2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia

A Review of Chaff Point Generation Methods for Fuzzy Vault Scheme

Bambang Pilu Hartato[1,2], Teguh Bharata Adji[1], Agus Bejo[1]
[1]Departement of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
[2]Departement of Informatics Engineering
STMIK Amikom Purwokerto
Purwokerto, Indonesia
[1,2]bambang.mti15@mail.ugm.ac.id, [1]adji@ugm.ac.id, [1]agusbj@ugm.ac.id

***Abstract*— Bio-encryption is a concept that aims to improve the traditional cryptographic security by combining biometrics and cryptography. Fuzzy Vault scheme is a bio-encryption method which is very famous for its ability to handle fuzzy biometric data while at the same time combines them with cryptographic mechanism. Moreover, Fuzzy Vault scheme is claimed as a bio-encryption method that can be deployed in System-on-Chip (SoC) based security devices. One of the requirements that must be fulfilled by a system that can be run on SoC based devices is less resource consuming. In other words, the system must have sufficient resource efficiency or must use the fairly efficient algorithms. One of the important phases in the Fuzzy Vault scheme is chaff point generation. Chaff points are analogized as a noise used for covering valuable information inside of Fuzzy Vault. Although the chaff point generation phase is a quite important phase, it would be the phase that has the highest computation level. Hence, research on chaff point generation method to find efficient methods while maintaining the security aspects of Fuzzy Vault Scheme became an interesting topic in bio-encryption field for some researchers. This paper will discuss and also will do some comparisons toward three current chaff point generation methods that have been proposed by some researchers. Thus, this paper is expected to contribute knowledge about the characteristics of these three methods.**

Keywords—bio-encryption; biometrics; chaff points; chaff point generation; cryptography; fuzzy vault

I. INTRODUCTION

There are at least four aspects that must be considered to ensure the information security, namely confidentiality, integrity, availability, and accountability [1]. In relation to information security issues, confidentiality aspect has a significant impact. That is because confidentiality is the main foundation of information security's pillars. In other words, if an attack successfully takes down the confidentiality aspect of a system, the attacker could easily attack the other security aspects. Hence, some particular methods are needed to build a quite strong confidentiality in order to maintain the security of information system.

One of the methods that can be used to strengthen the confidentiality of an information system is the using of cryptography [2]. So far, cryptography has evolved over time to strengthen its security. One of its latest evolution is bio-encryption, where its concept combines cryptographic scheme with biometric technology.

Combining the concept of cryptography with biometric technology is based on the idea to improve the security of cryptography by adding biometric features into cryptography. The biggest problem faced by the cryptography is the difficulty in providing the truly unique encryption keys, whereas biometrics has very unique data or features [3]. Hence, combining both of them will allow cryptographic scheme has the very unique keys and will enable cryptographic scheme can only be done by authorized people. Thus, the security level of a cryptographic scheme can be improved.

One of the bio-encryption techniques which are often developed because of their reliability is Fuzzy Vault Scheme [4]. This technique was introduced by Juels and Sudan in 2002 to renew the previous bio-encryption technique called Fuzzy Commitment proposed by Juels and Wattenberg [5]. Fuzzy Vault Scheme came with the capability that wasn't owned by Fuzzy Commitment Scheme, that was the ability to handle fuzzy and unordered biometric data. Biometric data are categorized as fuzzy data because of their unclearness. It is obtained from the fact that no two biometric samples taken from an individual will have 100% similarity [6].

Fuzzy Vault Scheme is a key binding and key-release type bio-encryption [7]. The encryption process on the Fuzzy Vault Scheme works by performing a secret key binding into biometric data, while the decryption process works by releasing a key which is bound on biometric data. The essence of Fuzzy Vault's security is how high the difficulty level to perform polynomial reconstruction. Meanwhile, for leveraging its security, Fuzzy Vault Scheme uses chaff points.

Chaff points are analogize as noises used for covering valuable information inside of Fuzzy Vault. The more chaff points used, the better security level that is owned by the Fuzzy Vault scheme [8]. Thus, chaff points generation's phase is a quite important phase in Fuzzy Vault scheme. However, this phase became one of the high computation phases in the

Fuzzy Vault encryption [3]. Hence, several researches to find particular methods in generating chaff points which have a fairly low level computation but still preserve Fuzzy Vault scheme's security have been done by some researchers.

There are at least three methods for generating chaff points that were recently proposed by some researchers. They are Image Celling [8], Square Boundary [3], and Non-Random Chaff Point Generator [9]. This paper will review and conduct some comparisons toward these three methods. Turning to the main structure of this paper, this paper will be structured into several sections. Section 1 discusses a brief review of background underlying the emergence of Fuzzy Vault Scheme. Section 2 discusses the principal concept of Fuzzy Vault bio-encryption and its chaff point generation's method. Section 3 discusses the three generating's methods that were previously mentioned. Section 4 discusses the comparisons of these three methods which are presented in a comparative table, and the conclusions are presented in Section 5.

II. Fuzzy Vault Scheme

A. The Concept of Fuzzy Vault Scheme

As described previously, Fuzzy Vault scheme is a bio-encryption technique which has key binding and key release type. In the encryption phase, Fuzzy Vault binds secret key into biometric data, while in the decryption phase, Fuzzy Vault releases the secret key which is bound into biometric data. Fingerprint is considered as the most appropriate biometric data with Fuzzy Vault scheme because of its practicality and its accuracy [3].

Fig. 1. Mechanism of Fuzzy Vault Encryption [9]

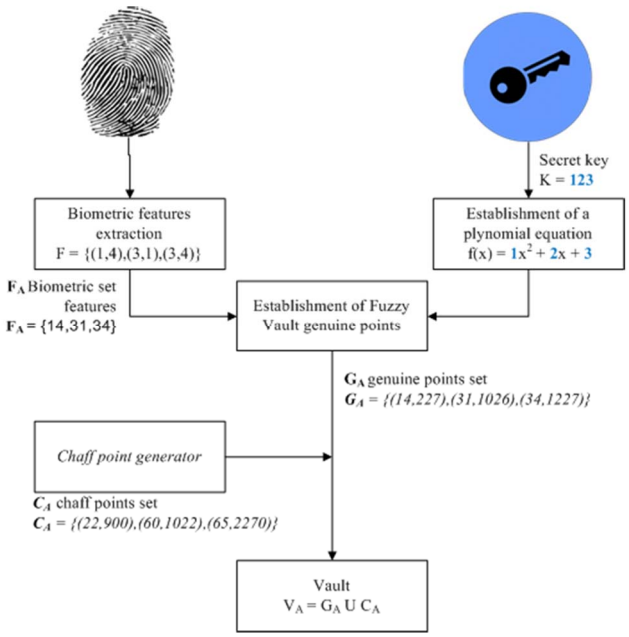


Fig. 1 shows the encryption mechanism which is carried out by the Fuzzy Vault Scheme. The encryption starts with the formation of an n degree polynomial equation, where n is the length of the secret key k minus 1. For example, if the length of the secret key k is 3 characters then the degree of polynomial p is 2 and each of the secret key's character becomes the coefficient of the polynomial equation p. The next step is conducting fingerprint features extraction. The results of this features extraction are minutiae's xy-coordinates. The minimum number of minutiae required in order to do successful encryption process is n+1. For example, if polynomial p has degree of 2 then it needs at least 3 minutiae to do successful encryption process. After xy-coordinate of all minutiae are obtained, every x-coordinate merged with its y-coordinate to form new values and save them into a set namely F_A. Each element of F_A will be projected as x value of polynomial p. Each of these polynomial operation's results will be paired with each of F_A set's elements to form the new coordinate pairs. Those coordinate pairs are then stored in a set namely G_A. If each element of G_A is projected into a Cartesian diagram then these elements will form a polynomial pattern with degree of n. To disguise that polynomial pattern, a set namely C_A containing random coordinates (chaff points) is generated and then combine it through G_A to form a vault set namely V_A. This set is what we called as the ciphertext in the Fuzzy Vault scheme.

Fig. 2. Mechanism of Fuzzy Vault Decryption [9]

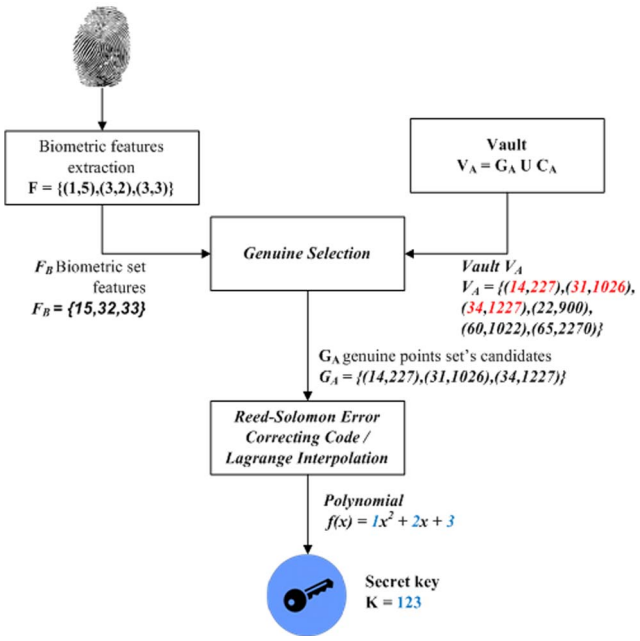


Fig. 2 shows decryption process of the Fuzzy Vault scheme. Decryption begins with the features extraction. Just like the encryption phase, the minimum number of minutiae needed in order to do successful decryption process is n+1, where n is the degree of polynomial p. After xy-coordinate of all minutiae are obtained, every x-coordinate merged with its y-coordinate to form new values and save them into a set namely F_B. Each element of the F_B is then projected on the V_A set to do genuine selection process. This process is a mechanism of selecting coordinate points of V_A set's member that has x-coordinate value equal or close enough to each element of F_B. Coordinate points that have been selected will

be stored in a set namely G_A. After all of the selected coordinates are stored in G_A, the next phase is a polynomial reconstruction. The purpose of this phase is to reconstruct or interpolate all of the elements contained in G_A to form a polynomial equation which has degree of n. There are two techniques that can be used to interpolate them, namely Reed-Solomon Error Correction Code and Lagrange Interpolation. After interpolation has been done successfully, it will produce a polynomial equation p' with degree of n. All of the coefficients belonged to polynomial equation p' are concatenated and arranged to form the secret key k. This secret key is what we called as the plaintext in Fuzzy Vault Scheme.

B. Earliest Concept of Chaff Point Generation

As previously described, chaff points are analogize as noisy points generated in such way to disguise the polynomial pattern existed in Fuzzy Vault. For the first time when the Fuzzy Vault was introduced to the public by Juels and Sudan [4], chaff point generation was done by pseudo random without regarding the efficiency aspect. In [4], there are only three conditions that must be fulfilled in order to generate chaff points:

1) X-coordinate of chaff point must be not equal to x-coordinate of each genuine point existed in the Vault.

2) Y-coordinate of chaff point must be not equal to the value of p(x), which means chaff points must not be located on the polynomial's path.

3) At least a chaff point is needed to disguise genuine points existed in the Vault.

Chaff point generation's technique began to evolve since the concept of Euclidean Distance was applied to the chaff point generation's process by Clancy et al. [10]. Euclidean Distance is used for measuring the optimum distance between coordinate of a chaff point's candidate and each minutiae's coordinates used by Vault. The basic idea of the technique proposed by Clancy et al. [10] is putting chaff point at the distance that is not too close to each genuine point (include minutiae and previous valid chaff points) existed in Vault. If the Euclidean Distance between a chaff point's candidate and at least one of the genuine points is below the predetermined threshold then this candidate will be ignored. In contrast, if the Euclidean Distance between a chaff point's candidate and all of the genuine points is equal to or above the threshold then this candidate will be considered as a valid chaff point. Furthermore, Clancy et al. [10] stated that in order to make a good disguising on the Fuzzy Vault scheme, the ratio between minutiae points and chaff points is 1:10. Thus, chaff point generation's method proposed by Clancy et al. [10] was claimed as the effective and efficient method for that time.

III. New Methods of Chaff Point Generation

A. Image Celling Method

This method was introduced for the first time in 2013 by Nguyen et al. [8]. This method was based on the weakness of the method proposed by Clancy et al. [10]. In Clancy's method, the computational time required in order to generate chaff points grow exponentially along with increasing number of chaff points needed. Moreover, in the case of a Fuzzy Vault using 20 minutiae and 200 chaff points to do an encryption mechanism, to generate the first chaff point, it takes at least 20 times the Euclidean Distance calculation. To generate the second chaff point, it takes at least (20+1) times the Euclidean Distance calculation. This step will be done continuously until the 200th candidate is confirmed as a valid chaff point.

Fig. 3. Illustration of Image Celling Method [8]

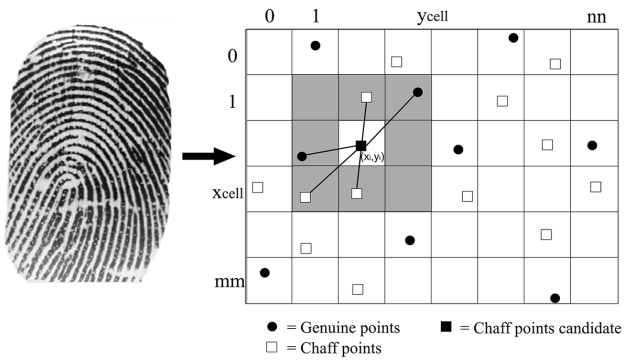


Fig. 3 shows how the chaff point generation method using Image Celling technique works. This technique begins with the fingerprint's features extraction to get minutiae's coordinates. These coordinates are then transformed into a particular coordinate system that locates the origin point at the upper left corner of its system. Unlike the Cartesian coordinate system, this system makes its vertical axis as the x-axis and its horizontal axis as the y-axis. After all of minutiae's coordinates are successfully transformed into the new coordinate system, this coordinate system is then divided into several cells such that each cell is only occupied by a minutiae and each cell will have a maximum of eight neighboring cells. The next step is choosing a cell randomly. If the selected cell is already occupied then the system will make the selection randomly once again. But, if the selected cell is unoccupied then the chaff point's candidate is put on this cell. Once the chaff point's candidate was put on an empty cell then the next step is measuring the Euclidean Distance between this candidate and its adjacent points. Unlike Clancy's method [10], Image Celling [8] just do a maximum of eight times Euclidean Distance calculation each iteration of chaff point generation. That is because each cell has only a maximum of eight adjacent cells and each cell contains only one point. If the chaff point's candidate doesn't have any neighboring points or if the Euclidean Distances made by this candidate and its neighboring points are higher than specified threshold then this candidate is considered as a valid chaff point and its coordinate will be recorded by the system. But, if there is at least one Euclidean Distance made by chaff point's candidate and its adjacent points below the specified threshold then generating process will be repeated once again from cell selection step. Those steps will be done continuously until the system has sufficient valid chaff point.

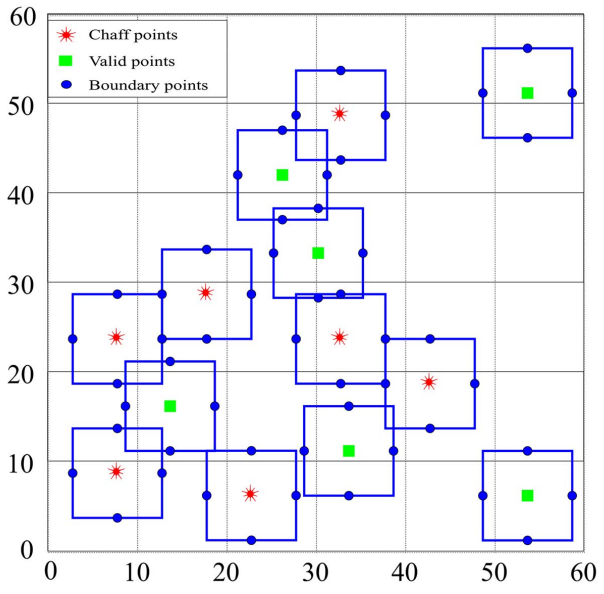
Performance analysis conducted by Nguyen et al. [8] shows that computational time's growth of the Image Celling method tends to be linear along with the increasing number of

chaff points needed. Thus, Image Celling [8] is more efficient than the Clancy's method [10] in the domain of computational time. However, Image Celling [8] still has a shortcoming in the use of its coordinate system. The coordinate system used by Image Celling method must have a certain ratio so that this coordinate system can be divided into square cells which have the same size. In addition, its origin point must be put in the upper left corner of this coordinate system and its vertical axis is turned into x-axis also its horizontal axis is turned into y-axis. Thus, particular coordinate transformation's techniques are needed, so that the minutiae's coordinates can be transformed into Image Celling's coordinate system. Of course, it is less practical, because when this method is applied to the device which has a different size of fingerprint scanner, it requires coordinate recalibration.

B. *Square Boundary Method*

This method was introduced for the first time in 2013 by Khalil-Hani et al. [3]. This method is inspired by the succession of the attack made by Chang et al. [11] toward chaff points generation proposed by Clancy et al. [10]. This attack was carried out by analyzing the Degree of Freedom (DoF) of each point contained in the Vault. Based on the analysis conducted by Chang et al. [11], they found that chaff points generated at a later stage have lower DoF than chaff points generated at an earlier stage. With this fact, the attacker will eliminate the points which have a fairly small DoF. Next, the attacker will perform a Brute Force [12] toward remaining points in order to get the secret key hidden in the Vault. Hence, the attacker will gain a higher probability to get the secret key than just using naïve Brute Force.

Fig. 4. Illustration of Square Boundary Method [3]



Not only criticized Clancy's method [10] vulnerabilities, but Khalil-Hani et al. [3] also criticized its complexity level. Khalil-Hani et al. [3] found that the complexity level of Clancy's method [10] is $O(n^3)$ and they have opinion that its complexity should be reduced, so that the chaff points generation can be run efficiently on the System-on-Chip based devices. To do so, Khalil-Hani et al. [3] proposed a method inspired by Circle Packing technique [13] and eliminate the mechanism of Euclidean Distance calculation.

Fig. 4 shows how the Square Boundary method works. Additional attributes such as boundary points are added to each point contained in the Vault, either genuine point or chaff point. When they are illustrated in a graph, those points will form a square boundary that surrounds the Vault's point as shown in Fig. 4. Principally, to generate a chaff point, Square Boundary method will select a coordinate randomly as a candidate of chaff point. Once the coordinate of the candidate has been selected, then the system will check whether this candidate inside the boundaries of the other points or not. If the candidate is outside the boundaries of all points existed in the Vault then this candidate will be considered as a valid chaff point and stored in the Vault. But, if the candidate is inside the boundary of at least a point existed in the Vault then this candidate will be ignored and the generation process will be repeated once again from the candidate selection's phase. Those steps will be done continuously until the system has sufficient valid chaff point.

Performance analysis conducted by Khalil-Hani et al.[3] shows that computational time's growth of the Square Boundary method tends to be linear along with the increasing number of chaff points needed. Moreover, this method has lower complexity ($O(n^2)$) than the method proposed by Clancy et al. [10]. This method is also said to be able to survive against Statistical Analysis attack because of its ability to distribute DoF randomly for each point contained in the Vault. However, this method is still using the similar verification method used by Clancy et al. [10]. For example, in the case of Fuzzy Vault utilizing 20 minutiae and 200 chaff points, to generate the first chaff point, it takes at least 20 times of the Euclidean Distance calculation. To generate the second chaff point, it takes at least (20+1) times of the Euclidean Distance calculation. This step will be done continuously until the 200th candidate is established as a valid chaff point. Thus, there are at least 200 x 20 + (1 + 2 + … + 199) examinations of Euclidean Distance, assuming that each iteration exactly produces a valid chaff point. It means that this method could potentially increase its computational time significantly, if the chaff points generated is more than 500 points.

C. *Non-Random Chaff Points Generator Method*

Non-Random Chaff Points Generator (NRCPG) was introduced for the first time in 2016 by Nguyen et al. [9]. The main idea of this method is generating chaff points with a structured pattern without involving random numbers' generation. Moreover, this method involves SHA hashing process to improve its security.

This method is inspired by the vulnerability of the Fuzzy Vault against Substitution Blend attack [14]. This attack works by modifying the coordinate information of points existed in the vault without changing the number of vault's members. Thus, this attack may increase False Rejection Rate (FRR) of the authentication process performed by the real user.

Fig. 5. Illustration of Non-Random Chaff Points Generator Method [9]

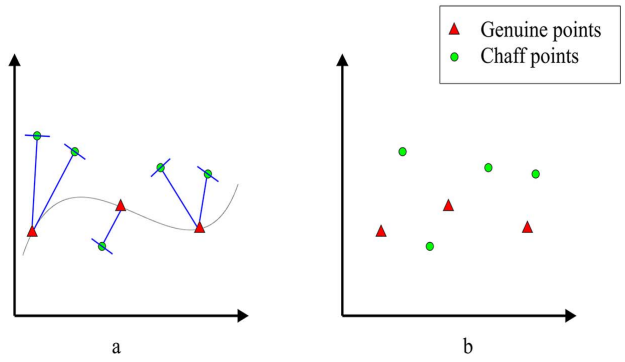


Fig. 5 shows an illustration of chaff points generation using NRCPG method. From this illustration, it appears that chaff points are formed by linear patterns created by the combination of genuine point's information and secret key's information. There are two significant differences between this method and the other generation methods [3],[8],[10]. Here are the differences:

1) NRCPG uses a pattern created by the hashed combination of genuine point's information and secret key's information in order to make a set of valid chaff points.

2) NRCPG does chaff point generation's mechanism not only at the enrollment phase (encryption), but also at the authentication phase (decryption).

Nguyen et al. [9] claimed that NRCPG is able to handle Blend Substitution attack on the Fuzzy Vault scheme. This method verifies all of chaff points existed in the vault when it enters the authentication phase. If there are significant differences between chaff points contained in the Vault and chaff points generated during authentication phase then this Vault is indicated to be suffered by Blend Substitution attack, assuming the biometric features used in the enrollment and the authentication phase are the same biometric features. However, NRCPG actually has a risk to leverage the likelihood of being suffered by Collusion and Key Inversion attack [9],[15] which could lead to leakage both of secret key and biometrics information contained in the Vault.

IV. Methods Comparison

In the previous section, concepts, objectives, advantages and disadvantages as well as illustrations of the three generation methods have been described separately. In this section, the comparisons of these three methods will be briefly summarized and presented in a comparative table. Thus, the differences among these methods can be obviously explained. Here is the comparative table of these methods:

TABLE I. COMPARATIVE TABLE

Methods	Domains/Objectives	Pros	Cons
Image Celling	• Computational time's efficiency	• Computational time's growth tends to be linear • This method needs only maximum 8 times of chaff point's validity checking for each iteration	• This method needs a particular coordinate system
Square Boundary	• Computational time's efficiency • Lowering complexity level • Enhance Fuzzy Vault's security	• Computational time's growth tends to be linear • Its complexity level is $O(n^2)$ • This method is able to survive against Statistical Analysis attack	• The number of chaff point's validity checking is continuously increase along with chaff points needed in the Fuzzy Vault scheme
Non-Random Chaff Points Generator	• Enhance Fuzzy Vault's security	• This method is able to survive against Blend Substitution attack	• This method has a risk to leverage the likelihood of being suffered by Collusion and Key Inversion attack

V. Conclusions

This paper discussed three methods of generating chaff points that have been proposed from 2013 until 2016. There are at least three objectives that become the focuses of these methods, namely high efficiency of computational time, low complexity level and high level of security. It is quite difficult to align all of these objectives in order to remain balanced. There is only Square Boundary method which is able to encompass all of these objectives, although not all of Fuzzy Vault's attacking methods can be handled by this method. In other words, there is no method that became the best solution to encompass all of Fuzzy Vault's problems. For cases that require the Fuzzy Vault to generate a lot of chaff points (e.g. more than 500 points) within stable computation time, Image Celling is better than others, but this method is less practical in the use of its coordinate system. For cases that require the Fuzzy Vault to be implemented on SoC devices which have limited memory resources, Square Boundary is better than others because of its low complexity level, but this method would have difficulty if the Fuzzy Vault is required to generate a lot of chaff points (e.g. more than 500 points). For cases that require the Fuzzy Vault to detect whether any illegal modification on the vault or not, NRCPG just the only one

among of those three methods which can do it, but this method needs higher memory resources because it includes hashing calculation into its generation procedure. Hence, if we need the chaff point generation technique that has stable computation time and low complexity level then we can combine Image Celling and Square Boundary technique to get a new technique of chaff point generation that has those requirements. Moreover, to improve its security, we can add concept of double phases chaff generation which is owned by NRCPG into this new technique, hence this new technique can validate whether vault template is illegally modified or not. However, we still have to consider about memory consumption to make sure that our new technique can be implemented on devices that have limited resources.

In the context of the attacks that threaten the Fuzzy Vault, there are at least two types of Fuzzy Vault's attacking methods that couldn't be handled by Image Celling, Square Boundary and NRCPG method, they are Key Inversion attack and Collusion attack. Hence, there are two tasks remaining that must be completed by research in the chaff point generation's field. First, find the generating method that truly able to align the three objectives previously mentioned. Second, find the generating method which is able to handle not only conventional attack but also Key Inversion and Collusion attack.

REFERENCES

[1] M. Pastore and E. Dulaney, *Security +*. San Fransisco: SYBEX, 2003.

[2] G. Coulouris, J. Dollimore, and T. Kindberg, *Distributed Systems: Concepts and Design*, vol. 4. 2012.

[3] M. Khalil-hani, M. N. Marsono, and R. Bakhteri, "Biometric encryption based on a fuzzy vault scheme with a fast chaff generation algorithm," *Futur. Gener. Comput. Syst.*, vol. 29, no. 3, pp. 800–810, 2013.

[4] A. Juels and M. Sudan, "A Fuzzy Vault Scheme," in *IEEE International Symposium on Information Theory*, 2002, p. 408.

[5] A. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme," in *CCS '99 Proceedings of the 6th ACM conference on Computer and communications security*, 1999, pp. 28 – 36.

[6] V. Matyas and Z. Riha, "Security of Biometric Authentication Systems– Extended Version," in *International Conference on Computer Information Systems and Industrial Management Applications, CISIM 2010 (2010)*, 2010, pp. 19–28.

[7] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: Issues and challenges," *Proc. IEEE*, vol. 92, no. 6, pp. 948–959, 2004.

[8] T. H. Nguyen, Y. Wang, Y. Ha, and R. Li, "Improved Chaff Point Generation for Vault Scheme in Bio-Cryptosystems," *IET Biometrics*, vol. 2, no. 2, pp. 48–55, 2013.

[9] M. T. Nguyen, Q. H. Truong, and T. K. Dang, "Enhance fuzzy vault security using nonrandom chaff point generator," *Inf. Process. Lett.*, vol. 116, no. 1, pp. 53–64, 2016.

[10] T. C. Clancy, N. Kiyavash, and D. J. Lin, "Secure Smartcardbased Fingerprint Authentication," in *Proceedings of the 2003 ACM SIGMM Workshop on Biometrics Methods and Applications*, 2003, pp. 45–52.

[11] E.-C. Chang, R. Shen, and F. W. Teo, "Finding the Original Point Set Hidden Among Chaff," in *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security*, 2006, pp. 182–188.

[12] P. Mihailescu, "The Fuzzy Vault for fingerprints is Vulnerable to Brute Force Attack," *CoRR*, vol. abs/0708.2, 2007.

[13] C. R. Collins and K. Stephenson, "A circle packing algorithm," *Comput. Geom.*, vol. 25, no. 3, pp. 233–256, 2003.

[14] W. J. Scheirer and T. E. Boult, "Cracking Fuzzy Vaults and Biometric Encryption," in *Biometrics Symposium, 2007*, 2007, pp. 1–6.

[15] H. T. Poona and A. Miria, "A Collusion Attack on the Fuzzy Vault Scheme," *ISC Int'l J. Inf. Secur. Bd*, vol. 1, no. 1, 2009.