Fuzzy Vault Security Enhancement avoid Statistical Biases

Sara Majbour, Morgan Barbier, Jean-Marie Le Bars

Fuzzy Vault Security Enhancement avoid Statistical Biases

Sara Majbour Morgan Barbier
Jean-Marie Le Bars

Normandie Univ, UNICAEN, ENSICAEN, CNRS,
GREYC, 14000 Caen, France

Abstract

We assess the fuzzy vault's security against the exploitation of statistical biases, conducting bias examination through features on a sample of biometric set. Our comparative analysis quantifies the scheme's vulnerability to security-compromising attacks, using three bases of feature templates derived from real biometric databases of various modalities, showcasing variable quality levels, and quantifying scheme weaknesses. This study shows a decrease in the scheme's security under such attacks and significantly contributes to understanding the fuzzy vault's limitations regarding biases in the stored set. Moreover, we propose the first solution without requiring additional information, preserving the security of the fuzzy vault against such attacks.

1 INTRODUCTION

Biometric authentication systems enhance security compared to traditional methods like passwords or keys. They refer to individual traits like fingerprints, facial features, retinal scans, and iris patterns for identification (Dargan and Kumar, 2020; Daugman, 2004). To authenticate, biometric data is captured, converted into a digital template, and compared to the enrolled template to confirm the individual's identity (Sharma et al., 2015). Since biometric templates are sensitive, various cryptographic methods are implemented to secure them (Uludag et al., 2004). A particularly effective solution in addressing the variability of biometric data involves adopting the cryptographic scheme of the fuzzy vault, developed by (Juels and Sudan, 2002). This scheme, incorporating error correction codes and an unordered set, enables error-tolerant authentication while preserving the confidentiality of the data. The standard fuzzy vault process hides a set by connecting it to a nonce with Reed-Solomon error correction codes. Authentication succeeds when a presented set closely matches the reference.

Despite significant progress in studying fuzzy vault schemes recently, each proposed approach is tailored to the context examined and the biometrics modalities used (Uludag et al., 2005; Nandakumar et al., 2007a; Rathgeb et al., 2023). Numerous studies have adapted the scheme to biometrics, focusing primarily on security (Benhammadi and Bey, 2014; Radha et al., 2010). The original article established an upper security bound based on the assumption of uniform distribution. However, deviations

from this assumption in real data distributions have been noted, which could lower the security level.

It has been recognized that biases exist, but the correlation between their nature, magnitude, and impact on fuzzy vault security is yet to be defined. Our study quantifies the impact of the lack of a uniform distribution on the fuzzy vault security deterioration, highlighting its overall significance by generally rendering the scheme unusable. Additionally, we underscore that the attacker model proposed by Juels and Sudan, possessing partial knowledge of the information, lacks realism and relevance. In the literature, some studies have proposed incorporating a password as a solution to mitigate bias issues and enhance the fuzzy vault's security against statistical attacks (Benhammadi and Bey, 2014; Radha et al., 2010). However, introducing passwords fundamentally establishes strong multi-factor authentication, differing from the original fuzzy vault scheme, and other vulnerabilities may arise with a multi-factor scheme. No secure single-factor fuzzy vault solution has been proposed yet. Our investigation specifically focuses on examining the single-factor fuzzy vault as initially proposed, to lead to a new secure single-factor system.

Our approach manages feature templates generic from various biometric modalities. We use three bases of biometric templates-fingerprint, face, and electrocardiogram, represented as feature vectors. A distinctive attribute of the fuzzy vault is its use of an unordered set of elements. In our work, the term biometric set refers to the outcome of transforming a biometric template of features, into a set of elements within a finite field, as illustrated Figure 1.

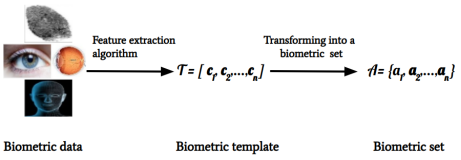


Figure 1: Biometric set construction for fuzzy vault

Our analysis focuses specifically on applying the fuzzy vault scheme as originally proposed, without incorporating additional information altering the authentication type. For the first time, we introduce the quantification of the significant advantage of an attack exploiting biases by features within a sample biometric set. Our results from different template bases show that the scheme as presented in the original article is ineffective and lacks relevance against such attacks. Taking into consideration these biases, we propose the first solution for applying the fuzzy vault scheme without passwords, using quantile methods during the transformation into a biometric set to ensure an equitable distribution of elements for each feature. The aim is to ensure the security of the single-factor fuzzy vault authentication scheme and neutralize attacks exploiting biases by features.

The remainder of this article follows this structure: Section 2 recalls the fuzzy vault concept and existing work on its application in biometrics. Then, Section 3, we explain our generic methodology for any biometric template base in a feature vector format. Within Section 4 we outline the conditions of the upcoming attack and the biometric templates used. Then Section 5, we present the obtained results. This understanding of biases allows us to propose Section 6 the first solution to secure the fuzzy vault without any additional information.

2 BACKGROUND

In this section, we aim to provide a more detailed explanation of the fundamental encoding and decoding process of the fuzzy vault scheme. Following that, we present an extensive review of prior research on applying this cryptographic scheme in a biometric domain.

2.1 Fuzzy Vault

The fuzzy vault concept is a versatile cryptographic approach that applies to various domains, like privacy-protected matching, personal entropy systems, and biometrics. Its strength lies in effectively handling differences between sets and having the capability to correct them, which is a standout feature of this approach. Its primary objective is to authenticate an individual based on a comparison between an authentication set $\mathcal{B}$, and another enrollment set $\mathcal{A}$, concealed within a vault using a nonce $\mathcal{K}$. Linear error correction codes, particularly Reed-Solomon codes (Juels and Sudan, 2002), are crucial to ensure reliable and efficient recovery of the nonce in this cryptographic system, allowing to handle variability in the enrollment and authentication sets effectively.

2.1.1 Enrollment Stage

The first stage, referred to as enrollment, involves recording user information into the system. This corresponds to a reference set $\mathcal{V}$ within the fuzzy vault, known as the vault. To construct it, a Reed-Solomon error correction code is employed, incorporating a biometric set $\mathcal{A}$ with n elements from a finite field $\mathbb{F}_q$, along with a random nonce vector $\mathcal{K}$ of length k. where each component corresponds to a coefficient of the polynomial $P \in \mathbb{F}_q[X]$. The choice of n exceeding k introduces redundancy to permit error detection and correction during decoding. Concurrently, to ensure confidentiality, additional chaff points are uniformly integrated, forming the vault $\mathcal{V}$ as illustrated Figure 2b. This stage consists of two steps: encoding and adding noise.

a) Encoding: the biometric set is encoded with the nonce $\mathcal{K}$ by associating the elements of $\mathcal{A}$ with their polynomial evaluations. Each element x in $\mathcal{A}$ is evaluated by applying the polynomial P, resulting in a value $P(x) = y$. These pairs (x, y) represent points in the fuzzy vault (Refer to Figure 2a).

b) Noise: at this step, random pairs $(x', y') \in \mathbb{F}_q^2$, known as chaff points, are introduced into the set $\mathcal{V}$. They are specifically chosen to not correspond to genuine points of $\mathcal{A}$ and do not follow the pattern of polynomial evaluations: $\forall x' \notin \mathcal{A},\ y' \neq P(x')$ (see Figure 2b). Adding these chaff points aims to make it challenging for attackers to discern genuine points from chaff points within set $\mathcal{V}$.

2.1.2 Authentication Stage

The subsequent stage, denoted as authentication, serves to verify the user's identity and grant access to appropriate resources. Authentication evaluation depends on the similarity between enrollment set $\mathcal{A}$ and authentication set $\mathcal{B}$, both of size n, irrespective of the order of their elements.

(a) Encoding

(b) Noise

Figure 2: Enrollment stage of the fuzzy vault scheme.

The retrieval of the random nonce $\mathcal{K}$ used during enrollment is contingent upon the similarity of these sets. This condition is ensured by a decoding algorithm associated with the codes used during the enrollment process, having the ability to correct up to e errors, referred to as the decoding radius. Therefore, to accurately recover the nonce $\mathcal{K}$, we check if the number of discrepancies between the two sets $\mathcal{A}$ and $\mathcal{B}$ is bounded by e.

The authentication stage can be divided into three steps: extraction, decoding, and verification.

a) Extraction: from the reference set $\mathcal{V}$, we extract pairs (x,y) where x belongs to the authentication set $\mathcal{B}$; thus, the set Q of size less than or equal to n is composed of the extracted pairs (see Figure 3a).

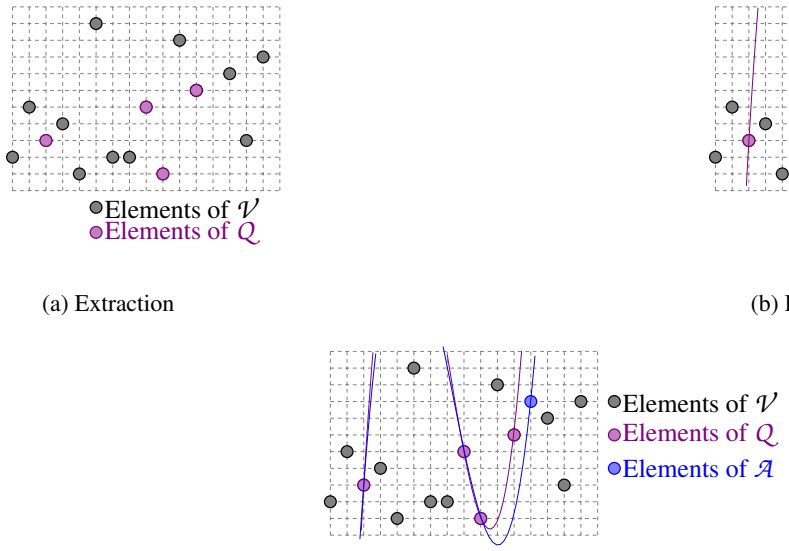
b) Decoding: in this phase, a Reed-Solomon decoding algorithm is employed, receiving the set Q and the length k as inputs. The algorithm produces a secret $\mathcal{K}'$ if there's adequate matching, facilitating potential error correction (see Figure 3b). Alternatively, it may return a null value if no polynomial aligns with the decoding of the received set Q.

c) Validation: authentication is successful, enabling user authentication, only if the candidate nonce $\mathcal{K}'$ is identical to the nonce $\mathcal{K}$ used during enrollment. This condition corresponds to the sufficient sharing of elements between two sets $\mathcal{A}$ and $\mathcal{B}$ (see Figure 3c).

2.2 Related Work

Original fuzzy vault scheme proposed by Juels and Sudan, is generic and can be applied to different purposes. When employed in a biometric context, it requires specific adjustments, primarily driven by considerations related to security, the choice of modality, the error correction codes parameters choice, and the concept of a set which is a fundamental characteristic of the fuzzy vault. This is why numerous different works have been done, to use it in a practical world. Highlighting the inability to directly implement this scheme.

The initial difficulty in implementing the fuzzy vault lies in transforming a biometric template into a biometric set, a process influenced by the biometric modality and extraction algorithm. Notably, the minutiae template, is extensively explored in the fuzzy vault implementations (Uludag et al., 2005; Nandakumar et al., 2007a; Poon



(a) Extraction

(b) Decoding

(c) Verification

Figure 3: Authentication stage of fuzzy vault scheme

and Miri, 2012; You and Wang, 2018). Each minutia is characterized by three values (x, y, θ). The set is achieved through the concatenation of bits derived from x, y, or x, y and θ. Minutiae template is a rare case where the concept of a set is inherently present. However, alignment issues complicate this process (Merkle et al., 2010; Nagar et al., 2008). Many studies have investigated the application of the fuzzy vault with biometric templates represented as feature vectors. The type of template we are interested in our study for different modalities. For example, in study (Lee et al., 2008), a secure fuzzy vault system employs local iris features and clustering techniques for accuracy, ensuring proximity between sets $\mathcal{A}$ and $\mathcal{B}$ in both biometric data and error-correcting codes. Another instance is seen in (Rathgeb et al., 2016), which introduces a multi-instance iris biometric cryptosystem, requiring a partitioning method for subsets based on each feature set.

The secondary difficulty pertains to the parameter selection of an error correction code and its contextual relevance. Various methodologies are employed in reconstructing the secret $\mathcal{K}$ from Reed-Solomon codes. Some researchers, faced with issues, opt for using a secret structured by a code such as Cyclic Redundancy Check (CRC), involving polynomial reconstruction through exhaustive search (Nandakumar et al., 2007a; You and Wang, 2018; V.S.Meenakshi and Padmavathi, 2010), aiming to avoid False Rejects. However, considerations regarding execution time are discussed (Khalil-Hani et al., 2013). Different studies have highlighted concerns about the efficiency and this approach's security, suggesting the use of decoding algorithms for Reed-Solomon codes (Velciu et al., 2015).

The scheme's security is a major concern, evaluated in various contexts and against different attack scenarios (Nandakumar et al., 2007b; Benhammadi and Bey, 2014; Radha et al., 2010). Studies focus on improving biometric data integrity, especially by assessing the False Acceptance Rate (FAR). Another approach involves evaluat-

ing min-entropy in protected models, offering insights into leaked information and the maximum probability of uncovering the secret (Merkle et al., 2010; Dodis et al., 2008). In their theoretical framework, Juels and Sudan do not explicitly consider biases in security analysis. They rely on a uniform distribution, establishing a security upper bound, and acknowledging the lack of justification for assuming uniformity. Other research supports this assertion by noting that non-uniformity in stored data can jeopardize the scheme's security (Nandakumar et al., 2007a; Merkle et al., 2010; Nagar et al., 2008). To overcome this limitation, some studies propose adding a password (Reddy and Babu, 2008; Nandakumar et al., 2007b; Benhammadi and Bey, 2014), known as a Hard fuzzy vault. The user is required to capture their biometric data and enter a password. Currently, no proposal for a secure fuzzy vault against such statistical attacks without additional information is known.

3 BIOMETRIC SET AND BIAS ASSESSMENT

This section details vault construction and its bias resistance, particularly against database attacks. We discuss the set creation function designed for feature models, explore theoretical measures for bias assessment in biometric sets, and review attacker models to evaluate fuzzy vault security.

3.1 Construction Biometric Set

The fuzzy vault scheme is characterized by its use of unordered sets. Our methodology employs a transformation approach tailored for feature templates from diverse biometric modalities, conceptually similar to the method in (Rathgeb et al., 2016) which encodes already features as unordered sets. However, our approach differs by taking into account individual variability and specific parameters, thereby enhancing authentication reliability and ensuring accuracy and compatibility with various biometric systems.

Feature values in biometric templates vary across captures from the same user. To ensure reliable identification and appropriate access, we use min-max normalization to address this variability. This technique, widely acknowledged for enhancing system reliability through precise and consistent data representation, is crucial for maintaining consistency in each feature's distribution relative to the initial data and effectively managing data variability (Zheng and Casari, 2018). Employing this approach also facilitates the security assessment of the fuzzy vault system by accounting for statistical biases in feature distributions within biometric samples.

This methodology enables the grouping of multiple feature values within a similar range. The normalization process converts the real values of each feature i into values ranging from 0 to 1 ($\star$), using the minimum min_i and maximum max_i values associated with each feature. Subsequently, the value is encoded using m_1 bits, representing the exponent of a prime number in our system. The selection of m_1 is based on the template base, aiming to choose the value that best distinguishes users. This selection results from experiments to determine the optimal number of bits to use, which may vary from one base to another.

Given a template containing n features, the normalized value f_{n_i} of each features f_i, encoded in m_1 bits, is obtained by applying the following formula:

$$f_{n_i} = \lfloor \frac{(f_i - min_i)}{max_i - min_i} \times 2^{m_1} \rfloor. \qquad (\star)$$

To construct a biometric set, we introduce the function S that maps elements encoded with m_1 bits into a biometric set within a finite field $\mathbb{F}_{2^p}$, where each element corresponds to specific features. Let S represent function ($\star\star$), which takes as input the m_2 bits of the index, encoded with the binary length of the biometric template size, and the m_1 bits of the feature value obtained from ($\star$). Through the concatenation of these input elements, this function produces an element within the finite field $\mathbb{F}_{2^p}$, where $p = m_2 + m_1$.

$$\begin{array}{rcccl} S & : & \{0,1\}^{m_2} \times \{0,1\}^{m_1} & \rightarrow & \mathbb{F}_{2^p} \\ & & (i, f_{n_i}) & \mapsto & e = i|f_{n_i} \ . \end{array} \qquad (\star\star)$$

3.2 Bias Quantification

Our analysis specifically focuses on statistical biases in biometric sets by features, aiming to understand how they affect the fuzzy vault system's security and quantitatively evaluate their effects. Using a training sample $TrainingS$ of the biometric template, we generate the biometric set from this sample using the specific transformation function ($\star\star$). We propose quantifying these biases in two distinct scenarios.

A. Scenario 1: in this context, we disregard features and calculate the frequencies of elements in the finite field $\mathbb{F}_{2^p}$ within biometric sets of $TrainingS$. This calculation is established by scaling the repetitions of each element $e \in \mathbb{F}_{2^p}$ (denoted as $rep(e)$) by the product of sample size $|TrainingS|$ and 2^{m_2}, representing the total number of elements (Rice, 2006):

$$\text{freq}(e) = \frac{\text{rep}(e)}{|TrainingS| \times 2^{m_2}}.$$

B. Scenario 2: in this context, the significance of the order of features is considered. We calculate the frequencies of elements f_{n_i} for each feature, from biometric sets of $TrainingS$. We compute the frequencies of the 2^{m_1} elements for each feature i, using the following formula:

$$\text{freq}(f_{n_i}) = \frac{\text{rep}(f_{n_i})}{|TrainingS|}.$$

To assess the overall distribution of these elements in each scenario, we employ Shannon entropy, a metric designed to measure the distribution of occurrences and its proximity to a uniform distribution.

Shannon entropy (Cover and Thomas, 2006), denoted $H_b(X,D)$, quantifies the uncertainty of a random variable X with outcomes in a finite field and their associated probabilities $D = (p_1, \ldots, p_n)$, where $p_i = \Pr(X = e_i)$. It is calculated using the base 2 logarithm to express information in bits and is defined as: $H_b(X,D) = -\sum_{i=1}^{n} p_i \log_2(p_i)$. Subsequently, for a uniform distribution D^U, where all outcomes are equally probable, the entropy simplifies to $H_b(X,D^U) = \log_2(n)$, serving as a benchmark to compare D against an ideal scenario of equal likelihood.

Finally, we propose assessing this outcome through the measure of statistical bias, denoted as $\mathcal{M}(D)$. This measure is defined as the ratio of Shannon entropy of the calculated distribution to the entropy of the uniform distribution, formulated as follows:

$$\mathcal{M}(D) = \frac{H_b(X,D)}{H_b(X,D^U)}.$$

The measure $\mathcal{M}$ generates a numerical value lower than 1. Any deviation from $\mathcal{M}$ compared to 1 indicates that the distribution is far from being uniform. This observation facilitates the quantification of the effect of statistical biases, thereby aiding in evaluating their influence on the security of the fuzzy vault.

3.3 Authentication Attacker Models

We evaluate the impact of statistical biases on security by introducing an attacker model that uses knowledge of biometric set distribution. This model contrasts with Juels and Sudan's, which is based on partial knowledge and is less clear. Additionally, we propose a model based on vault knowledge alone. To compare these models and understand how vault size and attacker knowledge influence security, we assess authentication efficacy across various vault sizes and attack scenarios.

A- Distribution-Knowledge Attacker (**DKA**): in this model, the attacker, having gained access to the vault, also possesses knowledge of the distribution of elements from the biometric set within the finite field. To generate n elements for an authentication set $\mathcal{B}$, the attacker draws from the vault according to this distribution. When the vault's size is smaller than the size of elements with probability in the sample, it becomes vital to concentrate solely on the present elements in the vault. To address this, we use a smoothing technique (Simonoff, 2012) to establish a new probability distribution. This method assigns a minimal, yet nonzero, probability to elements not found in r. This probability is contingent on the lowest probability within the sample, signifying that absent occurrences are improbable but not impossible. Using this updated distribution, the attacker randomly chooses elements from r to compose n elements within the authentication set.

B- Partial Knowledge Attacker (**PKA**): this model, proposed in (Juels and Sudan, 2002), formalizes the security framework by incorporating the concept of partial knowledge attributed to the adversary. Authors assume that the enrollment set is chosen according to a potentially non-uniform distribution, but they do not quantify security with biases, they assume a scenario where an attacker possesses knowledge of a part of the set. However, these are two distinct concepts. To illustrate, consider attempting to guess a password. Knowing the distribution of passwords provides a significant advantage, but it doesn't reveal specific letters in a given password from that distribution. Conversely, knowing specific letters offers an additional advantage. Thus, these two forms of knowledge appear complementary.

   To identify values of the partial set α that hold significance for the attacker and serve as a realistic prerequisite, we analyze this attack scenario to evaluate the vault's security across different knowledge levels. Our method entails being informed about different values of α elements from $\mathcal{A}$. To complete the remaining $n-\alpha$ elements, we use a uniform selection from the remaining elements of $\mathcal{V}$.

C- Uniform Attacker (**UA**): we intend to compare the advantage of the previous model with that observed when no prior information is available. In this context, the attacker possesses only knowledge about the vault itself. To generate the authentication set $\mathcal{B}$, the attacker uniformly selects n elements from the r elements contained in the vault. This implies that each element in the vault has an equal probability of being chosen to be part of the authentication set $\mathcal{B}$.

Studying and comparing the attacker's advantage in these three models provides valuable insights into assessing the impact on the security of the fuzzy vault. This analysis aids in identifying the most sensitive or vulnerable information within each model, facilitating adjustments in security measures and countermeasures accordingly.

4 FUZZY VAULT CONSTRUCTION

In this section, we introduce the three biometric template bases derived from different modalities, each exhibiting unique quality levels. Quality Quality assessment of each base relies on False Rejection Rates (FRR) and False Acceptance Rates (FAR). When these rates are equal, they represent the Equal Error Rate (EER). Minimizing these rates enhances the system's authentication performance. Next, we present the fuzzy vault parameters obtained for each base used.

4.1 Biometric Templates Bases

Our approach is dedicated to biometric template bases categorized by features, We use three bases derived from different modalities (see Table 1), each exhibiting varying levels of quality in terms of EER. These include the FVC fingerprint base, the PTB base of electrocardiograms used in (Gernot and Lacharme, 2022), and the higher-quality LFW base, used in (Dong et al., 2019). Each base contains **T** biometric templates for each of the **N** distinct individuals. An extraction algorithm applied to the image yields a feature vector of size n.

Table 1: Biometric templates bases

	FVC	**PTB**	**LFW**
Modality	Fingerprints	Electrocardiogram	Face
Image database	FVC2002 (Maio et al., 2002)	LFW (Bousseljot et al., 1995; Goldberger et al., 2000)	(Huang et al., 2008)
Extraction algorithm	Gabor filters (Belguechi et al., 2016)	ECG wave delineation (Martinez et al., 2004; Makowski et al., 2021)	deep network InsightFace (Dong et al., 2019)
N	100	158	158
T	8	7	10
n	512	990	512

4.2 Fuzzy Vault Parameters

This section provides a concise overview of the essential parameters for constructing the reference set $\mathcal{V}$, stored in the fuzzy vault. The specifications are tailored to the specificities of each biometric template base, as presented Table 2.

A. Biometric set size n: our set construction function establishes a mapping between each binary feature sequence of m_1 bits and an element of the finite field $\mathbb{F}_{2^p}$ associated with the biometric set. The set size corresponds to the number of features in the template specific to the biometric template base, denoted as n.

B. Secret length k: the secret length k is directly linked to e and influences the authentication algorithm based on Peterson-Berlekamp-Massey. This algorithm can correct up to $e = \lfloor \frac{n-k}{2} \rfloor$ errors. The minimal intersection I between the sets $\mathcal{A}$ and $\mathcal{B}$ is crucial, where $I = n - e = \lfloor \frac{n+k}{2} \rfloor$.

   To select k, we calculate the FAR and FRR rates for various threshold values I. The objective is to achieve a balance between security (minimized FAR) and user-friendly interaction (controlled FRR). The optimal threshold I is determined as the intersection between the discrete curves of FAR and FRR, corresponding to the (EER) with the biometric set. If achieving this equality proves impossible, a preferred threshold is selected to minimize FAR. Depending on each base, different thresholds are obtained. To maintain these levels of authentication as indicated Table 3, we select k with consideration to variable I, taking into account error corrections with the decoding algorithm, $k = 2I - n$.

C. Vault size r: the construction of the set V relies on the addition of chaff points, dependent on the function S and parameters (m_1, m_2) specific to each biometric template base, ensuring their indiscernibility from the biometric set's elements. The choice of variable m_1 is made to allow better differentiation between biometric sets of different users and to manage variability among different templates from the same user. This decision is also correlated with the FAR and FRR rates. Thus, following tests on our bases, the appropriate choice is $m_1 = 2$. As for m_2, it represents the smallest possible value satisfying the condition $n \leq 2^{m_2}$, with n being the number of features in the vector. Our construction of set V involves choosing sizes ranging from $n \times 2$ to $n \times 2^{m_1}$.

Table 2: Fuzzy vault parameters for each template base

	Fuzzy vault parameters			
Biometric template base	(m_1, m_2)	n	k	r
FVC	(2,9)	512	65	[1024,2048]
PTB	(2,10)	990	273	[1980,3960]
LFW	(2,9)	512	9	[1024,2048]

The objective of the fuzzy vault is to enable comparison for authentication. Our encoding method effectively protects data with minimal impact on authentication system performance, as evidenced by the obtained rates Table 3. The degradation in the authentication rate remains limited, with an observed impact of about 7%. For example, in the FVC base, an EER of 17% is obtained. However, in the PTB base, our method leads to an improvement in authentication and a reduction of approximately 3.8% in the EER. For the LFW base, no EER is obtained, and therefore, we choose a threshold of $k = 9$ for which FAR rates are minimal and do not impact security.

5 CONDUCTING AUTHENTICATION ATTACKS

Based on the framework from Section 4, this section evaluates the security and practicality of the fuzzy vault. We detail quantitative results on its resilience against statistical biases exploited through biometric set features. Multiple vaults are constructed for

Table 3: Authentication rates

	biometric template	**biometric set**		
Biometric template base	**EER**	**FAR**	**FRR**	**EER**
FVC	10%	17%	17%	17%
PTB	10.8%	7%	7%	7%
LFW	0.2%	9.5×10^{-4}%	4%	-

each base, followed by authentication tests against various attacker models previously introduced.

During each testing phase, a sample $TrainingS$, comprising 60% of individuals from the base, is used to evaluate biases in the biometric set, with one template per person. The remaining 40% sample, referred to as $TestS$, is reserved for authentication testing. This distribution is consistently maintained across all three template bases. The initial stage in the fuzzy vault process is enrollment. Considering sample $TestS$ with $|TestS| = N * 40\%$, we construct the associated vault by incorporating the previously presented parameters, generating vaults with sizes r ranging from $2n$ to $n \times 2^{m_1}$. During authentication stage, we create 50 authentication sets $\mathcal{B}$ for each vault. This authentication process is conducted using the three attacker models, depending on the specific scenario being investigated.

The validity of authentication is conditioned by the number of common elements between the authentication set $\mathcal{B}$ and enrolment set $\mathcal{A}$. We define two sets to be close if their intersection is at least equal to $\lfloor \frac{n+k}{2} \rfloor$, and then the decoding algorithm guarantees the correction of other errors. To calculate the authentication rate for each vault, we determine the ratio of valid sets $\mathcal{B}$ to the total number of constructed sets, which is $|TestS| \times 50$. The success rates computed for each attacker model will be compared to analyze the impact of the acquired knowledge, as visualized in the graphs below.

5.1 Results

To exploit biases, we focus on two attack scenarios. The first scenario relies on a global frequency analysis, where the assembly of the biometric set is performed without considering the order of elements. Conversely, the second scenario takes into account the specific distribution of each feature, thus adopting a more in-depth approach based on feature-based construction. Starting from each of the three template bases, we initially assess the biases inherent in the associated biometric sets and perform quantitative measurements $\mathcal{M}$. Subsequently, we conduct authentication by leveraging the quantified biases, as well as considering two other attacker models. The outcomes of the authentication rates are then presented in the form of curve graphs for each reference vault set of size r.

Every color in graph denotes the authentication rate associated with a specific attacker model, the red represents an attacker with knowledge of the distribution (**DKA**), while the black signifies a uniform attacker (**UA**). The green, purple, and blue shades represent attackers with partial knowledge, each having different levels of information (**PKA**α**%**).

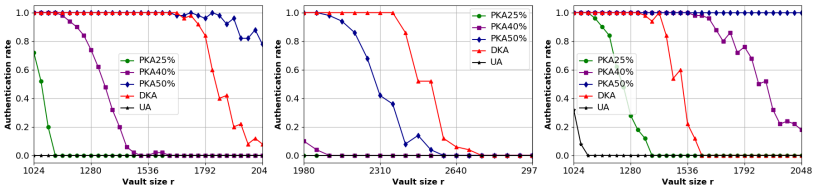
A. Scenario 1: in this context, the order of features is disregarded. From a sample of the biometric set, the frequency of each element $e \in \mathbb{F}_{2^p}$ is calculated, thereby determining the measure $\mathcal{M}$.

The specific results for the FVC, PTB, and LFW bases are 0.92 and 0.93, 0.94 respectively Table 4, indicating that they deviate significantly from a uniform distribution. Notably, there is a range of information concentration levels observed among the three bases, reflecting their divergence from uniformity. This observed diversity is associated with the quality of the base. A less pronounced dispersion is noted in the FVC database, which exhibits an EER rate of 17%, while the results for the LFW database, with a FAR of 4% and an FRR close to 0%, show greater dispersion.

Table 4: $\mathcal{M}(D)$ of biometric set.

template base	FVC	PTB	LFW
$\mathcal{M}(D)$	0.92	0.93	0.94

Overall, the measurements indicate pronounced diversity within the biometric set, with significant deviations from the uniform distribution. Specifically, pronounced disparities are observed in biases among biometric sets from different biometric template bases, particularly with the FVC showing a more marked deviation with a measurement of 0.92 compared to the uniform distribution, in contrast to the set from the base LFW with a measurement of 0.94. Through subsequent tests, we will seek to determine whether the slight differences observed among these measures will have a significant or negligible impact on the vault's security.



(a) FVC (b) PTB (c) LFW

Figure 4: Authentication for scenario 1

The results from Figure 4 indicate that, across all bases, the advantages of the three applied attacker models gradually decrease as the size of the vault r increases. With the FVC base, it is observed that knowledge of the set distribution, as depicted by the red curve, provides a significant advantage to the attacker compared to the other two bases where this advantage decreases for sizes starting from $3 * n$.

For the partial knowledge attacker model, we possess, in each case, knowledge of proportions of 25%, 40%, and 50% of the enrollment biometric set. The results obtained reveal that for all bases, knowledge of the distribution is always more important than knowledge of 25% of the information (represented in green), which still constitutes a significant bound. Even in the case of the FVC with multiple vault size r Figure 4a, the advantage obtained using the distribution is more significant than knowledge of 40%, and close to 50%. On the PTB base, we achieve the best rate with the red curve, surpassing the advantage gained with the knowledge of 50% of the enrollment set Figure 4b. However, it

is important to note that these knowledge percentages are unrealistic and do not represent a plausible attacker model, unlike knowledge of the distribution, which seems more practical.

The fifth black curve illustrates the advantage of a database-side attacker without any additional information, attempting to build elements of the set by drawing uniformly. It is observed that the vault's security remains intact, with the attacker frequently unable to gain any advantage. This supports the highest security threshold of the fuzzy vault under uniform conditions, consistent with assertions made by prior researchers. Comparing (**DKA**) and (**UA**) models, the red curve's notable advantage over the black indicates that exploiting set biases can weaken the fuzzy vault's security.

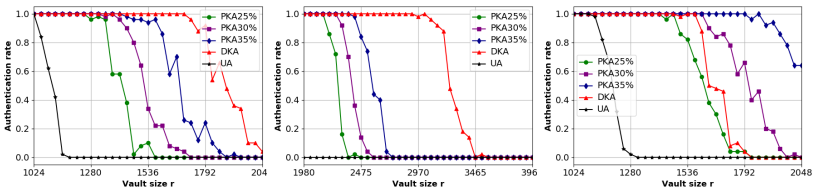
In connection with the base quality and measure $\mathcal{M}$, it is observed that the difference of 0.1 between the calculated measures Table 4 significantly impacts the vault's security against the attacker model considering the distribution. This is evident when comparing the degradation of the red curve across the three bases.

B. Scenario 2:

In this context, the relevance of knowledge through distribution becomes more pronounced, allowing for the provision of specific information regarding the distribution of each feature. We calculate the $\mathcal{M}_i$ metric for each feature i among the n in the template to assess the irregularity or dispersion of the feature values distribution.

The results from measuring feature biases reveal pronounced differences across the various bases. The majority of features for all three bases showed a $\mathcal{M}$ value ranging between 0.6 and 0.8, suggesting significant biases when compared to a uniform distribution. A few features, however, scored above 0.9, indicating less bias. Notably, in the FVC and PTB bases, some features recorded entropy measures below 0.5. A few feature groups even demonstrated distributions that were quite structured or condensed, with measures below 0.3. These findings will be used to assess their impact on the security of the vault.

In establishing the construction of the authentication set based on features, using each of the considered models. For the partial knowledge attacker model, we assume knowledge of 25%, 30%, and 35% proportions of the enrollment set.



(a) FVC (b) PTB (c) LFW

Figure 5: Authentication for scenario 2

The results presented Figure 5 highlight a significant deterioration in the vault's security through this feature-based approach. The three bases corroborate this observation. In general, authentication rates in this scenario are significantly

higher, attributable to additional specific information associated with each feature. With FVC Figure 5a, success rates approach 40% for most vault sizes and reach around 10% for sizes approaching the maximum value of r. In this scenario, compared to Figure 4b and 4c, we observe that more chaff points are needed to diminish the significance of the distribution advantage, as illustrated in Figure 5b and 5c.

Furthermore, with these bases of varying quality, a correlation is observed between the base quality and knowledge of the distribution: the higher the base quality, the less significant the impact of biases. As depicted Figure 5c, knowing the distribution is more advantageous than knowing 25% of the biometric set with the LFW base, in contrast to the PTB and FVC bases. In these cases, knowledge of the distribution proves to be more advantageous than knowing 35% of the set, with the distribution conferring a more substantial advantage over all other models.

5.2 Discussion

We have presented the results of our attack scenarios applied to three distinct template bases, characterized by different modalities and qualities. We assessed the resilience of the fuzzy vault against these three different attacker models.

Our initial observation reveals that the authentication rates obtained using the distribution model are substantial, affirming the scheme's inability to withstand such attacks, this casts doubt on its feasibility in such conditions and raises concerns about the fuzzy vault's security, especially in the case of feature-based construction, where we quantify biases related to features.

In comparison with the model suggested by Juels and Sudan, we illustrate that this model requires substantial knowledge to achieve a security level comparable to distribution knowledge, necessitating at least 30% of the enrollment set. This significant requirement is less apparent in the context of individual sensitive data.

In the scenario involving global data, the advantage is less pronounced than with feature-based construction. However, we observe a convergence of success rates with a non-realistic information knowledge model. This underscores the critical importance of the model used and its direct impact on security. It calls into question the scheme's ability to withstand this specific model, highlighting a potential vulnerability.

6 SECURE SINGLE-FACTOR FUZZY VAULT

The results of the previous section indicate that the original fuzzy vault cannot be considered secure, particularly when exploiting biases by features. This quantification, along with the suggested attacker model, helps us understand the dispersion of elements in the set. Differently from previous studies, we present the first solution while maintaining the original proposition of the fuzzy vault scheme, without modifying its inherent nature.

For our proposed solution, the goal is to avoid significant concentration within value ranges for each feature, thereby reducing biases in biometric sets by ensuring a balanced distribution for each feature. We rely on a quantile method, dividing the data for each feature into intervals of equal probability, where each integer appears with the same frequency.

6.1 Fuzzy Vault Parameters

Using the quantile method to mitigate biases in each feature decreases dominant feature values, consequently lowering the system's authentication level. Therefore, we reassess our selection of the value m_1 for the quantile method. Setting $m_1 = 1$ uses binary encoding for each feature based on its median, which minimizes data dispersion and aids in managing template value variability to achieve an EER with the biometric set. However, this parameter decreases the number of image elements of the function ($\star\star$), resulting in fewer chaff points for vault $\mathcal{V}$. Assuming the attacker knows all parameters, this compromises the vault's security. Despite reduced authentication from increased data dispersion, for security reasons, we choose $m_1 = 2$, providing four possible values corresponding to employing the quartile method. This divides the data into four equal parts, each representing 25%. Subsequently, an integer between 0 and $2^{m_1} - 1$ is associated with each interval, aiming to generate a sufficiently large finite field with function ($\star\star$). With this choice of m_1, we obtain the same values for n and r as those presented Table 2.

The correlation between the secret length k and authentication rates indicates that longer lengths are associated with reduced error correction. Employing the quartile method reveals variations in FAR and FRR rates. Regardless of specific k values, an increase in length is associated with higher FRR and lower FAR. During vault construction, emphasis is placed on selecting the smallest suitable value for k, tailored to the biometric template base. The following Table 5 presents the FAR and FRR results obtained for each base along with the corresponding k.

Table 5: Authentication rates

	biometric template	**biometric set**		
Biometric template base	**EER**	k	**FAR**	**FRR**
FVC	10%	5	1.3%	56%
PTB	10.8%	9	2%	6%
LFW	0.2%	4	0%	75%

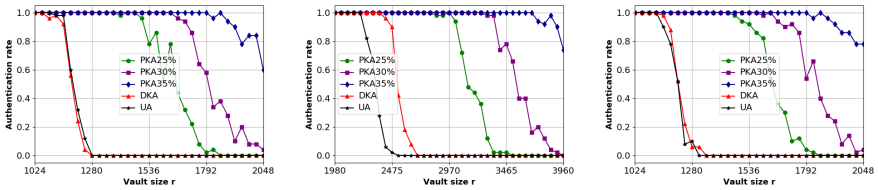
Overall, the three bases do not exhibit an EER, displaying FAR rates below 2% for all bases, thereby enhancing the fuzzy vault's security. However, the FRR rate has shown a significant increase for two databases, FVC and LFW, making usability less straightforward, requiring the user to authenticate repeatedly with a more precise representation of the biometric data until successful, while with PTB bases, no degradation is observed Table 5. Using our method removes biases from features, but considering the FRR results for some bases, one approach to reducing them could involve retaining certain biases, balancing system usability and security against statistical biases in features.

6.2 Results

Given that scenario 2, characterized by a feature-based attack, poses the most formidable threat compromising the integrity of the fuzzy vault and rendering it intricate to use. In this same scenario, we aim to quantify the biases of each feature obtained through the application of the quartile method. Subsequently, we intend to assess the advantages resulting from the exploitation of these biases and determine if this compromises the security of the fuzzy vault scheme once again.

The objective of our quartile method is to obtain features with more balanced values closer to uniformity. Reassessing the biases of features using the measure $\mathcal{M}$, as expected, reveals a predominance of features with a $\mathcal{M}$ measurement exceeding 0.9, contrasting with previous results where the majority were between 0.6 and 0.8.

To quantify the effectiveness of our proposed solution, we replicate the attack on the three bases using identical attacker models similar to those in scenario 2. The results are visually presented Figure 6.



(a) FVC (b) PTB (c) LFW

Figure 6: Authentication rate with quantile method

The advantage of an attacker exploiting biases decreases significantly. Based on the results from the three bases, in line with our method used to remove biases by features, we achieve an advantage of the red curve similar to that of the black curve, obtained through uniform sampling. Specifically, with vault sizes much smaller than $n*3$, it is observed that the vault becomes more secure without the need to add numerous chaff points. These conclusions also hold for Scenario 1, where there is no knowledge of biases specific to the features. Thus, the quartile method eliminates the effectiveness of bias attacks while preserving the nature of the proposed original scheme.

To conclude this study, we have proposed a first method based on eliminating biases from each feature during the transformation of a template into a biometric set, This approach employs the quantile method to ensure a fair presentation of each feature's values. Unlike methods to enhance the vault's security by adding passwords, our single-factor-based proposal maintains an equivalent level of security in the case of uniformity without requiring additional information. We have presented a solution to mitigate biases in each feature and mitigate attacks specifically aimed at these features. Nonetheless, the potential for an attack exploiting the correlation between features persists. This scenario could render biases exploitable, thus compromising the vault's security.

7 CONCLUSION

In this study, we have quantified the advantage of exploiting statistical biases in the biometric sets of the fuzzy vault, compromising its security and rendering its use by the initial proposal unfeasible. In response to these vulnerabilities, we have introduced a first method for a secure single-factor fuzzy vault authentication, aligned with the initial proposition. Quantiles were employed to achieve a balanced distribution of the values for each feature, eliminating the need for additional information and avoiding the use of multi-factor authentication.

We have obtained preliminary results supporting the effectiveness of the single-factor fuzzy vault, which is not sensitive to feature biases, highlighting a significant

correlation between its security and the construction function of the biometric sets. Currently, it is still possible to deduce the corresponding feature of the biometric template from an element of the biometric set. One method to prevent this is to add a secret on the server side to protect the parameters of the construction function, thereby enhancing security while maintaining the single-factor scheme and eliminating attacks based on features, thus avoiding a correlation attack between features. This approach also helps avoid constraints related to parameter choices, improving system performance.

REFERENCES

Belguechi, R., Hafiane, A., Cherrierand, E., and Rosenberger, C. (2016). Comparative study on texture characteristics for fingerprint recognition: application to the biohashing template protection scheme. *Journal of Electronic Imaging*.

Benhammadi, F. and Bey, K. B. (2014). Password hardened fuzzy vault for fingerprint authentication system. *Image and Vision Computing*.

Bousseljot, R., Kreiseler, D., and Schnabel, A. (1995). Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet.

Cover, T. M. and Thomas, J. A. (2006). Elements of information theory. *Wiley-Interscience*.

Dargan, S. and Kumar, M. (2020). A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications 143*.

Daugman, J. (2004). How iris recognition works. *IEEE TRANSACTIONS ON CIRCUITS and SYSTEMS FOR VIDEO TECHNOLOGY*.

Dodis, Y., Ostrovsky, R., Reyzin, L., and Smith, A. (2008). Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*.

Dong, X., Jin, Z., and Jin, A. T. B. (2019). A genetic algorithm enabled similarity-based attack on cancellable biometrics. *In IEEE Inter-BIBLIOGRAPHIE 119 national Conference on Biometrics : Theory, Applications and Systems (BTAS)*.

Gernot, T. and Lacharme, P. (2022). Biometric masterkeys. *Computers Security*.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet : Components of a new research resource for complex physiologic signals.

Huang, G. B., Mattar, M., Berg, T., and Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *in Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Juels, A. and Sudan, M. (2002). A fuzzy vault scheme. *Proceedings IEEE International Symposium on Information Theory*.

Khalil-Hani, M., Marsono, M. N., and Bakhteri, R. (2013). Biometric encryption based on a fuzzy vault scheme with a fast chaff generation algorithm. *Future Generation Computer Systems*.

Lee, Y. J., Park, K. R., Lee, S. J., Bae, K., and Kim, J. (2008). A new method for generating an invariant iris private key based on the fuzzy vault system. *IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics)*.

Maio, D., Maltonil, D., Cappelli, R., Wayman, J., and Jain, A. (2002). Fvc2002: Second fingerprint verification competition. *Proceedings of the 16th International Conference on Pattern Recognition (ICPR)*.

Makowski, D., Pham, T., Lau, Z., and Brammer, J. (2021). Neurokit2 : The python toolbox for neurophysiological signal processing. *https://github.com/neuropsychology/NeuroKit*.

Martinez, J., Almeida, R., Olmos, S., Rocha, A., and Laguna, P. (2004). A wavelet-based ecg delineator : evaluation on standard databases. *IEEE Transactions on Biomedical Engineering, 51(4) :570–581. 28*.

Merkle, J., Niesing, M., Schwaiger, M., Ihmor, H., and Korte, U. (2010). Security capacity of the fuzzy fingerprint vault. *International Journal on Advances in Security*.

Nagar, A., Nandakumar, K., and Jain, A. K. (2008). Securing fingerprint template: Fuzzy vault with minutiae descriptors. *19th International Conference on Pattern Recognition*.

Nandakumar, K., Jain, A. K., and Pankanti, S. (2007a). Fingerprint-based fuzzy vault: Implementation and performance. *IEEE Transactions on Information Forensics and Security*.

Nandakumar, K., Nagar, A., and Jain, A. K. (2007b). Hardening fingerprint fuzzy vault using password. *Advances in Biometrics International Conference ICB*.

Poon, H. T. and Miri, A. (2012). On efficient decoding for the fuzzy vault scheme. *11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*.

Radha, N., Karthikeyan, S., and P.Anupriya (2010). Securing retina fuzzy vault system using soft biometrics. *Global Journal of Computer Science and Technology*.

Rathgeb, C., Tams, B., Merkle, J., Nesterowicz, V., Korte, U., and Neu, M. (2023). Multi-biometric fuzzy vault based on face and fingerprints. *Computer Science*.

Rathgeb, C., Tams, B., Wagner, J., and Busch, C. (2016). Unlinkable improved multi-biometric iris fuzzy vault. *EURASIP Journal on Information Security*.

Reddy, E. S. and Babu, I. R. (2008). Performance of iris based hard fuzzy vault. *IEEE International Conference on Computer and Information Technology Workshops CIT*, 8.

Rice, J. A. (2006). Mathematical statistics and data analysis. *Duxbury Press*.

Sharma, A., Raghuwanshi, A., and Sharma, V. (2015). Biometric system-a review. *Int. J.Comput. Sci. Inf. Technol*.

Simonoff, J. (2012). Smoothing methods in statistics. *Springer Science Business Media*.

Uludag, U., Pankanti, S., and Jain, A. K. (2005). Fuzzy vault for fingerprints. *Audio- and Video-Based Biometric Person Authentication*.

Uludag, U., Pankanti, S., Prabhakar, S., , and Jain, A. K. (2004). Biometric cryptosystems: issues and challenges. *Proceedings of the IEEE, 92(6):948–960*.

Velciu, M.-A., Pătrascu, A., and Patriciu, V.-V. (2015). An evaluation of the reed-solomon error-correcting codes usage for bio-cryptographic algorithms. *10th Jubilee IEEE International Symposium on Applied Computational Intelligence and Informatics*.

V.S.Meenakshi and Padmavathi, D. G. (2010). Retina and iris based multimodal biometric fuzzy vault. *International Journal of Computer Applications*.

You, L. and Wang, T. (2018). A novel fuzzy vault scheme based on fingerprint and finger vein feature fusion. *Soft Computing*.

Zheng, A. and Casari, A. (2018). Feature engineering for machine learning: Principles and techniques for data scientists. *O'Reilly Media*.