Measuring Linkability of Protected Biometric Templates Using Maximal Leakage

Hatef Otroshi Shahreza, *Graduate Student Member, IEEE*, Yanina Y. Shkel, *Member, IEEE*, and Sébastien Marcel, *Senior Member, IEEE*

***Abstract*— As the applications of biometric recognition systems are increasing rapidly, there is a growing need to secure the sensitive data used within these systems. Considering privacy challenges in such systems, different biometric template protection (BTP) schemes were proposed in the literature, and the ISO/IEC 24745 standard defined a number of requirements for protecting biometric templates. While there are several studies on evaluating different requirements of the ISO/IEC 24745 standard, there have been few studies on how to measure the linkability of biometric templates. In this paper, we propose a new method for measuring linkability of protected biometric templates. The proposed method is based on maximal leakage, which is a well-studied measure in information-theoretic literature. We show that the resulting linkability measure has a number of important theoretical properties and an operational interpretation in terms of statistical hypothesis testing. We compare the proposed measure to two other linkability measures: one previously introduced in the literature, and a similar measure based on differential privacy. In our experiments, we use the proposed measure to evaluate the linkability of biometric templates from different biometric characteristics (face, voice, and finger vein), which are protected with different BTP schemes. The source codes of our proposed measure and all experiments are publicly available.**

***Index Terms*— Biometrics, biometric template protection, linkability, maximal leakage, differential privacy, statistical hypothesis testing, template.**

I. Introduction

Biometric recognition systems generally establish the identity of users based on their physiological (e.g., face, finger vein, fingerprint, iris, etc.), behavioral (e.g., voice, gait, signature, etc.), or chemical (e.g., DNA, etc.) attributes, which are unique to individuals. As biometric authentication and identification systems offer great convenience for users and also provide fast and accurate recognition, applications of biometric recognition systems tend towards ubiquity, from personal (e.g., smart phone unlocking with face[1] or fingerprint[2] recognition, etc.) to large-scale applications (e.g., face[3], fingerprint[4], and iris[5] recognition in national identity system, or face recognition for passport control at borders and airports[6] etc.). In such systems, biometric templates (a.k.a., features) are often extracted from biometric data and are stored in the system's database during the enrolment stage. Later, during the recognition stage, a new biometric template is extracted and compared with the templates in the database. Since biometric templates convey important information about the user's identity, data protection regulations, such as the European Union General Data Protection Regulation (GDPR) [1], consider biometric templates as sensitive data and impose legal obligations to protect biometric templates.

To protect biometric templates and address privacy issues in biometric recognition systems, several schemes are proposed in the literature [2], [3], [4]. For each biometric template protection (BTP) scheme, the ISO/IEC 24745 standard [5] defines four criteria. First, the protection scheme should not significantly degrade the accuracy of the biometric recognition system (i.e., performance preservation). Second, the protected template should be irreversible, meaning it should be computationally infeasible to reconstruct the original template from the protected template. (i.e., irreversibility). Third, if a protected template is compromised, it should be possible to revoke that protected template and generate a new protected template (i.e., revocability/renewability). Fourth, if two or more protected templates are leaked, it should not be feasible to determine whether they are from the same subject or different subjects (i.e., unlinkability).

Notwithstanding standardized metrics to evaluate and report the recognition performance of biometric systems (e.g., ISO/IEC 19795-1 standard [6]), no standardized measure has been included in the ISO/IEC 30136 standard [7] for evaluating the irreversibility and unlinkability of protected templates. In addition, while a lot of research has been devoted to the irreversibility evaluation of protected templates [8], there

Manuscript received 20 October 2022; revised 28 February 2023; accepted 24 March 2023. Date of publication 10 April 2023; date of current version 18 April 2023. This work was supported by the Horizon 2020 TRaining in Secure and PrivAcy-preserving biometricS Early Training Networks Marie Skłodowska-Curie Early Training Network under Agreement 860813. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Walter J. Scheirer. *(Corresponding author: Hatef Otroshi Shahreza.)*

Hatef Otroshi Shahreza is with the École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, and also with the Biometrics Security and Privacy Group, Idiap Research Institute, 1920 Martigny, Switzerland (e-mail: hatef.otroshi@epfl.ch).

Yanina Y. Shkel is with the École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

Sébastien Marcel is with the Biometrics Security and Privacy Group, Idiap Research Institute, 1920 Martigny, Switzerland, and also with the School of Criminal Justice, Universite de Lausanne (UNIL), 1015 Lausanne, Switzerland.

Digital Object Identifier 10.1109/TIFS.2023.3266170

[1] https://apple.co/3mLGCYV
[2] https://bit.ly/3cTJ7Gp
[3] https://bbc.in/3QeIsO2
[4] https://bit.ly/3SkvAbi
[5] https://uidai.gov.in
[6] https://cnet.co/3sG8qSd

TABLE I
SUMMARY OF GENERIC METHODS IN THE LITERATURE FOR EVALUATING THE LINKABILITY OF PROTECTED BIOMETRIC TEMPLATES

Ref.	Measure Basis	Quantify Linkability Degree of system	Assumptions
[10]	Accuracy (EER)	✗	-
[11]	Accuracy (ROC)	✗	-
[12]	Accuracy (ROC)	✗	-
[13]	Accuracy (combined)	✗	-
[14]	Accuracy (CMC)	✗	-
[15]	Score distribution	✗	-
[16]	Score distribution	✗	closed-set
[17]	Score distribution	✓	prior probabilities

have been few works proposing measures for evaluating the linkability of protected templates.

According to the ISO/IEC 30136 standard [7], the more precise definition of unlinkability is: "***unlinkability** is the difficulty of **distinguishing** between Auxiliary Data (AD)s and/or Pseudonymous Identifiers (PIs) of two Renewable Biometric References (RBRs) generated from **the same subject's** characteristic and ADs and/or PIs of two RBRs generated from **different subjects'** characteristics*" [emphasis added]. In the context of BTP, we can extend the definition of *mated* and *non-mated* pairs in the ISO/IEC 2382-37 standard [9] as:

- *mated*: two protected templates are mated if they correspond to *the same subject* (they can be either from the same sample or different samples) and with different keys.
- *non-mated*: two protected templates are non-mated if they correspond to *different subjects* with different keys.

Therefore, to gain the unlinkability criterion, the protected templates should be such that an adversary would not be able to distinguish mated and non-mated protected pairs.

Table I summarizes the previous works in the literature which have used a generic method to evaluate the linkability of protected biometric templates. Buhan et al. [10] considered a biometric cryptosystem and compared the recognition accuracy of the system in terms of Equal Error Rates (EER) in two scenarios: i) templates protected with a single key (i.e., regular recognition accuracy analysis), ii) templates protected with different keys (i.e., unlinkability analysis). While the increase of EER implies some degree of unlinkability, the unlinkability is not quantified in their work. Kelkboom et al. [11] considered similar scenarios and compared the recognition performance of the system in terms of the Receiving Operating Characteristic (ROC). Then, if the recognition accuracy shown by the ROC curve decreases, the system is considered to be unlinkable. However, the unlinkability can neither be quantified in this approach. Similarly, Nagar et al. [12] found the ROC curve of matching templates with different keys to evaluate the unlinkability of the system.

Piciucco et al. [13] used a similar approach to [10], [11], and [12], but combined the results of regular analysis and unlinkability analysis. They plotted the True Match Rate (TMR) in the unlinkability analysis[7] versus the system's False Non-Match Rate (FNMR) in the regular analysis. Their method does not evaluate the True Match Rate (TMR) in the unlinkability analysis, and the degree of general unlinkability is also not quantified in their method. Along the same lines, Rua et al. [14] found the probability that the adversary can determine the correct identity in a top-N list and plotted this probability similar to Cumulative Match Curves (CMC). Then, as an evaluation of the unlinkability of the system, they compared this plot with the curve corresponding to the probability of random guesses being correct (i.e., full unlinkability). However, their method does not provide a single number to quantify the general unlinkability of the system.

In contrast to [10], [11], [12], [13], and [14] which have evaluated unlinkability based on accuracy metrics, [15], [16], [17] considered score distributions in their unlinkability evaluations. In [15], Ferrara et al. calculated three distributions of scores, including scores of templates with different keys from: 1) the same sample, 2) different samples of the same subject, and 3) samples of different subjects. Then, according to visual comparisons of these distributions, they evaluate the unlinkability of templates. Wang and Hu [16] used the latter two score distributions only and evaluated unlinkability by visual comparison of these distributions. Gomez-Barrero et al. [17] proposed two quantitative measures (local and global) based on score distributions. Similar to [16], they considered two distributions of scores for mated and non-mated pairs. Then, as their local measure for each score, they consider the difference in conditional probabilities of the hypothesis of being mated and the hypothesis of being non-mated. To calculate their local measure, they use the likelihood ratio of mated and non-mated hypotheses and the ratio of prior probabilities. For their global measure, they considered the conditional expectation of their local measure over score values. The global measure ($\mathrm{D}_{\leftrightarrow}^{sys}$) proposed in [17] was the first quantitative evaluation that measures the degree of unlinkability of the biometric systems. It is also properly defined and bounded in the [0, 1] interval. However, it has several drawbacks that we discuss in Section III-B.

In addition to prior work on linkability, there is ample work on general privacy measures in information theory and computer science communities [18], [19]. The most prominent notions of privacy are ϵ-differential privacy and (ϵ, δ)-differential privacy which were developed for the database release problem [20], [21], [22]. The main idea behind this approach is to control the influence of a single database entry on the output of differentially private queries. BTP schemes have been studied from the differential privacy perspective in [23] where a differentially private distributed face-recognition system is proposed. A hypothesis testing perspective on differential privacy has been introduced in [24] and extended in [25]. In particular, [25] show that (ϵ, δ)-differential privacy guarantees could be interpreted as bounds on the ROC curves of appropriately defined hypothesis tests.

Another recent measure of interest is maximal leakage which seeks to control the adversary's ability to refine his or her estimate of any function of data [26], [27]. Maximal leakage has been recently discussed in the context of hypothesis testing: Privacy-utility trade-offs using maximal leakage as a privacy metric and the type II (false alarm) error exponent

[7] refered as Renewable Template Matching Rate (RTMR) in their work.

as the utility metric have been studied in [28]; In [29] the so-called "noiseless privacy" is related to hypothesis testing and to maximal leakage; And, maximal leakage is used to bound generalization errors of learning algorithms in [30].

In this paper, we propose a new measure for evaluating the linkability of protected biometric templates. Our proposed measure combines the work on maximal leakage from information-theoretic literature [26], [27] with the perspective on global linkability introduced in [17]. Since our proposed measure is based on a well-studied information measure, it inherits many of the theoretic properties of this measure. In addition, we show that the proposed linkability measure has an appealing operational interpretation in terms of hypothesis testing that the adversary could perform on a pair of protected templates. This hypothesis testing interpretation of our proposed measure makes it consistent with the definition of linkability in the ISO/IEC 30136 standard [7]. We further compare our proposed measure to a similar measure based on differential privacy [22] and show that the differential privacy-based measure is too strict for the linkability application. Finally, the experimental implementation of our proposed measure shows that it gives intuitively correct linkability scores across different BTP schemes, biometric characteristics, and scoring functions.

The remainder of the paper is organized as follows. In Section II we define our proposed measure, as well as discuss its operational interpretations and its properties. In Section III we compare the proposed measure to two other linkability measures: the global measure introduced in [17] and a similar measure based on differential privacy. In Section IV, we evaluate the unlinkability of different biometric recognition systems based on different biometric characteristics and protected with different BTP schemes. Finally, the paper is concluded in Section V.

II. Proposed Measure

In this section, we propose a new measure of linkability for biometric templates. In Section II-A, we introduce our notation and overview the maximal leakage information measure. In Section II-B, we define our measure of linkability as a maximal leakage of information about the mated and non-mated hypothesis, as well as review its properties. We end by interpreting the new measure in terms of statistical hypothesis testing in Section II-C.

A. Paper Notation and Maximal Leakage

Throughout the paper, we use capital letters to denote random variables, calligraphic letters to denote support sets of these random variables (and sets in general), and lower case letters to denote realizations of these random variables. For example, X is a random variable taking values on $\mathcal{X}$ while $x \in \mathcal{X}$ is a possible realization of this random variable. We use the notation $X \leftrightarrow Y \leftrightarrow Z$ to denote that X, Y, and Z form a Markov chain. We use p_X to denote the probability mass function (if $\mathcal{X}$ is discrete) or the probability density function (if $\mathcal{X}$ is continuous) of X. If the associated random variable is clear from context, we omit the subscript: for example, $p(y|s)$. We use sanserif font to indicate functions, for example $\mathsf{f}\colon \mathcal{X} \to \mathcal{Y}$ denotes a function from $\mathcal{X}$ to $\mathcal{Y}$. Finally, all the logarithms in this paper will be assumed to have base two.

Maximal leakage is an information leakage measure introduced in [26] and [27]. Specifically, [26] defined this measure as follows. Let X and Y be two jointly-distributed random variables, where X represents some secret information which may be of interest to an adversary, while Y represents the actual observations of an adversary. The maximal leakage of information from X to Y is defined as

$$\mathcal{L}(X \to Y) = \sup_{U \leftrightarrow X \leftrightarrow Y \leftrightarrow \hat{U}} \log \frac{\mathbb{P}\left(U = \hat{U}\right)}{\max_{u \in \mathcal{U}} p_U(u)} \tag{1}$$

where U, $\hat{U}$ are random variables over some common finite alphabet. The auxiliary random variable U in Eq.1 denotes some, possibly random, mapping of secret information X, while $\hat{U}$ denotes the best guess an adversary could make about U. Thus, the ratio $\frac{\mathbb{P}\left(U=\hat{U}\right)}{\max_{u\in\mathcal{U}} p_U(u)}$ captures how much an adversary's ability to guess any hidden mapping of data U improves by observing Y. The whole quantity in Eq. 1 measures multiplicative improvement of the adversary's ability to guess any possible function of the secret X.

Maximal leakage was independently introduced in [27] where it was defined as

$$\mathcal{L}(X \to Y) = \sup_{p_X} \log \frac{\mathbb{P}\left(X = \hat{X}\right)}{\max_{x \in \mathcal{X}} p_X(x)} \tag{2}$$

where $X \leftrightarrow Y \leftrightarrow \hat{X}$. When X has full support, both definitions in Eq. 1) and Eq. 2 are equivalent [26].

Although it is not immediately clear that Eq. 1) and Eq. 2 are computable, it is shown in [26, Theorem 1] that, for discrete (X, Y), maximal leakage could be evaluated via the following simple formula

$$\mathcal{L}(X \to Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}\colon\, p_X(x) > 0} p_{Y|X}(y|x). \tag{3}$$

This result could be extended to more general settings [26, Theorem 7]. For example, a setting that will be of interest to us is when $\mathcal{Y}$ is continuous, $\mathcal{X}$ is discrete, and the probability density functions $p_{Y|X}(y|x)$ exist. In this case, the maximal leakage reduces to

$$\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \max_{x \in \mathcal{X}} p_{Y|X}(y|x) \mathsf{d}\mathsf{y}. \tag{4}$$

Finally, it is shown in [26] that

$$\mathcal{L}(X \to Y) = I_\infty(X; Y) \tag{5}$$

where $I_\infty(X; Y)$ denotes the *Sibson's mutual information* of order infinity [31], [32]. In other words, $\mathcal{L}(X \to Y)$ could be viewed as a generalization of Shannon's mutual information in the same way that Rényi entropy is a generalization of Shannon's entropy [33].

Because maximal leakage is a well-defined information measure, it has a number of mathematical properties. We highlight some of the most important properties here:

- First, maximal leakage is non-negative, that is
$$\mathcal{L}(X \to Y) \geq 0. \tag{6}$$
It is zero if and only if X and Y are statistically independent.
- Secondly, it satisfies the *data processing inequality* which states that
$$\mathcal{L}(X \to Z) \leq \mathcal{L}(X \to Y) \tag{7}$$
where $X \leftrightarrow Y \leftrightarrow Z$ form a Markov chain.
- Finally, for a discrete random variable X,
$$\mathcal{L}(X \to Y) \leq \log|\mathcal{X}|. \tag{8}$$

Proofs of these properties and additional properties of maximal leakage could be found in [26].

B. Maximal Linkabilty of Biometric Templates

The proposed linkability metric uses maximal leakage to measure the amount of information revealed by two templates about the two possible hypotheses: the templates are mated, and the templates are not mated. Specifically, given two biometric systems, let $\mathcal{T}_1$ be the space of all possible protected templates that could be produced by the first system and $\mathcal{T}_2$ be the space of all possible protected templates that could be produced by the second system. Given two templates $(t_1, t_2) \in \mathcal{T}_1 \times \mathcal{T}_2$ we can define the following hypothesis:

$$h_m = \{\text{templates } t_1 \text{ and } t_2 \text{ belong to mated instances}\}$$
$$h_{nm} = \{\text{templates } t_1 \text{ and } t_2 \text{ belong to non-mated instances}\}.$$

Moreover, let (T_1, T_2) be random variables each taking values on $\mathcal{T}_1 \times \mathcal{T}_2$ and let H be a random variable taking values on $\mathcal{H} = \{h_m, h_{nm}\}$. In other words, H denotes the true hypotheses about templates T_1 and T_2.

Definition 1 (Maximal Linkability): Maximal linkability of two systems producing templates (T_1, T_2) is defined as

$$\mathrm{M}_{\leftrightarrow}^{sys} = \mathcal{L}(H \to (T_1, T_2)) \tag{9}$$
$$= \log \sum_{(t_1,t_2)\in\mathcal{T}_1\times\mathcal{T}_2} \max\{p(t_1, t_2|h_m), p(t_1, t_2|h_{nm})\}. \tag{10}$$

We can make two observations about maximal linkability in light of Eq. 10. First, since maximal linkability depends only on the conditional distributions $p(t_1, t_2|h_m)$ and $p(t_1, t_2|h_{nm})$, it is independent of the distribution of the hypothesis H. This is a desirable property for a linkability measure since it means that $\mathrm{M}_{\leftrightarrow}^{sys}$ depends on the BTP scheme itself, and not on any assumptions on the distributions of mated and non-mated pairs of templates.

Secondly, from an information-theoretic perspective, it is important to define $\mathrm{M}_{\leftrightarrow}^{sys}$ as we do in Definition 1. This measure is the 'true' linkability score of the system. That is, as we will see in Lemma 2, this score gives us the most general guarantees with fewest assumptions on the behaviour of the adversary. However, to compute $\mathrm{M}_{\leftrightarrow}^{sys}$, it is necessary to know $p(t_1, t_2|h_m)$ and $p(t_1, t_2|h_{nm})$ for all possible values of $(t_1, t_2) \in \mathcal{T}_1 \times \mathcal{T}_2$. This means that if $\mathrm{M}_{\leftrightarrow}^{sys}$ is to be estimated from data, we need to generate a number of samples on the order of $|\mathcal{T}_1||\mathcal{T}_2|$ and this is prohibitive in most practical settings. To circumvent this issue, we follow [17] and propose a linkability measure based on a similarity function. That is, we assume that there is a similarity function

$$\mathsf{s}\colon \mathcal{T}_1 \times \mathcal{T}_2 \to \mathcal{S} \tag{11}$$

which captures the relevant information about the similarity of the two templates. This similarity function could then be used to approximate the linkability score proposed in Definition 1. To this end, we define another linkability measure with respect to a fixed similarity function.

Definition 2 (Maximal s*-Linkability): Let $S = \mathsf{s}(T_1, T_2)$ be a similarity score for templates T_1 and T_2, and a similarity function* s*. Maximal* s*-linkability of two systems producing templates (T_1, T_2) is defined as*

$$\mathrm{M}_{\leftrightarrow}^{\mathsf{s}} = \mathcal{L}(H \to S). \tag{12}$$

Then, for discrete S,

$$\mathrm{M}_{\leftrightarrow}^{\mathsf{s}} = \log \sum_{s\in\mathcal{S}} \max\{p(s|h_m), p(s|h_{nm})\}, \tag{13}$$

and for continuous S,

$$\mathrm{M}_{\leftrightarrow}^{\mathsf{s}} = \log \int_{\mathcal{S}} \max\{p(s|h_m), p(s|h_{nm})\}\,\mathsf{d}\mathsf{s}. \tag{14}$$

Maximal s-linkability generalizes maximal linkability in the following sense. It measures the amount of information revealed by the similarity score S about the two possible hypotheses: the templates are mated, and the templates are not mated. If s is taken to be the identity function, maximal s-linkability reduces to maximal linkability. Thus, just like in [17], the linkability of the system should be evaluated for several similarity functions and the worst-case score should be considered.

Lemma 1: Let s *be any similarity function on $\mathcal{T}_1 \times \mathcal{T}_2$. Then*

$$0 \leq \mathrm{M}_{\leftrightarrow}^{\mathsf{s}} \leq \mathrm{M}_{\leftrightarrow}^{sys} \leq 1. \tag{15}$$

Proof: Eq. 15 follows from Eq. 6, 7, and 8. Specifically, the first inequality follows from Definition 2 and from Eq. 6. In other words, since $\mathrm{M}_{\leftrightarrow}^{\mathsf{s}}$ is an information measure, it cannot be negative. The second inequality follows from the data processing inequality (i.e., Eq. 7) since we have a Markov chain $H \leftrightarrow (T_1, T_2) \leftrightarrow S$. Finally, the last inequality follows from Definition 1 and Eq. 8 since H is a binary-valued random variable. ■

Just like the linkability measure proposed in [17], our measure is supported on [0, 1]. If $\mathrm{M}_{\leftrightarrow}^{sys} = 0$ then the system is completely unlinkable. That is, templates T_1 and T_2 reveal nothing about the hypothesis h_m and h_{nm}. On other hand, $\mathrm{M}_{\leftrightarrow}^{\mathsf{s}} = 1$ means that the system is completely linkable and the adversary could always determine the correct hypothesis after observing T_1 and T_2.

C. Maximal Linkability and Hypothesis Testing

In this section, we interpret $\mathrm{M}_{\leftrightarrow}^{sys}$ and $\mathrm{M}_{\leftrightarrow}^{\mathsf{s}}$ in terms of Neyman-Pearson hypothesis testing. Recall that in this framework, the goal is to design a hypothesis test based on the available data while trading-off two types of errors: *false alarm* error and *missed detection* error. In the present case, the adversary's goal is to distinguish between two hypotheses $\{h_m, h_{nm}\}$, while keeping the two errors small. In the biometrics literature, the false alarm error is also known as *false match rate* (FMR), while the missed detection error is also known as the *false non-match rate* (FNMR). The maximal linkability metrics provide impossibility bounds on the adversary's ability to design well-performing hypothesis tests. If an adversary has access to the protected templates (T_1, T_2), the relevant bound is derived in terms of $\mathrm{M}^{sys}_{\leftrightarrow}$. On the other hand, if an adversary has access to similarity score $S = \mathsf{s}(T_1, T_2)$ only, the relevant bound is derived in terms of $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$. These impossibility bounds are formalized in the following lemmas.

Lemma 2: Suppose $\hat{H}$ is a decision rule for the hypothesis H based on observing (T_1, T_2) and taking values on $\{h_m, h_{nm}\}$. In other words, $H \leftrightarrow (T_1, T_2) \leftrightarrow \hat{H}$. Let

$$FMR = \mathbb{P}\left[\hat{H} = h_m | H = h_{nm}\right]$$
$$\text{and } FNMR = \mathbb{P}\left[\hat{H} = h_{nm} | H = h_m\right]$$

be the False Match and False Non-match Rates for this decision rule. Let $\mathrm{M}^{sys}_{\leftrightarrow}$ be the maximal linkability score of the system. Then

$$(1 - FMR) + (1 - FNMR) \leq 2^{\mathrm{M}^{sys}_{\leftrightarrow}}. \tag{16}$$

The proof of Lemma 2 is given in the Appendix A. The Proof of the following Lemma 3 is identical to the proof of Lemma 2 with the key difference being that the adversary's hypothesis testing is assumed to be done on the similarity score S and not on the protected templates (T_1, T_2).

Lemma 3: Suppose $\hat{H}$ is a decision rule for the hypothesis H based on observing $S = \mathsf{s}(T_1, T_2)$ and taking values on $\{h_m, h_{nm}\}$. In other words, $H \leftrightarrow S \leftrightarrow \hat{H}$. Let

$$FMR = \mathbb{P}\left[\hat{H} = h_m | H = h_{nm}\right]$$
$$\text{and } FNMR = \mathbb{P}\left[\hat{H} = h_{nm} | H = h_m\right]$$

be the False Match and False Non-match Rates for this decision rule. Let $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ be the maximal s-linkability score of the system. Then

$$(1 - FMR) + (1 - FNMR) \leq 2^{\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}}. \tag{17}$$

We see from Lemma 2 that a low value of $\mathrm{M}^{sys}_{\leftrightarrow}$ guarantees that an adversary cannot perform any meaningful hypothesis testing on observed templates T_1 and T_2 to decide if they are mated or non-mated. Likewise, we see from Lemma 3 that a low value of $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ guarantees that an adversary cannot perform any meaningful hypothesis testing on an observed similarity score S to decide if it comes from mated or non-mated templates. These results give an operational interpretation to $\mathrm{M}^{sys}_{\leftrightarrow}$ and $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ an addition to those already provided in [26], see Figure 1.

Figure 2 further illustrates different examples of synthetic scores with Gaussian distributions, and the corresponding ROC curves. For almost overlapping distributions (e.g., Figure 2a) our measure returns a low value (i.e, near zero), while for distributions with less overlap (e.g., Figure 2d) our measure returns a higher value. In addition, we see in all four cases that our measure provides a good upper bound on the true ROC curve of an optimal hypothesis test performed by the adversary.

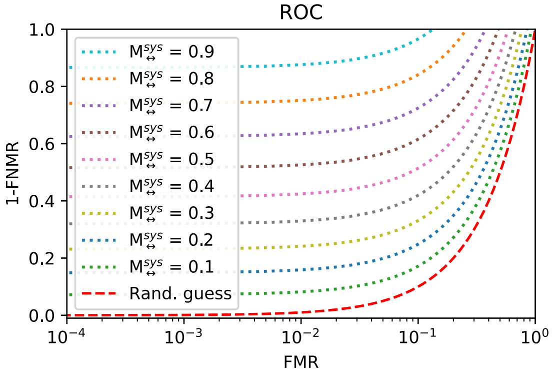


Fig. 1. Bounds on adversary's ability to perform hypothesis testing for different maximal linkability scores. For example, for $\mathrm{M}^{sys}_{\leftrightarrow} = 0.1$, a ROC curve for any hypothesis test that could be performed by an adversary on (T_1, T_2) will be between the dashed red (random guess) and the dotted green ($\mathrm{M}^{sys}_{\leftrightarrow} = 0.1$) curves.

III. Comparison With Other Measures

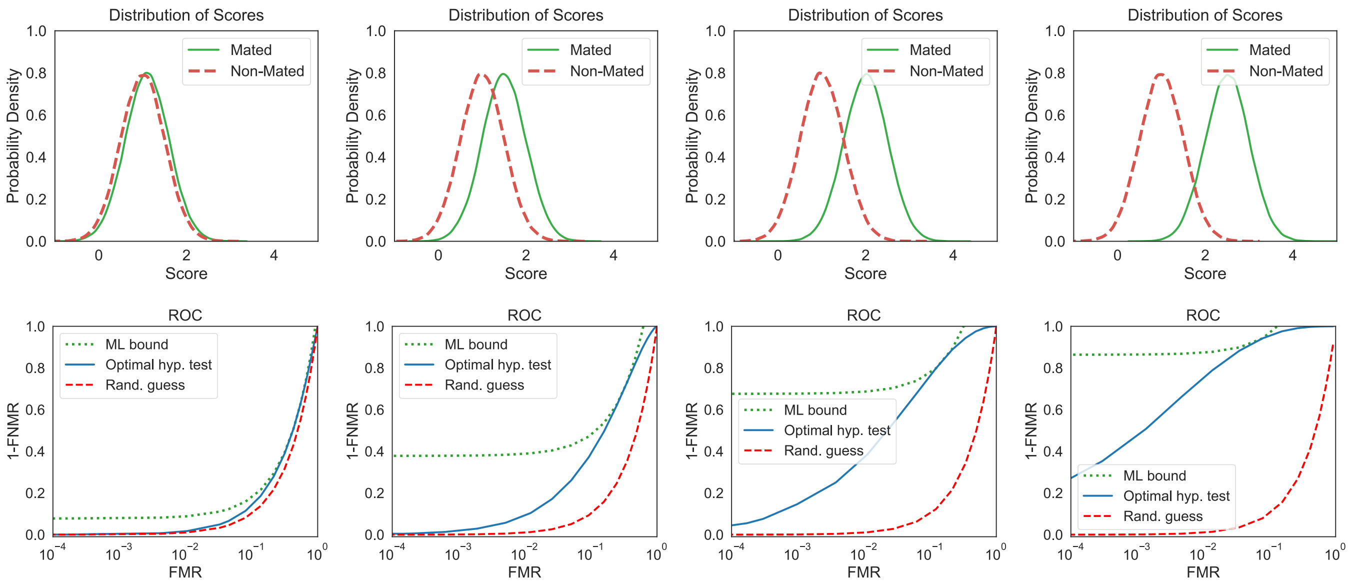
In this section, we compare the proposed measure to other approaches to measuring linkability. In Section III-A, we discuss the implications of using differential privacy as an information measure in the definition of linkability. In Section III-B, we compare our proposed measure to the one from [17], as the most relevant linkability measure in the literature for protected biometric templates.

A. On Linkability via Differential Privacy

The main insight behind the proposed linkability measure is to measure the amount of information leaked by a pair protected biometric templates about whether these templates are mated or not mated. Definitions 1 and 2 use maximal leakage as a measure of such information leakage. This raises the question of whether other measures of privacy loss could be used instead of maximal leakage. In this section, we consider the most prominent such measure: differential privacy [22].

We will show that for ϵ-differential privacy the resulting linkability measure does not differentiate between the four distinct examples in Figure 2. That is, it assigns the value of infinity to all four examples and classifies all four systems as completely linkable. Another possible approach is to apply a common relaxation of ϵ-differential privacy known as (ϵ, δ)-differential privacy. We will show as well, from the example of Figure 2, that this approach does not provide us with a single linkability measurement. Instead, it provides us with a curve trading off between the ϵ and the δ privacy parameters.

1) ϵ-Differential Privacy: Differential privacy is the most prominent approach to privacy that was designed for a private data release problem [22]. In this discussion, we view



(a) (b) (c) (d)

Fig. 2. Synthetic distributions of mated and non-mated scores (first row) and their corresponding ROC plots (second row): (a) mated: $\mathcal{N}(1.1, 0.5)$, non-mated: $\mathcal{N}(1, 0.5)$, and $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow} = 0.1077$, (b) mated: $\mathcal{N}(1.5, 0.5)$, non-mated: $\mathcal{N}(1, 0.5)$, and $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow} = 0.4626$, (c) mated: $\mathcal{N}(2.0, 0.5)$, non-mated: $\mathcal{N}(1, 0.5)$, and $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow} = 0.7450$, (d) mated: $\mathcal{N}(2.5, 0.5)$, non-mated: $\mathcal{N}(1, 0.5)$, and $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow} = 0.8980$. In the ROC plots, the green dotted curves indicate the maximal likability bound for the adversary hypothesis test, the solid blue curves show the optimal possible hypothesis test by the adversary, and the dashed red curves depict the random guess accuracy.

ϵ-differential privacy as an information measure between our true hypothesis H and an observed template pair T_1, T_2, and apply it in the manner similar to Definition 1. In other words, we seek to measure how differentially private the mapping H to (T_1, T_2) is. In this way, we can define a new measure of linkability:

$$\mathcal{DP}(H \to (T_1, T_2)) = \max_{\substack{(t_1,t_2)\in\mathcal{T}_1\times\mathcal{T}_2,\\ h,\hat{h}\in\{h_m,h_{nm}\}}} \log \frac{p(t_1, t_2|h)}{p(t_1, t_2|\hat{h})}. \tag{18}$$

Likewise, for a given similarity function s with continuous scores S, we can define a measure of linkability:

$$\mathcal{DP}(H \to S) = \sup_{\substack{s\in\mathcal{S},\\ h,\hat{h}\in\{h_m,h_{nm}\}}} \log \frac{f(s|h)}{f(s|\hat{h})}. \tag{19}$$

where $f(s|h)$ denotes the probability density function of S given $h \in \{h_m, h_{nm}\}$.

As it turns out, these definitions do not distinguish between any of the cases in Table II and instead classify all of them as fully linkable. In other words, ϵ-differential privacy is too pessimistic for the linkability application. For example, let the score distribution of mated and non-mated templates be any of the four normally distributed pairs in Table II. Then

$$\mathcal{DP}(H \to S) = \infty. \tag{20}$$

This is because the four synthetic distributions in Table II are all examples of a Gaussian additive mechanism applied to a database $\{h_m, h_{nm}\}$. These do not satisfy ϵ-DP according to [22, Theorem A.1].To be more precise, we can take $\mathsf{f}\colon \{h_m, h_{nm}\} \to \{\mu_m, \mu_{nm}\}$ where, for example, $\mu_m = 1.1$ and $\mu_{nm} = 1$ as in Figure 2a. Setting $\delta = 0$ in [22, Theorem A.1] we see that $\epsilon = \infty$.

TABLE II

Linkability of Synthetic Distributions of Scores for Mated and Non-Mated Templates in Figure 2 Using the Measure in [17] (i.e, $\mathrm{D}^{sys}_{\leftrightarrow}$ as in Eq. 24) With Different Values of ω and Our Measure (i.e, $\mathrm{M}^{sys}_{\leftrightarrow}$ as in Eq. 10)

Figure	Mated	Non-Mated	[17] measure			Proposed measure
			$\omega = 0.1$	$\omega = 1$	$\omega = 10$	
Fig. 2a	$\mathcal{N}(1.1, 0.5)$	$\mathcal{N}(1, 0.5)$	0	0.0434	0.8188	0.1077
Fig. 2b	$\mathcal{N}(1.5, 0.5)$	$\mathcal{N}(1, 0.5)$	0.0063	0.2890	0.8357	0.4626
Fig. 2c	$\mathcal{N}(2.0, 0.5)$	$\mathcal{N}(1, 0.5)$	0.2265	0.6111	0.8941	0.7450
Fig. 2d	$\mathcal{N}(2.5, 0.5)$	$\mathcal{N}(1, 0.5)$	0.6027	0.8310	0.9507	0.8980

2) (ϵ,δ)-Differential Privacy: (ϵ, δ)-Differential privacy is a well-studied relaxation of differential privacy which introduces a second parameter δ. We could also consider treating this as an information measure between our true hypothesis H and an observed template pair (T_1, T_2), and apply it in the manner similar to Definition 1. Or, we could consider treating this as an information measure between our true hypothesis H and an observed similarity score S, and apply it in the manner similar to Definition 2. However, in both of these cases we would need to estimate two parameters: ϵ and δ. In general, a BTP scheme will not satisfy (ϵ, δ)-differential privacy for a single pair (ϵ, δ), but would instead satisfy it for an (ϵ, δ) curve.

As an example, take the score distribution of mated and non-mated templates be normally distributed $\mathcal{N}(1.1, 0.5)$ and $\mathcal{N}(1, 0.5)$ as in Figure 2a. Let $c \in [0, \infty]$ be any non-negative constant. Then, mapping from H to S induced by the BTP scheme satisfies (ϵ, δ)-differential privacy with

$$\epsilon > \frac{0.1c}{\sqrt{0.5}} \quad \text{and} \quad \delta > 1.25\ e^{-0.5c^2}. \tag{21}$$

This is again an examples of a Gaussian additive mechanism applied to a database $\{h_m, h_{nm}\}$ where we take $\mathsf{f}\colon \{h_m, h_{nm}\} \to \{\mu_m, \mu_{nm}\}$ with $\mu_m = 1.1$ and $\mu_{nm} = 1$ as in Figure 2a. Applying [22, Theorem A.1] with $\Delta_1 f = 0.1$ and $\sigma = \sqrt{0.5}$ we obtain lower bounds on ϵ and δ in terms of $c \in [0, \infty]$.

As we see from the above discussion, differential privacy does not appear to be an appropriate information measure for the linkability application. In the case of ϵ-differential privacy, it does not differentiate between the simple synthetic examples in Table II and labels all of them completely linkable. On the other hand, in the case of (ϵ, δ)-differential privacy, it is not clear how to obtain a single linkability score.

B. Comparison With Gomez-Barrero et al. [17] Measure

Recall that the first quantitative measure of linkage was introduced in [17]. The main idea of [17] is to base the measure on the distributions of mated and non-mated hypotheses conditioned on a similarity score.

1) Overview of Gomez-Barrero et al. [17] Measure: As mentioned in Section I, Gomez-Barrero et al. [17] proposed two quantitative measures (local and global) based on score distributions. They considered a similarity function s to find the score $s = \mathsf{s}(t_1, t_2) \in \mathcal{S}$ between two templates t_1 and t_2, and found distributions of mated and non-mated pairs. Next, they defined their local measure for each score s in [17, Eq. 4] as:

$$\mathrm{D}_{\leftrightarrow}(s) = p(h_m|s) - p(h_{nm}|s). \tag{22}$$

With some assumptions and simplification, they define their local unlinkability measure in [17, Eq. 14] as:

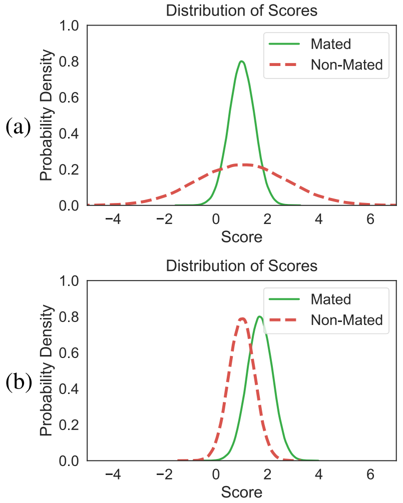
$$\mathrm{D}_{\leftrightarrow}(s) = \begin{cases} 0 & \text{if } LR(s).\omega \le 1 \\ 2\dfrac{LR(s).\omega}{1 + LR(s).\omega} - 1 & \text{if } LR(s).\omega > 1 \end{cases}, \tag{23}$$

where $LR(s) = p(s|h_m)/p(s|h_{nm})$ is the likelihood ratio and $\omega = p(h_m)/p(h_{nm})$ denotes the ratio between the prior probabilities of the mated and non-mated samples. The value of $\omega = 1$, i.e, $p(h_m) = p(h_{nm})$, is proposed as the worst-case scenario. Finally, the global measure $\mathrm{D}^{sys}_{\leftrightarrow}$ is found by calculating the conditional expectation of the local measure $\mathrm{D}_{\leftrightarrow}(s)$ over all comparison scores in [17, Eq. 19] as:

$$\mathrm{D}^{sys}_{\leftrightarrow} = \int p(s|H_m)\mathrm{D}_{\leftrightarrow}(s)\mathsf{d}s. \tag{24}$$

The global measure $\mathrm{D}^{sys}_{\leftrightarrow}$ was the first quantitative evaluation that measures the degree of unlinkability of the biometric systems. In addition to the mathematical definition of $\mathrm{D}^{sys}_{\leftrightarrow}$, [17, Section V] proposes a general protocol for evaluating linkability from data.

2) Comparison With Maximal Linkability: Both $\mathrm{D}^{sys}_{\leftrightarrow}$ (as in Eq. 24) and $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ are based on the similarity score of biometric templates. As discussed in Section II-B, the true linkability of the system is given by $\mathrm{M}^{sys}_{\leftrightarrow}$. However, as this is computationally infeasible in most real-world biometric systems, we follow [17] and focus on computing $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ as proxies for the true linkability. Just like in [17], it is thus important to compute $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ for a number of different similarity scores.



mated: $\mathcal{N}(1, 0.5)$
non-mated: $\mathcal{N}(1, 1.75)$
$\mathrm{M}^{sys}_{\leftrightarrow}$ [ours]: 0.6177
$\mathrm{D}^{sys}_{\leftrightarrow}$ [17]: 0.3902

mated: $\mathcal{N}(1.7, 0.5)$
non-mated: $\mathcal{N}(1, 0.5)$
$\mathrm{M}^{sys}_{\leftrightarrow}$ [ours]: 0.5988
$\mathrm{D}^{sys}_{\leftrightarrow}$ [17]: 0.4289

Fig. 3. Distributions of scores for mated and non-mated templates: (a) mated: $\mathcal{N}(1, 0.5)$, non-mated: $\mathcal{N}(1, 1.75)$, with linkability value of 0.6177 (i.e., somewhat linkable) by our measure (i.e, $\mathrm{M}^{sys}_{\leftrightarrow}$ as in Eq. 10) and 0.3902 (i.e., somewhat unlinkable) by the measure in [17] (i.e, $\mathrm{D}^{sys}_{\leftrightarrow}$ as in Eq. 24). (b) mated: $\mathcal{N}(1.7, 0.5)$, non-mated: $\mathcal{N}(1, 0.5)$, with linkability value of 0.5988 (i.e., somewhat linkable) by our measure (i.e, $\mathrm{M}^{sys}_{\leftrightarrow}$ as in Eq. 10) and 0.4289 (i.e., somewhat unlinkable) by the measure in [17] (i.e, $\mathrm{D}^{sys}_{\leftrightarrow}$ as in Eq. 24). Note that these two systems are ranked differently by our measure and the one in [17].

In addition, maximal linkabilities $\mathrm{M}^{sys}_{\leftrightarrow}$ and $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ as well as $\mathrm{D}^{sys}_{\leftrightarrow}$, are bounded in [0, 1], where 0 indicates full unlinkability and 1 indicates fully linkability. However, maximal linkability and $\mathrm{D}^{sys}_{\leftrightarrow}$ do have a number of significant differences which are highlighted next.

First, while the values of both measures are bounded in [0, 1], the value of maximal linkability is always higher. This result is formalized in the following lemma.

Lemma 4: Assume that $\mathrm{D}^{sys}_{\leftrightarrow}$ is computed using similarity function s and $\omega \le 1$. Then

$$0 \le \mathrm{D}^{sys}_{\leftrightarrow} \le \mathrm{M}^{\mathsf{s}}_{\leftrightarrow} \le \mathrm{M}^{sys}_{\leftrightarrow} \le 1. \tag{25}$$

The proof for $\mathrm{D}^{sys}_{\leftrightarrow} \le \mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ is given in Appendix A, while the other inequalities follow from Lemma 1 and [17]. We highlight that even though $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ is always higher than $\mathrm{D}^{sys}_{\leftrightarrow}$, it is possible for the two measures to give different rankings to two biometric systems. As an example, consider distributions of scores for mated and non-mated pairs as depicted in Figure 3. In this example, the linkability of mated and non-mated templates is 0.6177 by our measure (i.e, $\mathrm{M}^{sys}_{\leftrightarrow}$ as in Eq. 10) and 0.3902 by the measure in [17] (i.e, $\mathrm{D}^{sys}_{\leftrightarrow}$ as in Eq. 24) for system (a). For system (b), the linkability of mated and non-mated templates is 0.5988 by our measure and 0.4289 by the measure in [17].

Secondly, according to Lemmas 2 and 3, maximal linkability has a clear operational interpretation in terms of hypothesis testing capabilities of an adversary. This makes it consistent with the definition of unlinkability in the ISO/IEC 30136 standard [7] presented in Section I. The measure $\mathrm{D}^{sys}_{\leftrightarrow}$ does not appear to have such a hypothesis testing interpretation. Considering again the example in Figure 3, we see that from the hypothesis testing perspective of Lemmas 2 and 3 it is correct to label system (a) as more linkable than system (b). The rational for labeling system (b) as more linkable than system (a) (as is done by $\mathrm{D}_{\leftrightarrow}^{sys}$) is less apparent. In addition, unlike maximal linkability, $\mathrm{D}_{\leftrightarrow}^{sys}$ has a built-in asymmetry where it prioritizes the linkability of mated templates in its definition. While according to the definition of unlinkability in the ISO/IEC 30136 standard [7] given in Section I, a linkability measure should take into account the difficulty of arriving at both, mated and non-mated, hypotheses. From an information-theoretic perspective, understanding that two templates are non-mated could also leak information to the adversary and should not be overlooked by a linkability measure. A closely related issue is that, to prevent (23) from being negative, it is rounded up to zero in certain cases. This rounding again leads to a similar loss of information.

A third difference is that maximal linkability appears to be numerically more stable. For example, to estimate $\mathrm{M}_{\leftrightarrow}^{\mathrm{S}}$ we simply need to estimate the area under the curve of the maximum of mated and non-mated probability density function as in Eq. 14. On the other hand, to calculate $\mathrm{D}_{\leftrightarrow}(s)$ in Eq. 23, it is necessary to estimate the likelihood ratio $LR(s) = p(s|h_m)/p(s|h_{nm})$, which is numerically unstable for low values of $p(s|h_{nm})$. In addition, for estimating $LR(s)$ in practical evaluation in the case of $p(s|h_{nm}) = 0$, the authors considered $LR(s) = 1$ in their open-source implementation[8] which is theoretically incorrect.

Finally, maximal linkability is independent of the prior probabilities of mated and non-mated hypotheses. By contrast, $\mathrm{D}_{\leftrightarrow}^{sys}$ requires the ratio of prior probabilities of the mated and non-mated samples (ω). We further discuss the effect of this assumption in Section III-B.3.

3) Different Values of ω: As mentioned in Section III-B.1, the measure in [17] requires the ratio of prior probabilities of the mated and non-mated samples (i.e., $\omega = p(H_m)/p(H_{nm})$). If we vary the value of ω in this measure, we get counter intuitive results. For small values of ω, clearly linkable systems are characterized as unlinkable. On the other hand, for large values of ω, clearly unlinkable systems are characterized as linkable. Table II reports the linkability measurement of synthesized distributions in Figure 2 using the measure in [17] with different values of ω and our measure. As this table shows, while our linkability measure is independent of prior probabilities, the linkability measure $\mathrm{D}_{\leftrightarrow}^{sys}$ is sensitive to the value ω and thus depends on the prior distributions of mated and non-mated template pairs. This may be an issue for two reasons. First, estimating this prior probability could, in general, be hard. While the authors in [17] considered $\omega = 1$ as the worst-case scenario, such an assumption is not necessarily realistic in many practical cases. In particular, the adversary might have some knowledge about the prior probabilities. For instance, in many practical cases, it is reasonable to assume that non-mated pairs have a higher probability than mated pairs. Secondly, a linkability measure should depend on the BTP scheme and not on the prior belief about the distribution of the hypothesis. Arguably, it makes sense to consider measures that do not depend on the prior probability of H.

TABLE III

SUMMARY OF BTP SCHEMES

BTP scheme	output	score function
BioHashing [34]	binary	Hamming distance*
MLP-Hash [35]	binary	Hamming distance*
Bloom Filters [36]	binary	Normalized† Hamming distance*
IoM-GRP [37]	integer	number of collisions
IoM-URP [37]	integer	number of collisions
HE [38]	ciphertext	Euclidean distance* (in ciphertext)

*To have similarity values, distance functions are multiplied by -1.
†Normalized by the total number of ones in two templates.

IV. EXPERIMENTS

In this section, we describe the experimental results of evaluating the linkability of protected biometric templates using the proposed measure. First, we describe our experimental setup in Section IV-A. Next, we analyze the numerical results of linkability measurement for different BTP schemes, different scoring functions, different characteristics, different feature extractors, and also examples of linkable templates in Section IV-B. Finally, we discuss our experiments in Section IV-C.

A. Experimental Setup

In our experiments, we evaluate the linkability of different BTP schemes on different characteristics (face, voice, and finger vein). We also considered DNN-based (face and voice) and hand-crafted (finger vein) feature extractors in our experiments.

1) BTP Schemes: We measure the linkability of biometric templates, which are protected using different BTP schemes, including BioHashing [34], Multi-Layer Perceptron (MLP) Hashing [35], Bloom Filters [36], two methods based on Index-of-Maximum (IoM) Hashing [37] (i.e., Gaussian random projection-based hashing, shortly GRP, and uniformly random permutation-based hashing, shortly URP), and Homomorphic Encryption (HE) based on Brakerski/Fan-Vercauteren (BFV) [38] algorithm. Table III summarizes the list of BTP schemes and compares their outputs and corresponding scoring functions.

2) Biometric Characteristics: In our experiments, we use different biometric characteristics, including face, voice, and finger vein. We build different biometric recognition systems based on the aforementioned characteristics as follows. Table IV summarises different biometric recognition systems used in our experiments.

a) Face recognition: For face recognition, we use ArcFace-InsightFace [39], ElasticsFace [40], and FaceNet [41] models as different feature extractors and generate mated and non-mated templates from MOBIO [42] dataset. The MOBIO dataset is a bimodal dataset including face and voice data taken with mobile and laptop devices from 150 individuals, captured in 12 sessions (6-11 samples in each session) for each subject. To generate mated scores, we consider all possible combinations of samples for different subjects. For non-mated comparisons, we use the first 10 samples for each subject, and then we consider all possible pairs of samples from different subjects.

b) Voice (speaker) recognition: For voice (speaker) recognition, we use ECAPA-TDNN model [43] as the feature

[8] Available at https://github.com/dasec/unlinkability-metric

TABLE IV

SUMMARY OF BIOMETRIC RECOGNITION SYSTEMS

Characteristic	Feature Extractor	Feature Type	Dataset	# Subjects	# Sessions	# Mated	# Non-mated
Face	ArcFace, ElasticFace, FaceNet	DNN-based	MOBIO (face)	150	12	1,516,300	2,235,000
Voice	ECAPA-TDNN	DNN-based	MOBIO (voice)	150	12	1,516,300	2,235,000
Finger Vein	WLD	Hand-crafted	UTFVP	360	2	216,000	2,067,840

extractor, and use voice data in MOBIO [42] dataset to generate mated and non-mated templates. To generate mated and non-mated scores, we use the same protocol as we use for face recognition.

c) Finger vein recognition: For finger vein recognition, we use Wide Line Detector (WLD) [44] as the feature extractor on the UTFVP [45] finger vein dataset. The UTFVP dataset includes 1440 finger vein images from 60 individuals captured in two identical sessions. For each subject, the vascular patterns of the middle, index, and ring fingers of both hands were collected twice at each session. In our experiments, we consider different fingers for each user as a different data subject (i.e., 6 data subjects corresponding to each individual). For mated comparisons, we generate 10 different protected templates from each unprotected template using different keys. Then, we consider all possible combinations of protected templates for each subject. For non-mated comparisons, we consider all possible pairs of samples from different subjects.

3) Implementation Details and Source-Code: In our experiments, we use the Bob[9] toolbox [46], [47] to both build the biometric recognition systems and generate mated and non-mated pairs. In addition, we use the open-source implementations (in Bob) of the BioHashing, MLP-Hash, IoM-GRP, IoM-URP, Bloom Filters, and HE schemes [35], [48], [49], [50], [51], [52]. For the implementation of HE, we use its implementation in Bob [52] using the SEAL-Python[10] wrapper on Python 3.8 for the C++ SEAL library [53]. The source code of all our experiments is publicly available to help researchers reproduce our results as well as to allow them to use our method to measure the linkability of their own protected templates.[11]

B. Analyze

In this section, we describe our experiments on different biometric recognition systems. We evaluate the linkability of protected templates with different BTP schemes (in Section IV-B.1), based on different scoring functions (in Section IV-B.2), across different characteristics (in Section IV-B.3), and from different feature extractors (in Section IV-B.4). In each experiment, we try to fix all biometrics modules, except only one module.[12] We also evaluate the linkability of exemplary linkable templates in Section IV-B.5, including linkable protected templates (Section IV-B.5.a) and linkable unprotected templates (Section IV-B.5.b).

TABLE V

LINKABILITY OF DIFFERENT BTP SCHEMES FOR ARCFACE TEMPLATES (VALUES IN THE PARENTHESES INDICATE THE RANK OF THE CORRESPONDING BTP SCHEME COMPARED TO OTHER SCHEMES)

BTP scheme	[17] measure	Proposed measure
BioHashing	0.0058 (6)	0.0162 (6)
MLP-Hash	0.0034 (5)	0.0096 (5)
Bloom Filters	0.0002 (1)	0.0007 (1)
IoM-GRP	0.0009 (3)	0.0027 (3)
IoM-URP	0.0008 (2)	0.0022 (2)
HE	0.0018 (4)	0.0053 (4)

1) Linkability Measurement of Different BTP Schemes: In this experiment, we consider the features extracted from face images using the ArcFace model, and apply different BTP schemes, including BioHashing, MLP-Hashing, Bloom Filters, IoM-GRP, IoM-URP, and HE. Table V reports the linkability measurement of protected templates using the measure in [17] and our proposed measure. As this table shows, protected templates by these BTP schemes are almost unlinkable. This table also compares the rank of each BTP scheme compared to other schemes in terms of unlinkability by both measures (ranks are reported in parentheses). As this table shows, both methods rank these schemes the same in terms of the unlinkability of protected templates. However, the values of the measure [17] do not have any interpretation, and it is not clear how significant is the difference in unlinkability of these BTP schemes based on their unlinkability values by measure [17]. Whereas the values of our measure can be interpreted by Lemma 3 by providing an upper bound given the unlinkability value which guarantees that the adversary cannot perform any better hypothesis test than that upper bound. Therefore, each of these BTP schemes leads to a different upper bound for the accuracy of the adversary's hypothesis testing (similar to the upper bounds illustrated in the ROC plots of Figure 1 and Figure 2).

2) Linkability Measurement With Different Scoring Functions: Recall that our proposed measure and the one proposed in [17] are both based on score distributions of mated and non-mated templates. Therefore, as also discussed in Section II, different scoring functions can provide different levels of linkability for protected templates. To evaluate the effect of the scoring function, in this experiment, we generate BioHash-protected templates from the features extracted by

[9] Available at https://www.idiap.ch/software/bob/

[10] Available at https://github.com/Huelse/SEAL-Python

[11] Source code: https://gitlab.idiap.ch/bob/bob.paper.tifs2023_linkability_ml

[12] We consider BioHash-protected templates in our experiments in Sections IV-B.2-IV-B.5, since BioHashing is the simplest BTP scheme in Table III. Similarly, we use face templates in our experiments in Sections IV-B.1, IV-B.2, IV-B.4, and IV-B.5 since face is one of the most popular biometric characteristics. However, we should note that similar experiments with other BTP schemes and other biometric characteristics can be implemented using our open-source paper package.

TABLE VI
LINKABILITY OF BIOHASH-PROTECTED TEMPLATES OF ACRFACE WITH DIFFERENT SCORING FUNCTIONS

Function	[17] measure	Proposed measure
Hamming distance	0.0058	0.0162
Euclidean distance	0.0058	0.0163
Cosine distance	0.0088	0.0245
Kulsinski distance	0.0098	0.0270
Russell-Rao distance	0.0102	0.0280
Sokal-Michener distance	0.0059	0.0166
Correlation distance	0.0059	0.0165

TABLE VII
LINKABILITY OF BIOHASH-PROTECTED TEMPLATES ACROSS DIFFERENT BIOMETRIC CHARACTERISTICS

Characteristic	Feat. Extractor	[17] measure	Proposed measure
Face	ArcFace	0.0058	0.0162
Voice	ECAPA-TDNN	0.0053	0.0149
Finger Vein	WLD	0.0054	0.0153

TABLE VIII
LINKABILITY OF BIOHASH-PROTECTED TEMPLATES FOR DIFFERENT FEATURE EXTRACTORS

Feat. Extractor	[17] measure	Proposed measure
ArcFace	0.0058	0.0162
ElasticFace	0.0049	0.0139
FaceNet	0.0109	0.0302

the ArcFace model from face images. Then, we apply different scoring functions,[13] including Hamming distance, Euclidean distance, Cosine distance, Kulsinski distance, Russell-Rao distance, Sokal-Michener distance, and Correlation distance. Table VI represents the linkability measurement of BioHash-protected templates using the measure in [17] and our proposed measure based on different scoring functions. This table shows that the different scoring functions can lead to different levels of linkability of templates. Therefore, it is necessary to consider different scoring functions when measuring the linkability of protected templates.

3) Linkability Measurement Across Different Biometric Characteristics: To explore the application of our measure on different biometric characteristics, in this experiment, we evaluate the linkability of BioHash-protected templates across different biometric characteristics, including face (ArcFace), voice (ECAPA-TDNN), and finger vein (WLD). Table VII compares the linkability measurement of BioHash-protected templates using the measure in [17] and our proposed measure across different biometric characteristics. This experiment confirms that our measure can be applied to templates with different biometric characteristics, and Table VII show that BioHash-protected templates are almost unlinkable across different biometric characteristics.

4) Linkability Measurement for Different Feature Extractors: To evaluate the effect of the feature extractor, in this experiment, we evaluate the linkability of BioHash-protected templates of face data extracted using different feature extractors, including ArcFace, ElasticFace, and FaceNet. Table VIII compares the linkability measurement of BioHash-protected templates using the measure in [17] and our proposed measure for different feature extractors. As this table shows, BioHash-protected templates are almost unlinkable for these feature extractors.

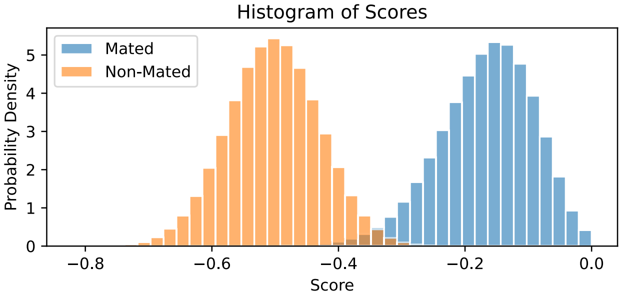


Fig. 4. Histogram of mated and non-mated scores for linkable protected templates (FaceNet templates protected by BioHashing scheme using user-specific keys). The linkability of the mated and non-mated scores in this example is 0.9765 and 0.9574 by our and [17] measure, respectively.

5) Linkability Measurement of Linkable Templates: In our experiments in Sections IV-B.1-IV-B.4, we measured the linkability of protected biometric templates using different BTP schemes across different biometric recognition systems. Our experiments indicate that the protected templates with the aforementioned BTP schemes are almost fully unlinkable. In this section, we consider two examples of linkable protected templates and linkable unprotected templates:

a) Linkable protected templates: As an example of linkable protected templates, we consider FaceNet features protected by the BioHashing scheme using *user-specific* keys. Note that in our experiments in Sections IV-B.1-IV-B.4, we considered *sample-specific* keys for generating protected templates. While considering *user-specific* keys in this experiment may be assumed as a hypothetical scenario, it can reflect the situation where templates with the same key[14] for each user are leaked. For instance, consider a biometric recognition system where multiple protected templates are stored for each user in the system's database (i.e., multiple reference templates). Then, an adversary gains access to all (or a portion of) the templates stored in the system's database, and aims to distinguish mated and non-mated pairs. In such a situation, since mated templates are generated using the *same* key corresponding to the user (i.e., *user-specific* key), there should be a high link between protected templates. Figure 4 depicts the histogram of scores for mated and non-mated templates for FaceNet features protected by the BioHashing scheme. The linkability of mated and non-mated templates in this example is 0.9765 and 0.9574 by our proposed measure and the measure in [17], respectively. Therefore, as also expected from the histogram of scores, these templates are almost fully linkable.

b) Linkability of unprotected templates: In this experiment, we consider an unprotected system, and because no key is applied to generate templates in such systems, we expect to observe a high distinguishability between mated and non-mated templates (as expected from the normal operation

[13] Implementations of all these scoring functions are available in the SciPy package: https://scipy.org

[14] As in the typical operating of protected biometric systems.

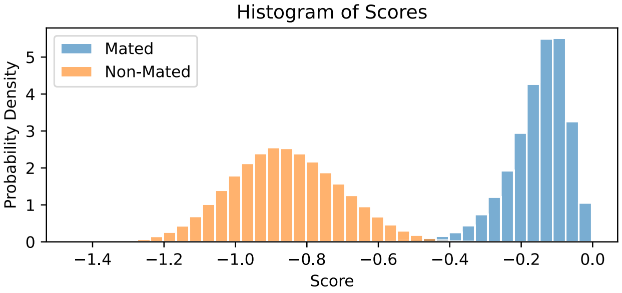


Fig. 5. Histogram of mated and non-mated scores for unprotected templates (FaceNet). The linkability of the mated and non-mated scores in this example is 0.9912 and 0.9669 by our and [17] measure, respectively.

of a biometric recognition system). As an example of such a case, we consider FaceNet features in this experiment. Figure 5 illustrates the histogram of scores for (unprotected) mated and non-mated templates. The linkability of templates for this case is 0.9912 by our proposed measure and 0.9669 by the one in [17]. Therefore, this experiment confirms that unprotected templates are almost fully linkable.

C. Discussions

In our experiments in Sections IV-B.1-IV-B.4, we evaluated the linkability of protected biometric templates. In Section IV-B.1, we observed that our proposed measure and the one proposed in [17] return low values for linkability, and therefore the protected templates with different BTP schemes are almost unlikable based on both measures. Comparing the values for different BTP schemes in Table V, both methods rank the evaluated BTP schemes similarly. While the values for different BTP schemes in each of these measures are close, there is theoretically no interpretation possible for the values of measure [17] and the significance of the difference between the two values in this measure. In contrast, the values of our measure can be interpreted according to Lemma 3, which provides an upper bound for the adversary's hypothesis testing (similar to the upper bounds depicted in the ROC plots of Figure 1 and Figure 2). For example, to compare the linkability of BioHashing and Bloom Filters, we have different values for the linkability measurement of these schemes in Table V, and therefore we have different upper bounds according to Lemma 3. Comparing the corresponding bounds, we can say that if an adversary can gain access to BioHash-protected templates instead of templates protected with Bloom Filters, then the adversary can achieve up to $2^{0.0162} - 2^{0.0007} = 0.0108(\approx 1.1\%)$ more accuracy when performing hypothesis test (i.e., up to 1.1% more accuracy in distinguishing mated and non-mated templates). However, such an exercise cannot be done with [17] because there is no practical interpretation for the linkability values in [17].

The experiment in Section IV-B.2 showed that different scoring functions can provide different levels of linkability for protected templates. This is reasonable since each scoring function compares two given templates differently, and thus provides different information from the similarity of the two templates. Hence, since our proposed measure and the one in [17] are based on score distributions of mated and non-mated templates, different scoring functions lead to different linkability values. Therefore, it is important to consider different scoring functions when evaluating the linkability of protected templates.

In our experiments in Section IV-B.3 and Section IV-B.4, we measured the linkability of BioHash-protected biometric templates across different biometric characteristics and for different feature extractors, respectively. These experiments show that the BioHash-protected biometric templates from different biometric characteristics and from different feature extractors are almost fully unlinkable. This experiment also confirms the application of our measure across different biometric characteristics and for different feature extractors.

In our experiments in Section IV-B.5, we measured the linkability of two systems that we expect to be linkable. In Section IV-B.5.a we considered an example of linkable protected templates where we assumed that *user-specific* keys are used to generate protected templates. Since keys to generate protected templates for each user are the same in this scenario, we should have high linkability between templates, which is also confirmed by our results. As another example of linkable templates, we considered unprotected templates in Section IV-B.5.b. Similarly, in this case, we expect that the templates from the same user be similar and differ from templates of other users, which means a high level of linkability. The result of our linkability measurement also confirms that unprotected templates are almost fully linkable.

All in all, our experiments confirm that our proposed method can be deployed to measure the linkability of protected templates, and the results are intuitively correct. We evaluated the linkability of protected templates using our measure for different BTP schemes, scoring functions, biometric characteristics, and feature extractors. Furthermore, we evaluated two examples of linkable templates, where our measure also showed a high level of linkability. As discussed in Section II our measure has a solid theoretical background, and also the values of our measure have a practical interpretation according to Lemma 3, where our proposed measure can provide an upper bound for the accuracy of the adversary's hypothesis testing given score distributions for mated and non-mated templates.

V. Conclusion

In this paper, we proposed a new method for measuring the linkability of protected biometric templates. We used maximal leakage, which is a well-studied measure in information-theoric literature. Our proposed measure is based on hypothesis testing using the distributions of similarity scores of mated and non-mated protected templates.

The proposed measure is consistent with the definition of linkability in the ISO/IEC 30136 standard and quantifies the linkability degree of protected templates. In particular, we showed that our measure can provide an upper bound on the accuracy of the adversary's hypothesis test given distributions of scores, and guarantees that an adversary cannot achieve better performance than the provided upper bound. The value of our measure is bounded in the [0, 1] interval, where a higher value indicates more linkability (i.e., 0 shows fully unlinkable and 1 shows fully linkable). The proposed method is also computationally stable and does not require any assumptions on prior probabilities of mated or non-mated hypotheses.

We also investigated the application of differential privacy to measure the linkability of protected biometric templates and showed that the differential privacy-based measure is too strict for the linkability application. Last but not least, in our experiments, we used the proposed measure to evaluate the linkability of biometric templates from different biometric characteristics (face, voice, and finger vein), different feature extractors, and protected with different BTP schemes. The experimental implementation of our proposed measure showed that it gives intuitively correct linkability scores across different BTP schemes, biometric characteristics, and scoring functions.

We conclude the discussion with some comments on an important question: how to estimate, $\mathrm{M}^{sys}_{\leftrightarrow}$, the true linkability of the system. In this paper, we adopted the approach of using $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow}$ as proxies for $\mathrm{M}^{sys}_{\leftrightarrow}$. As we see in Lemma 1, the value of $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow}$ is always lower than the value of $\mathrm{M}^{sys}_{\leftrightarrow}$ and it is therefore important to take the highest available value of $\mathrm{M}^{\mathsf{S}}_{\leftrightarrow}$ across different similarity scores. Other approaches to this problem include stronger theoretical analysis of Eq. 7, as well as a more extensive analysis of how well different similarity functions estimate $\mathrm{M}^{sys}_{\leftrightarrow}$. Understanding how to better estimate the true linkability of a system is thus an important direction for future work.

Appendix

A. Proofs for Sections II and III

Proof: [Lemma 2] From Eq. 2 and Eq. 9 we obtain

$$\mathrm{M}^{sys}_{\leftrightarrow} = \sup_{p_H} \log \frac{\mathbb{P}\left[H = \hat{H}\right]}{\max\left\{p_H(h_m),\, p_H(h_{nm}\right\}}. \tag{26}$$

Fixing a distribution $p_H(h_m) = p_H(h_{nm}) = 0.5$ on $\{h_m, h_{nm}\}$,

$$\begin{aligned} \mathrm{M}^{sys}_{\leftrightarrow} &\geq \log \frac{\mathbb{P}\left[H = \hat{H}\right]}{\max\left\{p_H(h_m),\, p_H(h_{nm}\right\}} && (27)\\ &= \log\left(\mathbb{P}\left[H = \hat{H} | H = h_m\right] + \mathbb{P}\left[H = \hat{H} | H = h_{nm}\right]\right) && (28)\\ &= \log\left((1 - \mathrm{FMR}) + (1 - \mathrm{FNMR})\right) && (29) \end{aligned}$$

where Eq. 28 is obtained by applying the law of total probability

$$\begin{aligned} \mathbb{P}\left[H = \hat{H}\right] = \mathbb{P}\left[H = h_m\right]\mathbb{P}\left[H = \hat{H} | H = h_m\right] \\ + \mathbb{P}\left[H = h_{nm}\right]\mathbb{P}\left[H = \hat{H} | H = h_{nm}\right]. \end{aligned} \tag{30}$$

■

Proof: [Lemma 4] The first inequality is shown in [17], while the third and fourth are shown in Lemma 1. It remains to show that $\mathrm{D}^{sys}_{\leftrightarrow} \leq \mathrm{M}^{\mathsf{S}}_{\leftrightarrow}$. Let $\mathcal{F} = \{s \colon p(h_m|s) \geq p(h_{nm}|s\}$ be the set of all the scores for which the mated hypothesis is at least as likely as the non-mated one. Then

$$\begin{aligned} \mathrm{D}_{\leftrightarrow}(s) &= p(h_m|s) - p(h_{nm}|s) && (31)\\ &= p(s|h_m)\frac{p(h_m)}{p(s)} - p(s|h_{nm})\frac{p(h_{nm})}{p(s)} && (32)\\ &= \frac{p(s|h_m)\omega - p(s|h_{nm})}{p(s|h_m)\omega + p(s|h_{nm})} && (33)\\ &\leq \frac{p(s|h_m) - p(s|h_{nm})}{p(s|h_m)} && (34) \end{aligned}$$

where the last line holds only for $\omega \leq 1$. Then

$$\begin{aligned} \mathrm{D}^{sys}_{\leftrightarrow} &= \int p(s|h_m)\mathrm{D}_{\leftrightarrow}(s)\mathrm{d}s && (35)\\ &= \int_{\mathcal{F}} p(s|h_m)\mathrm{D}_{\leftrightarrow}(s)\mathrm{d}s && (36)\\ &\leq \int_{\mathcal{F}} \left[p(s|h_m) - p(s|h_{nm})\right]\mathrm{d}s && (37)\\ &\leq \int_{\mathcal{F}} p(s|h_m)\mathrm{d}s + \int_{\bar{\mathcal{F}}} p(s|h_{nm})\mathrm{d}s - 1 && (38)\\ &= \tilde{D} - 1 && (39) \end{aligned}$$

where we defined $\tilde{D} = \int_{\mathcal{F}} p(s|h_m)\mathrm{d}s + \int_{\bar{\mathcal{F}}} p(s|h_{nm})\mathrm{d}s$. Note that

$$\mathrm{M}^{\mathsf{S}}_{\leftrightarrow} = \log(\tilde{D}) \tag{40}$$

and thus,

$$\mathrm{M}^{\mathsf{S}}_{\leftrightarrow} \geq \log(1 + \mathrm{D}^{sys}_{\leftrightarrow}) \geq \mathrm{D}^{sys}_{\leftrightarrow}, \tag{41}$$

where recall that the logarithm has base two. ■

Acknowledgment

The authors would like to thank Vedrana Krivokuća Hahn (Idiap Research Institute, Switzerland) for discussions and her valuable comments.

References

[1] *Regulation of the European Parliament and of the Council on the Protection of Individuals With Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*, European Council, Brussels, Belgium, Apr. 2016.

[2] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP J. Inf. Secur.*, vol. 2011, no. 1, pp. 1–25, Dec. 2011.

[3] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 88–100, Sep. 2015.

[4] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 54–65, Sep. 2015.

[5] *Information Technology—Security Techniques—Biometric Information Protection, International Organization for Standardization*, Standard ISO/IEC 24745, 2022.

[6] *Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework, International Organization for Standardization and International Electrotechnical Committee*, Standard ISO/IEC 19795-1, 2021.

[7] *Information Technology—Security Techniques—Performance Testing of Biometric Template Protection Schemes, International Organization for Standardization*, Standard ISO/IEC 30136, 2018.

[8] M. Gomez-Barrero and J. Galbally, "Reversing the irreversible: A survey on inverse biometrics," *Comput. Secur.*, vol. 90, Mar. 2020, Art. no. 101700.

[9] *Information Technology—Vocabulary—Part 37: Biometrics, International Organization for Standardization*, Standard ISO/IEC 2382-37, 2022.

[10] I. Buhan, J. Breebaart, J. Guajardo, K. D. Groot, E. Kelkboom, and T. Akkermans, "A quantitative analysis of indistinguishability for a continuous domain biometric cryptosystem," in *Data Privacy Management and Autonomous Spontaneous Security*. Berlin, Germany: Springer, 2009, pp. 78–92.

[11] E. J. C. Kelkboom, J. Breebaart, T. A. M. Kevenaar, I. Buhan, and R. N. J. Veldhuis, "Preventing the decodability attack based cross-matching in a fuzzy commitment scheme," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, pp. 107–121, Mar. 2011.

[12] A. Nagar, K. Nandakumar, and A. K. Jain, "Biometric template transformation: A security analysis," in *Proc. SPIE*, vol. 7541, pp. 237–251, Jan. 2010.

[13] E. Piciucco, E. Maiorana, C. Kauba, A. Uhl, and P. Campisi, "Cancelable biometrics for finger vein recognition," in *Proc. 1st Int. Workshop Sens., Process. Learn. Intell. Mach. (SPLINE)*, Jul. 2016, pp. 1–5.

[14] A. Rua, E. Maiorana, J. Castro, and P. Campisi, "Biometric template protection using universal background models: An application to online signature," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 1, pp. 269–282, Jan. 2012.

[15] M. Ferrara, D. Maltoni, and R. Cappelli, "A two-factor protection scheme for MCC fingerprint templates," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2014, pp. 1–8.

[16] S. Wang and J. Hu, "Design of alignment-free cancelable fingerprint templates via curtailed circular convolution," *Pattern Recognit.*, vol. 47, no. 3, pp. 1321–1329, Mar. 2014.

[17] M. Gomez-Barrero, J. Galbally, C. Rathgeb, and C. Busch, "General framework to evaluate unlinkability in biometric template protection systems," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 6, pp. 1406–1420, Jun. 2018.

[18] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–38, Jun. 2018, doi: 10.1145/3168389.

[19] M. Bloch et al., "An overview of information-theoretic security and privacy: Metrics, limits and applications," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 5–22, Mar. 2021.

[20] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Conf. Theory Cryptogr.* Berlin, Germany: Springer-Verlag, 2006, pp. 265–284, doi: 10.1007/11681878_14.

[21] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology—EUROCRYPT* (Lecture Notes in Computer Science), S. Vaudenay, Ed. Berlin, Germany: Springer, 2006, pp. 486–503.

[22] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014, doi: 10.1561/0400000042.

[23] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy preserving face recognition utilizing differential privacy," *Comput. Secur.*, vol. 97, Oct. 2020, Art. no. 101951.

[24] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *J. Amer. Statist. Assoc.*, vol. 105, no. 489, pp. 375–389, 2010.

[25] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 4037–4049, Jun. 2017.

[26] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1625–1657, Mar. 2020.

[27] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," *Electron. Notes Theor. Comput. Sci.*, vol. 249, pp. 75–91, Aug. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1571066109003077

[28] J. Liao, L. Sankar, F. P. Calmon, and V. Y. F. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 779–783.

[29] F. Farokhi, "Noiseless privacy: Definition, guarantees, and applications," *IEEE Trans. Big Data*, vol. 9, no. 1, pp. 51–62, Feb. 2023.

[30] A. R. Esposito, M. Gastpar, and I. Issa, "Generalization error bounds via Rényi-, f-divergences and maximal leakage," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 4986–5004, Aug. 2021.

[31] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969.

[32] S. Verdu, "α-mutual information," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2015, pp. 1–6.

[33] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. 1. Berkeley, CA, USA: University of California Press, 1961, pp. 547–561.

[34] A. T. B. Jin, D. N. C. Ling, and A. Goh, "BioHashing: Two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognit.*, vol. 37, no. 11, pp. 2245–2255, Apr. 2004.

[35] H. O. Shahreza, V. K. Hahn, and S. Marcel, "MLP-hash: Protecting face templates via hashing of randomized multi-layer perceptron," 2022, *arXiv:2204.11054*.

[36] C. Rathgeb, F. Breitinger, and C. Busch, "Alignment-free cancelable iris biometric templates based on adaptive bloom filters," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–8.

[37] Z. Jin, J. Y. Hwang, Y.-L. Lai, S. Kim, and A. B. J. Teoh, "Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 2, pp. 393–407, Feb. 2018.

[38] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic encryption," *Cryptol. ePrint Arch.*, 2012. [Online]. Available: https://eprint.iacr.org/2012/144

[39] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4690–4699.

[40] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "ElasticFace: Elastic margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1578–1587.

[41] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.

[42] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel, "Session variability modelling for face authentication," *IET Biometrics*, vol. 2, no. 3, pp. 117–129, Sep. 2013.

[43] B. Desplanques, J. Thienpondt, and K. Demuynck, "ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification," in *Proc. Interspeech*, Oct. 2020, pp. 3830–3834.

[44] B. Huang, Y. Dai, R. Li, D. Tang, and W. Li, "Finger-vein authentication based on wide line detector and pattern normalization," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 1269–1272.

[45] B. T. Ton and R. N. J. Veldhuis, "A high quality finger vascular pattern dataset collected using a custom designed capturing device," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–5.

[46] A. Anjos, L. El-Shafey, R. Wallace, M. Gunther, C. McCool, and S. Marcel, "Bob: A free signal processing and machine learning toolbox for researchers," in *Proc. 20th ACM Int. Conf. Multimedia*, Oct. 2012, pp. 1449–1452.

[47] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, S. Marcel, "Continuously reproducing toolchains in pattern recognition and machine learning experiments," in *Proc. ICML Reproducibility Mach. Learn. Workshop*, 2017, pp. 1–8. [Online]. Available: https://openreview.net/forum?id=BJDDItGX-

[48] H. O. Shahreza and S. Marcel, "Towards protecting and enhancing vascular biometric recognition methods via biohashing and deep neural networks," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 3, pp. 394–404, Jul. 2021.

[49] H. O. Shahreza, V. K. Hahn, and S. Marcel, "On the recognition performance of BioHashing on state-of-the-art face recognition models," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2021, pp. 1–6.

[50] H. O. Shahreza and S. Marcel, "Deep auto-encoding and BioHashing for secure finger vein recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2585–2589.

[51] H. O. Shahreza et al., "Benchmarking of cancelable biometrics for deep templates," 2023, *arXiv:2302.13286*.

[52] H. O. Shahreza, C. Rathgeb, D. Osorio-Roig, V. K. Hahn, S. Marcel, and C. Busch, "Hybrid protection of biometric templates by combining homomorphic encryption and cancelable biometrics," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2022, pp. 1–10.

[53] (Nov. 2020). *Microsoft SEAL (Release 3.6)*. Microsoft Research, Redmond, WA, USA. [Online]. Available: https://github.com/Microsoft/SEAL

Hatef Otroshi Shahreza (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the University of Kashan, Iran, in 2016, and the M.Sc. degree in electrical engineering from the Sharif University of Technology, Iran, in 2018. He is currently pursuing the Ph.D. degree with École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is also a Research Assistant with the Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland, where he received the H2020 Marie Skłodowska-Curie Fellowship (TReSPAsS-ETN) for his doctoral program. During his Ph.D., he also experienced being a Visiting Scholar with the Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany, for six months. His research interests include deep learning, machine learning, computer vision, biometrics, and biometric template protection.

Yanina Y. Shkel (Member, IEEE) received the B.S. degree in mathematics and computer science and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Wisconsin–Madison, Madison, WI, USA, in May 2005, August 2010, and December 2014, respectively. She is currently a Scientist with École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, where she is affiliated with the Information Processing Group (IPG). Before joining the Graduate School, she was a Database Developer with Morningstar Inc. In 2013, she spent her time with 3M Corporate Research, as an Intern. From 2014 to 2019, she was a Post-Doctoral Researcher with Princeton University, Princeton, NJ, USA; and the University of Illinois at Urbana–Champaign, Champaign, IL, USA. Her research interests include theoretical aspects of data science, information learning, coding theory, statistics, and cryptography, with a particular focus on applications to privacy and secrecy. She was a recipient of the 2015 NSF Center for Science of Information (CSoI) Post-Doctoral Fellowship and the 2022 Swiss NSF Starting Grant.

Sébastien Marcel (Senior Member, IEEE) received the Ph.D. degree in signal processing from CNET, Université de Rennes I, France, in 2000, and the Research Center of France Telecom (currently Orange Laboratories). He heads the Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland, where he conducts research on face recognition, speaker recognition, vein recognition, attack detection (presentation attacks, morphing attacks, and deepfakes), and template protection. He is a Professor with the School of Criminal Justice, University of Lausanne, and a Lecturer with École Polytechnique Fédérale de Lausanne. He is also the Director of the Swiss Center for Biometrics Research and Testing, which conducts certifications of biometric products.