



Contents lists available at ScienceDirect

# Journal of King Saud University – Computer and Information Sciences

journal homepage: [www.sciencedirect.com](http://www.sciencedirect.com)

## A cancelable biometric identification scheme based on bloom filter and format-preserving encryption

Vidhi Bansal<sup>a</sup>, Surabhi Garg<sup>b,\*</sup><sup>a</sup> Indira Gandhi Delhi Technical University for Women, India<sup>b</sup> IIT-Delhi, India

### ARTICLE INFO

#### Article history:

Received 23 June 2021

Revised 3 November 2021

Accepted 21 January 2022

Available online 10 February 2022

#### Keywords:

Multi-biometrics

Format-preserving encryption

Bloom filter

Biometric template protection

Identification

### ABSTRACT

Biometric based authentication systems are being prominently used everywhere. The biometric data, popularly known as a biometric template, is generally stored on the database server in its unprotected form. Unlike passwords, once compromised, biometric data can never be recovered. An ideal biometric system should provide accessibility, acceptability, availability, high security, and high biometric performance to the user. Current deployments generally relax in one or more requirements, resulting in lingering concerns about the privacy and security of individuals' biometric data. Our proposed work introduces a cancelable biometric template protection scheme based on the format-preserving encryption and Bloom filters. The format-preserving encryption encrypts the biometric template, which then maps to the Bloom filter based template that represents the cancelable template. The use of format-preserving encryption along with Bloom filters helps to achieve the security of the input biometric template and identification with good recognition performance. We achieve 0.2% FRR at 0.01% FAR for IITD-CASIA virtual dataset in the uni-biometric scenario. A comparison with the existing schemes shows that our proposed scheme exhibits high recognition performance for both uni-biometric and multi-biometric datasets, while simultaneously, the security of the overall system is preserved.

© 2022 The Authors. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Biometric-based authentication systems are being prominently used these days. A biometric system relies on several unique features extracted from the iris, fingerprint, face, etc., to identify an individual on a real-time basis. These unique features extracted from the biometric characteristics construct the biometric templates. A biometric authentication system constitutes two phases enrolment and authentication. During the enrolment phase, the user provides a biometric template which is stored on the database server. A query template is provided in the authentication phase, which is compared with the stored biometric templates. If there is a match, the user is authenticated. The authentication can be further classified in two ways- verification, a one to one comparison, and identification that requires one-to-many comparisons. In comparison to verification, identification is a computationally intensive task. A biometric authentication system can be a uni-biometric or a multi-biometric system. A uni-biometric system uses a single biometric input to authenticate a user, whereas a multi-biometric sys-

tem relies on two or more biometric inputs fused based on score-level, decision level or feature level (Li et al., 2015; Sudhamani et al., 2014; Chang et al., 2021; Nagar et al., 2011; Dwivedi and Dey, 2018). A multi-biometric system is more reliable as it is difficult to spoof multiple biometric features simultaneously (Ross and Poh, 2009).

Generally, the biometric data of the enrolled users is stored in its unprotected, plain form on the database server during the enrolment phase. Even if the data is stored in encrypted form, the encryption requires decryption during the authentication phase which can result in compromising the biometric data. Further, the European Union (EU) General Data Protection Regulation 2016/679 (Regulation, 2016) has classified biometric data as sensitive information which should be covered under the right to privacy. An ideal biometric system should prevent the misuse of biometric data while retaining its biometric performance. A biometric system should ensure that genuine users can access the system without any difficulty, and all the impostors are identified. An ideal system should aim for a low false acceptance rate and low false rejection rate. In addition to achieving high recognition performance, a biometric system should fulfil the requirements stated in ISO/IEC IS 24745 : 2011 (ISO, 2011):

\* Corresponding author.

E-mail address: [surabhig@iitd.ac.in](mailto:surabhig@iitd.ac.in) (S. Garg).

**Unlinkability:** It should be infeasible to determine if two or more templates are derived from the same or different subjects to prevent cross-matching of templates across different databases. The correlation of different protected templates of the same subject should be infeasible.

**Irreversibility:** Given the protected template and another secret parameter(s) used in creating the protected template, it should be infeasible to find out the original biometric data.

**Renewability:** Issue of the new template should be possible while revoking the old one from the same biometric instance to prevent misuse of a template in case the database is being compromised.

The above-mentioned requirements and the security and privacy concerns prompt the need to design a biometric protection scheme. Issues of biometric sample recovery from a protected template and user linkability across various databases should also be taken care of (Pagnin and Mitrokovtsa (2017) and Nandakumar and Jain (2015)). The existing biometric template protection schemes (Nagar et al., 2011; Patel et al., 2015; Chang et al., 2020; Chang et al., 2021; Yasuda et al., 2013) can be broadly classified into biometric cryptosystems, cancelable biometrics and homomorphic encryption schemes. These schemes protect the input biometric template by using some transformation or encryption mechanism such that the protected template does not reveal any information about the original biometric template. Section 3 presents a few of the notable works with their limitations for all three categories.

**Motivation.** Biometric data leakage may lead to severe threats to the privacy and security of the user. The risk associated with the compromise of biometric data requires the biometric data to be stored in a secured format. At the same time, the security aspect should not degrade the biometric performance rate, where identification of biometric templates is an essential and computationally intensive task. Identification involves one-to-many comparisons that should be done securely along with high recognition performance. Most of the existing schemes lack in one or more aspects (Jegade et al., 2017; Pagnin and Mitrokovtsa, 2017). Some of the existing Bloom filter based cancelable biometric schemes (Drozdzowski et al., 2018; Gomez-Barrero et al., 2016) achieve fast identification (while assuming the underlying security operation to be done in the offline mode). However, they degrade the recognition performance as compared to the baseline (Daugman, 2003), and state-of-the-art Bloom filter based approaches (Rathgeb et al., 2013; Rathgeb et al., 2014). Moreover, these existing, baseline and state-of-the-art Bloom filter-based approaches lack security. Additionally, other biometric template protection schemes (Chang et al., 2020; Rathgeb et al., 2021; Chang et al., 2021) are proved to be secure; however, either the identification is not feasible in real-time, or they degrade the recognition performance.

To mitigate the limitations of existing schemes, we introduce a cancelable biometric scheme where we use format-preserving encryption to encrypt the user's biometric data before it is mapped to the Bloom filter based templates. Format-preserving encryption preserves the format and length of the biometric data. No additional bit errors are introduced during the encryption of original biometric data. The encrypted biometric template is then mapped to the corresponding Bloom filter based template (Rathgeb et al., 2014) which supports an efficient identification. Since Bloom filters are extensively used to protect the biometric templates while achieving the identification aspect, in our work, we primarily focus on improving the existing and state-of-the-art Bloom filter based schemes (Rathgeb et al., 2014; Drozdzowski et al., 2018). We denote the state-of-the-art Bloom filter based approaches as a baseline Bloom filter scheme.

**Our Contributions.** The contributions of our proposed scheme are as follows:

- We propose a cancelable biometric template projection scheme for identification that provides the security of biometric data and the preservation of the security of the overall system. Our proposed scheme works for both uni-biometric as well as multi-biometric scenarios.
- Our proposed scheme encrypts the original, input biometric data of the user using format-preserving encryption and transforms the encrypted data into a Bloom filter based template which is then stored onto the database server. The recognition performance of our proposed scheme is equivalent to the baseline Bloom filter based scheme (Rathgeb et al., 2013; Rathgeb et al., 2014).
- Both theoretical and experimental analysis of the security and performance of the scheme has been provided. We also provide a comprehensive comparative analysis of our proposed approach on both uni-biometric and multi-biometric systems with some of the other existing prominent schemes. The proposed scheme pivots around the core principles of perfect secrecy and high biometric performance.
- The results in Fig. 7–10 show the scope of deployment of our proposed scheme in real-time scenarios. We provide a thorough security analysis that shows that our proposed scheme satisfies all three security and privacy properties- irreversibility, unlinkability and renewability.

**Organization:** The rest of the paper is organized as follows. Section 2 provides background on some fundamental concepts. Various other template protection schemes are provided in Section 3. The system model and its participants are explained in Section 4. A detailed explanation of the proposed scheme is given in Section 5. Design rationale is presented in Section 6. Datasets used, evaluation mechanism, and other related information about the experiments conducted and results obtained is provided in Section 7. Detailed security analysis is provided in Section 8. Further, conclusions and future work are in Section 9.

## 2. Preliminaries

In this section, we provide some terms and notations used throughout the paper.

### 2.1. Notations

The original unprotected biometric template is denoted as  $B$ ,  $\bar{B}$  is the encrypted template, and  $C$  is the protected template or the cancelable template. It is obtained by mapping of  $\bar{B}$  to the Bloom filter based templates. Here,  $C$  denotes the set of  $m$  Bloom filters represented as  $\{c_1, c_2, \dots, c_m\}$ .  $K$  is the symmetric key.

### 2.2. Format-preserving Encryption (FPE)

Format-preserving encryption (FPE) is the encryption of plaintext into a ciphertext such that it preserves the format and length of the ciphertext. For example, encrypting Aadhaar, a set of 12 decimal digit numbers where each digit  $\in [0 - 9]$  using FPE gives a ciphertext of 12 decimal digits where each digit  $\in [0 - 9]$  (Bellare et al., 2009). FPE can encrypt the plaintext of any domain. Data is encrypted using a symmetric key  $K$ . FPE uses a permutation-based deterministic algorithm for encrypting the plaintext. The permutation function obscures the plaintext information. Feistel-based encryption algorithm takes in a key  $K$ , plaintext  $X$  of length  $n$  from a set of the domain  $D$ , and a tweak  $T$  as inputs to generate ciphertext  $Y$  where  $Y \in D^n$ . Tweak enhances the security of the encryption algorithm, especially against dictionary attacks. It is a user-specific, public value (Dworkin, 2019). Recently, FF3 mode is

officially deprecated by NIST due to the cryptanalysis of FF3 which has reduced its security level below 128 bits with a key of size 128 bits (Hoang et al., 2018; Durak and Vaudenay, 2017). In the proposed work, we have used the updated mode, FF3-1 (Dworkin, 2019) for format-preserving encryption recommended by NIST which is so far considered to be secure against the attack.

### 2.3. Bloom Filters

A Bloom filter is represented in the form of a bit array whose bits are initialized to 0. Given the input biometric template represented as a two-dimensional binary vector of width  $W$  and height  $H$ ,  $m$  equal sized blocks are created such that each block size  $= W/m$ . The column is of height  $h \leq H$ , where  $h$  represents word size. Each block is then transformed to a corresponding Bloom filter (Rathgeb et al., 2014) using binary to integer mapping. Each column of each block is mapped to an integer value that represents the index to the corresponding Bloom filter. Resultant protected biometric template  $C$  is the set of  $m$  Bloom filters represented as  $\{c_1, c_2, \dots, c_m\}$  that denotes the cancelable template. To compare two protected templates, average pairwise template hamming distance  $HD$  is calculated between corresponding Bloom templates. A match score  $S$  between two Bloom filters  $C, C'$  (Drozdowski et al., 2018) is given as  $S(C, C') = 1 - \frac{\sum_{i=1}^m (HD(c_i, c'_i) / |c_i| + |c'_i|)}{m}$ .

## 3. Related work

In this section we present some of the existing prominent biometric template protection schemes.

### 3.1. Biometric cryptosystems

Biometric cryptosystems, also known as biocryptosystems, generate biometric dependent public information known as helper data from the biometric template. Helper data does not reveal any significant information from the biometric template. Authentication is done indirectly by deriving keys from the template received (Dodis et al., 2004; Juels and Wattenberg, 1999). Fuzzy vault and Fuzzy commitment are the two popular biocryptosystems. Fuzzy commitment scheme for biometric systems is first explored by Dodis et al. (2004). Rathgeb and Uhl (2010), Rathgeb and Uhl (2010), Rathgeb and Uhl (2011) proposed an adaptive fuzzy commitment scheme based on iriscodes error analysis. Juels and Sudan (2006) introduced fuzzy vault scheme. Biometric features are provided as input to the system where they are encoded on a polynomial with the secret key used as the coefficient to the polynomial. A fuzzy vault is generated as a helper data which is stored on the database server (Nandakumar et al., 2007). The key is recovered during authentication if a similar biometric feature set overlaps with the enrolled set. In Nagar et al. (2011), an implementation of feature-level fusion framework for both fuzzy vault and fuzzy commitment schemes is provided. A fuzzy vault constructed from multiple biometric features of a user provides high security and recognition performance than the vault constructed from a single biometric feature. BIOFUSE (Chang et al., 2021) is a framework for multi-biometric fusion which uses format-preserving encryption for combining fuzzy vault and fuzzy commitment. The approach provides high security but relies on the underlying fuzzy commitment and fuzzy vault's architecture to provide high recognition performance. In general, the major limitation of the biometric cryptosystems is that they require error correction code to correct the bit-error in the biometric template. The error correction code can correct the bit-errors in biometric templates only to a fixed limit. Additionally, efficient identification is not feasible in such systems.

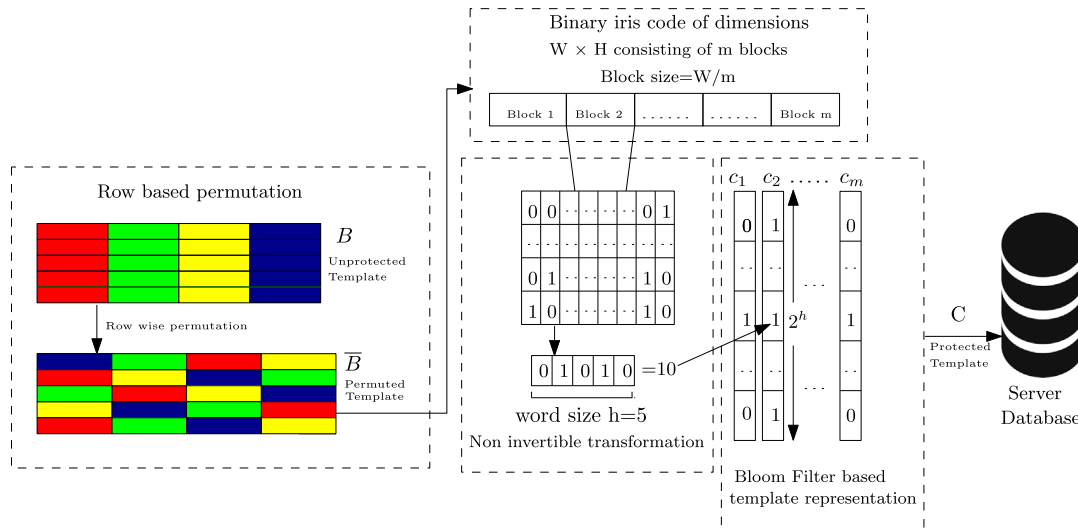
### 3.2. Cancelable biometrics

Cancelable Biometrics work by distorting the original biometric template. The comparisons during the authentication phase are performed on the protected or cancelable biometric templates such that no information regarding the original biometric templates is revealed. Cancelable biometrics are broadly classified into Salting and Non-invertible techniques. In Salting, a cancelable template is created by adding an artificial random pattern to the biometric template. GRAY-SALT and BIN-SALT (Zuo et al., 2008) are the two salting approaches. A synthetic pattern is mixed with input iris image using pixel-wise addition or multiplication in the GRAY-SALT. BIN-SALT is a similar approach that works on binarized iris images. The strength of noise can affect the performance rate, and security (Patel et al., 2015). Jin et al. (2004) introduced Biohashing. It uses two-factor authentication and combines unique tokens generated from a hash key and user biometric data to secure the biometric data. However, if that unique token is revealed, the security of the system is compromised. Jegede et al. (2017) discusses various challenges and open research issues in cancelable biometrics as well as in the combination of cancelable biometrics with biocryptosystems. In Chang et al. (2020), the bit-wise encryption scheme and fuzzy extractor are combined to generate a cancelable template. High security is provided on the assumption that obtaining access to one biometric template by an attacker is equivalent to getting both biometric templates of the user.

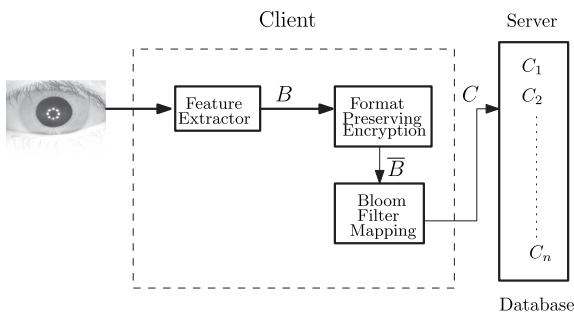
Bloom filters have been extensively used in the literature in the domain of biometric security. The application of Bloom filters in biometric template protection is first introduced in Rathgeb et al. (2013) and Rathgeb et al. (2014) where the Bloom filters based templates are created from the input biometric template. Gomez-Barrero et al. (2018) proposed a multi-biometric fusion based on the Bloom filter approach where the biometric templates are converted to Bloom filters. A boolean OR operation is performed between the corresponding Bloom filter arrays. An unlinkable and irreversible template protection scheme is proposed in Gomez-Barrero et al. (2016) which is based on the Bloom filters. Based on a similar approach, Drozdowski et al. (2018) proposed a row-wise permutation of iriscodes using a system-generated key. The biometric template is permuted row-wise using a system-generated key as shown in Fig. 1. Indexing using binary trees is done on the Bloom filter based templates. The major drawback of the scheme is that the performance rate degrades in comparison to the baseline Bloom filter based techniques (Rathgeb et al., 2014; Rathgeb et al., 2013) due to the row-wise permutation. Moreover, the permutation is done in offline mode, making it an infeasible scheme for several real-time deployments. We primarily focus on improving the existing Bloom filter based approaches in our proposed scheme.

### 3.3. Homomorphic encryption schemes for biometric template protection

Homomorphic encryption allows computations on encrypted data. Homomorphic encryption function permits to operate on the encrypted data without decrypting it (Rivest et al. (1978)). Cheon et al. (2016) proposed the use of homomorphic encryption and message authentication code. SIMD operations are used for developing matching functions and one-time MAC for homomorphic evaluations. Gomez-Barrero et al. (2017) proposed a multi-biometric template protection scheme based on Homomorphic Encryption. The data, whether stored in a database or transit between server and client, is encrypted with three fusion levels: feature level, score level, and decision level. Yasuda et al. (2013) proposed a packed Homomorphic Encryption based on ideal



**Fig. 1.** Row-wise permutation: The 2D biometric template is permuted row wise using a system generated key. A  $W \times H$  sized template is transformed to template of  $m \times 2^h$ ,  $h$  is the word size.  $C$  is the set of  $m$  Bloom filters represented as  $\{c_1, c_2, \dots, c_m\}$ .



**Fig. 2.** Schematic model of proposed model for a uni-biometric system: Biometric template  $B$  of the user is encrypted using format-preserving encryption. Encrypted biometric template is denoted as  $\bar{B}$ .  $\bar{B}$  is mapped to Bloom filters to obtain a protected template  $C$ . All this is performed at client's site. The protected template  $C$  is then stored on the server database.

lattices and its application in biometric security. In Zhou and Ren (2018), a user-centric biometric authentication scheme PassBio is proposed that enables a client to encrypt templates with the proposed light-weighted encryption scheme based on matrix multiplications.

#### 4. Models and settings

In this section, we present the system model along with various threats.

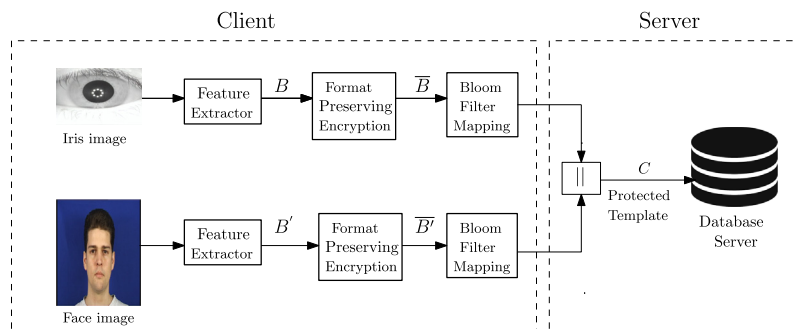
##### 4.1. System model and participants

The proposed system employs a client-server architecture. A user's biometric data or template is extracted at the client's site using a device capable of extracting biometric data, e.g. desktop computer, tablet, smartphone. A client-side user can be genuine or an impostor. The server is assumed to be honest but curious. The proposed scheme is evaluated for both uni-biometric and multi-biometric characteristics. Feature extraction, data encryption, and transformation into Bloom templates are done at the client site. Biometric authentication is performed on the server-side. The schematic models of our proposed scheme are presented in Fig. 2 for the uni-biometric system and the multi-biometric system in Fig. 3.

##### 4.2. Threat model

To ensure the security of the user's biometric data, we focus on the following threats:

- An attacker may be the passive attacker who can access the database. In such a scenario, the attacker can get access to any of the protected biometric templates.
- An attacker at a client site or server-side may try to achieve reversibility. In such a case, the attacker could try to invert the protected template stored on the database server to get the original biometric template.



**Fig. 3.** Schematic model of proposed model for a multi-biometric system: Biometric templates  $B$  and  $B'$  of the user is encrypted using format-preserving encryption. Encrypted biometric template is denoted as  $\bar{B}$  and  $\bar{B}'$  respectively.  $\bar{B}$  and  $\bar{B}'$  are mapped to their respective Bloom filters and are concatenated to obtain a protected template  $C$ . The protected template  $C$  is then stored on the server database.

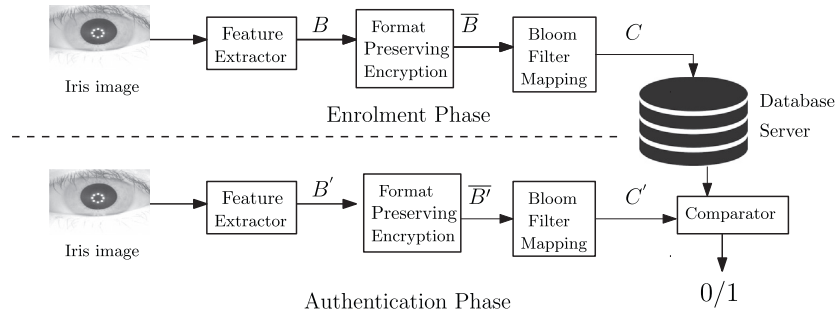
- An attacker may try to achieve linkability of biometric templates of a user stored across different databases.

## 5. Proposed work

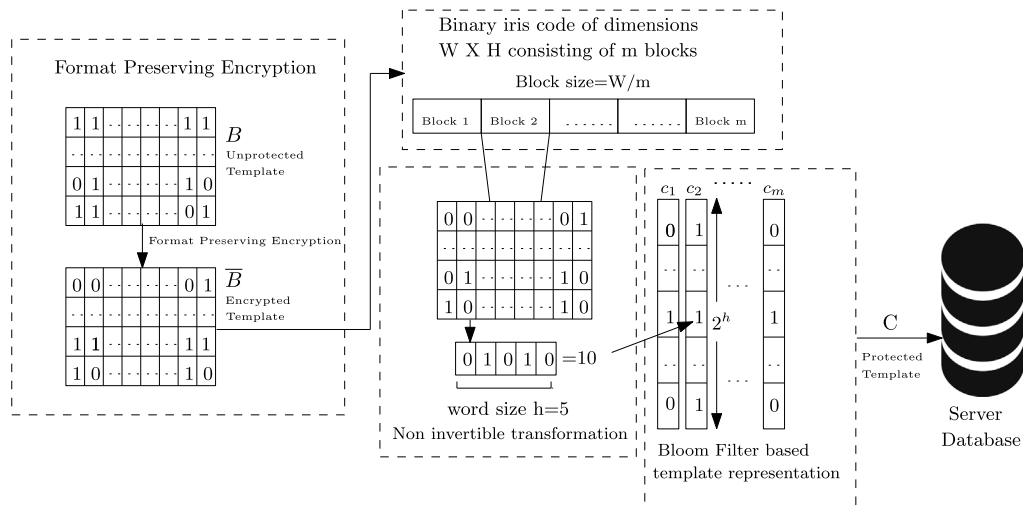
In this section, we present the working of our proposed system. It requires encryption of the biometric data using format-preserving encryption with the help of symmetric key  $K$ . The encrypted template is transformed into a Bloom filter and stored on the server. System overview of the proposed model is presented in Fig. 4, and implementation details are presented in Fig. 5. The following three modules are used.

1. **Key Generation:** A randomly generated symmetric key is used to encrypt the biometric template of the user. The random key is generated from a pseudorandom number generator. To ensure the security of the key, we generate the key  $K$  such that,  $|K| \geq 128$  bits. The key is treated as a system's generated key and would remain the same for all the enrolled users of the system. It is assumed to be securely stored in the hardware security module such as a trusted platform module to avoid any case of key compromise.
2. **Tweak Generation:** A 56 bits tweak is generated using a pseudorandom number generator. The tweak is a user-specific public value that is unique for every user.

3. **Format-preserving Encryption:** The biometric template  $B$  extracted from the user's input biometric characteristics is encrypted using format-preserving encryption. FPE preserves the format and length of the input. It further means that the length and format of the output are the same as that of the input. Given the input template, we encrypt every column of the template, considering a single column at a time as a plaintext. A system generated 128 bits key is used for encryption. A public, unique tweak is used for every individual that provides access to independent pseudo-random permutations of the original biometric template. FF3-1 requires a tweak of 56 bits (Dworkin, 2019; Durak and Vaudenay, 2017). The output of format-preserving encryption would represent the corresponding column in the encrypted biometric template,  $\bar{B}$ . Similarly, all the columns of the input template are encrypted to the corresponding columns in the encrypted template. Since the Bloom filter based mapping is done in a column-wise manner, we encrypt the template in a similar (column-wise) manner so that if there is a bit-error in a particular column of the input biometric template, the same would be retained in the corresponding column and is not propagated to other columns after the encryption using format-preserving encryption. It helps to preserve the recognition performance of the encrypted template.
4. **Bloom filter based Template Generation:** Given the encrypted biometric template  $\bar{B}$  as input of dimensions  $W \times H$ , the Bloom



**Fig. 4.** Overview of the proposed scheme: During enrolment, the input biometric data  $B$ , is encrypted using format-preserving encryption  $\bar{B}$ .  $\bar{B}$  is mapped to Bloom filters to obtain a protected template  $C$  which is stored on the server. During authenticating a user  $B'$  is extracted and encrypted using format-preserving encryption to generate  $\bar{B}'$  which is mapped to Bloom filters.



**Fig. 5.** Proposed system: The biometric template of the user data  $B$  is encrypted to get  $\bar{B}$  using format-preserving encryption. Domain and size of the input is retained in ciphertext also.  $m$  Bloom filters are created each of size  $2^h$  from  $\bar{B}$  which is of dimensions  $W \times H$  using a non invertible function.  $C$  is the set of  $m$  Bloom filters represented as  $\{c_1, c_2, \dots, c_m\}$  and denotes cancelable templates.  $C$  is stored on the server database.



filter based template is generated which represents the cancelable template.  $m$  Bloom filters are created, each of size  $2^h$  from the encrypted biometric template with dimensions  $W \times H$ , such that  $C = \{c_1, c_2, \dots, c_m\}$ . A non invertible function (in our case, integer to binary representation) is used to map encrypted biometric template  $\bar{B}$  to Bloom filters.

### 5.1. Enrolment phase

During the enrolment phase, a user provides its input biometric characteristics. The input can be uni-biometric or multi-biometric based on the system's deployment. Considering the uni-biometric case, the features are extracted from the biometric characteristics using a feature extractor to construct the biometric template,  $B$ . The biometric template  $B$  is encrypted using the format-preserving encryption such that each column in the input biometric template  $B$  is encrypted to the corresponding column in the encrypted template  $\bar{B}$ . The encrypted template is mapped to the Bloom filter based template  $C$  using the approach discussed in Section 2. The Bloom filter based template for each user is then enrolled and stored on the server database. In the case of multi-biometric as shown in Fig. 3, the Bloom filter based templates are constructed for both the input templates  $B$  and  $B'$ . The concatenation of both Bloom filter based templates is stored on the database server.

### 5.2. Authentication phase

On providing the query biometric characteristics, the biometric template  $B'$  is constructed from it. It is encrypted similarly using format-preserving encryption as done during the enrolment phase. The encrypted biometric template  $\bar{B}'$  is then mapped to the Bloom filter based template to get  $C'$  as the protected or cancelable template, which is compared with the enrolled templates to check if there is a match or not.

## 6. Design rationale

### 6.1. Role of format-preserving encryption

In format-preserving encryption, both the plaintext and the ciphertext have the same format and same length. It means that the encrypted biometric template would be of the same length and format as that of the unprotected, original biometric template. Considering the operation to be done in a column-wise manner, i.e. encrypting one column at a time, the number of bit-errors in the original biometric template  $B$  would be completely preserved after the encryption to the encrypted biometric template  $\bar{B}$ . Thus, performance will not be degraded during the encryption process. The mapping of an encrypted biometric template to the Bloom filter based template would provide the recognition performance equivalent to the mapping of an unprotected biometric template to the Bloom filter based template (Rathgeb et al., 2014) (denoted as baseline Bloom). However, in the case of the existing scheme discussed in Drozdowski et al. (2018), the permuted biometric template does not retain the bit-errors after permutation. In such a case, additional column-wise bit errors may be introduced during the permutation of the original biometric template to the permuted biometric template. The performance is degraded compared to the baseline Bloom filter based performance due to the introduction of errors after permutation. The same can be reflected in the results shown in Fig. 7–10. The following example depicts the error propagation in our proposed scheme, and row-wise permutation scheme (Drozdowski et al., 2018) for a given demo 2D template

$B$ . Given,  $K = 2D4B6150645367566B59703373367639792442264528482B4D6251655468576D$ .

$T = 2B4B6250655368$

**Case 1: Our proposed scheme:** During the enrolment phase, the user provides the input biometric template  $B$  which is encrypted using Format-preserving Encryption with Key  $K$ , and Tweak  $T$  to generate  $\bar{B}$ . For simplicity, we consider the mapping of the first three columns (as an example) of  $\bar{B}$  to the corresponding Bloom filter with the following index positions being set to 1.

<table> <tr><td>1</td><td>0</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> </table> <div><math>B</math></div>	1	0	1	..	..	0	1	0	1	0	..	..	1	1	1	1	1	..	..	0	0	0	0	0	..	..	1	1	1	1	0	..	..	1	0	<table> <tr><td>0</td><td>1</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>1</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> </table> <div><math>\overline{B}</math></div>	0	1	1	..	..	0	1	1	0	1	..	..	1	0	0	0	1	..	..	1	1	1	0	0	..	..	1	0	1	0	0	..	..	0	0	<table> <tr><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr> </table> <div>= 11</div>	0	1	0	1	1
1	0	1	..	..	0	1																																																																							
0	1	0	..	..	1	1																																																																							
1	1	1	..	..	0	0																																																																							
0	0	0	..	..	1	1																																																																							
1	1	0	..	..	1	0																																																																							
0	1	1	..	..	0	1																																																																							
1	0	1	..	..	1	0																																																																							
0	0	1	..	..	1	1																																																																							
1	0	0	..	..	1	0																																																																							
1	0	0	..	..	0	0																																																																							
0	1	0	1	1																																																																									
		<table> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table> <div>= 16</div>	1	0	0	0	0																																																																						
1	0	0	0	0																																																																									
		<table> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td></tr> </table> <div>= 28</div>	1	1	1	0	0																																																																						
1	1	1	0	0																																																																									
		Mapping to Bloom filters																																																																											

During the authentication phase,  $B'$  is the query biometric template with a one bit-error in the second column of the template (denoted by red color). It is encrypted using FPE to get  $\bar{B}'$ . Due to FPE, the bit-error is propagated in  $\bar{B}'$  only to the column (number 2) for which the corresponding column in  $B'$  has the single bit-error. Therefore, mapping to the Bloom filter will only change one index position.

<table> <tr><td>1</td><td>0</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>1</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> </table> <div><math>B'</math></div>	1	0	1	..	..	0	1	0	1	0	..	..	1	1	1	0	1	..	..	0	0	0	0	0	..	..	1	1	1	1	0	..	..	1	0	<table> <tr><td>0</td><td>1</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> </table> <div><math>\overline{B'}</math></div>	0	1	1	..	..	0	1	1	1	1	..	..	1	0	0	1	1	..	..	1	1	1	1	0	..	..	1	0	1	0	0	..	..	0	0	<table> <tr><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr> </table> <div>= 11</div>	0	1	0	1	1
1	0	1	..	..	0	1																																																																							
0	1	0	..	..	1	1																																																																							
1	0	1	..	..	0	0																																																																							
0	0	0	..	..	1	1																																																																							
1	1	0	..	..	1	0																																																																							
0	1	1	..	..	0	1																																																																							
1	1	1	..	..	1	0																																																																							
0	1	1	..	..	1	1																																																																							
1	1	0	..	..	1	0																																																																							
1	0	0	..	..	0	0																																																																							
0	1	0	1	1																																																																									
		<table> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>0</td></tr> </table> <div>= 30</div>	1	1	1	1	0																																																																						
1	1	1	1	0																																																																									
		<table> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td></tr> </table> <div>= 28</div>	1	1	1	0	0																																																																						
1	1	1	0	0																																																																									
		Mapping to Bloom filters																																																																											

Thus, it could be inferred from the example that using format-preserving encryption; we can achieve the equivalent recognition performance as compared to the baseline Bloom filter based performance (Rathgeb et al., 2013; Rathgeb et al., 2014).

**Case 2: Row-wise permutation (Drozdowski et al., 2018):** During enrolment, provided  $B$  as input biometric template, we permute it row-wise (Drozdowski et al., 2018) to get the permuted template  $\bar{B}$ . Similar to Case 1, we map only first three columns of the permuted template (as an example) to the Bloom filter.

<table> <tr><td>1</td><td>0</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>..</td><td>..</td><td>1</td><td>0</td></tr> </table> <div><math>B</math></div>	1	0	1	..	..	0	1	0	1	0	..	..	1	1	1	1	1	..	..	0	0	0	0	0	..	..	1	1	1	1	0	..	..	1	0	<table> <tr><td>1</td><td>1</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>1</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>..</td><td>..</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>..</td><td>..</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>..</td><td>..</td><td>0</td><td>1</td></tr> </table> <div><math>\bar{B}</math></div>	1	1	1	..	..	0	1	1	0	1	..	..	0	0	0	0	1	..	..	1	1	0	1	0	..	..	0	0	1	1	1	..	..	0	1	<table> <tr><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </table> <div>Mapping to Bloom filters</div>	1	1	0	0	1	1	0	0	1	1	1	1	1	0	1
1	0	1	..	..	0	1																																																																																	
0	1	0	..	..	1	1																																																																																	
1	1	1	..	..	0	0																																																																																	
0	0	0	..	..	1	1																																																																																	
1	1	0	..	..	1	0																																																																																	
1	1	1	..	..	0	1																																																																																	
1	0	1	..	..	0	0																																																																																	
0	0	1	..	..	1	1																																																																																	
0	1	0	..	..	0	0																																																																																	
1	1	1	..	..	0	1																																																																																	
1	1	0	0	1																																																																																			
1	0	0	1	1																																																																																			
1	1	1	0	1																																																																																			
		$= 25$																																																																																					
		$= 19$																																																																																					
		$= 29$																																																																																					

Now, during authentication, on providing  $B'$  as the query biometric template with one bit-error in the second column, we permute it to get  $\bar{B}'$ . Here, the bit-error from one column is now propagated to multiple columns (in our case, it is propagated to 2 columns due to the underlying row-wise permutation), which can be reflected in the mapping of the first 3 columns to the Bloom filter.

1 0 1 .. .. 0 1	1 1 1 .. .. 0 1	1 1 1 0 1 = 29
0 1 0 .. .. 1 1	1 0 1 .. .. 0 0	1 0 1 1 1 = 23
1 0 1 .. .. 0 0	1 1 1 .. .. 1 1	1 1 1 0 1 = 29
0 0 0 .. .. 1 1	0 1 0 .. .. 0 0	
1 1 0 .. .. 1 0	1 1 1 .. .. 0 1	

$B'$                        $B'$                       Mapping to Bloom filters

It is observed that in the row-wise permutation-based scheme, the number of columns with bit-errors is increased after permutation, whereas, in our proposed scheme, the number of columns with bit-error would remain the same after encryption.

## 6.2. Role of bloom filter

Bloom filter has been extensively used in the literature (Rathgeb et al., 2014; Drozdowski et al., 2018; Gomez-Barrero et al., 2016) to protect the biometric templates. Given an output as a Bloom filter based template, it is infeasible to recover the original biometric template of a particular individual. An attacker can see only the indexes of Bloom filters as part of the protected template. To reconstruct the corresponding block of biometric template from Bloom filter C, an arrangement of the corresponding  $|c| \leq m$  different words to a block of size  $m$  in the biometric template is required where  $|c|$  denotes active indexes of Bloom filter for a particular block. Given  $n$  total number of possible sequences resulting in the block of biometric template,  $n$  is given by Gomez-Barrero et al. (2016):

$$n = \sum_{i=1}^{|c|} (-1)^{|c|-i} \binom{|c|}{i} i^m \quad (1)$$

$n$  in Eq. 1 is very large even if  $|c|$  is small. This shows that reconstruction of the biometric template from Bloom filters is infeasible; hence, Bloom filters-based transformation is irreversible. Thus, Bloom filters, though, degrades the recognition performance to a small extent compared to the original unprotected system's recognition performance, provide high security and privacy when used with a secret parameter, key.

## 7. Experiments and results

This section presents the experiments and results of our proposed scheme and its comparison with other state-of-the-art schemes. Several experiments are performed on the iris and face features using publicly available databases. The example images are shown in Fig. 6. IITD (Kumar and Passi, 2010) and CASIA-Iris-Thousand were combined and denoted as IITD-CASIA. All the 20000 samples of CASIA-Iris-Thousand<sup>1</sup> are considered impostors. For a multi-biometric system, various virtual databases are created by combining samples from 2 different instances of a particular user. For one-to-one correspondence in the virtual database, we delete extra samples. Table 1 presents the genuine and impostor distribution in the selected databases.

Several open-source libraries and software are used for iris and face feature extraction. For iriscode generation, we use OSIRIS (Othman et al., 2016) and University of Salzburg Iris Toolkit v1.0 (Rathgeb et al., 2016). Feature extraction is done using Daugman-like 1D-Log Gabor (LG) (Masek, 2003) to generate feature vector, representing iriscode of size  $512 \times 20 = 10240$  bits. For face, FaceRecLib of the free signal and image processing toolbox Bob<sup>2</sup> (Anjos et al., 2017; Anjos et al., 2012) is used to

obtain a cropped  $4 \times 8$  sub-image of  $32 \times 2400 = 76800$  bits. We generate 16 Bloom filters for the iris while mapping the encrypted iriscode to Bloom filter based template. Each block from iriscode is selected with height 10 and width 32. For the face, 960 Bloom filters are generated, each of 16 bits. To compensate for the binary misalignment, circular shifts can be done (Daugman, 2003). Hamming distance is computed for each shift position. In our case, we tilted it by  $\pm 4$  bits. Minimum hamming distance is chosen for further computations. We implemented the format-preserving encryption scheme (Dworkin, 2019) using the python library for FF3-1 mode of implementation<sup>3</sup>. We suggest using 128,256 bits for key size. We use 128 bit key  $K$ . The tweak of 56 bits is randomly generated for each user using a pseudo-random number generator. For FF3-1, the size of the tweak is fixed to 56 bits (Dworkin, 2019).

## 7.1. Recognition performance evaluation

To evaluate the recognition performance on the proposed scheme, we plot the DET curves (Detection Error Trade-off). It demonstrates the false match rates against the false non-match rates (F NMR). For the multi-biometric system, the scores are normalized before fusion (He et al., 2010). Uni-biometric results are presented in Fig. 7 and multi-biometric in Fig. 8–10. Following schemes are implemented for fusion.

1. **Score-Level:** Scores from individual biometric characteristics are summed (using sum-rule He et al., 2010) to generate the final score value (Nagar et al., 2011; Dwivedi and Dey, 2018). Score level fusion results are presented in Fig. 8.
2. **Decision-OR:** Decision-OR fusion performs a Boolean OR operation between the individual scores of each component against the given threshold values to generate a final decision output (Li et al., 2015). Decision-OR fusion results are presented in Fig. 9.
3. **Decision-AND:** Decision-AND fusion performs a Boolean AND operation between the individual scores of each component against the given threshold values to generate a final decision output (Sudhamani et al., 2014). Decision-AND fusion results are presented in Fig. 10.

We used the following approaches for comparison:

1. **Baseline unprotected** (Daugman, 2003): In the baseline method, the biometric template of the user is stored in raw format without any protection. The recognition performance rate of this scheme is high.
2. **Baseline Bloom Filter** (Rathgeb et al., 2013; Rathgeb et al., 2014): From the biometric data, Bloom filters are created using the approach discussed in Fig. 5.
3. **Permuted template + Bloom filter** (Drozdowski et al., 2018; Gomez-Barrero et al., 2016): A row-wise permutation of the biometric template extracted from the input biometric characteristics is done using a system's generated key before its conversion into Bloom filters. Fig. 1 shows row-wise permutation on the input biometric template.
4. **Proposed:** Our proposed scheme is based on format-preserving encryption of biometric template of biometric information before its conversion to Bloom filters based template as discussed in Section 5.

Following are the observations from the Fig. 7, 8, 9, 10

<sup>1</sup> <http://biometrics.idealtest.org/>

<sup>2</sup> <http://idiap.github.io/bob/>

<sup>3</sup> <https://pypi.org/project/ff3/>

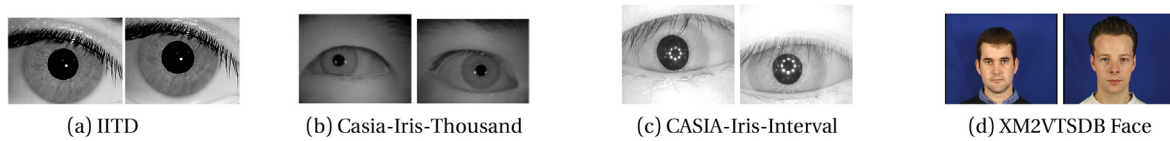


Fig. 6. Example images from the selected databases.

Table 1

Virtual Databases with Genuine and Impostor Distribution. Face denotes XM2VTSDB database.

Uni-biometric database			Multi-biometric database		
Database	Genuine	Impostors	Database	Genuine	Impostors
IITD-CASIA	256	2192	IITD-CASIA Left Iris & Right Iris	128	1096
CASIA-Iris-Interval <sup>1</sup>	128	209	CASIA-Iris-Interval Left Iris & Right Iris	128	36
Face (Ortega-Garcia et al., 2009)	128	167	Face Rotation 1 & Face Rotation 2	128	167
			IITD-CASIA & CASIA-Iris-Interval	256	81
			IITD-CASIA & Face	256	39
			CASIA-Iris-Interval & Face	256	39

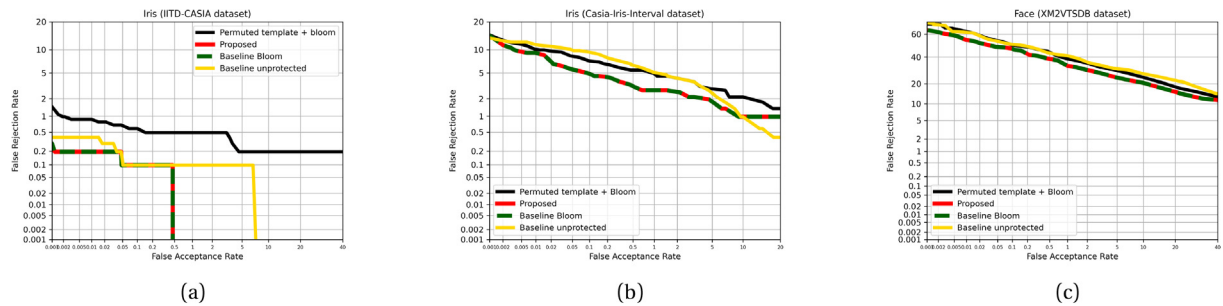


Fig. 7. DET curve for uni-biometric databases: a) Iris (IITD-CASIA) b) Iris (Casia-Iris-Interval) c) Face (XM2VTSDB).

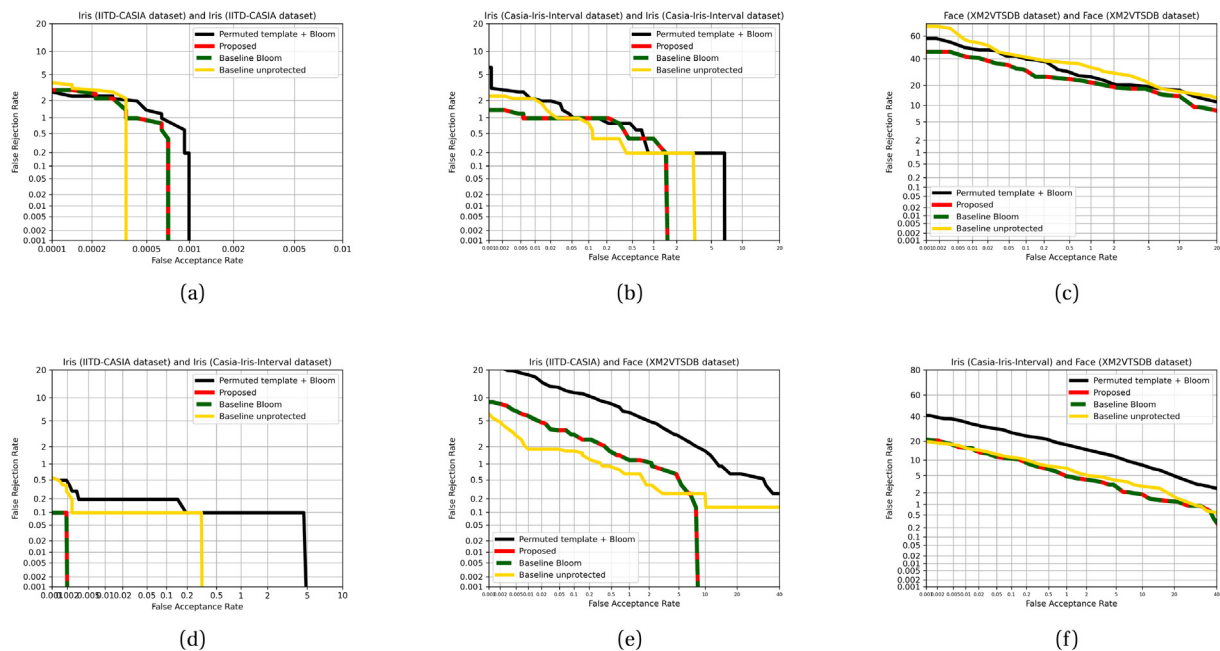


Fig. 8. DET curve for multi-biometric databases – Score Level Fusion: a) Iris (IITD-CASIA) and Iris (IITD-CASIA) b) Iris (Casia-Iris-Interval) and Iris (Casia-Iris-Interval) c) Face (XM2VTSDB) and Face (XM2VTSDB) d) Iris (IITD-CASIA) and Iris (Casia-Iris-Interval) e) Iris (IITD-CASIA) and Face (XM2VTSDB) f) Iris (Casia-Iris-Interval) and Face (XM2VTSDB).



- Our proposed scheme using format-preserving encryption gives recognition performance equivalent to the baseline Bloom filter based scheme (Rathgeb et al., 2014) while providing an additional layer of security on it.
- The recognition performance of our proposed scheme outperforms the performance of the existing Permuted template + Bloom filter based scheme (Drozdzowski et al., 2018) for all the databases used in our work.
- Though the difference between recognition performance of our proposed work and the work done in Drozdzowski et al. (2018) is not very significant, our proposed scheme does not require time consuming row-wise permutations of the original biometric template, which is assumed to be done in the offline mode (not always practically feasible). We use format-preserving encryption on the original biometric templates which is much faster and is comparable with the block cipher modes of encryption, the real-time deployment of our proposed scheme is feasible.
- Baseline unprotected scheme stores the template data in unprotected and raw format. Even though the recognition performance of our proposed scheme is marginally low as compared to the baseline unprotected scheme (Daugman, 2003); yet, our scheme is significant because of the additional security it provides to the biometric templates while providing a comparable recognition performance.
- In the uni-biometric system, our proposed scheme gives the best results for all the datasets used. Our proposed scheme achieves 0.2% FRR at 0.01% FAR and 0.1% FRR at 0.1% FAR for IITD-CASIA virtual dataset.
- In a multi-biometric system with score level fusion and Decision OR fusion, our proposed scheme gives result equivalent or better than unprotected baseline scheme for all the given datasets as shown in Fig. 8 and 9. Our proposed scheme achieves 0% FRR at 0.01% FAR and 0.1% FAR for IITD-CASIA virtual dataset.
- In a multi-biometric system with Decision AND fusion, our proposed scheme gives result equivalent or better than baseline

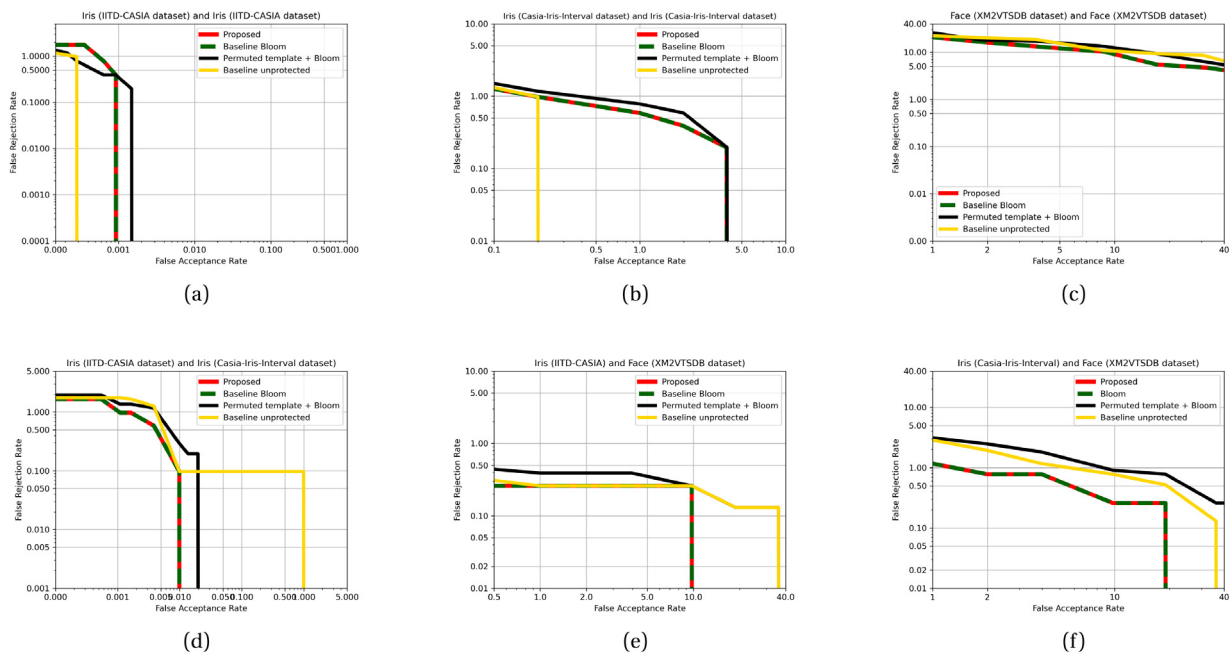
unprotected scheme for all the datasets as shown in Fig. 10. Our proposed scheme achieves almost 0.23% FRR at 0.01% FAR and 0.1% FAR for IITD-CASIA virtual dataset.

## 8. Security analysis

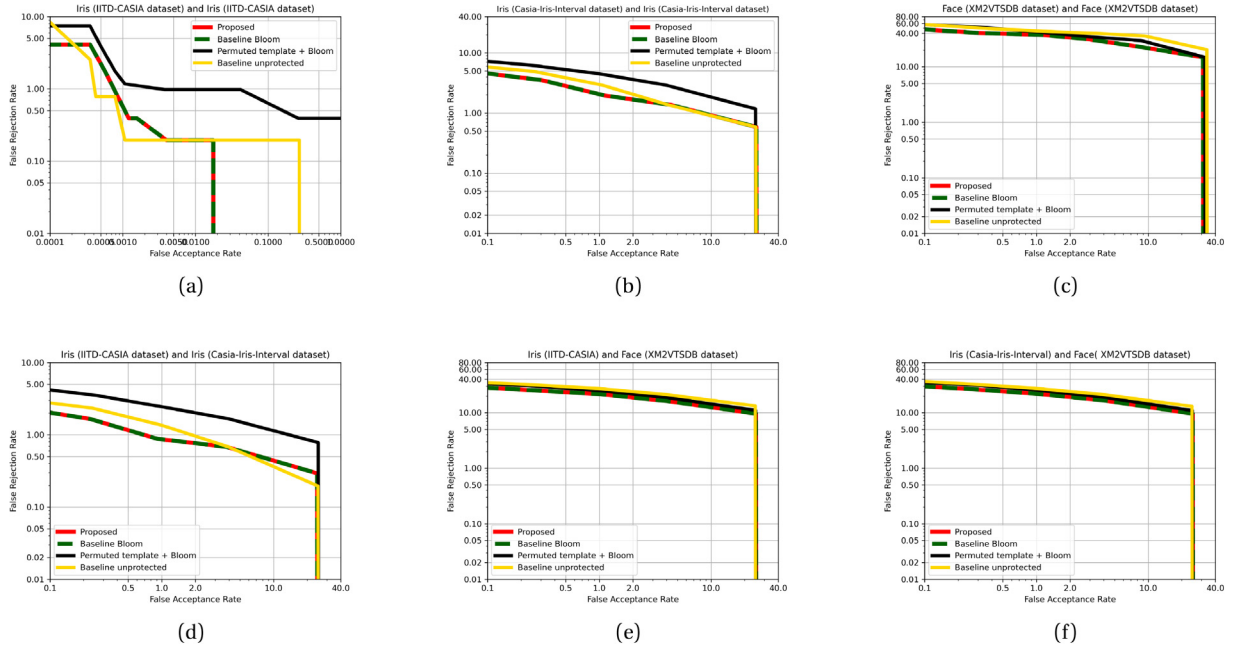
The proposed scheme fulfils irreversibility, unlinkability, and renewability.

### 8.1. Unlinkability

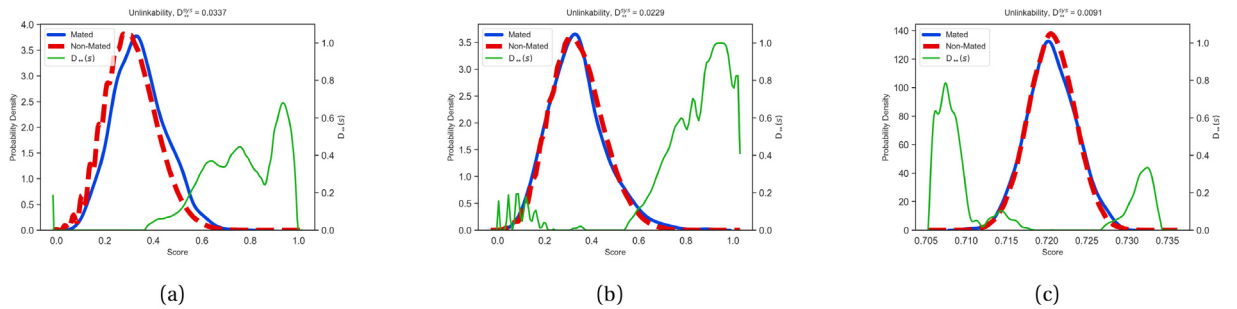
Unlinkability means given different samples, an attacker could not determine if they belong to the same instance or different instances. Unlinkability is computed on Mated and non-Mated samples distribution. Mated samples denote the two samples obtained from the same subject, i.e., enrolled and the genuine samples. The two samples are generated using the two different keys across two different applications. Different keys ensure that one is not able to link if the two samples belong to the same subject or not. Non-mated samples represent the two samples obtained from two different subjects, i.e., the enrolled and the impostor samples were taken across different applications. Here, the different applications use different keys or security parameters for template protection. We calculate the hamming distance based scores between the mated samples to generate mated distribution. Further, hamming distance based scores are calculated between the non-mated samples to generate non-mated distribution. In Fig. 11, the distribution of mated and non-mated scores is shown. We use local measure  $D_{\text{loc}}(s)$  and a global measure  $D_{\text{sys}}$  proposed in Gomez-Barrero et al. (2018) to compute linkability.  $D_{\text{loc}}(s)$  measures linkability on a score-wise basis locally, while  $D_{\text{sys}}$  perform a global analysis of the entire system independent of the score. As shown in Fig. 11, mated and non-mated distributions are significantly overlapped for our proposed scheme. It means, given a score value obtained by comparing any two samples, one cannot deduce whether the score belongs to a mated distribution (the two samples belong to the same user) or a non-mated distribution (the



**Fig. 9.** DET curve for multi-biometric databases – Decision OR Fusion: a) Iris (IITD-CASIA) and Iris (IITD-CASIA) b) Iris (Casia-Iris-Interval) and Iris (Casia-Iris-Interval) c) Face (XM2VTSDB) and Face (XM2VTSDB) d) Iris (IITD-CASIA) and Iris (Casia-Iris-Interval) e) Iris (IITD-CASIA) and Face (XM2VTSDB) f) Iris (Casia-Iris-Interval) and Face (XM2VTSDB).



**Fig. 10.** DET curve for multi-biometric databases – Decision AND Fusion: a) Iris (IITD-CASIA) and Iris (IITD-CASIA) b) Iris (Casia-Iris-Interval) and Iris (Casia-Iris-Interval) c) Face (XM2VTSDB) and Face (XM2VTSDB) d) Iris (IITD-CASIA) and Iris (Casia-Iris-Interval) e) Iris (IITD-CASIA) and Face (XM2VTSDB) f) Iris (Casia-Iris-Interval) and Face (XM2VTSDB).



**Fig. 11.** Unlinkability analysis (Gomez-Barrero et al., 2018): a) Iris (IITD-CASIA) b) Iris (Casia-Iris-Interval) c) Face (XM2VTSDB).

two samples belong to different users).  $D_{\leftrightarrow sys} = 0.0337$  for IITD-CASIA database. For Casia-Iris-Interval database  $D_{\leftrightarrow sys} = 0.0229$  and for XM2VTSDB database  $D_{\leftrightarrow sys} = 0.0091$ . For all cases  $D_{\leftrightarrow sys}$  is close to 0. It shows that unlinkability is well preserved in our proposed scheme.

## 8.2. Renewability

Renewability indicates that if a protected template data is compromised, it should be possible to revoke the existing template and issue a new one. In our proposed work, it is ensured by using a large keyspace. The key size is always  $\geq 128$  bits. Hence renewability in our proposed scheme is ensured.

## 8.3. Irreversibility

Reversibility is required to be achieved by the attacker at two different levels:

1. Reversing the encrypted biometric template  $\bar{B}$  to get the original biometric template  $B$ :  
The original biometric template is encrypted to the encrypted biometric template using format-preserving encryption in our

proposed scheme. The encryption is independently performed on each column of the original template such that a column of height  $h$  (denoted as word size) is encrypted to the corresponding column in the encrypted template with the same height  $h$  as shown in Fig. 5. The attacker can reveal the original biometric template from the encrypted template in two different modes:

**Mode 1: Brute-force attack on the encrypted template:** Considering the case when  $h$  is small (in our case  $h = 10$ ), an attacker guess a single column in the original biometric template in  $2^h$  number of trials.

Further, to guess the complete original biometric template, the number of trials (in terms of bits) would effectively become equal to the number of bits in the original biometric template, i.e., in the case of iriscode would be 10240 bits. In other words, the total number of trials in such a case is given as  $N = 2^{hW}$ , where  $W$  denotes the width of the biometric template (1024 in our case when the height of a single column, also known as word size  $h = 10$ ).

**Mode 2: Guessing system's key  $K$ :** Given an encrypted biometric template, it is computationally infeasible to decrypt it to get the original biometric template without knowing key  $K$ . The key  $K$  can be guessed in  $2^{128}$  trials by the attacker if the encryption is

done by using block cipher modes of operation (Dworkin, 2001). In such a case, an attacker can check the format of the corresponding ciphertext or plaintext to ensure if the guessed key is correct or incorrect. If the guessed key is correct, the corresponding output would be a correctly formatted value. Otherwise, the output would be a pseudorandom value. However, in the case of format-preserving encryption, since the format of plaintext and ciphertext is wholly preserved, an attacker cannot guess a correct key by checking the corresponding ciphertext or plaintext format. Thus, the key-guessing attack is not feasible in our proposed approach, irrespective of the key length. Still, we ensure that the system's key size used during the format-preserving encryption in our proposed scheme is  $|K| \geq 128$  bits which would help provide a larger key space to achieve unlinkability and renewability properties. We have used FF3-1 mode (Dworkin, 2019) of format preserving encryption that is not vulnerable to the attack proposed for FF3 mode due to which the security level of FF3 mode has been reduced below 128 bits for a key size 128 bits.

2. Reversing the Bloom filter based template to get the encrypted template or the original biometric template (in case if key  $K$  is compromised): Given the Bloom filter based template as an output, an attacker would aim to recover the original biometric template or reveal some sensitive information from it. In the case of iris characteristics, the neighbouring iris features have a high correlation, leading to multiple identical columns in the iriscode (Rathgeb et al., 2013). As a result, multiple columns in the 2-D iriscode are mapped to the same index in the Bloom filter for a particular block of iriscode. Now, since the Bloom filter index positions are public, i.e. known to an attacker, the attacker could try to recover the encrypted iriscode by reversing the Bloom filter based mapping. The reconstruction of a block of iriscode requires an arrangement of  $|c| \leq m$  different indexes to a block of length  $m$ , where  $|c|$  denotes the number of bits set to 1 in the Bloom filter array. Using Eq. 1, we get the probability  $P$  to get a single block of input biometric template for a given Bloom filter  $c$  with  $|c|$  active indexes (Gomez-Barrero et al., 2016).

$$P = \frac{1}{\sum_{i=1}^{|c|} (-1)^{|c|-i} \binom{|c|}{i} i^m} \quad (2)$$

As shown in Eq. 2, the success probability of guessing a single block of the encrypted biometric template, given that the protected template or the cancelable template is disclosed to the attacker is almost negligible. Thus, reversibility, even of the individual blocks of the encrypted biometric template is computationally infeasible.

## 9. Conclusions and future work

We introduced a biometric template protection scheme that provides an efficient identification based on the format-preserving encryption and Bloom filter. Format-preserving encryption preserves the column-wise errors in the encrypted template such that the recognition performance of the proposed scheme is equivalent to the baseline Bloom (Rathgeb et al., 2013; Rathgeb et al., 2014) while adding a layer of security over the original scheme. Bloom filter helps to achieve identification over a large dataset. We perform experiments for recognition performance on several datasets. Results show high recognition performance for both uni-biometric and multi-biometric datasets. We achieve 0.2% FRR at 0.01% FAR for IITD-CASIA virtual dataset in a uni-biometric system. Similarly, for a multi-biometric system, on IITD-CASIA virtual dataset, we achieve a 0% FRR at 0.01% FAR for Score level and Decision OR fusion level and 0.23% FRR at 0.01%

FAR for Decision-AND level. Further, the thorough security analysis depicts that the ISO/IEC IS 24745:2011 requirements of unlinkability, irreversibility, and renewability are fulfilled by our proposed scheme. The empirical results and the security analysis show that our proposed scheme accomplishes achieving high security and high performance needed to deploy the biometric systems on a broader scale.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Anjos, A., Günther, M., de Freitas Pereira, T., Korshunov, P., Mohammadi, A., Marcel, S., 2017. Continuously reproducing toolchains in pattern recognition and machine learning experiments. In: International Conference on Machine Learning (ICML). URL: [http://publications.idiap.ch/downloads/papers/2017/Anjos\\_ICML2017-2\\_2017.pdf](http://publications.idiap.ch/downloads/papers/2017/Anjos_ICML2017-2_2017.pdf).
- Anjos, A., Shafey, L.E., Wallace, R., Günther, M., McCool, C., Marcel, S., 2012. Bob: a free signal processing and machine learning toolbox for researchers. In: 20th ACM Conference on Multimedia Systems (ACMMM). Nara, Japan. URL: [https://publications.idiap.ch/downloads/papers/2012/Anjos\\_Bob\\_ACMMM12.pdf](https://publications.idiap.ch/downloads/papers/2012/Anjos_Bob_ACMMM12.pdf).
- Bellare, M., Ristenpart, T., Rogaway, P., Stegers, T., 2009. Format-preserving encryption. In: Jacobson, M.J., Rijmen, V., Safavi-Naini, R. (Eds.), *Selected Areas in Cryptography*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 295–312.
- Chang, D., Garg, S., Ghosh, M., Hasan, M., 2021. Biofuse: A framework for multi-biometric fusion on biocryptosystem level. *Inf. Sci.* 546, 481–511.
- Chang, D., Garg, S., Hasan, M., Mishra, S., 2020. Cancelable multi-biometric approach using fuzzy extractor and novel bit-wise encryption. *IEEE Trans. Inf. Forensics Secur.* 15, 3152–3167.
- Cheon, J.H., Chung, H., Kim, M., Lee, K.W., 2016. Ghostshell: Secure biometric authentication using integrity-based homomorphic evaluations. *IACR Cryptol. ePrint Arch.* 2016, 484.
- Daugman, J., 2003. The importance of being random: statistical principles of iris recognition. *Pattern Recogn.* 36, 279–291.
- Dodis, Y., Reyzin, L., Smith, A., 2004. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *International conference on the theory and applications of cryptographic techniques*, Springer, 523–540.
- Drozdzowski, P., Garg, S., Rathgeb, C., Gomez-Barrero, M., Chang, D., Busch, C., 2018. Privacy-preserving indexing of iris-codes with cancelable bloom filter-based search structures. In: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, pp. 2360–2364.
- Drozdzowski, P., Rathgeb, C., Busch, C., 2018. Bloom filter-based search structures for indexing and retrieving iris-codes. *IET Biometrics* 7, 260–268.
- Durak, F.B., Vaudenay, S., 2017. Breaking the ff3 format-preserving encryption standard over small domains. *Annual international cryptology conference*, Springer, 679–707.
- Dwivedi, R., Dey, S., 2018. Score-level fusion for cancelable multi-biometric verification. *Pattern Recogn. Lett.*
- Dworkin, M., 2001. Recommendation for block cipher modes of operation, methods and techniques. Technical Report. National Inst of Standards and Technology Gaithersburg MD Computer security Div.
- Dworkin, M., 2019. Recommendation for block cipher modes of operation: Methods for format-preserving encryption. Draft NIST Special Publication 800. 38G Revision 1.
- Gomez-Barrero, M., Galbally, J., Rathgeb, C., Busch, C., 2018. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Trans. Inf. Forensics Secur.* 13, 1406–1420. <https://doi.org/10.1109/TIFS.2017.2788000>.
- Gomez-Barrero, M., Maiorana, E., Galbally, J., Campisi, P., Fierrez, J., 2017. Multi-biometric template protection based on homomorphic encryption. *Pattern Recogn.* 67, 149–163.
- Gomez-Barrero, M., Rathgeb, C., Galbally, J., Busch, C., Fierrez, J., 2016. Unlinkable and irreversible biometric template protection based on bloom filters. *Inf. Sci.* 370, 18–32.
- Gomez-Barrero, M., Rathgeb, C., Li, G., Ramachandra, R., Galbally, J., Busch, C., 2018. Multi-biometric template protection based on bloom filters. *Inf. Fusion* 42, 37–50.
- He, M., Horng, S.J., Fan, P., Run, R.S., Chen, R.J., Lai, J.L., Khan, M.K., Sentosa, K.O., 2010. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recogn.* 43, 1789–1800.
- Hoang, V.T., Tessaro, S., Trieu, N., 2018. The curse of small domains: New attacks on format-preserving encryption. *Annual International Cryptology Conference*, Springer, 221–251.
- ISO, 2011. Information technology – Security techniques – Biometric information protection. ISO/IEC 24745:2011(en). International Organization for Standardization.

- Jegede, A., Udzir, N.I., Abdullah, A., Mahmod, R., 2017. Cancelable and hybrid biometric cryptosystems: current directions and open research issues.
- Jin, A.T.B., Ling, D.N.C., Goh, A., 2004. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recogn.* 37, 2245–2255.
- Juels, A., Sudan, M., 2006. A fuzzy vault scheme. *Des. Codes Crypt.* 38, 237–257.
- Juels, A., Wattenberg, M., 1999. A fuzzy commitment scheme. In: *Proceedings of the 6th ACM conference on Computer and communications security*, pp. 28–36.
- Kumar, A., Passi, A., 2010. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recogn.* 43, 1016–1026.
- Li, C., Hu, J., Pieprzyk, J., Susilo, W., 2015. A new biocryptosystem-oriented security analysis framework and implementation of multibiometric cryptosystems based on decision level fusion. *IEEE Trans. Inf. Forensics Secur.* 10, 1193–1206.
- Masek, L., 2003. Recognition of human iris patterns for biometric identification.
- Nagar, A., Nandakumar, K., Jain, A.K., 2011. Multibiometric cryptosystems based on feature-level fusion. *IEEE Trans. Inf. Forensics Secur.* 7, 255–268.
- Nandakumar, K., Jain, A.K., Pankanti, S., 2007. Fingerprint-based fuzzy vault: Implementation and performance. *IEEE Trans. Inf. Forensics Secur.* 2, 744–757.
- Nandakumar, K., Jain, A.K., 2015. Biometric template protection: Bridging the performance gap between theory and practice. *IEEE Signal Process. Mag.* 32, 88–100.
- Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., Galbally, J., Freire, M.R., Gonzalez-Rodriguez, J., Garcia-Mateo, C., Alba-Castro, J.L., Gonzalez-Agulla, E., Otero-Muras, E., et al., 2009. The multiscenario multienvironment biosecure multimodal database (bimdb). *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1097–1111.
- Othman, N., Dorizzi, B., Garcia-Salicetti, S., 2016. Osiris: An open source iris recognition software. *Pattern Recogn. Lett.* 82, 124–131.
- Pagnin, E., Mitrokovtsa, A., 2017. Privacy-preserving biometric authentication: challenges and directions. *Secur. Commun. Networks*.
- Patel, V.M., Ratha, N.K., Chellappa, R., 2015. Cancelable biometrics: A review. *IEEE Signal Process. Mag.* 32, 54–65.
- Rathgeb, C., Breiting, F., Busch, C., 2013. Alignment-free cancelable iris biometric templates based on adaptive bloom filters. In: *2013 international conference on biometrics (ICB)*, IEEE, pp. 1–8.
- Rathgeb, C., Breiting, F., Busch, C., Baier, H., 2014. On application of bloom filters to iris biometrics. *IET Biometrics* 3, 207–218.
- Rathgeb, C., Merkle, J., Scholz, J., Tams, B., Nesterowicz, V., 2021. Deep face fuzzy vault: Implementation and performance. *arXiv preprint arXiv:2102.02458*.
- Rathgeb, C., Uhl, A., 2010. Adaptive fuzzy commitment scheme based on iris-code error analysis. *European Workshop on Visual Information Processing (EUVIP)*, IEEE, pp. 41–44.
- Rathgeb, C., Uhl, A., 2011. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP J. Inf. Secur.* 2011, 1–25.
- Rathgeb, C., Uhl, A., Wild, P., Hofbauer, H., 2016. Design decisions for an iris recognition sdk. *Handbook of Iris Recognition*. Springer, 359–396.
- Regulation, G.D.P., 2016. Regulation (eu) 2016/679 of the european parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. *Official Journal of the European Union (OJ)* 59, 294.
- Rivest, R.L., Adleman, L., Dertouzos, M.L., et al., 1978. On data banks and privacy homomorphisms. *Found. Secure Comput.* 4, 169–180.
- Ross, A., Poh, N., 2009. Multibiometric systems: Overview, case studies, and open issues. *Handbook Remote Biometrics*, 273–292.
- Sudhamani, M., Venkatesha, M., Radhika, K., 2014. Fusion at decision level in multimodal biometric authentication system using iris and finger vein with novel feature extraction. In: *2014 Annual IEEE India Conference (INDICON)*, IEEE, pp. 1–6.
- Yasuda, M., Shimoyama, T., Kogure, J., Yokoyama, K., Koshiba, T., 2013. Packed homomorphic encryption based on ideal lattices and its application to biometrics. *International Conference on Availability, Reliability, and Security*, Springer, 55–74.
- Zhou, K., Ren, J., 2018. Passbio: Privacy-preserving user-centric biometric authentication. *IEEE Trans. Inf. Forensics Secur.* 13, 3050–3063.
- Zuo, J., Ratha, N.K., Connell, J.H., 2008. Cancelable iris biometric. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, IEEE, pp. 1–4.