Uncoupling Biometrics from Templates for Secure and Privacy-Preserving Authentication

Aysajan Abidin, Enrique Argones Rúa, and Roel Peeters
imec-COSIC KU Leuven, Belgium
firstname.lastname@esat.kuleuven.be

ABSTRACT

Biometrics are widely used for authentication in several domains, services and applications. However, only very few systems succeed in effectively combining highly secure user authentication with an adequate privacy protection of the biometric templates, due to the difficulty associated with jointly providing good authentication performance, unlinkability and irreversibility to biometric templates. This thwarts the use of biometrics in remote authentication scenarios, despite the advantages that this kind of architectures provides. We propose a user-specific approach for decoupling the biometrics from their binary representation before using biometric protection schemes based on fuzzy extractors. This allows for more reliable, flexible, irreversible and unlinkable protected biometric templates. With the proposed biometrics decoupling procedures, biometric metadata, that does not allow to recover the original biometric template, is generated. However, different biometric metadata that are generated starting from the same biometric template remain statistically linkable, therefore we propose to additionally protect these using a second authentication factor (e.g., knowledge or possession based). We demonstrate the potential of this approach within a two-factor authentication protocol for remote biometric authentication in mobile scenarios.

CCS CONCEPTS

•Security and privacy → Cryptography; Biometrics; Multi-factor authentication; Privacy-preserving protocols;

KEYWORDS

Biometrics; multi-factor authentication; template protection; unlinkability; irreversibility;

ACM Reference format:
Aysajan Abidin, Enrique Argones Rúa, and Roel Peeters. 2017. Uncoupling Biometrics from Templates for Secure and Privacy-Preserving Authentication. In *Proceedings of SACMAT'17, Indianapolis, IN, USA, June 21-23, 2017,* 9 pages.
DOI: http://dx.doi.org/10.1145/3078861.3078863

SACMAT'17, June 7-10, 2017, *Indianapolis, IN, USA*

978-1-4503-4702-0/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3078861.3078863

1 INTRODUCTION

Biometrics is a way to measure personal features, having the potential to authenticate individuals with a high degree of assurance while offering convenience to the user. Consequently, the use of automated biometric-based frameworks has become increasingly popular in both governmental and commercial services for user authentication. However, the use of biometrics also poses serious threats both to privacy and security of the users. It was shown by Pagnin *et al.* [25] that leakage of biometric data in a remote setting is hard to avoid. Leakage of biometric data may lead to the disclosure of personal information, e.g., demographics (for instance age, as shown by Han *et al.* [14]) or medical information (for instance as shown by Bolling [8] and Penrose [22] for the case of iris and fingerprint patterns, respectively). Furthermore, the leaked biometric data may be used to create spoofed biometric samples, thus thwarting the secure use of biometrics for authentication. These difficulties have slowed down the adoption of biometrics in remote authentication schemes, where the risks of information leakage are high.

The main goal of designing a biometric protection scheme is to provide irreversibility and unlinkability, as defined in the ISO/IEC 24745:2011 standard [16]. Irreversibility means that it should be computationally hard to reconstruct the original biometric template from the stored reference data, i.e., the protected biometric template, while it should be easy to generate the protected biometric template. Unlinkability means that different versions of protected biometric templates can be generated based on the same biometric data (renewability), while protected biometric templates should not allow cross-matching (diversity).

In literature, several approaches were proposed for protecting biometric data, including the use of biometric template protection schemes, such as those based on fuzzy extractors [12], or through the use of homomorphic encryption schemes. However, most of the proposed biometric template protection schemes suffer from degradation of verification performance, partial reversibility or linkability. Recently, research moved towards a new direction, taking the wider context of authentication into account and leverage multi-factor user authentication. In multi-factor user authentication, the user is verified by a combination of different authentication factors that belong to one of the following categories: (a) *possession factor*, e.g., a token stored on the user's mobile phone; (b) *knowledge factor*, e.g., a PIN or password that the user remembers; and (c) *inherent factor*, e.g., a biometric attribute. This new direction does not only result in increased security, since the user has to provide more than one type of authentication evidence, but it can also be used for improving the privacy protection of biometric data.

In this work, we introduce the first sound procedure to decouple the biometrics from their binary representation *before* using

biometric protection schemes. Furthermore, we build further on the latest direction in biometric protection research by proposing a privacy-preserving and secure two-factor authentication protocol for remote scenarios.

1.1 Related Work

Recently, biometric protection research is moving into the direction of combing biometrics with other (knowledge-based or possession-based) authentication factors. Given the shear amount of related work, we limit ourselves to the most relevant related works.

Zheng [32], Zhu [33] and Wua *et al.* [31] presented different approaches that rely on the encryption of the biometric template. However, these schemes need to decrypt the biometric template, which is solely protected by means of the encryption, and do not deal with unlinkability of template.

Syta *et al.* [29] proposed a combination of two factors, but the privacy of biometrics would be disclosed if the token is obtained by the authentication server. Fan *et al.* [13] designed a privacy-preserving scheme using three-factor authentication. However, biometric unlinkability is not achieved if the token is disclosed. Meenakshi and Padmavathi [23] suggested a password-hardened fuzzy vault. The password is directly combined with the biometric binary representation and a uniform distribution of the samples is assumed in order to avoid any information disclosure. Nevertheless, the disclosure of the password would expose the biometrics, and renewability is only achievable by continuously changing the user's password, which is not a real biometric renewability. More recently, Abidin *et al.* [1, 2, 4] proposed privacy-preserving biometric authentication protocols secure against malicious (as opposed to semi-honest) adversaries. These protocols also use a second authentication factor which is used to enhance the privacy of biometric templates, but do not consider revocability and/or renewability of biometric templates.

Abidin *et al.* [3] presented a modification of an earlier protocol presented by Bringer *et al.* [9], which relies on homomorphic encryption. In their modified version, the authors improve the protocol to achieve security against malicious but not colluding insider adversaries, utilising additional secret keys. As in the original protocol, the Abidin *et al.* protocol also stores the reference biometric templates in the clear. Although this approach is secure against malicious adversaries, it comes with more complexity and requires the storage of additional cryptographic keys, and it only protects the link between biometric data and their original owners.

BioHash [17], introduced by Andrew Teoh Beng *et al.*, was a first attempt at decoupling the biometric representation from the original biometric sample. Even though this feature transformation approach asymptotically preserves distances between genuine and impostor biometric samples in the original and transformed domains, the statistical assumptions on the biometric sources (intra-class and inter-class scatter matrices must be known) and the size of the obtained metadata can thwart its practical adoption. Moreover, the use of the stored information (the transformed template) and the metadata (can be stored locally) allows for reconstruction of the biometric representation, posing a serious threat to privacy and security.

Monrose *et al.* [24] presented an approach for generating long-term secret keys from passwords and keystroke dynamics. This approach is similar to ours in the sense that it encrypts a metadata table with the password. The approach proposed by Monrose *et al.* only takes into account the most reliable biometric features provide biometric-dependent information for reconstructing the secret key and it is not taking advantage of the information provided by features with low reliability, which makes the system not useful when the biometric trait exclusively provides a huge number of unreliable feature. The scheme is adaptive in the sense that the changes in the reliability of the features are smoothly taken into account. However, in order to provide the same secret key, all the reliable features have to agree, which results in a trade-off between a high entropy of the secret key and a low false rejection rate. Moreover, the stored metadata related to non-reliable features contains information that can be used to reconstruct the secret key even without access to an original biometric sample.

1.2 Contributions

We propose a procedure that decouples the original biometric samples from their binary representation and use this result to construct a new biometric protection method for two-factor authentication in remote settings. Furthermore, the security and privacy properties of the resulting method are formally analysed. Concretely, our contributions are:

- In Sect. 2, we present our adversarial model and security definitions. We assume malicious external adversaries for security and even allow for an malicious Service Provider (the party where the user authenticates to) for privacy. This adversarial model is stronger than the ones commonly used in related work, i.e. honest-but-curious.
- Our proposed biometric decoupling procedure is presented in Sect. 3, with a concrete example for the case of IrisCodes. The choice of iris biometrics for a mobile user authentication application is motivated by the existence of several commercial and academic systems using this modality (e.g., Wang and Liu [30]) and the camera as a biometric capture device being generally and openly available. The latter in contrast with other commonly used mobile biometrics such as fingerprints, where the biometric capture behaviour is fixed, and its security usually relies on industrial secrecy [5, 26]. We also discuss the application of our biometric decoupling procedure to other biometric modalities in Sect. 3.2, including behavioural and dynamic biometrics. The proposed decoupling procedure provides increased reliability of the binary features by using user-specific information, thus improving security and biometric authentication performance. It also enables us to choose an arbitrary binary representation for the biometrics, thus facilitating unlinkability.
- The feasibility of the proposed decoupling procedure in the context of remote authentication is shown in Sect. 4, by presenting a protocol for secure and privacy-preserving remote user authentication, based on the biometric treat and a second factor based on either knowledge or possession. The metadata, generated by the biometric decoupling

procedure, is encrypted using a cryptographic key, which is derived from an independent authentication factor to avoid linkability between metadata. It should also be noted that we do not rely on homomorphic encryption (such as, e.g., [9] and [3]), resulting in reduced complexity.
- In Sect. 5, we show that the proposed protocol is secure against malicious external adversaries and that the overall construction provides privacy both against malicious Service Provider and external adversaries.

2 BACKGROUND

Cryptographic schemes relies on keys, which are chosen uniformly at random and then remain fixed. This means that, in order to use biometrics as keys for cryptographic primitives, one first needs to transform these inherently noisy sources into a stable string which is indistinguishable from a random distribution. A common approach for doing this is using *fuzzy extractors*. A fuzzy extractor is a construction that allows to characterise noisy information sources with fixed random strings. Juels [18] presented a practical construction, which is based on the use of Error Correcting Codes, PseudoRandom Generators and Hash functions. A formal definition can be found in [12]. In our case, we use a binary Error Correcting Code ECC $(n, k, t) \subset \{0,1\}^n$, where n is the length of the codewords, k is the length of messages, $k < n$, and t is the number of errors that can be corrected in the received codewords. The two associated functions are denoted as $\text{ECC}_{\text{encode}}$ and $\text{ECC}_{\text{decode}}$.

For the proposed protocol we make use of a secure key derivation function (KDF) to derive cryptographic keys using a secret input data, e.g., a password, with sufficient min-entropy. In particular, we use a KDF to generate keys for an IND-CPA-secure symmetric encryption and a strongly-unforgeable digital signature scheme. We denote by ENC and DEC the symmetric key encryption and decryption algorithms, and by SIGN and VER the public key digital signature signing and verification algorithms, respectively.

Definition 2.1 (Adapted from [21]). A key derivation function is said to be secure with respect to a source of input with a min-entropy greater than or equal to λ if no probabilistic polynomial time (PPT) attacker $\mathcal{A}$ can distinguish its output from a random output of equal length, except with a non-negligible probability $\text{negl}(\lambda)$, where negl is a negligible function.

2.1 Adversarial model

We consider two types of adversaries: adversaries that aim to break security and those that aim to break privacy. We assume that the adversary is in full control of all communication between Service Providers (SP) and user devices (D) and can hence eavesdrop, modify, re-order, replay, inject and drop messages at will. We assume that the user and the user's device are fully trusted, i.e., users abide by the protocol specifications when authenticating as themselves and trust their own device only in the sense that the device is not compromised.

The security adversary will try to impersonate an uncompromized user to SP potentially having access, in addition to all protocol transcripts, to all available input of parties involved with the exception of at least one of the authentication factors of the legitimate user it is trying to impersonate.

Definition 2.2 (Security). Π = (Enroll, Authenticate) is a secure multi-factor authentication system if no PPT adversary $\mathcal{A}$ can successfully authenticate itself to the verifier as the legitimate user it impersonates, even when given all protocol transcripts and all inputs of the verifier and all provers with the exception of at least one authentication factor of the user it tries to impersonate.

The privacy adversary will try to link users across enrollments (possibly at different possibly colluding service providers) for which we only consider the linkage of information that is derived from the user's biometric template Q as this cannot (easily) be changed (as opposed to other knowledge-based and possession-based authentication factors). The adversary provides two biometric templates, Q_0 and Q_1, from which one will be used for enrolling user U, after which the adversary can authenticate poly-many times as this user using any input as the other authentication factor(s). We do not consider privacy in the sense that multiple authentication attempts by the same user (for the same enrollment) might be linked.

Definition 2.3 (Privacy). For Π = (Enroll, Authenticate) as before, consider the following game played between a PPT adversary $\mathcal{A}$ and a challenger:

$\text{Exp}^{\text{Priv}}_{\Pi,\mathcal{A}}(\lambda)$:
$(Q_0, Q_1) \leftarrow \mathcal{A}(\lambda)$
$b \xleftarrow{R} \{0, 1\}$
$\alpha \leftarrow$ Enroll(U, Q_b, $auth\_factor^*$), with
α the storage at the verifier and * meaning 1 or more
For $i = 1, \cdots, \text{poly}(\lambda)$:
$\beta_i, \leftarrow$ Authenticate(U, Q_b, $auth\_factor^*$), with
β_i the protocol transcript i
$b' \leftarrow \mathcal{A}(\lambda, Q_0, Q_1, \alpha, (\beta_i)_{i=1}^{\text{poly}(\lambda)})$
Return 1 if $b == b'$, 0 otherwise

We define the adversary's advantage in this game as

$$\text{Adv}^{\text{Priv}}_{\Pi,\mathcal{A}}(\lambda) = \left| \Pr\left\{ \text{Exp}^{\text{Priv}}_{\Pi,\mathcal{A}}(\lambda) = 1 \right\} - \frac{1}{2} \right|.$$

Π is privacy-preserving if $\text{Adv}^{\text{Priv}}_{\Pi,\mathcal{A}}(\lambda) \leq \text{negl}(\lambda)$, for all PPT adversary $\mathcal{A}$.

3 DECOUPLING BIOMETRIC REPRESENTATIONS

The fuzzy commitment approach does not provide any mechanism for guaranteeing unlinkability of the protected templates. It is easy to determine that two templates have been generated using the same biometric sample, as shown by Simoens et al. in [27], therefore this cryptosystem does not provide unlinkability when the secret (the binary biometric representation) must remain the same.

Furthermore, the number and robustness of biometric features are not usually appropriate for being protected using a fuzzy commitment scheme, and this is a critical issue. On one hand, the fuzzy commitment scheme needs that the number of protected bits coincides with the length of the used error correcting code. And on the other hand, the reliability of these bits (the probability to remain unchanged among different biometric samples from the same individual) determines the required error correcting capability of the code: the higher this reliability, the lower the required error correcting capability, and the higher the message length, or equivalently the security parameter of the fuzzy commitment scheme.

We propose to decouple the binary representation from the biometric data, producing a target binary representation with increased

robustness from the biometric features, thus circumventing these problems in the fuzzy commitment. The target binary representation can be chosen during the enrollment stage, making possible that the final binary string is completely independent from the biometric data, thus guaranteeing unlinkability between protected templates generated from the same biometric features. This is achieved by using two intermediate functions. The first one takes as inputs the biometric enrollment samples and the desired random binary representation, and returns a set BM of binarization metadata as output, $BM = f(Q^E, b)$.

This is the binarization metadata extractor function, and must be used during the enrollment phase. The second function takes as inputs the biometric verification samples and the set of binarization metadata produced by the first function. Its output is the binary representation of the verification biometrics. This intermediate function is called the parameterised binarization function, defined as $\widetilde{\mathbf{b}} = g(Q^v, BM)$.

It is important to underline that the binarization metadata establishes a relationship between the biometric samples and the binary representation, therefore this binarization metadata have to be protected in order to avoid a leak of information. As a remark, in this work we do not present a general approach for computing these functions, since the paper is focused on the possibilities that such an approach offers for designing authentication protocols providing both the highest security and privacy protection. However, in the next section an example of these two functions for the IrisCode case is illustrated, and the application of this approach to other biometric modalities is then discussed.

3.1 Specific Case: IrisCode

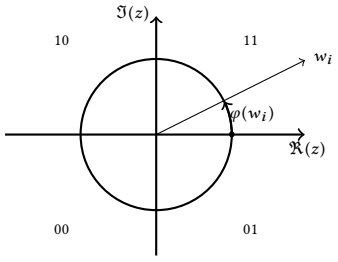

Figure 1: Binary encoding of phase information in IrisCode.

3.1.1 Feature extraction. In the specific example of iris, bits conforming the Daugman's IrisCode [10] encode the phase of complex Gabor wavelets' responses as a vector $\mathbf{w} = [w_1, ..., w_N]^t$. Biometric measurements can be written before the binary encoding step as a vector of phase measurements:

$$\begin{aligned} Q = \boldsymbol{\varphi}(\mathbf{w}) &= [\varphi(w_1), ..., \varphi(w_N)]^t \\ &= [\varphi_1, ..., \varphi_N]^t, \end{aligned}$$

where $\varphi_i = \arctan(\Im(w_i)/\Re(w_i))$, $i \in \{1, ..., N\}$. Two bits are extracted from each phase value, as portrayed in Fig. 1, and formally defined by the following IrisCode coding function,

$$\begin{aligned} \mathbf{b_i} &= \mathrm{IC}(\varphi_i) \\ &= \begin{cases} [1,1]^t, \text{ if mod}\{\varphi_i, 2\pi\} \in [0, \pi/2) \\ [1,0]^t, \text{ if mod}\{\varphi_i, 2\pi\} \in [\pi/2, \pi) \\ [0,0]^t, \text{ if mod}\{\varphi_i, 2\pi\} \in [\pi, 3\pi/2) \\ [0,1]^t, \text{ if mod}\{\varphi_i, 2\pi\} \in [3\pi/2, 2\pi). \end{cases} \end{aligned}$$

3.1.2 Proposed randomized binarization process. The bit values $\mathbf{b_i}$ can be decoupled from the corresponding biometric phase measurement φ_i by using shifted versions $\phi_i = \varphi_i + \theta_i$ in the binarization in Eq. 3.1.1, where θ_i are phase shift terms. We could simply choose these phase shift terms for each phase measurement as:

$$\theta_i = k_i \frac{\pi}{2}, \tag{1}$$

with $k_i \in \{0, \ldots, 3\}$, obtaining a binarization metadata $BM = [\theta_1, \ldots, \theta_N]$ completely independent from the original biometric representation Q. However, phase measurements near the decision thresholds will provide low reliability, i.e., new biometric samples from the same individual could produce different bits. A further improvement can be done for increasing the reliability (resilience to change) of the bits in the resulting template, making the phase $\phi_i = \varphi_i + \theta_i$ to be centred in its corresponding quadrant, i.e. $\phi_i = k_i\pi/2 + \pi/4 + j(\pi/2)$, with $k_i \in \{0, \ldots, 3\}$ and $j \in \mathbb{Z}$. Let us define $j_{\varphi_i} = \text{minarg}_{j \in \mathbb{Z}} \{|j(\pi/2) + \pi/4 - \varphi_i|\}$. Then, the phase shifts producing robust binary features can be defined as:

$$\theta_i = \text{mod}\left\{k_i \frac{\pi}{2} + \left[\frac{\pi}{4} - \varphi_i + j_{\varphi_i} \frac{\pi}{2}\right], 2\pi\right\}, \tag{2}$$

with k_i chosen from the set $\{0, \ldots, 3\}$.

Given a binary representation for the i-th phase term $\mathbf{b_i} = [b_i^0, b_i^1]^t$, we remind that it holds the IrisCode binarization, i.e. $\mathbf{b_i} = \mathrm{IC}(\phi_i) = \mathrm{IC}(\varphi_i + \theta_i)$. Therefore, the k_i indexes can be computed as $k_i = \{k \in \{0, \ldots, 3\} \mid \mathbf{b_i} = \mathrm{IC}(\varphi_i + \theta_i)\}$, with θ_i computed using Eq. 1, or Eq. 2 for increased binary reliability.

The complete *metadata extractor function* is defined as the randomised function $BM = f(Q, \mathbf{b}) = [\theta_1, ..., \theta_N]^t$, where θ_i is computed using Eq. 1, or alternatively Eq. 2. For the binarization of a biometric phase measurement $Q' = [\varphi_1, \ldots, \varphi_N]$, the phase shifts stored in BM are applied to Q'. This is done by the *binarization function*, defined as $g(Q', BM) = [IC(\varphi'_1 + \theta_1)^t, ..., IC(\varphi'_N + \theta_N)^t]^t$. It can be checked that $g(Q, f(Q, \mathbf{b})) = \mathbf{b}$, in both cases when Eq. 1 or Eq. 2 are used for calculating the phase shift terms.

The phase shifts obtained using this approach carry information about the original representation, since $\text{mod}\{\theta_i, \pi/2\}$ is the distance of the original phase measurement to the closest phase bin centre. However, this metadata does not carry information about the original bin, and therefore it does not pose any risk to security, keeping a 2-bit uncertainty on the corresponding biometric phase. As long as these metadata are encrypted, the privacy risk posed by the distances to phase bin centre can be assumed for the sake of both (a) improved feature reliability, and thus increased security parameter in the biometric template protection scheme and improved biometric authentication performance, and (b) protected template unlinkability.

Further increasing robustness. The procedure described in the previous section can be modified to also reduce the binary representation length n and further improve binary features reliability. This length must not depend on the number of biometric features N. Instead, n can be imposed by other criterion, such as available code lengths when using a fuzzy commitment scheme. In the previously explained procedure, $n = 2N$. We present here how to proceed for $n < 2N$. Some biometric features will be contributing to the same couple of bits, thus increasing the reliability of these bits. This is specially useful in the case of iris, where occlusions of large parts of the iris pattern is a very usual situation.

Let us define the set of biometric features contributing to a given couple of bits $\mathbf{b}_i$ as $\mathcal{F}_i = \{\varphi_{i_1}, \ldots, \varphi_{i_{|\mathcal{F}_i|}}\}$, with $i_j \in \{1, \ldots, N\}$, and $j \in \{1, \ldots, |\mathcal{F}_i|\}$. Then, all the corresponding phase shifts are calculated using the same target bit values $\mathbf{b}_i$, i.e.

$$k_{i_j} = \left\{k \in \{0, \ldots, 3\} \mid \mathbf{b_i} = \mathrm{IC}\left(\varphi_{i_j} + \theta_{i_j}\right)\right\},$$

and θ_{i_j} is calculated using Eq. 1 or Eq. 2 alternatively. The sets describing which phase measurements contribute to each bit should be mutually exclusive to avoid undesired correlations between bits, and become part of the binarization metadata:

$$BM = \{\{\mathcal{F}_1, \ldots, \mathcal{F}_n\}, \boldsymbol{\theta}\}.$$

Describing an exact procedure for partitioning the phase measurements into these sets in an optimal way is out of the scope of this paper, though we provide here an intuition. This partition must be aimed at minimising the mean number of bit errors in the verification phase for genuine users. Therefore, it must take into account the following factors:

- Inter-session user-dependent distribution of the biometric features. This will allow for designing partitions where the resultant bits are equally reliable.
- Feature occlusion model, to distribute the occluded features in an uniform way through all the bits.

In the case of the binarization function, it relies on the available biometric measurements in each set. If the user-dependent distribution model assumes the same inter-session noise distribution and independence between the phase features, this function can simply be the arithmetic mean.

3.2 Application to other biometric modalities

The derivation of the parameterised binarization and metadata extractor uncoupling functions depends on the nature of the biometric features. Similar derivations to the one presented in the previous section could be done for other biometric modalities, as long as biometric features are presented in a fixed-length vectorial form. In general, the proposed approach can be applied to any modality where a fuzzy extractor can be derived, with the advantage that user-specific information can be easily integrated in the uncoupling process, as shown in the iris case, for increasing biometric authentication performance. This covers most of the biometric modalities:

(1) Face recognition, where textural descriptors and eigenspace representations are usually fixed-length. Examples of the successful application of fuzzy extraction schemes to face biometrics can be found for instance in [19] and [28].
(2) Speaker recognition, where the state-of-the-art i-vector representation presented by Dehak et al. in [11] is also a fixed-length vector representation. Feasibility of fuzzy extraction for the speech modality has been shown by Billeb et al. in [7].
(3) Dynamic biometrics, such as online signature recognition, where fixed-length representations can be obtained using eigen-model representations and successfully used for building fuzzy extractors, as shown by Argones Rúa et al. in [6]. Other fixed length descriptions for dynamic biometrics, such as the global features presented by Ibrahim et al. in [15] for online signature, or the ones used by Monrose et al. in [24] for keystroke dynamics, are also well suited.
(4) Textural-based fingerprint recognition, such as the scheme presented by Khalil et al. in [20].

4 USER AUTHENTICATION

We now present our two-factor secure and privacy-preserving user authentication system, consisting of an enrollment and an authentication protocol. One factor is the biometric data, from which we extract an uncoupled random binary representation as explained in the previous section. The second factor is used to protect the binarization metadata. The random biometric binary representation is transformed into a signing key sk with a corresponding verification key pk. Authentication is done by signing a challenge from the verifier. Intuitively, our authentication system is secure because in order to impersonate a legitimate user, an attacker needs to correctly sign the challenge, where the signing key is protected by both factors.

Authentication takes place between a user through his or her device and a service provider. The device mainly acts as a proxy for the user, being able to do the necessary computations and setting up communication with the service provider. The protocols are designed such that no storage is required on the device. This has two major benefits: (1) users can use any trusted device to authenticate to the service provider, and (2) losing the device does not lead to security or privacy issues as there is no secret information stored. On the downside, it is generally acknowledged that users tend to choose passwords with low min-entropy. This has an impact on the maximal achievable security and privacy, where security can be maintained at the same level as long as the adversary has no access to the biometric data of the user, the privacy-preserving properties of the system go down to the min-entropy of the password. Another reason for opting for a possession-based second factor is user convenience as the user is not required to type the password. For this reason, we leave both options open to which second authentication factor to use.

4.1 Enrollment Protocol

The enrollment protocol is illustrated in Fig. 2 and involves the following:

- The device asks the user to provide a (unique) username for the service provider and in a second phase a set of biometric samples together with a second authentication factor. In case of a possession-based second authentication factor, this could be generated by the device itself.

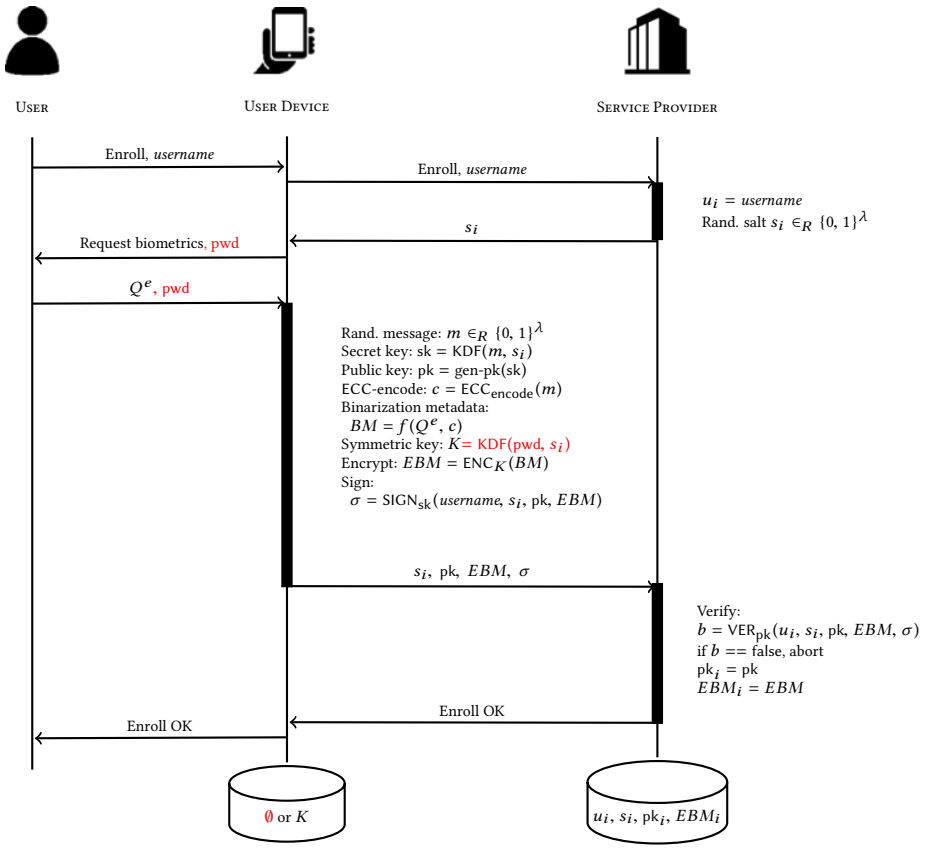

Figure 2: Two factor user enrollment protocol.

- The service provider supplies the device with a unique salt for the given username. This salt will be used for deriving both the symmetric key (when having a knowledge-based second authentication factor) and the signing key, ensuring that there is sufficient randomness as input for the key derivation function for both keys. It also ensures that if the database of the server gets compromised, one cannot efficiently precompute the symmetric keys derived from the most frequently used passwords.
- The device generates a random message that serves as input to the KDF for generating the signing key and corresponding public verification key. To ensure that one can later-on recover this message from the randomised binary representation, it is first encoded with an error correcting code.
- The binarization metadata is generated and protected by a symmetric key encryption scheme to ensure the privacy of the user.
- The device sends the public key and the encrypted binarization metadata back to the service provider, together with a signature over these data and the username. With this signature, the device proves knowledge of the secret key corresponding to the public key. The signature also effectively ties the entire protocol transcript together, ensuring matching conversations between device and service provider. The salt being a unique identifier, is also used as session identifier for the service provider to link the messages.

4.2 Authentication protocol

The verification protocol is illustrated in Fig. 3 and involves the following:

- The device asks the user to provide his or her username for the service provider and in a second phase a fresh biometric sample together with a second authentication factor. In case of a possession-based second authentication factor, the user is only asked for a biometric sample.
- The service provider supplies the device with the salt, encrypted binarization metadata for the given username. Additionally the service provider also supplies a challenge to the device to ensure freshness of the authentication.
- The device reconstructs the message that was chosen at random during the enrollment protocol from the received inputs. Thanks to the error correcting code, one can decode errors due to the biometric sample being slightly different from the ones supplied during the enrollment protocol.

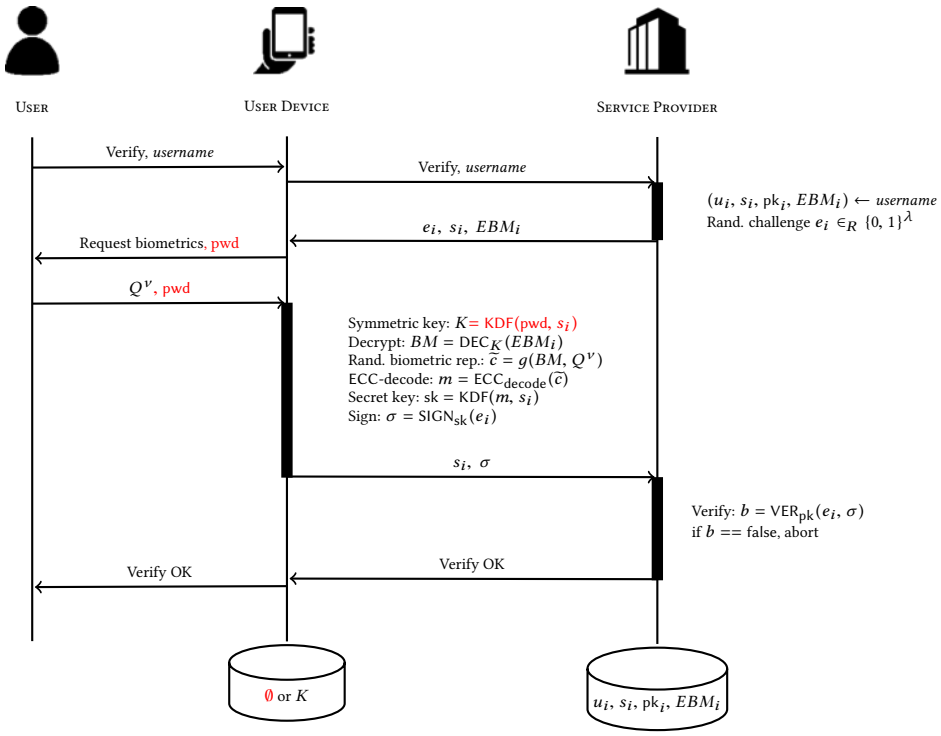

Figure 3: Two-factor user authentication protocol.

From the decoded message the signing key is derived as in the enrollment protocol.

- The device signs the service provider's challenge and sends the signature back to the service provider. Again the salt is used as a session identifier.
- If the signature verifies with the stored public key, the user is authenticated.

5 SECURITY AND PRIVACY ANALYSIS

We show that the proposed scheme is secure and privacy-preserving. The security and privacy analysis are performed for the worst case, i.e. assuming a knowledge-based second authentication factor with min-entropy $\ell << \lambda$. When using a possession-based second authentication factor, $\ell = \lambda$.

5.1 Security

Theorem 5.1 (Security). *The proposed user authentication scheme is a secure two-factor authentication scheme in the security parameter ℓ according to Definition 2.2, assuming a secure KDF, IND-CPA-secure encryption scheme and an universal unforgeable signature scheme.*

Proof. The adversary $\mathcal{A}$ is given a complete copy of the server's database and one authentication factor of the user it tries to impersonate. The proof consists of two cases: in the first $\mathcal{A}$ is given the user's password pwd while in the second it is given the user's biometric feature vector Q^v.

The adversary succeeds in impersonating a user if it can produce a valid signature on a given challenge. Without knowledge of the user's signing key sk, this implies $\mathcal{A}$ breaking universal unforgeability of the signature scheme of which the probability of success is negl(λ). We will now show for both cases, that the adversary cannot recover the message m and hence not sk. m can be recovered using the binarization metadata BM_i and a valid biometric data Q^v by means of the function g and the error-correction code.

- *Case 1:* Given the user's password pwd, $\mathcal{A}$ can decrypt EBM to obtain the binarization metadata BM. However, without knowledge on the biometric data Q^v, $\mathcal{A}$'s probability to recover $\widetilde{c}$, and hence m, is negl(λ).
- *Case 2:* Given the user's biometric data Q^v, $\mathcal{A}$ needs BM to recover m. Without knowledge of the secret key K, this implies $\mathcal{A}$ breaking IND-CPA security of the encryption scheme of which the probability of success is negl(λ). The KDF is secure with respect to the source min-entropy, being ℓ. Hence $\mathcal{A}$'s probability of success of recovering K, hence BM and m, is negl(ℓ).

□

5.2 Privacy

Theorem 5.2. *The proposed user authentication scheme is privacy-preserving in the security parameter ℓ according to Definition 2.3, assuming a secure KDF and IND-CPA-secure encryption scheme.*

Proof. The proof is based on a series of hybrid games.

- Game$_0$: This is the original privacy game $\text{Exp}^{\text{Priv}}_{\Pi,\mathcal{A}}(\ell)$. Let X_0 be the event that $b' = b$ in this game.

- Game_1: This is the same as Game_0, except that now the symmetric encryption key generated from $\text{KDF}(\text{pwd}, s_b)$ is replaced by a uniformly distributed random key of equal length. Let X_1 be the event that $b' = b$ in this game.
- Game_2: This is the same as Game_1, except that the secret signing key generated from $\text{KDF}(m, s_b)$ is replaced by a uniformly distributed random key of equal length. Let X_2 be the event that $b' = b$ in this game.

Claim 1: $|\Pr\{X_0\} - \Pr\{X_1\}|$ is $\text{negl}(\ell)$.

Proof: (by reduction) An adversary $\mathcal{A}$ with an advantage $\epsilon > \text{negl}(\ell)$, can be used to construct another adversary $\mathcal{B}$ against the security of the KDF with advantage ϵ for the source's min-entropy of ℓ, which contradicts the definition of a secure KDF.

Claim 2: $|\Pr\{X_1\} - \Pr\{X_2\}|$ is $\text{negl}(\lambda)$.

Proof Similar to the proof of claim 1.

Claim 3: $|2\Pr\{X_2\} - 1|$ is $\text{negl}(\ell)$.

Proof In Game_2, both the symmetric encryption key and the signing key are replaced by uniformly random keys. Therefore, the claim follows from the assumption that the symmetric encryption scheme is IND-CPA secure (in the security parameter λ). □

5.3 Unlinkability and Irreversibility

In our protocol, we achieve unlinkability by uncoupling the biometrics from its binary representation (cf. Sect. 4 for details). So if the binary representation is somehow compromised, then the user can just revoke it and re-enroll, knowing that a new random binary representation will be generated. A compromised binary representation cannot be linked to a user biometrics. Regarding linkability due to the disclosure of the stored metadata, this is avoided by encrypting it using a second authentication factor.

Regarding irreversibility, the binarization metadata provides some information about the original biometric features in the case that this metadata is designed to increase the reliability of the binary features. The binary representation is not present in the stored metadata, impeding the recovering of the original biometric features. As our protocol protects both the binarization metadata and the binary biometric representation, no information is disclosed about the original biometric template to the considered adversaries.

6 CONCLUSIONS

We presented a method for decoupling biometrics from their protected binary representation. The usefulness of this construction for biometric template protection is demonstrated by incorporating it in a multi-factor authentication protocol in a remote scenario, based on an inherent factor (biometric) combined with either a possession factor (token stored on the smart phone) or a knowledge factor (password inputted by user). The proposed protocol uses the decoupling primitives for providing unlinkability to biometrics and relies on a minimum of two authentication factors to provide resistance against malicious adversaries both regarding security and privacy. The proposed solution provides irreversibility and unlinkability of the biometric template. Comprehensive security and privacy analysis demonstrate the effectiveness and robustness of the design.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable feedback. The work was supported by the European Commission FP7 project "EKSISTENZ" grant number: 607049. In addition, this work was supported in part by the Research Council KU Leuven: C16/15/058, and by imec through ICON Diskman.

REFERENCES

[1] Aysajan Abidin. 2017. *On Privacy-Preserving Biometric Authentication.* Springer International Publishing, Cham, 169–186. https://doi.org/10.1007/978-3-319-54705-3_11

[2] Aysajan Abidin, Abdelrahaman Aly, Enrique Argones Rúa, and Aikaterini Mitrokotsa. 2016. *Efficient Verifiable Computation of XOR for Biometric Authentication.* Springer International Publishing, Cham, 284–298. https://doi.org/10.1007/978-3-319-48965-0_17

[3] Aysajan Abidin, Kanta Matsuura, and Aiketerini Mitrokotsa. 2014. Security of a Privacy-Preserving Biometric Authentication Protocol Revisited. In *International Conference on Cryptology & Network Security (LNCS)*, Vol. 8813. Springer, 291–304.

[4] Aysajan Abidin, Enrique Argones Rúa, and Bart Preneel. 2016. *An Efficient Entity Authentication Protocol with Enhanced Security and Privacy Properties.* Springer International Publishing, Cham, 335–349. https://doi.org/10.1007/978-3-319-48965-0_20

[5] Apple. 2015. *KeychainTouchID: Using Touch ID with Keychain and LocalAuthentication.* https://developer.apple.com/library/ios/samplecode/KeychainTouchID/Introduction/Intro.html

[6] E. Argones Rúa, E. Maiorana, J. L. Alba Castro, and P. Campisi. 2012. Biometric Template Protection Using Universal Background Models: An Application to Online Signature. *IEEE Transactions on Information Forensics and Security* 7, 1 (Feb 2012), 269–282. https://doi.org/10.1109/TIFS.2011.2168213

[7] S. Billeb, C. Rathgeb, H. Reininger, K. Kasper, and C. Busch. 2015. Biometric template protection for speaker recognition based on universal background models. *IET Biometrics* 4, 2 (2015), 116–126. https://doi.org/10.1049/iet-bmt.2014.0031

[8] J. Bolling. 2000. A window to your health. *Jacksonville Medicine, Special Issue: Retinal Diseases* 51 (2000).

[9] Julien Bringer and Hervé Chabanne. 2008. An Authentication Protocol with Encrypted Biometric Data. In *AFRICACRYPT '08 (LNCS)*, Vol. 8813. Springer, 109–124.

[10] John Daugman. 1998. Recognizing people by their iris patterns. *Inf. Sec. Techn. Report* 3, 1 (1998), 33–39.

[11] Najim Dehak, Réda Dehak, Patrick Kenny, Niko Brümmer, Pierre Ouellet, and Pierre Dumouchel. 2009. Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification. In *INTERSPEECH '09*. 1559–1562.

[12] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. 2008. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. *SIAM J. Comput.* 38, 1 (2008), 97–139. https://doi.org/10.1137/060651380

[13] Chun-I Fan and Yi-Hui Lin. 2009. Provably secure remote truly three-factor authentication scheme with privacy protection on biometrics. *IEEE Transactions on Information Forensics and Security* 4, 4 (2009), 933–945. https://doi.org/10.1109/TIFS.2009.2031942

[14] Hu Han, Charles Otto, and Anil K. Jain. 2013. Age estimation from face images: Human vs. machine performance. In *International Conference on Biometrics - ICB 2013*. IEEE, 1–8.

[15] M. T. Ibrahim, M. Kyan, and L. Guan. 2009. On-line signature verification using global features. In *Electrical and Computer Engineering, 2009. CCECE '09. Canadian Conference on*. 682–685. https://doi.org/10.1109/CCECE.2009.5090216

[16] ISO/IEC 24745:2011. 2011. Information technology – Security techniques – Biometric information protection. (2011).

[17] Andrew Teoh Beng Jin, Alwyn Goh, and David Ngo Chek Ling. 2006. Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 12 (2006), 1892–1901.

[18] Ari Juels and Martin Wattenberg. 1999. A Fuzzy Commitment Scheme. In *ACM CCS'99*. ACM Press, 28–36.

[19] Tom A. M. Kevenaar, Geert Jan Schrijen, Michiel van der Veen, Anton H. M. Akkermans, and Fei Zuo. 2005. Face Recognition with Renewable and Privacy Preserving Binary Templates.. In *AutoID*. IEEE Computer Society, 21–26.

[20] Mohammed S. Khalil, Dzulkifli Muhammad, and Qais AL-Nuzaili. 2009. Fingerprint Verification Using the Texture of Fingerprint Image. *Machine Vision, International Conference on* 0 (2009), 27–31. https://doi.org/10.1109/ICMV.2009.18

[21] Hugo Krawczyk. 2010. Cryptographic extraction and key derivation: The HKDF

scheme. In *Advances in Cryptology–CRYPTO 2010*. LNCS, Vol. 6223. Springer, 631–648.

[22] L. S. Penrose. 1965. Dermatoglyphic Topology. *Nature* 205 (February 1965), 544 – 546. https://doi.org/doi:10.1038/205544a0

[23] V.S. Meenakshi and Dr.G. Padmavathi. 2010. Security analysis of password hardened multimodal biometric fuzzy vault with combined feature points extracted from fingerprint, iris and retina for high security applications. *Procedia Computer Science* 2 (2010), 195 – 206. https://doi.org/10.1016/j.procs.2010.11.025

[24] Fabian Monrose, Michael K. Reiter, and Susanne Wetzel. 1999. Password Hardening Based on Keystroke Dynamics. In *ACM CCS '99*. ACM, 73–82. https://doi.org/10.1145/319709.319720

[25] Elena Pagnin, Christos Dimitrakakis, Aysajan Abidin, and Aikaterini Mitrokotsa. 2014. On the Leakage of Information in Biometric Authentication. In *INDOCRYPT 2014 (LNCS)*, Vol. 8885. Springer, 265–280.

[26] Samsung. 2016. *Pass Programming Guide*. http://developer.samsung.com/resources/pass

[27] Koen Simoens, Pim Tuyls, and Bart Preneel. 2009. Privacy Weaknesses in Biometric Sketches. In *IEEE Symposium on Security and Privacy 2009*. 188–203.

[28] Yagiz Sutcu, Qiming Li, and Nasir Memon. 2009. Design and analysis of fuzzy extractors for faces. *Proc. of SPIE* 7306 (2009), 73061X–73061X–12. https://doi.org/10.1117/12.820571

[29] Ewa Syta, Michael J. Fischer, and Abraham Silberschatz. 2012. *Strong Theft-Proof Privacy-Preserving Biometric Authentication*. Technical Report. Yale/DCS/TR-1455.

[30] Shuo Wang and Jing Liu. 2011. Biometrics on mobile phone. In *Recent Application on Biometrics*. InTech.

[31] Fan Wua, Lili Xu, Saru Kumari, and Xiong Li. 2015. A novel and provably secure biometrics-based three-factor remote authentication scheme for mobile client-server networks. *Computers and Electrical Engineering* 45 (2015), 274–285.

[32] Jian De Zheng. 2011. A Framework for Token and Biometrics Based Authentication in Computer Systems. *JCP* 6, 6 (2011), 1206–1212. https://doi.org/10.4304/jcp.6.6.1206-1212

[33] Hongfeng Zhu. 2015. One-time identity-password authenticated key agreement scheme based on biometrics. *Security and Communication Networks* 8, 13 (2015), 2350–2360. https://doi.org/10.1002/sec.1182