Privacy-preserving speaker verification system based on binary I-vectors

Aymen Mtibaa¹ | Dijana Petrovska-Delacrétaz¹ | Jérôme Boudy¹ | Ahmed Ben Hamida²

¹Institut Polytechnique de Paris, Télécom SudParis, Paris, France
²Ecole Nationale d’Ingénieurs de Sfax, Sfax University, ATMS, Sfax, Tunisia

Correspondence
Aymen Mtibaa, Institut Polytechnique de Paris, Télécom SudParis, Paris, France.
Email: aymen.mtibaa@telecom-sudparis.eu

Funding information
European Union’s Horizon 2020 research and innovation program, Grant/Award Numbers: 653586, 789872

Abstract
Speaker verification is a key technology in many services and applications like smartphones and intelligent digital assistants. These applications usually require users to transmit their recordings, features, or models from their voices over untrusted public networks which stored and processed them on cloud-based infrastructure. Furthermore, the voice signal contains a great deal of the speaker’s personal and private information which raises several privacy issues. Therefore, it is necessary to develop speaker verification systems that protect the user’s voice against such threats. Herein, the cancellable biometric systems have been introduced as a privacy-preserving solution. A cancellable method for speaker verification systems is proposed using speaker i-vector embeddings. This method includes two stages: (i) i-vector binarisation and (ii) the protection of the binary i-vector with a shuffling scheme derived from a user-specific key. Privacy evaluation of this method according to the standards of biometric information protection (ISO/IEC 24745) shows that the proposed cancellable speaker verification system achieves the revocability, unlinkability, and irreversibility requirements. Moreover, the cancellable system improves biometric performance compared with the unprotected system and makes it resistant to different attack scenarios. Additionally, we demonstrate that this method can also operate to protect deep neural network speaker embeddings such as x-vectors.

1 | INTRODUCTION

Maintaining security and privacy is a priority for people who safeguard their personal information. Authentication methods such as passwords and PINs are no longer reliable and efficient. Automatic biometric verification systems are becoming a popular technology that enables the recognition of individuals based on their unique biometric traits which cannot be forgotten or lost [1]. Its process is composed of two steps: enrolment and verification. During the enrolment phase, the system collects biometric data samples from a user to create his/her enrolment biometric reference (as templates or models). Then, during the verification phase, the authentication is performed by comparing the biometric sample provided by the user and his/her biometric reference. Such verification systems provide greater security and convenience than traditional methods of authentication.

Today, among biometric verification systems available, speaker verification systems are increasingly ubiquitous. They are used to authenticate individuals and control access across a wide range of different services and devices. Authentication of the user by his/her voice is more convenient than entering passwords. It consists of automatically verifying who is speaking using his/her voice characteristics captured by a recording device. However, biometric systems such as speaker recognition are not full-proof and have their vulnerabilities [2, 3]. The process of using a speaker authentication system requires that the system stores the speakers’ models and has access to the recordings or features derived from the speakers’ voice. This process poses threats to privacy and security. In fact, speakers’ models, features, or recordings could be stolen by an adversary who can use them to create fake recordings and gain unauthorised access to the system. Also, in the case of speaker verification systems, we are in front of additional privacy concerns because from stolen speech data, several sensitive information related to the speaker’s identity, gender [4], age [5], or health status [6] could be extracted.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. IET Biometrics published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.
Moreover, unlike passwords, biometric characteristics are not revocable. In the case of a text-independent speaker verification system, where no prior constraints are considered for the spoken phrases by the speaker, once a non-target user succeeds to prerecord or synthesise the voice of the target speaker, the voice sample is rendered useless in terms of security because the new speaker model generated from this voice sample will be the same as the compromised. For the text-dependent case, where a predefined pass-phrase is employed for verification, one possible solution is to replace the passphrase. However, in some services and applications based on speaker verification, we are confronted with a limited choice of passphrases. For example with Google assistant, we have the choice between only _ok Google_ or _hey Google_. In addition, speaker models stored in different applications are linkable because they are extracted from the same biometric trait. Once, one of the speaker models is compromised, an adversary can exploit it to cross-match across the different applications.

Since 2019, strong customer authentication for financial transactions is requested by the EU payments service directive [7] which requires customer authentication with at least two factors. Currently, some banks use biometric systems as a second factor to authenticate an online credit card payment. However, in case the biometric data is stolen, this second factor becomes useless, because biometric characteristics are not revocable. Moreover, with the privacy concerns of biometric data addressed in the EU General Data Protection Regulation (GDPR) [8], the privacy-preserving of biometric systems is becoming essential to ensure that sensitive biometric data like voice recordings or speaker models are properly protected. A biometric system is considered privacy-preserving according to the standard ISO/IEC 24745 [9] for Biometric Information Protection (BIP) when the following requirements are achieved:

- **Revocability**: When the target biometric reference is leaked, the reference data should be able to be revoked and renewed from the same biometric trait. In addition, the new biometric reference should not match the old one.
- **Performance**: When applied a BIP method to protect the biometric reference, the performance should be maintained compared with the unprotected system.
- **Non-invertibility**: It should be computationally infeasible to recover the biometric data related to the target user from the protected biometric reference.
- **Unlinkability**: Given the same biometric data, it must be feasible to generate different protected biometric references in a way that they cannot be linked to each other or to the subject from which they were derived.

In this context, cancellable biometrics [10] was proposed as a solution to guarantee the privacy of biometric information. It is a process of protecting user biometric reference based on transformations that allow us to perform the biometric comparison in a protected domain. Although various methods of cancellable biometric have been proposed to preserve privacy such as facial, iris, or fingerprint recognition systems [11, 12], common and standardised evaluations are still missing. In addition, cancellable biometrics usually involve some accuracy degradation, and there are few methods proposed to protect privacy of speaker verification systems. Most existing cancellable solutions are not easily adaptable to the characterisation of the speaker, since speaker representation is based on speaker model, such as Gaussian Mixture model (GMM) [13]. In this case, instead of protecting biometric templates, privacy-preserving methods must be applied to biometric models and be robust to the voice variability.

Herein, we present a cancellable method to protect speaker verification systems based on speaker embeddings. The privacy is achieved by binarising speaker embeddings and its protection with BIP scheme named shuffling. This method was employed to protect speaker i-vectors [14] and demonstrated its feasibility on speaker embedding x-vector [15] extracted from Deep Neural Networks (DNNs).

In addition, we make a step forward towards the evaluation of privacy-preserving methods on common and standardised requirements by the evaluation of the proposed cancellable system using i-vector according to the requirements described in the standard ISO/IEC 24745 [9] for biometric information protection. Moreover, security analysis is performed based on the evaluation methodology described in Ref. [16].

Our experimental evaluations on public databases show that the protection of binary speaker embedding with the shuffling scheme achieves the revocability, unlinkability, irreversibility, and performance requirements. Furthermore, the cancellable system improves the biometric performance compared with the unprotected system.

Herein, in Section 2, we present some related works on privacy-preserving speaker recognition systems. The proposed cancellable method for the protection of speaker embeddings is described in Section 3. Experimental protocols and results are presented in Section 4. Conclusions and perspectives are given in Section 5.

## 2 RELATED WORKS

Various researches have contributed to the development of privacy-preserving biometric systems. Most biometric protection methods proposed in the literature are devoted to preserve privacy for biometric systems based on face, iris, and fingerprint modalities where biometric data are represented with templates. However, such methods cannot be applied in the case of the voice modality. The human voice usually varies from one session to another, depending on the pronounced phrase, noise, and frequency. Also, in speaker verification systems, the speaker is represented with models rather than templates; models take into account the variability of the speech signal in the form of variable contents and signal qualities. Thus, in the case of a speaker verification system, instead of protecting biometric templates, BIP methods should protect biometric models. This protection must be done without degrading the biometric performance compared with the baseline speaker recognition system. A recent survey of existing BIP methods that address the privacy-preserving in
the context of speaker recognition system is presented in Ref. [17]. These methods can be classified on approaches based on cancellable biometrics and approaches based on the encryption of biometric data, where techniques such as homomorphic encryption (HE) and secure two-party competition (STPC) are used to achieve the privacy requirements.

Pathak et al. [18, 19] adapted the Gaussian Mixture Model-based speaker recognition to reach privacy requirements using Paillier encryption. This work employs HE and STPC techniques to perform biometric comparison, without exposing the speaker features and models to the system. This method achieves privacy requirements while maintaining the accuracy of the baseline GMM system. However, the shortcoming of this method is the huge computational overhead as compared with the baseline GMM speaker verification system, which makes it impractical to be used in real-life applications. In Ref. [20], to reduce the computational overhead, an alternative scheme based on Locality Sensitive Hashing (LSH) was proposed. This method represents the speech provided by the speaker using super-vectors [21] following with LSH to transform it into bit strings. The bit strings are protected using a cryptographic hash function, and biometric comparisons are performed based on perfect matches of hashed bit strings derived from the speech and the stored templates. Although this approach reduces computation time, the LSH transformation degrades system performance compared with the baseline system.

Homomorphic encryption was also used as a privacy-preserving method for speaker verification based on i-vector embeddings in Ref. [22]. During enrollment, the speaker's i-vector is encrypted and stored in the form of encrypted data. Then, during the verification, the encrypted speaker's i-vector is sent to the user device to compute the verification score in the encrypted domain. The score computed is then sent to an authentication server, which decrypts the encrypted score to take the verification decision.

The use of HE encryption schemes allows to preserve privacy while maintaining the biometric performance obtained with the unprotected system. However, the size of the encrypted data and the huge number of operations required in the encrypted domain, result in overheads of computation and communication, which slows down the verification process. HE-based solutions rely on noise to hide the plaintext. This noise grows during processing speech data in the encrypted domain due to the number of operations (addition, multiplication) required. As a result, the calculation will be performed with larger data than the actual plaintext and the noise will eventually overflow. Hence, an expensive operation named bootstrapping [23] is introduced to reduce it, making the computational overhead too heavy. Thus, the integration of these schemes while keeping verification time low enough for real-time applications is very challenging, especially when considering computationally limited devices such as mobile phones.

Nautsch et al. [24] addressed the issue of computational overhead. They present a solution for computationally manageable privacy-preserving speaker recognition with cohort score normalisation. This solution proposes a cohort pruning scheme based on secure multi-party computation that operates with binary voice representations to reduce the computation time for biometric comparisons in the encrypted domain.

Cancellable biometrics approaches are also employed to achieve privacy for speaker verification systems. These approaches refer to schemes in which the biometric data is protected based on transformations and the biometric comparison is carried out in the transformed domain. Compared with homomorphic encryption schemes, these transformations do not require huge demands on computations, communications, and data storage. However, most of cancellable schemes lead to a degradation in the verification performance.

Teoh and Chong [25] provided a cancellable GMM speaker verification system based on probabilistic random projections [26]. This scheme protects the speaker's model by hiding the features through a random subspace projection process and its parameters are stored in a subject-specific key. This method achieves the revocability and unlinkability requirements and it is shown that the cancellable system maintains the biometric performance even in the stolen-token scenario.

More recent approaches propose cancellable speaker verification methods based on i-vector embeddings. Portelo et al. [27] proposed a cancellable system that performs speaker verification without exposing speaker data to the server by transforming speaker i-vectors to bit strings. This transformation uses an hashing scheme known as Secure Binary Embeddings (SBE) [28]. The SBE is a scheme for privacy-preserving based on LSH, that use a quantised random projection. Results reported with hashed i-vectors show that the speaker verification performance depends on parameters fixed for the SBE which include the number of bits $M$ in the hashed i-vector and the amount of data leakage from the speaker representation. With the best configuration of these parameters to achieve high privacy, the proposed system does not maintain the performance obtained with the unprotected i-vector system. Besides, this scheme was not evaluated according to the BIP requirements, and there is no guarantee that a non-target user is not able to infer information about the plaintext i-vectors when he succeeds to obtain the secrets parameters of the SBE, or when he has some prior knowledge about the plaintext i-vectors.

Choo et al. [29] proposed a cancellable scheme, named Random Binary Orthogonal Matrices Projection (RBOMP) hashing, to protect i-vectors in speaker verification systems. This scheme is inspired by Winner-Takes-All (WTA) hash and further strengthened by the integration of a prime factorisation function. The RBOMP method projects the i-vector using random binary orthogonal matrices from linear space to ordinal space to achieve irreversibility requirements. However, as the WTA focuses on the index of the projected i-vectors instead of the value of the features itself, an adversary may obtain the order of the features and reconstruct the original i-vector. Hence, the prime factorisation is used to conceal the returned index with the help of a user-specific random token. This cancellable system shows good resistance against irreversibility and attack-via-record multiplicity. However, the
verification performance of the protected system degrades compared with the baseline i-vector system.

Regarding the BIP methods used for the protection of facial, iris and fingerprint biometric recognition systems, most of these methods require a binary representation of features or templates. To exploit these methods to protect the speaker verification systems, binarisation schemes of the speaker’s models or features are employed in some research.

Paulini et al. [30] proposed a binarisation method for voice features known as multi-bit allocation based upon the GMM-UBM (Universal Background Model) paradigm. The proposed method is designed to extract discriminative compact binary feature vectors to be exploited in a voice biometric information protection algorithm. Their binarisation acts over GMM super-vectors estimated over Mel Frequency Cepstral Coefficients (MFCC) features. The feature space is divided into intervals, which are encoded with multiple bits using a Grey code. Experimental evaluation shows that the binary representation of voice features causes a negligible decrease in biometric performance compared with the baseline system.

Li et al. [31] investigated the use of binary embeddings for speaker recognition. They studied two binarisation approaches, one is based on LSH and the other is based on Hamming distance learning to transform i-vectors to binary vectors. Evaluations show that binary speaker embeddings deliver competitive results on speaker recognition and reduce the computation cost.

Billeb et al. [32] proposed a binarisation method based on GMM-UBM, that is used to extract high-entropy binary voice reference from speaker models. Speaker binary references are then protected with a fuzzy commitment scheme that uses error correction list decoding to overcome the high intra-class variance in the voice samples. Experimental evaluation has shown that the system achieves privacy requirements. However, the biometric performance degrades due to the binarisation step.

Based on binary speaker representation proposed in [33], a cancellable speaker verification system using the GMM model was presented in [34]. A binary speaker vector is extracted from a speaker model (GMM) and is then transformed using a shuffling scheme [2]. This architecture achieves privacy requirements. However, it needs that the system knows the GMM model of the speaker in plain, leaking a characterisation of the speaker’s voice.

From the above-cited research, we can observe that most of the biometric information protection methods are not evaluated according to the complete set of requirements defined in the standard ISO/IEC 24745. Moreover, most of cancellable methods used to protect the speaker verification systems degrade the biometric performance.

Remaining sections address the privacy issue for speaker verification system using i-vectors. We propose a cancellable solution that performs speaker verification without revealing the speaker’s voice information to the system, either during enrollment or during the verification phase, while maintaining biometric performance. More precisely, the following contributions are provided:

- We propose a cancellable speaker verification system to mitigate privacy and security issues based on two steps; the i-vector binarisation by thresholding the i-vector with its median value, and then the transformation of the binary i-vector with the shuffling scheme.
- We make a step forward towards the evaluation of privacy-preserving methods on common and standardised cancellable requirements by the evaluation of the proposed cancellable system according to the requirements described in the standard ISO/IEC 24745 [9] for biometric information protection.
- This cancellable system reaches better biometric performance than the baseline i-vector system contrary to existing privacy protection methods.
- This system achieves the biometric information protection requirements [9], and shows a good level of security against different attack scenarios.
- We also demonstrate that this cancellable solution could operate on the state-of-the-art of speaker verification systems based on Deep Neural Network (DNN) speaker embeddings.

3 | PRIVACY-PRESERVING SPEAKER VERIFICATION BASED ON BINARISED I-VECTORS AND A SHUFFLING SCHEME

In this section, we describe the cancellable speaker verification system based on the binarisation of i-vector and its transformation with the shuffling scheme. We begin by outlining the i-vector approach for speaker verification, then we describe the proposed cancellable i-vector solution.

3.1 | Speaker verification system based on i-vectors

The i-vector system proposed by Ref. [14] provides a way to generate a low-dimensional fixed-length representation of a speech utterance that preserves speaker-specific information. This technique was inspired by the Joint Factor Analysis framework presented in Ref. [35]. The i-vector system maps a sequence of features such as MFCC obtained from a speech utterance to a fixed-length low-dimensional vector. A Universal Background Model UBM with N component Gaussian Mixture is used to collect Baum-Welch statistics from the speech utterance. Then, the speaker-and channel-dependent GMM supervector $M$ is constructed by appending together the first-order statistics for each mixture component that can be represented via a single total variability subspace as follows:

$$ M = m + Tw $$

where $m$ is the speaker- and channel-independent supervector extracted from the UBM, $T$ is a low-rank matrix named total variability matrix spanning the subspace with speaker-specific information variability, and $w$ is a standard normal distributed
vector. The posterior mean of \( \mathbf{w} \) is the corresponding i-vector. Since the i-vector comprises both speaker and channel variability, in the i-vector framework for speaker verification some sort of channel compensation or channel modelling technique usually follows the i-vector extraction process. Regarding channel compensation, Linear Discriminant Analysis (LDA) or Within-Class Covariance Normalisation (WCCN), are typically applied to compensate for channel nuisance in the i-vector space [36]. For the verification phase, first, the cosine scoring is computed between the target speaker i-vector \( \mathbf{w}_{\text{target}} \) and the test utterance i-vector \( \mathbf{w}_{\text{test}} \):

\[
\text{score}(\mathbf{w}_{\text{target}}, \mathbf{w}_{\text{test}}) = \frac{\mathbf{w}_{\text{target}} \cdot \mathbf{w}_{\text{test}}}{\|\mathbf{w}_{\text{target}}\| \cdot \|\mathbf{w}_{\text{test}}\|}
\]

Later, the probabilistic linear discriminant analysis (PLDA) [37, 38] was introduced as a back-end scoring.

For the proposed cancellable i-vector system, we address the protection of i-vector system using cosine scoring as back-end. As a baseline system, we exploit the total variability matrix to extract a speaker representation with a fixed length followed with a length normalisation and LDA as channel compensation.

### 3.2 Binary i-vector extraction

To extract a binary representation from the speaker’s i-vectors, we apply thresholding. The use of the mean or median of the i-vectors as threshold gives close results on binary i-vectors since i-vectors distribution is close to the normal distribution. For the proposed system, we use the median to be sure that independently of the speaker, each binary i-vector contains an equal number of ones and zeros. This is useful in the reversibility and irreversibility analysis when we need to compute the number of possible permutations. Therefore, after speaker’s i-vector generation, the elements having a higher value than the median are converted to one, while the remaining are converted to zero. From an i-vector of dimension 400 \( \mathbf{X} = (x_1, \ldots, x_{400}) \), we obtain a binary vector \( \mathbf{X}_{\text{bin}} = (b_1, \ldots, b_{400}) \) by comparing each component to the median value of the i-vector.

\[
b_i = \begin{cases} 
0, & \text{if } x_i \leq \text{median}(X) \\
1, & \text{otherwise}
\end{cases} \quad \text{for } i \in (1, \ldots, 400)
\]

### 3.3 Cancellable i-vector system

After i-vector binarisation, we transform the binary i-vector with a cancellable scheme named shuffling, to obtain the protected binary i-vector. The concept of shuffling scheme was introduced in Ref. [2]. The shuffling scheme requires a binary shuffling key \( K_{sh} \) of length \( L_{sh} \) equal or smaller than the binary i-vector dimension. This key can be stored on a secure token or it can be obtained using a password. The binary vector extracted from the i-vector is divided into \( L_{sh} \) blocks each of which has the same length. To start the shuffling, these \( L_{sh} \) blocks of the feature vector are aligned with the \( L_{sh} \) bits of the shuffling key \( K_{sh} \). In the next step, two distinct parts containing biometric features are created: part 1 comprises all positions’ blocks where the shuffling key bit value is one and part 2 comprises all positions’ blocks where the shuffling key bit is zero. These two parts are concatenated to form the shuffled i-vector which is treated as the protected i-vector. When two binary i-vectors are shuffled using the same shuffling key, the absolute positions of the blocks change but this change occurs in the same way for both of the representations. As a result, the Hamming distance between them does not change. On the other hand, if they are shuffled using two different keys, the result is a randomisation of the representations and the Hamming distance increases.

The shuffled binary i-vector which is treated as a cancellable template is the result of combining the biometric sample (i-vector) and the shuffling key. Therefore, once the protected i-vector is leaked, it can be revoked and a new template can be generated by changing the shuffling key. In this work, we take a block size of 1 because it gives the best biometric performance with a shuffling key size equal to the dimension of i-vector. A specific shuffling key is assigned to each user during enrolment and he/she has to provide that same key during each subsequent verification. The pseudo-code of the shuffling scheme is shown in Algorithm 1.

**Algorithm 1** Shuffling scheme pseudo-code

```
shuffling (data, K_{sh});
Input: data: Binary i-vector. K_{sh}: shuffling key
Output: shuffled binarized i-vector

Initialization
part 1 = [] Define empty vector
part 2 = [] Define empty vector
L_{sh} = Length of the shuffling key (K_{sh})
L_{data} = Length of the binary i-vector (data)
block_size = L_{data} / L_{sh}

j ← 1
for i = 1 : L_{sh} do
    if K_{sh}(i) = 1 then
        | part 1 ← [part1, data(j; j + block_size - 1)]
        else
            | part 2 ← [part2, data(j; j + block_size - 1)]
    end
    j ← j + block_size
end

Shuffled binarized i-vector ← concatenate [part1 ; part2]
```

Figure 1 illustrates the architecture of the proposed cancellable i-vector. According to the ISO/IEC 24745, the system falls under the category Model G. This model employs data separation through distributed storage of data elements. We propose the following protocol for the proposed cancellable system. As input, we assume that the server already has the total variability matrix \( \mathbf{T} \) and the shuffling key of the user is stored in the token. During the enrollment phase, the user provides the enrollment voice samples to the client-side that extracts the MFCC features and generates the binary i-vector.
using the total variability matrix received from the server. Then, the client-side transforms the binary i-vector using the user's shuffling key received from the token and sends it to the server. As an output, the server receives the protected i-vector called the pseudonymous identifier \( PI^* \).

During the verification phase, as an input, the user provides the probe voices samples to extract the MFCC features for the test and the server has the total variability matrix \( T \) and the pseudonymous identifier \( PI \). The server sends the \( T \) to the client-side to extract the test binary i-vector. Then the token sends the shuffling key to the client-side that transforms the binary i-vector with the shuffling key to generate the protected i-vector for test \( PI^* \) and transfers it to the server. Then, the server calculates the Hamming distance between the stored \( PI \) and the \( PI^* \) to decide based on the predefined threshold the outcome of the verification. Based on this protocol, the server never has access to the voice recorded by the user, and it does not possess a model of the speaker's voice that it could be misused. The server stores only the protected speaker i-vector generated during enrollment and the total variability matrix \( T \), that it does not reveal sensitive information.

4 | EXPERIMENTAL EVALUATION AND RESULTS

In this section, we evaluate the cancellable i-vector system according to the requirements described in the standard for the biometric information protection (performance, revocability, irreversibility, and unlinkability). Also, we demonstrate the feasibility of the proposed cancellable scheme to protect DNN speaker embeddings such as x-vectors. Furthermore, a security analysis of the cancellable i-vector system against different attack scenarios is reported.

4.1 | Databases and experimental setups

To evaluate the proposed cancellable scheme for text-dependent and text-independent speaker verification systems, we report the biometric performance evaluation of the cancellable i-vectors system using the RSR2015 [39] text-dependent database and using the VoxCeleb [40] text-independent database for the cancellable x-vectors system. Even if the i-vector approach is not well-suited for text-dependent speaker verification system, the privacy-preserving evaluation of the cancellable i-vector system using RSR2015 is competitive enough. For text-dependent scenario, we will explain that due to the revocability property of the cancellable system, in case the passphrase of the target user is compromised, instead of selecting a new one, we can generate a new speaker representation from the same compromised passphrase.

The RSR2015 database comprises speech recorded from 300 speakers, including 143 females and 157 males. For our evaluation, part1 of RSR2015 is used. This part focuses on a text-dependent speaker verification task where each speaker pronounces 30 fixed sentences with nine sessions. The duration of each sentence varies between 2 and 3 s. In the experiments, we use the 300 speakers divided into three partitions of background, development, and evaluation. For comparable results, the protocol described in [39] is followed. From the nine sessions of each speaker, three sessions are used for enrollment while the rest of the sessions are used for the test. RSR2015 provides four types of trials that can be considered given that the test utterance is spoken by the target user or
not and that the spoken utterance is the correct passphrase or not.

**Target-correct (tar-c):** where the target speaker pronounces the expected pass-phrase.

**Target-wrong (tar-w):** where the target speaker pronounces a wrong pass-phrase (a phrase i.e. different from the enrollment one).

**Impostor-correct (imp-c):** where a non-target speaker pronounces the expected pass-phrase.

**Impostor-wrong (imp-w):** where a non-target speaker pronounces a wrong pass-phrase (a phrase i.e. different from the enrollment one).

Target correct trials are considered as target trials, while the others are considered as non-target trials. The impostor-correct tests are more challenging, as the non-target user pronounces the expected passphrase that is used to train the target speaker model.

For the baseline i-vector system, gender-independent GMM-UBM containing 1024 Gaussians is trained using all male and female data of background partition (approximately 26,140 sentences). The GMM-UBM training data are reused for the training of the total variability matrix of rank 400 using 10 iterations of the expectation-maximisation algorithm. For the i-vector based system, an i-vector of dimension 400 is extracted for each sentence. During enrollment, an average i-vector computed using the three sessions dedicated to represents each speaker. For the Linear Discriminant Analysis (LDA) training, we take the training data used for the UBM-GMM. Sentences having the same passphrase are treated as belonging to the individual speaker class. This gives a total of (50 male + 47 female) * 30 = 2910 speaker-passphrase class. In our work, we reduce the dimensionality of the i-vector from 400 to 200 (named i-vector-LDA200). For spectral analysis, the feature vector is composed of 20 MFCC coefficients, with their first and second derivative and the log energy, leading to a 63-dimensional feature vector using the MSR Identity toolbox [41].

For the baseline x-vector system, we use the recipe available on Kaldi\(^1\) using the VoxCeleb database.

### 4.2 Biometric performance evaluation of the cancellable i-vector system

The proposed cancellable i-vector system involves two-factors, the biometric data and the shuffling key. In real-world applications, three different scenarios need to be evaluated:

**Legitimate key scenario:** In this zero-effort scenario, the target user employs his probe biometric sample with his shuffling key to be authenticated, and the non-target user will use his probe biometric sample with a random shuffling key to impersonate the target user. We consider that target users never lose their shuffling keys.

**Stolen biometric scenario:** In this scenario, an adversary accesses the biometric data of the target user and transforms it with a random key to impersonate the target user.

**Stolen key scenario:** In this scenario, an adversary steals the target shuffling key and used it to transform its own biometric data to gain access.

The system performance is reported in terms of Equal Error Rate (EER). The EER is the rate at which the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) are equal. One of the main requirements of BIP described in ISO 24745 is the fact that the cancellable biometric scheme should not degrade the performance of the baseline biometric system. Therefore, for a fair comparison, first the biometric verification performance of the baseline i-vector system is reported followed by the performance of the proposed cancellable i-vector system.

Regarding the results obtained for the baseline and binary i-vector system in Tables 1 and 2, we observe that better biometric performance is obtained with target-correct/impostor-wrong trials than with the target-correct/impostor-correct trials. This was expected since the impostor-correct trials are more challenging, as the non-target user pronounces the expected passphrase used by the target user to authenticate. Moreover, results show that the LDA performs better when applied to the i-vectors extracted from male speakers than from female. This could be explained by the fact that the number of speech utterances for male used to train the LDA was higher than the female ones.

For the cancellable i-vector system, we report the biometric performance of legitimate key scenario using the target and non-target distributions as shown in Figures 2 and 3, where the dissimilarity measure is used to measure the discrimination. For

| TABLE 1 | Biometric performance of the proposed cancellable system and the baseline i-vector on the RSR2015 female evaluation subset for the impostor correct and impostor wrong trials in terms of EER (%) |
|----------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| System   | Test                                           | Baseline i-vector | Binary i-vector | Cancellable i-vector |
|----------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| i-vector  | tar-c/imp-c                                     | 2.55                                             | 7.27                                             | 1.12                                             |
|          | tar-c/imp-w                                     | 0.37                                             | 2.48                                             | 1.01                                             |
| i-vector-LDA200 | tar-c/imp-c                             | 3.39                                             | 9.43                                             | 0.08                                             |
|          | tar-c/imp-w                                     | 0.203                                            | 1.45                                             | 0.07                                             |

| TABLE 2 | Biometric performance of the proposed cancellable system and the baseline i-vector on the RSR2015 male evaluation subset for the impostor correct and impostor wrong trials in terms of EER (%) |
|----------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| System   | Test                                           | Baseline i-vector | Binary i-vector | Cancellable i-vector |
|----------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| i-vector  | tar-c/imp-c                                     | 6.22                                             | 10.35                                            | 2.35                                             |
|          | tar-c/imp-w                                     | 2.015                                            | 4.84                                             | 2.5                                              |
| i-vector-LDA200 | tar-c/imp-c                             | 5                                                | 10.15                                            | 0.22                                             |
|          | tar-c/imp-w                                     | 0.04                                             | 0.92                                             | 1.04                                             |

\(^1\)https://github.com/kaldi-asr/kaldi/tree/master/egs/voxceleb/v2
the non-target distribution, an adversary will use his binary test i-vector and a random shuffling key to bypass the target cancellable template. The performance of the biometric systems is related to the overlap between the two distributions. The smaller the overlap between the two distributions, the better the system performs. Through the distributions in Figure 3, it can be shown that the shuffling scheme preserves the target Hamming distances and increase the non-target Hamming distances. When applying the shuffling scheme, the mean of the target distribution is preserved exactly just like in the binary i-vector level before performing the shuffling transformation. Contrarily, the mean of non-target distribution is augmented when the shuffling scheme is applied and distribution is right-shifted. This leads to the separation of target and non-target distribution, which implies an improvement of the biometric performances.

Tables 1 and 2 report the biometric performance obtained from the cancellable i-vector and the baseline i-vector system. It can be seen that the cancellable i-vector system improves the biometric performance compared with the unprotected i-vector system. Also, the best performance of the cancellable i-vector is obtained when the shuffling scheme is applied to the i-vector reduced with the LDA. This could be explained through Figure 3, where it is shown that the overlap between the target and non-target scores distributions of cancellable i-vector with LDA is smaller than the one obtained without LDA. In fact, the mean of the non-target distribution of binary i-vector with LDA moves from 0.35 to 0.49 shifted by 0.14 due
to the shuffling. However, the mean of binary-i-vector without LDA is only shifted by 0.04 from 0.45 to 0.49.

We report in Figure 4 about the DET curves obtained for the cancellable i-vector-LDA200 system and the baseline i-vector on the female evaluation subset of RSR2015 for the target-correct/impostor-correct trials. As shown the cancellable i-vector reaches better results with an EER = 0.08% compared with the baseline i-vector system with EER = 3.39%.

### 4.3 Biometric performance evaluation of the cancellable x-vector system

The proposed cancellable method shows its effectiveness on the i-vector speaker embeddings. To demonstrate its feasibility on the state-of-the-art speaker verification systems, we used this cancellable method to protect x-vectors DNN speaker embeddings. As a baseline x-vectors system, we adopt the recipe available on Kaldi where 512-dimensional x-vector speaker embeddings are extracted using a Time Delay Neural Network [15] trained on VoxCeleb1,2 [40]. For the back-end scoring, we use a simple cosine scoring without dimensionality reduction. For the cancellable system, we binarise the 512-dimensional speaker’s x-vector using the median as described in Section 3.2 and then we transform it with the shuffling scheme. The comparison is performed with Hamming distance. The results of this system are reported on the test set of VoxCeleb1 text-independent database since we have already shown that the proposed system operates on text-dependent (RSR2015) speaker verification. Results in Table 3 validate that the shuffling scheme improves the biometric performance. As shown, cancellable x-vectors performs better with EER = 0.05% than the baseline x-vectors with cosine scoring as backend (EER = 8.18%) and even better than x-vectors results reported in Kaldi recipe with PLDA as backend scoring (EER = 3.18%). This improvement in biometric performance for the cancellable x-vectors compared with the baseline x-vectors system is due to the use of the shuffling scheme as a second factor.

The above evaluations show that the cancellable system improves the biometric performance compared with the unprotected speaker verification systems based on i-vectors or x-vectors as speaker embeddings and the cosine distance scoring as backend. Also, as reported in Table 3, the cancellable system does not degrade the performance compared with the speaker verification system using PLDA as backend scoring. However, the cancellable method is not dedicated to protect such a system based on log-likelihood scores because it does not take into consideration the protection of PLDA model parameters. Otherwise, during the latest NIST SRE’19 speaker recognition evaluation, the x-vectors extracted from residual networks using cosine distance scoring performed the best on the VAST database avoiding the need for PLDA [42]. We believe that the proposed cancellable method could be applied to protect such state-of-the-art systems.

### 4.4 Security analysis of the cancellable i-vector system

The proposed cancellable i-vector is based on two-factor authentication, the i-vector and the shuffling key. In this section, we report the robustness of this cancellable system against different attack scenarios. For that, we follow the methodology proposed in Ref. [16], proposing different attacks to evaluate the security of such a system:

- **Zero effort attack A1:** Non-target user x provides its voice features and a random shuffling key to impersonate a target user.
- **Brute force attack A2:** Non-target user tries to be verified by trying different random values of a protected template.
- **Stolen token attack A3:** Non-target user has stolen the shuffling key of the target user and tries different random values of a binary vector to generate the protected template.
- **Stolen biometric data attack A4:** Non-target user has stolen the voice features of the target user and tries different random shuffling keys to generate the protected template.
- **Worst case attack A5:** Non-target user provides its own biometric feature and has also stolen the shuffling key of the target user to generate the protected template.

### Table 3 Biometric performance of the cancellable and the baseline x-vector systems on the test set of VoxCeleb text-independent database in terms of EER (%)

| System      | Back-end scoring | Baseline x-vector | Binary x-vector | Cancellable x-vector |
|-------------|------------------|-------------------|-----------------|---------------------|
| x-vector    | Cosine           | 8.18              | 9.68            | 0.05                |
| Kaldi PLDA  |                  | 3.18              | –               | –                   |

**FIGURE 4** DET curves for the baseline and cancellable i-vector systems along with the possible security threats using target-correct/impostor-correct trials on the female evaluation subset of RSR2015.
For all these attack scenarios, we compute the False Acceptance Rates (FAR) for each attack scenario \( Ai \), when the EER threshold of the cancellable i-vector system \( \varepsilon_{\text{EER}} \) is taken as the decision threshold. A high value of \( Ai \) implies that the system is not resistant to this attack scenario. Table 4 presents the values of \( Ai \) obtained if the proposed cancellable i-vector system is attacked with such a scenario. We notice that for most attacks, we obtain a FAR equal to 0 except for the worst case scenario \( A5 \). These results validate the performances reported in the DET curve Figure 4. The cancellable i-vector system resists to the stolen biometric scenario with EER = 0.09%. However, for the worst case scenario, the performance obtained is equal to that obtained with binary i-vector before transformation with the shuffling scheme. Figure 5 presents the evolution of the FAR for each attack scenario related to the EER threshold of the cancellable i-vector. As we can see, the proposed cancellable i-vector based on shuffling key transformation is robust for all presented attack scenarios. For the worst case scenario, when binary i-vector scores obtained from the target-correct/impostor-correct trials are considered, we obtain a FAR = 70%. In a real use case, this value can be improved by tuning the threshold of the cancellable system according to the FAR and FRR. For example, when we take the threshold equal to 0.35 where the FAR = 0.003% and the FRR = 0.6% on the cancellable system, the FAR in the worst case attack is reduced to 47%. Moreover, in this evaluation, we have reported the FAR in the worst case attack using binary i-vector scores obtained from the target-correct/impostor-correct trials where the performance at binary level degrades further comparing to impostor wrong trials. Using binary i-vector scores obtained from the target-correct/impostor-wrong trials, the FAR obtained is equal to 24% as shown in Figure 6.

| Attack scenario         | FAR (\( \varepsilon_{\text{EER}} \)) |
|-------------------------|--------------------------------------|
| Zero effort attack \( A1 \) | 0                                     |
| Brute force attack \( A2 \)   | 0                                     |
| Stolen token attack \( A3 \)   | 0                                     |
| Stolen biometric data attack \( A4 \) | 0                                   |
| Worst case attack \( A5 \)     | 0.714                                 |

The security of the shuffling key is very important. We believe that in real use cases such security can be guaranteed with the novel technologies as the Embedded Secure Element [43] or the secure chip which provides a secure space to store and manage personal data.

4.5 Revocability analysis of the cancellable i-vector system

For the cancellable biometric system, the protected biometric reference should be able to be revoked and renewed to replace the compromised reference. Revocability is evaluated by calculating the pseudo-impostor scores. The pseudo-impostor score is the comparison of a protected i-vector of a user with other protected i-vectors of the same user generated with the same biometric sample and transformed with different shuffling keys. For this, we shuffled one speaker binary i-vector with 480,000 randomly generated shuffling keys. The first shuffled speaker binary i-vector is compared with the remaining shuffled templates to compute the pseudo-impostor scores. This process is repeated with 30 different users. As shown in Figure 7, the distribution of the pseudo-impostor scores resembles the non-target distribution which means that the generated protected i-vectors are indistinguishable to each other, although they are generated from the same voice sample. As a result, in case of compromise, a cancelation is possible and a new protected i-vector can be generated from the same voice sample by changing the shuffling key. As an example, when the passphrase is compromised in a text-dependent speaker verification system, instead of selecting a new one, we can generate a new speaker reference from the compromised passphrase.

For the protection of i-vectors using the shuffling scheme transformation, the maximum number of the protected i-vector or Pseudonymous Identifier PI that can be generated from the same voice is given using the number of possible permutations. Moreover, because the decision in the proposed system is based on a threshold comparison, we must not account for the possible templates that when compared with the enrollment speaker template, give a score within the range of target distribution scores. For that, we estimate the maximum number of templates using the Hamming bound [44]. We assume that the target speaker template is the centre of a sphere with a radius of \( r \), known as a Hamming sphere. \( r \)
represents the maximum number of non-matching bits obtained when comparing two templates belonging to the same speaker. \( r \) is equal to \( (t \cdot l) \) where \( t \) is the threshold of the cancellable system and \( l \) is the length of cancellable i-vector. Then, the possible templates, that their distance compared with the speaker template are less than the radius \( r \) (meaning they are within the sphere) are not taken into account. Using a threshold \( t = 0.4 \), for i-vector representations of length 400, we get almost 212 possible protected i-vector PI for each user as given in (4).

\[
PI = \frac{\text{Number of possible permutation}}{\text{Volume of Hamming spheres}}
\]

\[
= \frac{400!}{(200!)^2} \sum_{k=0}^{(400)} \frac{400!}{k!} \approx 2^{12}
\]  

(4)

### 4.6 Irreversibility analysis of the cancellable i-vector system

The irreversibility refers to the security of the biometric data from which the protected biometric reference was generated [9]. The reversibility analysis depends on whether the attacker has information about the transformation shuffling key or not. Given the shuffling key, an attacker can revert to the original binary vector. We note that due to the binarisation process of the i-vector before transformation, it is not feasible to recover the original i-vector or the speaker feature. However, without having information about the shuffling key and prior knowledge about the distribution of the non-shuffled i-vectors it is computationally not feasible to revert to the original binary i-vector as the number of permutations to be tested is too big. In the proposed system, if the adversary wants to guess the correct value of the shuffled binary vector with a length of 400 and knowing that the number of bits equal to 1 is the same number of bits equal to 0 since we binarise the i-vector by applying the median, the guessing complexity is equal to \( 2^{395} \) the number of possible permutation, given by (5) as follows:

\[
\text{Number of possible permutation} = \frac{400!}{(200!)^2} \approx 2^{395}
\]  

(5)

### 4.7 Unlinkability analysis of the cancellable i-vector system

As defined in [9] the unlinkability is ‘a property of two or more biometric references that they cannot be linked to each other or to the subject(s) from which they were derived’. The goal of this evaluation is to determine if from two protected i-vectors T1 and T2 enrolled in different applications, we can know whether they are generated from the same voice or not. For this, we use the framework defined in Ref. [45]. Two types of score distributions will be analysed for the assessment of the unlinkability provided by the protected i-vectors:

- **Mated instances**: scores computed by comparing protected i-vectors extracted from different samples of the same subject using different shuffling keys.
Non-mated instances: scores computed by comparing protected i-vectors generated from samples of different subjects using different shuffling keys.

As described in Ref. [45], the global metric $D^{sys} \leftrightarrow$ gives an estimation of the global linkability of the system. If a system has $D^{sys} \leftrightarrow = 1$, where both score distributions (mated and non-mated) have no overlap means that the system is fully linkable. If a system has $D^{sys} \leftrightarrow = 0$, where both score distributions (mated and non-mated) are totally overlapped means that the system is fully unlinkable for the whole score range. As observed in Figure 8, the distribution of mated and non-mated scores are overlapped with global linkability $D^{sys} \leftrightarrow$ equal to 0 rendering the system fully unlinkable.

5 | CONCLUSIONS

Herein, we proposed a cancellable scheme for privacy-preserving speaker verification systems based on i-vectors. This is achieved by first binarising the speaker embedding and then its transformation with the shuffling scheme. We also demonstrate that this cancellable scheme could operate to protect speaker verification systems based on deep neural network speaker embeddings such as x-vectors. The cancellable scheme was evaluated using the RSR2015 text-dependent database for the system based on i-vectors and using the VoxCeleb text-independent database for the system based on x-vectors and it has shown its effectiveness.

Experimental evaluations according to the ISO/IEC 24745 requirements show that the cancellable system made it possible to simultaneously achieve privacy requirements and preserve the biometric verification performance. Due to the shuffling scheme, the cancellable speaker template is revocable. In case the biometric or the key is stolen, different cancellable speaker templates could be generated from the same voice sample without the possibility to be linked. Besides, compared with the majority of ongoing works on voice biometric protection, the proposed cancellable system improves the biometric performance compared with the unprotected system. Furthermore, it resists several attacks and even if the biometric data is stolen, the system’s EER remains lower than the unprotected biometric system. For future work, a novel approach for the transformation of speaker embeddings into binary vectors while maintaining biometric performance will be proposed to improve the resistance of the cancellable system against the worst-case attack scenario.

ACKNOWLEDGEMENT

This work was partially supported by the SpeechXRays and Empathic projects that have received funding from the European Union’s Horizon 2020 research and innovation program, under Grant Agreements Nos. 653586 and 769872.

REFERENCES

1. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. IEEE Trans. Circ. Syst. Video Technol. 14(1), 4–20 (2004)
2. Kanade, S.G., Petrovska Delacêtrê, D., Dorizzi, B.: Enhancing information security and privacy by combining biometrics with cryptography. Synth. Lect. Inf. Secur. Priv. Trust. 3(1), 1–140 (2012)
3. Biggio, B., et al.: Adversarial biometric recognition: a review on biometric system security from the adversarial machine-learning perspective. IEEE Signal Process. Mag. 32(5), 31–41 (2015)
4. Harb, H., Chen, L.: Voice-based gender identification in multimedia applications. J. Intell. Inf. Syst. 24(2–3), 179–198 (2005)
5. Gómez-García, J.A., et al.: Automatic age detection in normal and pathological voice. In: Sixteenth Annual Conference of the International Speech Communication Association, (2015)
6. Jeancolas, L., et al.: Comparison of telephone recordings and professional microphone recordings for early detection of Parkinson’s disease, using mel-frequency cepstral coefficients with Gaussian mixture models. Proc. Interspeech. 2019, 3033–3037 (2019)
7. Románová, I., et al.: The payment services directive 2 and competitiveness: the perspective of European Fintech companies. Eur. Res. Stud. J. 21(2), 5–24 (2018)
8. Bréner, A.: Payment service directive II and its implications. In: Disrupting Finance., pp. 103–119. Palgrave Pivot, Cham (2019)
9. ISO/IEC JTC1 SC27 Security Techniques. ISO/IEC 24745:2011. Information technology - security techniques - biometric information protection (2011)
10. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. IBM Syst. J.. 40(3), 614–634 (2001)
11. Ruthgeb, C., Ubl, A.: A survey on biometric cryptosystems and cancellable biometrics. EURASIP J. Inf. Secur. (1), 3 (2011)
12. Patel, V.M., Ratha, N.K., Chellappa, R.: Cancellable biometrics: a review. IEEE Signal Process. Mag. 32(5), 54–65 (2015)
13. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker verification using adapted Gaussian mixture models. Digit. Signal Process. 10(1–3), 19–41 (2000)
14. Dehak, N., et al.: Front-end factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. 19(4), 788–798 (2011)
15. Snyder, D., et al.: X-vectors: robust dnn embeddings for speaker recognition. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5329–5333. IEEE (2018).
16. Rosenberger, C: Evaluation of biometric template protection schemes based on a transformation. In: ICISPP, pp. 216–224 (2018).
17. Nautsch, A., et al.: Preserving privacy in speaker and speech characterisation. Comput. Speech Lang. 58, 441–480 (2019)
18. Pathak, M.A., Raj, B.: Privacy preserving speaker verification using adapted gmm. In: Twelfth Annual Conference of the International Speech Communication Association (2011)
19. Pathak, M.A., Raj, B.: Privacy-preserving speaker verification and identification using Gaussian mixture models. IEEE Trans. Audio Speech Lang. Process. 21(2), 397–406 (2012).
20. Pathak, M.A., Raj, B.: Privacy-preserving speaker verification as password matching. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1849–1852. IEEE (2012)
21. Campbell, W.M., et al.: SVM based speaker verification using a gmm supervector kernel and nap variability compensation. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, p. I, vol. 1. IEEE (2006)
22. Nautsch, A., et al.: Homomorphic encryption for speaker recognition: protection of biometric templates and vendor model parameters. In: Proceedings of Odyssey 2018 the Speaker and Language Recognition Workshop, pp. 16–23 (2018).
23. Gentry, C: Fully homomorphic encryption using ideal lattices. In: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, pp. 169–178 (2009).
24. Nautsch, A., et al.: Privacy-preserving speaker recognition with cohort score normalisation. arXiv (2019). arXiv–1907.
25. Teoh, A.B.J., Chong, L.Y.: Secure speech template protection in speaker verification system. Speech Commun. 52(2), 150–163 (2010).
26. Teoh, A.B.J., Yang, C.T.: Cancellable biometrics realization with multi-space random projections. IEEE Trans. Syst. Man Cybern. B Cybern. 37(5), 1096–1106 (2007).
27. Portelo, J., et al.: Privacy-preserving speaker verification using secure binary embeddings. In: 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1268–1272. IEEE (2014).
28. Boufounos, P., Rane, S.: Secure binary embeddings for privacy preserving nearest neighbors. In: 2011 IEEE international Workshop on Information Forensics and security, pp. 1–6. IEEE (2011).
29. Chee, K.Y., et al.: Cancellable speech template via random binary orthogonal matrices projection hashing. Pattern Recogn. 76, 273–287 (2018).
30. Paulini, M., et al.: Multi-bit allocation: Preparing voice biometrics for template protection. In: Odyssey, pp. 291–296 (2016).
31. Li, L., et al.: Binary speaker embedding. In: 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), pp. 1–4. IEEE (2016).
32. Billeh, S., et al.: Biometric template protection for speaker recognition based on universal background models. IET Biom. 4(2), 116–126 (2015).
33. Bonastre, J.F., et al.: Speaker modeling using local binary decisions. In: Twelfth Annual Conference of the International Speech Communication Association (2011).
34. Mtibaa, A., Petrovska-Delacrétaz, D., Hamida, A.B.: Cancellable speaker verification system based on binary Gaussian mixtures. In: 2014 4th International Conference on Advanced Technologies for Signal and Image Processing (ATISP), pp. 1–6. IEEE (2018).
35. Kenny, P., et al.: A study of interspeaker variability in speaker verification. IEEE Trans. Audio Speech Lang. Process. 16(5), 980–988 (2008).
36. Kanagasundaram, A., et al.: I-vector based speaker recognition on short utterances. In: Proceedings of the 12th Annual Conference of the International Speech Communication Association, pp. 2341–2344 (2011).
37. Kenny, P.: Bayesian speaker verification with heavy-tailed priors. In: Odyssey, vol. 14 (2010).
38. Garcia-Romero, D., Espy Wilson, C.Y.: Analysis of i-vector length normalization in speaker recognition systems. In: Twelfth Annual Conference of the International Speech Communication Association (2011).
39. Larcher, A., et al.: Test-dependent speaker verification: classifiers, databases and rsr2015. Speech Commun. 60, 56–77 (2014).
40. Nagrani, A., Chung, J.S., Zisserman, A.: Voxceleb: a large-scale speaker identification dataset. Telephony. 3, 33–039 (2017).
41. Sadjadi, S.O., Slaney, M., Heck, L.: Msr identity toolbox v1.0: a matlab toolbox for speaker-recognition research. Speech Lang. Process. Tech. Committee Newslett. 1(4) (2013).
42. López, J.A.V., et al.: Advances in speaker recognition for telephone and audio-visual data: the jhu-mit submission for NIST SRE19 (2020).
43. Tremlet, C.: Embedded secure element for authentication, storage and transaction within a mobile terminal. (Google Patents, 2016). US Patent. 9,436,940, 2016.
44. MacWilliams, F.J., Sloane, N.J.A.: The theory of error correcting codes, vol. 16. Elsevier (1977).
45. Gomez.Barbero, M., et al.: General framework to evaluate unlinkability in biometric template protection systems. IEEE Trans. Inf. Forensics Secur. 13(6), 1406–1420 (2017).

How to cite this article: Mtibaa A, Petrovska-Delacrétaz D, Boudy J, Ben Hamida A. Privacy-preserving speaker verification system based on binary I-vectors. IET Biome. 2021;10:233–245. https://doi.org/10.1049/bme2.12013