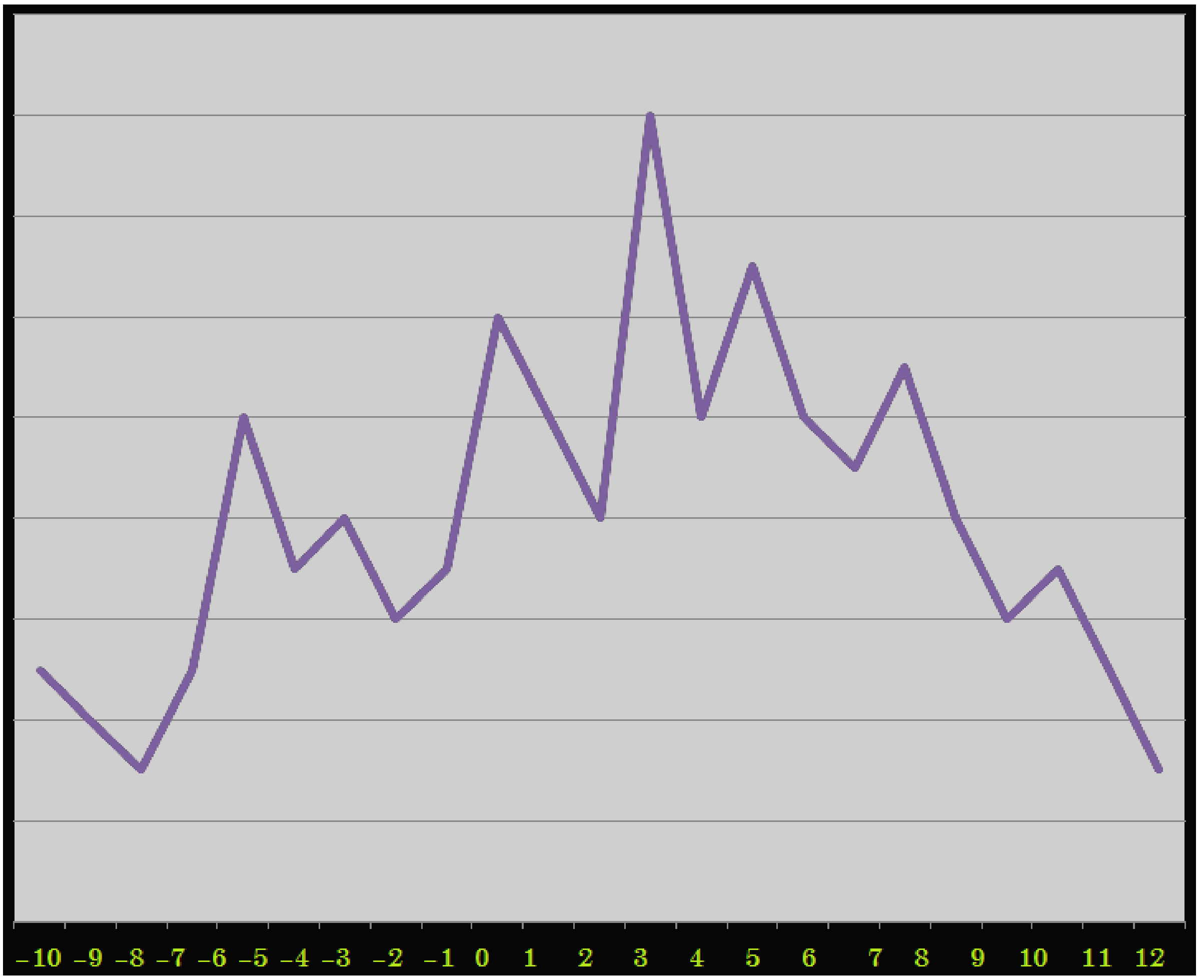

Figure 9: Accumulations after permuting the templates.

Security Analyses

The security of the system is analyzed from two points of view. At first, it was found that the proposed system has fairly low FAR (which can be made as low as 0.01%). This is due to two-stepped verification; first, system aligns templates on the secondary image and counts the accuracy of the fitting. If this value is less than predefined threshold value, this person is not authenticated. In the second phase algorithm generates a password based on the template locations. These passwords should be exact; otherwise, the user is again rejected.

Another important security measure is the size of the generated password. In this work, the size of the password is 84 bits. After locating all templates (in this work it is eight templates), system generates seven X and seven Y coordinates. The first template is the starting point in the coordinate system. Each coordinate can waive between

-24 to 24, thus each requires 6 bit memory. The size of the password equals 14*6 = 84 bit, which is the smallest general purpose-level key size [10].

The size of the password can be increased by keeping more templates, but this affects the accuracy of the system. The dependency of the error from the number of the templates is illustrated in Figure 10 (the best number of templates is found to be 8).

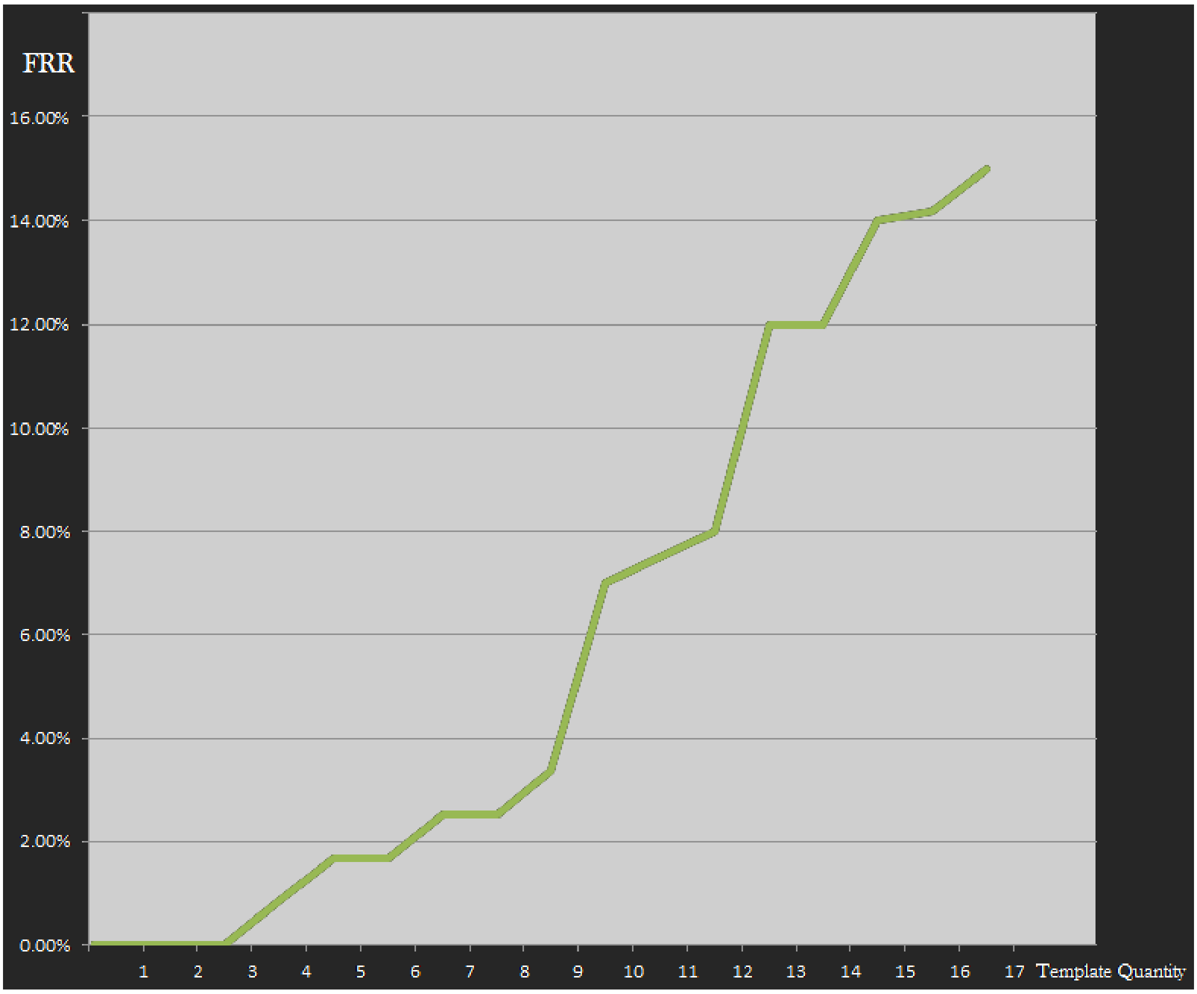

Figure 10: Error dependency from the template quantity.

For generating longer passwords template matching and coordinate approximation techniques should be made much more sensitive. To further enhance the security of the system, biometric passwords can be combined with other additional information (for example passwords or passphrases).

The problem of the fingerprint faking is also considered. In order to take over these issues only registered scanners should be used in the system, which can ensure if the fingerprint is covered with artificial layer or not (e.g. performing skin distortion [11] and odor analyses [12]).

Comparisons with Other Reference Systems

The advantage of the proposed method is that biometric data is not stored locally. Stealing the reference data (or smart card) becomes meaningless as it does not give access to the secret key and password without the finger of legitimate user. Trojan-horse attacks are also excluded, since algorithm makes decryption of the encrypted secret key.

In contrast to minutiae-based systems, this method demonstrates higher reliability in the field of password generation. In the minutiae-based systems, despite the fact, that only reliable minutia can be chosen as a source for password generation [13, 14] the probability that some of minutiae will not be observed or wrong minutiae will be extracted is still high. Unlike minutiae-based techniques, this method is able to handle low-quality images with missing and spurious minutiae.

The main disadvantage of the previous works was the demand for high computational power and low accuracy. For an image with sizes 240x280 and template's size 30x30 pixels, (240-30)*(280-30)*30*30=47*106 XOR operations were required to locate a template on the secondary image. As such, the enrollment phase required 210*250*(47*106) = 2.5*1012 XOR operations, and the verification phase 8*(47*106) = 378*106 XOR operations.

In the current work, enrollment requires 2.5 to 6*109 XOR processes for the same fingerprint (500 times faster). Verification is also improved and takes 18 to 23*106 XOR operations (16 times faster).

The accuracy and sensitivity of the system are also improved; FRR equals 3.35%, FAR is 0.01% (the EER of previous system was 7.98% [5, 6]). These improvements are the results of better template shape selection and more accurate enrollment and verification procedures.

Conclusion

The main target of this work was to develop a reliable password generation algorithm. Passwords should have been generated exactly the same each time user was verified. Nevertheless, the consistency of the passwords was one of the primary challenges.

The entropy of the generated passwords is also analyzed and is enough to resist brute force attacks. The result of analyses can be different for a very large database of fingerprints, but as these biometric passwords are used for encryption of a randomly generated secret key, a little change in entropy cannot affect the security of the system.

The security of the passwords' size is considered to be the smallest general-purpose level (84 bit key provides long-term protection against small organizations). However, it can provide only short-term protection against agencies [10]. The size of the password can be increased by processing another biometrics (e.g. palm prints) and by developing more sensitive averaging and alignment algorithms.

Acknowledgement

This research has been carried within the project *Application of Security to Biometrics and Communications*, sponsored by the Volkswagen Foundation.

Bibliography

[1] E. Spafford. Observations on reusable password choices. In Proceedings of the3rd USENIX Security Symposium, September 1992.

[2] T. Wu. A real-world analysis of Kerberos password security. In Proceedings of the 1999 Network and Distributed System Security Symposium, February 1999.

[3] R. Morris and K. Thompson. Password security: A case history. Communications of the ACM, 22(11):594–597, November 1979.

[4] Stavroulakis, P., Stamp, M.: Handbook of Information and Communication Security. Springer, Heidelberg (2010).

[5] M. Maslennikov, Practical Cryptography, Saint Petersburg, (2003).

[6] A.M. Bazen, G.T.B. Verwaaijen, S.H. Gerez, L.P.J. Veelenturf, B.J. van der Zwaag: A correlation-based fingerprint verification system, 11th Annual Workshop on Circuits Systems and Signal Processing (2000).

[7] R. Deriche, Recursively implementing the Gaussian and its derivatives, V. Srinivasan, Ong S.H., Ang Y.H. (Eds.), Proc. Second Int. Singapore Conf. on Image Proc. (Singapore, Sept.7–11, 1992), 263–267, 1992.

[8] T.Romen Singh, Sudipta Roy, O.Imocha Singh, Tejmani Sinam and Kh.Manglem Singh," A New local Adaptive Thresholding Technique in Binarisation", IJCSI-Vol 8, issue 6 No. 2 pp. 271-277 (Nov, 2011).

[9] U.are.U 4500 Fingerprint Reader

http://www.digitalpersona.com/Biometrics/Hardware-Products/U-are-U-4500-Reader/4500-Reader/

[10] Cryptographic Key Length Recommendation, ECRYPT II Recommendations (2011)

http://www.keylength.com/en/3/

[11] Antonelli, A., Cappelli, R., Maio, D., and Maltoni, D. (2006), "Fake Finger Detection by Skin Distortion Analysis," IEEE Transactions on Information Forensics and Security 1(3), 360–373 (2006).

[12] D. Baldisserra, A. Franco, D. Maio, and D. Maltoni, "Fake fingerprint detection by odor analysis," in Proc. Int. Conf. on Biometric Authentication (ICBA06) (2006).

[13] N. J. Short, A. L. Abbott, M. S. Hsiao, and E. A. Fox, "A Bayesian Approach to Fingerprint Minutia Localization and Quality Assessment using Adaptable Templates". In Proceedings of the International Joint Conference on Biometrics, 2011.

[14] Min Wu, A. Yong, Tong Zhao and Tiande Guo, "A Systematic Algorithm for Fingerprint Image Quality Assessment". In International Joint Conference on Biometrics, 2011.

Authors' Information

***Gurgen Khachatrian** – Professor, American University of Armenia, Full member of Armenian National Academy of Sciences*

e-mail: gurgenkh@aua.am

Major Fields of Scientific Research: Cryptography, Error-control coding

***Hovik Khasikyan** – Researcher, American University of Armenia*

e-mail: hovik_khasikyan@edu.aua.am

Major Fields of Scientific Research: Computer Vision, Image Processing, Biometrics, Security, Privacy