Some Observations on Fuzzy Vault Based Biometric Protocols

DURBET Axel
Université Clermont-Auvergne,
CNRS, Mines de Saint-Étienne,
LIMOS France

THIRY-ATIGHEHCHI Kévin
Université Clermont-Auvergne,
CNRS, Mines de Saint-Étienne,
LIMOS France

***Abstract*—Biometric protocols are increasingly used for their practicality. However, biometric data are highly sensitive personal data since they are identifying data that cannot be easily revoked. Moreover, their changing nature makes them hardly usable in password-based authentication protocols. One solution proposed in the literature is the Fuzzy Vault based on Shamir's secret sharing. In this poster, we present this scheme, highlight its theoretical shortcomings, and propose ways to improve it.**

***Index Terms*—Biometric, Fuzzy Vault, Protocol**

I. Introduction

Biometric authentication is seeing widespread use due to the common integration of fingerprint sensors and cameras on many smart objects. Since biometrics is more convenient and quicker to use, and biometric characteristics cannot be lost or forgotten, biometric authentication solutions are in general preferred over their password or physical token counterparts. Despite their many advantages, biometric solutions are not exempt from vulnerabilities. In fact, Biometric data are as prone to exhaustive search attacks as passwords, but unlike passwords, they cannot be efficiently revoked. Thus, biometric databases are prime targets for cyberattackers. These cyberattacks have the potential to be harmful on the long term if they lead to the theft of biometric data. Therefore, a biometric data may actually be vulnerable to impersonation attacks and privacy leakage. Moreover, these data may disclose information like genetic information [1] and diseases [2], [3]. Biometric data are categorized as highly sensitive personal data covered by the GDPR. The essential security and performance criteria that must be fulfilled by biometric recognition systems are identified in ISO/IEC 24745 [4] and ISO/IEC 30136 [5]: *irreversibility*, *unlinkability*, *revocability* and *performance preservation*. Most of the time, to ensure the aforementioned criteria, a Biometric Templates Protection is used. For more details on BTP schemes, the reader is referred to the surveys [6]–[8]. However, the main problem remains to manage the variability of the template. Thus, in recent years, various methods to achieve this have appeared. For the most famous, we have Error Correcting Code [9]–[11], Fuzzy Extractor [12]–[14] and Fuzzy Vault [15], [16]. In recent years, the community has been particularly interested in Fuzzy Vault and some protocols based on it has been proposed [17]–[19].

ANR-20-CE39-0005 (project PRIVABIO).

Contributions.: In this paper, we propose a detailed mathematical explanation of how fuzzy vaults work. We highlight the shortcomings of these and propose solutions.

II. Background

A. Lagrange interpolation

Lagrange interpolation [20] is a mathematical technique to find a the unique polynomial of degree d from $d+1$ distinct evaluations of the polynomial. This technique works on any field including finite field.

Theorem 2.1 (Lagrange interpolation): Let f be a polynomial of degree d and
$P = \{(x_i, f(x_i)) \in (\mathbf{D}_f \times f(\mathbf{D}_f)), i \in \{1, \ldots, d-1\}\}$ a set of $d-1$ evaluations of f. Then,

$$f(x) = \sum_{i=1}^{t} f(x_i) \prod_{j=1, i\neq j}^{t} \frac{x - x_j}{x_i - x_j} \tag{1}$$

B. Shamir's Secret Sharing.

Shamir's secret sharing [21] allows the sharing of a secret between n users in such a way that if at least $k \leq n$ users collaborate, they can find the initial secret. The idea is to hide the secret s in a polynomial f of degree k such that $f(0) = s$. Then each person is given a different evaluation of the polynomial. If the users collaborate, using Lagrange 2.1 interpolation, they are able to retrieve the polynomial and evaluate it in 0 so that the secret is recovered. This algorithm is a perfect (k, n)-threshold scheme. In other words, if less than k people among the n who have a share of the secret collaborate, it is unlikely that they retrieve the secret.

Theorem 2.2 (Shamir's secret sharing): Let s in $\mathbb{F}_q$ be the secret, q a prime number, $n \ll q$ the number of participants and $k \leq n$ the minimum number of shares needed to find s. Then, the secret sharing proceeds as follows:

1) Draw randomly a polynomial of degree $k-1$ such that $f(0) = s$. Draw n points p_i two by two distinct and different from zero.
2) The i-th participant receives his share of the secret $D_i = (p_i, f(p_i))$.

The reconstruction of the secret proceed as follows:

1) k users put their D_i contributions in common.

2) With the D_i, they compute $f(x)$ using the Lagrange interpolation and find the secret $s = f(0)$.

Proposition 2.1: The most important properties are the following:

- Secure: The secret is protected.
- Minimal: The size of each part of the secret does not exceed the size of the secret.
- Extensible: When k does not change, one can add or remove parties without affecting the others.
- Dynamic: The security can be easily improved without changing the secret. It is enough to change the polynomial and to rebuild the parties.
- Flexible: In organizations where the hierarchy is important, each participant can be assigned a different number of parties according to his importance within the organization.

III. Fuzzy Vault

Fuzzy Vaults (FV) [15], [16] are cryptological primitives which have the vocation to produce a key from a noisy source. This particularity makes it quite usable in the field of biometrics, and it makes it a good candidate for cryptological primitives in biometrics, as shown in [22], [23]. Moreover, this kind of primitive is used to implement some biometric protocol such as [24]–[26]. The FV construction is a pair of algorithms FV_{Gen} and FV_{open} based on the Lagrange Interpolation and Shamir's Secret Sharing. Fuzzy vaults (FV) are cryptographic schemes that take as input a key with noise, in our case, a biometric data and that returns a Helping Value (HD). The purpose of this helper value is to recover the initial secret when using a sufficiently close secret. In the case of FV, the biometric data hide a hey. HD is a helping value, which alone does not allow finding the original data nor the secret. However, having access to the helping value allows exhaustive searches if the verification of the secret can be performed.

A. FV_{Gen} Algorithm

The FV_{Gen} algorithm constructs the Helping Data HD which will allow to recover the secret with FV_{OPEN}.

The public parameters are:

- p: Prime number which generates $\mathbb{F}_p$.
- m: Size of the biometric data.
- n: Number of chaff points.
- d: Number of parts needed to reconstruct the secret.
- w: The threshold relative to d to say if yes or no two data are close.

Definition 3.1 (FV_{GEN}): FV_{GEN} run as follows:

1) Read the biometric data and extract the informations $BT = (b_1, \ldots, b_m)$ such that $\forall (i,j); i \neq j, dist(b_i, b_j) > w$.
2) Choose a random secret k and draw a polynome $f \in \mathbb{F}_q$ of degree $d-1$ such that $f(0) = k$.
3) Evaluate the b_i with f as in Shamir's Secret Sharing.
4) Draw randomly $CP = (r_1, \ldots, r_n)$ the chaff points such that $\forall (i,j), dist(b_i, r_j) > w$ and such that the r_i are indistinguishable from b_i.
5) Evaluate the r_i with a random function f'.
6) Let the data D be: $((b_1, f(b_1)), \ldots, (b_m, f(b_m), (r_1, f'(r_1)), \ldots, (r_n, f(r_n)))$.
7) Choose P a random permutation and set the helping data $HD = P(D)||H(k)$.

B. FV_{OPEN} Algorithm

The FV_{OPEN} algorithm allow recovering the secrete using a close biometric data. The public parameters are the same as before.

Definition 3.2 (FV_{OPEN}): To retrieve the secret with HD, FV_{OPEN} proceed as follows:

1) Read the fresh biometric data $BT' = (b'_1, \ldots, b'_m)$ and get the helping value HD.
2) For each index j such that $dist(\bar{b}_i, b'_j) \leq w$, get $(\bar{b}_i, f(\bar{b}_i))$ with $\bar{b}_i = r_i$ or b_i.
3) If BT and BT' are close enough, at least d pairs are correct.
4) For each subset of d couples, perform the Lagrange interpolation as in Shamir's Secret Sharing to get $f(x)$ and verify it by testing $H(f(0)) = H(k)$.

C. Current Related Problems

With the FV constructions, some problems can be highlight:

1) Extract enough well distinct minutiae: In fact, in 1 of FV_{Gen}, we ensure that the pairwise minutiae does not lie in the same ball. In other word, the minutiae must be pairwise well distinct. This condition must be fulfilled otherwise, the step 2 of FV_{OPEN} provides additional informations. More precisely, for each minutia which belong on the same ball, the shared secrete needs one less participant to be found.
2) Generation of chaff points that are statistically indistinguishable from real minutia [27]. Moreover, the above problem stay here and is amplified cause the chaff minutiae must be pairwise well distinct from the real minutiae.
3) Linkability of templates on several services if the servers are collaborating. As only the chaff point are randomized, the intersection of two distinct enrollement of the same person gives the genuine minutiae and both secret can be reconstruct.
4) Handling several distance metrics for different biometric modalities. In fact, in biometric authentication, the community use several distances such as Hamming Distance, $\mathcal{L}_1$ distance, Manhattan distance, Euclidean distance, etc. Moreover, each modality comes with its own specific distance.
5) Multiple interpolations make the protocol impractical for large vectors or large threshold. For a client who get $k > d$ points, the cost is $O(\binom{k}{d}(d\ln(d)))$ [28].
6) Good preservation of the recognition accuracy. Because of the previous point, polynomials of smaller degrees are used which impacts the performances. Ideally, we would like to have polynomials of sufficiently large degrees to minimize the loss of recognition performance.

D. Current Idea to Patch the Issues

Current ideas to patch some of the issues are:

- Issue 2: Use a Generative Adversarial Network (GAN) [29]. More precisely, the idea would be to use a GAN to generate fingerprint images and extract the minutiae. If the network is trained enough, the minutiae will be indistinguishable from real minutiae and all that will be left to do is to choose those that are well separated from the ones we already have.
- Issue 3: Introduce some variability on the minutia with a salted hash function. Indeed, each minutiae could be hashed with a different one to make them uniformly distributed. The problem is the management of the distance between these hashes. One solution would be to use a property-preserving hash functions to preserve the distance as presented in [30], [31].
- Issue 4: Use an error correcting codes (BCH or RS). Error correcting code are well known to handle the distances. However, as in biometric authentication there is several distances for several data representation, all error correcting code will not fit. In fact, for some biometric systems (IrisCode [32], FingerCode [33]), the Hamming distance is enough thus a Reed Solomon Code [11] fit the problem. For the other system, a code on finite field with Bose–Chaudhuri–Hocquenghem codes [10] with a scale and round process as proposed in [34].

Acknowledgment

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-20-CE39-0005 (project PRIVABIO).

References

[1] L. Penrose, "Dermatoglyphic topology," *Nature*, vol. 205, no. 4971, pp. 544–546, 1965.

[2] A. Imran, J. Li, Y. Pei, J.-J. Yang, and Q. Wang, "Comparative analysis of vessel segmentation techniques in retinal images," *IEEE Access*, vol. 7, pp. 114 862–114 887, 2019.

[3] A. Ross, S. Banerjee, and A. Chowdhury, "Deducing health cues from biometric data," *Computer Vision and Image Understanding*, p. 103438, 2022.

[4] "ISO/IEC24745:2011: Information technology – Security techniques – Biometric information protection," International Organization for Standardization, Standard, 2011.

[5] "ISO/IEC30136:2018(E): Information technology – Performance testing of biometrictemplate protection scheme," International Organization for Standardization, Standard, 2018.

[6] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine*, vol. 32, pp. 88–100, 2015.

[7] I. Natgunanathan, A. Mehmood, Y. Xiang, G. Beliakov, and J. Yearwood, "Protection of privacy in biometric data," *IEEE Access*, vol. 4, pp. 880–892, 2016.

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

[9] R. W. Hamming, "Error detecting and error correcting codes," *The Bell system technical journal*, vol. 29, no. 2, pp. 147–160, 1950.

[10] "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68–79, 1960.

[11] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the society for industrial and applied mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[12] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: A brief survey of results from 2004 to 2006," *Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting*, pp. 79–99, 2007.

[13] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM journal on computing*, vol. 38, no. 1, pp. 97–139, 2008.

[14] Y. Wen, S. Liu, and S. Han, "Reusable fuzzy extractor from the decisional Diffie–Hellman assumption," *Designs, Codes, and Cryptography*, vol. 86, no. 11, pp. 2495–2512, 2018. [Online]. Available: https://doi.org/10.1007/s10623-018-0459-4

[15] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237–257, 2006.

[16] V. S. Baghel, S. Prakash, and I. Agrawal, "An enhanced fuzzy vault to secure the fingerprint templates," *Multimedia Tools and Applications*, vol. 80, no. 21, pp. 33 055–33 073, 2021.

[17] R. K. Mahendran and P. Velusamy, "A secure fuzzy extractor based biometric key authentication scheme for body sensor network in internet of medical things," *Computer Communications*, vol. 153, pp. 545–552, 2020.

[18] C. Rathgeb, J. Merkle, J. Scholz, B. Tams, and V. Nesterowicz, "Deep face fuzzy vault: Implementation and performance," *Computers & Security*, vol. 113, p. 102539, 2022.

[19] S. Albermany and F. M. Baqer, "Eeg authentication system using fuzzy vault scheme," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 25, no. 8, pp. 2405–2410, 2022.

[20] E. Waring, "Vii. problems concerning interpolations," *Philosophical transactions of the royal society of London*, no. 69, pp. 59–67, 1779.

[21] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[22] K. Nandakumar and A. K. Jain, "Multibiometric template security using fuzzy vault," in *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*. IEEE, 2008, pp. 1–6.

[23] K. Nandakumar, A. K. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: Implementation and performance," *IEEE transactions on information forensics and security*, vol. 2, no. 4, pp. 744–757, 2007.

[24] P. Bauspieß, T. Silde, A. Tullot, A. Costache, C. Rathgeb, J. Kolberg, and C. Busch, "Improved biometrics-authenticated key exchange," Cryptology ePrint Archive, Paper 2022/1408, 2022.

[25] R. Canetti, B. Fuller, O. Paneth, L. Reyzin, and A. Smith, "Reusable fuzzy extractors for low-entropy distributions," *Journal of Cryptology*, vol. 34, pp. 1–33, 2021.

[26] J. H. Cheon, J. Jeong, D. Kim, and J. Lee, "A reusable fuzzy extractor with practical storage size: Modifying canetti et al.'s construction," in *Information Security and Privacy: 23rd Australasian Conference, ACISP 2018, Wollongong, NSW, Australia, July 11-13, 2018, Proceedings 23*. Springer, 2018, pp. 28–44.

[27] H. Kim, X. Cui, M.-G. Kim, and T. H. B. Nguyen, "Fingerprint generation and presentation attack detection using deep neural networks," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2019, pp. 375–378.

[28] H.-J. Stoss, "The complexity of evaluating interpolation polynomials," *Theoretical Computer Science*, vol. 41, pp. 319–323, 1985.

[29] O.-A. Ugot, C. Yinka-Banjo, and S. Misra, "Biometric fingerprint generation using generative adversarial networks," in *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*. Springer, 2021, pp. 51–83.

[30] N. Fleischhacker and M. Simkin, "Robust property-preserving hash functions for hamming distance and more," Cryptology ePrint Archive, Paper 2020/1301, 2020.

[31] K. Minematsu, "Property-preserving hash functions and combinatorial group testing," Cryptology ePrint Archive, Paper 2022/478, 2022.

[32] J. Daugman, "How iris recognition works," in *The essential guide to image processing*. Elsevier, 2009, pp. 715–739.

[33] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE transactions on Image Processing*, vol. 9, no. 5, pp. 846–859, 2000.

[34] S. Ali, K. Karabina, and E. Karagoz, "Formal accuracy analysis of a biometric data transformation and its application to secure template generation," in *SECRYPT 2020*, 2020, pp. 485–496.