IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 July 2024, accepted 18 July 2024, date of publication 25 July 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3433536

RESEARCH ARTICLE

On Measuring Linkability of Multiple Protected Biometric Templates Using Maximal Leakage

HATEF OTROSHI SHAHREZA[1,2], (Graduate Student Member, IEEE), YANINA Y. SHKEL[2], (Member, IEEE), AND SÉBASTIEN MARCEL[1,3], (Senior Member, IEEE)

[1]Idiap Research Institute, 1920 Martigny, Switzerland
[2]École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[3]Université de Lausanne (UNIL), 1015 Lausanne, Switzerland

Corresponding author: Hatef Otroshi Shahreza (hatef.otroshi@epfl.ch)

This work was supported in part by the H2020 Marie Skłodowska-Curie TRaining in Secure and PrivAcy-preserving biometricS Early Training Networks (TReSPAsS-ETN) under Grant 860813, and in part by the Swiss NSF under Grant 211337.

ABSTRACT With the rapid development of biometric recognition systems, users can be simultaneously enrolled in multiple biometric recognition systems, either with a single or multiple biometric characteristics (e.g., face, voice, etc.). With such a growth of biometric systems, it is important to secure the sensitive information used within these systems. In particular, considering the privacy issues in such systems, several biometric template protection schemes are proposed in the literature. According to the ISO/IEC 24745 standard, each template protection scheme should satisfy the unlinkability property. While previous measures to evaluate unlinkability were based on two protected templates, the adversary may have access to more information. Such information can correspond to multiple templates from different biometric systems, a single multi-modal biometric system, or even a single unimodal biometric system. In this paper, we focus on measuring the linkability of multiple protected biometric templates, and define maximal linkability in the presence of multiple similarity scores. We define different scenarios where the adversary gains access to multiple similarity scores and evaluate the linkability of protected templates in each scenario. We investigate the theoretical properties of the maximal linkability measure, and compare the theoretical prediction with the calculated linkability of the compositive systems in our experiments. To our knowledge, this is the first work on measuring the linkability of multiple protected biometric templates. The source codes of our measure and all experiments are publicly available.

INDEX TERMS Biometrics, biometric template protection, feature extraction, linkability, maximal leakage, multi-modal, multiple templates, statistical hypothesis testing, template.

I. INTRODUCTION

Biometric recognition systems establish the identity of people based on their physiological (e.g., face, finger vein, iris, fingerprint, etc.), behavioral (e.g., voice, keystroke, signature, gait, etc.), or chemical (e.g., DNA, etc.) traits. Since these traits are unique to each person, biometric recognition systems are considered a reliable alternative to traditional authentication systems (e.g., PIN, passwords, tokens, etc.). Over the last few years, biometric systems have become one of the most popular authentication tool, and their applications range from personal (e.g., smartphone unlocking[1,2]) to large-scale authentications (e.g., national identity systems[3],[4] border controls at airports,[5] etc.). As a matter of fact, the widespread application of biometric systems is due to simultaneously offering a high level of security as well as convenience to their users. The ubiquity of biometric systems is such that nowadays many

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

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

people may be concurrently enrolled in multiple biometric systems with the same or different biometric modalities. Multi-modal biometric recognition systems (also known as multi-biometric systems) are also proposed to provide further security by using multiple biometric modalities in the automatic biometric authentication process [1], [2], [3], [4], [5], [6], [7].

Biometric recognition systems typically extract biometric templates (also known as features or embeddings) from the captured biometric data and store the extracted templates in the system's database during enrollment. During the recognition stage, the newly extracted templates are compared with the ones in the system's database, and recognition is made based on the similarity score. Therefore, biometric templates convey important information about users' identity in a biometric system. Along the same lines, data protection regulations[6] consider biometric data as sensitive information and impose legal obligations to protect biometric data [10]. To this end, several biometric template protection (BTP) schemes have been proposed in the literatures [11], [12], [13], and [14] to protect biometric templates.

The ISO/IEC 24745 standard [15] defines four requirements for each BTP scheme: First, the protected templates should not considerably degrade the accuracy of the biometric recognition system (i.e., recognition accuracy preservation). Second, the protected templates should be non-invertible. In other words, it should be computationally infeasible to reconstruct the original unprotected templates from the protected templates (i.e., irreversibility). Third, it should be possible to cancel a compromised protected template and replace it with a new protected template (i.e., renewability/revocability). Fourth, if two or more protected templates are compromised, it should not be possible to link the protected templates and find if they are from the same subject or different subjects (i.e., unlinkability).

While there are standardized methods for evaluating the recognition accuracy of biometric systems (e.g., ISO/IEC 19795-1 standard [16]), no measure has been standardized in the ISO/IEC 30136 standard [17] for evaluating the irreversibility and unlinkability of protected biometric systems. Several papers have focused on the irreversibility evaluation of protected templates [18], [19]. In the case that the adversary gains access to *multiple* protected templates, there are also some works on the inversion of *multiple* protected templates which investigated if the adversary can estimate the original raw template from *multiple* protected templates [20], [21]. However, there has been less work on evaluating the linkability of protected templates, and to our knowledge, no previous work investigated the linkability of *multiple* protected templates in biometric systems. In this paper, we focus on measuring the linkability of *multiple* biometric protected templates.

[6]Such as the European Union General Data Protection Regulation (EU-GDPR) [8], the Illinois Biometric Information Privacy Act 740 ILCS 14 (BIPA) [9], etc.

Previous works for evaluating the linkability of protected biometric templates considered the scenario where the adversary has *two* templates and aim to determine if these two templates are for the same subject (i.e., *mated* templates) or for the different subjects (i.e., *non-mated* templates) [22]. In [23], [24], [25], [26], and [27] the recognition performance of the protected system in different scenarios is used to evaluate the linkability of protected templates. Buhan et al. [23] and Kelkboom et al. [24] considered two scenarios: i) regular analysis, where templates are protected with a single key, ii) unlinkability analysis, where templates are protected with different keys. Then, to evaluate the linkability of the system, Buhan et al. [23] compared the recognition accuracy of the system in terms of Equal Error Rates (EER) and Kelkboom et al. [24] compared the recognition performance of the system in terms of the Receiving Operating Characteristic (ROC). In both of these methods, if the recognition performance degrades in the unlinkability analysis compared to the regular recognition performance, the system is considered to be unlinkable to some degree. Similar to [24], Nagar et al. [25] used the ROC plot of matching templates with different keys (i.e., second scenario), but they merely used this ROC plot to evaluate the unlinkability of the system. In contrast, Piciucco et al. [26] combined the results of regular analysis (first scenario) and unlinkability analysis (second scenario), and plotted the True Match Rate (TMR) from the unlinkability analysis[7] versus the system's False Non-Match Rate (FNMR) from the regular analysis. Rua et al. [27] also plotted the probability that the adversary can find the correct identity in a top-N list (similar to Cumulative Match Curves (CMC)) and compared the resulting plot with the one corresponding to random guess, as a fully unlinkable situation. However, similar to [23], [24], [25], and [26], they did not quantify the general unlinkability of the system as a single number.

In contrast to accuracy-based methods [23], [24], [25], [26], [27], recent works used score distributions to evaluate the unlinkability of protected templates [22], [28], [29], [30]. Ferrara et al. [28] calculated three distributions of scores for the comparison of two templates protected with different keys, including scores for the comparison of two templates from 1) the same sample (same subject), 2) different samples of the same subject, and 3) samples of different subjects. Then, they evaluated the unlinkability of the templates by visually comparing the overlap of distributions. Wang and Hu [29] only used the last two score distributions and similarly evaluated unlinkability by visual comparison of the score distributions. Similarly, Gomez-Barrero et al. [30] considered two distributions of scores, called mated and non-mated scores, and proposed two quantitative measures (local and global) based on score distributions. As their local measure for each score, they considered the difference in conditional probabilities of the hypothesis that the given score is mated or non-mated. They calculated the local measure using the

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

likelihood ratio of mated and non-mated hypotheses and the ratio of prior probabilities. For their global measure, they calculated the conditional expectation of the local measure over score values. Their proposed global measure was properly defined and bounded in the $[0, 1]$ interval. However, as discussed in [22], it has several drawbacks. First, the proposed global measure does not have any operational interpretation, which makes difficult to understand the level of linkability of a system after calculating the global measure in [30]. In particular, when comparing two systems, it is not clear the significance of the difference in linkability of the systems given the values of measure. Second, the method [30] is dependent on the prior probabilities of mated and non-mated hypotheses. Finally, it is necessary to estimate likelihood ratios in this method [30], which makes it numerically unstable for low values of probability of being non-mated (as it appears in the denominator of the likelihood ratio).

In [22], score distributions of mated and non-mated templates was used with maximal leakage from information-theoretic literature [31] to propose a measure (called maximal linkability) for evaluating the linkability of protected biometric templates. The proposed measure in [22] is also properly defined and bounded in the $[0, 1]$ interval. In particular, the proposed measure has a number of important theoretical properties and an appealing operational interpretation in terms of statistical hypothesis testing. More precisely, it is shown in [22] that the proposed measure gives a theoretical upper bound on the adversary's hypothesis test and guarantees that an adversary cannot achieve higher accuracy than the resulting bound.

In this paper, we build upon the previous measure proposed in [22] for measuring the linkability of protected biometric templates and extend it to situations where the adversary has access to multiple (two or more) similarity scores. We define different scenarios in which the adversary gains access to two or more protected templates within a single biometric system or across multiple biometric systems with the same enrolled user. Then, the adversary will have more than a single score value for the hypothesis testing task of determining if the templates are mated or not. In general, this means that the adversary has more information and the associated linkability score should be higher. We investigate how this more general setting degrades the linkability guarantees of the system in terms of theoretical properties of the maximal linkability measure. Then, we compare the theoretical prediction with the actual linkability of the compositive systems. In our experiments, we use different BTP schemes, different biometric modalities (face and voice), and state-of-the-art deep neural network feature extractors, to evaluate the linkability of protected templates with multiple similarity scores. We explore the linkability in biometric systems considering leakage from different biometric systems, in a multi-modal biometric system, or in different stages of a single biometric system.

In summary, the contributions of the paper are as follows:

- This paper presents the first study on measuring the linkability of multiple protected biometric templates. We define different scenarios where the adversary gains access to two or more protected templates within a single biometric system or across multiple biometric systems.
- We extend the definition of maximal linkability [22] to the scenario where the adversary gains access to multiple similarity scores, and evaluate the linkability of protected systems in such scenarios.
- We investigate the theoretical properties of the maximal linkability measure for multiple similarity scores, and compare the theoretical prediction with the calculated linkability of the compositive systems in our experiments.

The remainder of this paper is organized as follows. In Section II, we present the notations used in the paper and define the problem of linkability evaluation based on multiple similarity scores. In Section III, we propose a method for measuring the linkability with multiple similarity scores based on our previous measure proposed in [22]. In Section IV, we report our experiments for different scenarios defined in Section II for biometric systems and evaluate the linkability of the protected templates using our method. Finally, the paper is concluded in Section V.

II. PROBLEM DEFINITION AND FORMULATION

A. PAPER NOTATION

Throughout the paper, we denote a biometric sample (e.g., face image or voice signal, etc.) captured by a sensor with C and a feature extractor with $\mathsf{e}(\cdot)$. We also denote the features extracted from C by the feature extractor with $\mathsf{e}(\cdot)$ as unprotected templates with U. In a protected biometric system, a template protection scheme $\mathsf{p}(\cdot, \cdot)$ is applied to each unprotected template U to generate protected template $T = \mathsf{p}(U, k)$, where k is a key. Let us also denote a protected biometric system with $\mathsf{b} = \mathsf{p} \circ \mathsf{e}$, which generates a protected template $T = \mathsf{b}(C, k)$ from the biometric sample C and key k. We also denote a scoring function with $\mathsf{s}(\cdot, \cdot)$ to compare two protected templates T_1 and T_2 and find the similarity score $S = \mathsf{s}(T_1, T_2)$. We distinguish different templates of the same subject with different indices. For example, $T_{1,1}$ and $T_{1,2}$ indicate two templates of subject 1.

B. DIFFERENT SCENARIOS WITH MULTIPLE SIMILARITY SCORES

Fig. 1 shows a general block diagram of a protected biometric recognition system. Based on this block diagram, we consider different scenarios (denoted with Sc.), in which the adversary may have multiple scores from templates leaked from different points in biometric systems and aims to find the linkability of the protected templates.

SC. 1: DIFFERENT BIOMETRIC MODALITIES

We have two biometric modalities, e.g., face and voice. We have two samples captured for the face ($C_{1,f}$ and $C_{2,f}$)

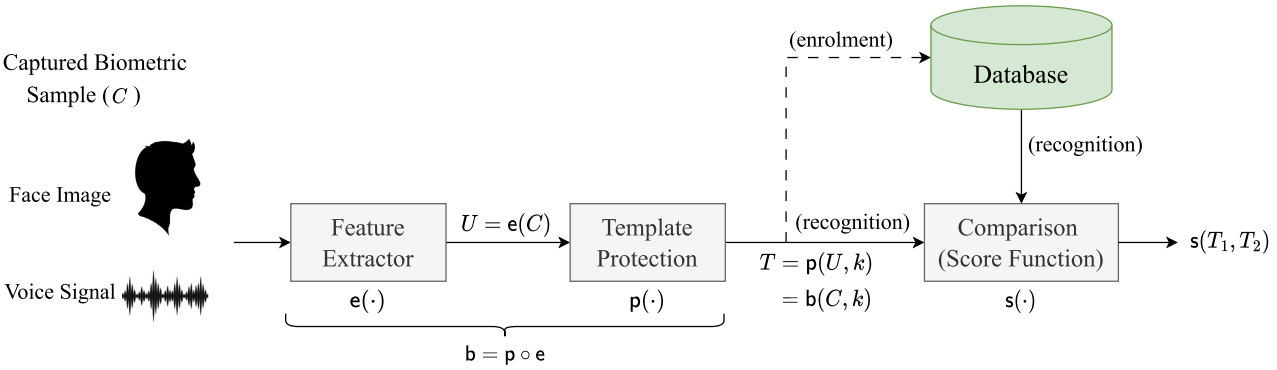


FIGURE 1. General block diagram of a biometric recognition system.

and two samples captured for voice ($C_{1,v}$ and $C_{2,v}$) and know that $C_{1,f}$ and $C_{1,v}$ are from the same person (subject 1), and also $C_{2,f}$ and $C_{2,v}$ are from the same person (subject 2). These samples are used in two biometric recognition systems[8] $\mathsf{b}_f(\cdot)$ (i.e., face recognition) and $\mathsf{b}_v(\cdot)$ (i.e., speaker recognition), yielding $T_{1,1} = \mathsf{b}_f(C_{1,f}, k_{1,1})$ and $T_{1,2} = \mathsf{b}_v(C_{1,v}, k_{1,2})$ as well as $T_{2,1} = \mathsf{b}_f(C_{2,f}, k_{2,1})$ and $T_{2,2} = \mathsf{b}_v(C_{2,v}, k_{2,2})$. We use two scoring functions for these two biometric systems $\mathsf{s}_f(\cdot,\cdot)$ and $\mathsf{s}_v(\cdot,\cdot)$ and have $S_1 = \mathsf{s}_f(T_{1,1}, T_{2,1})$ and $S_2 = \mathsf{s}_v(T_{1,2}, T_{2,2})$. We would like to determine whether an adversary can say if $C_{1,f}$ and $C_{2,f}$ (and similarly $C_{1,v}$ and $C_{2,v}$) are for the same person or not, given S_1 and S_2? (i.e., subjects 1 and 2 are the same person or not?)

SC. 2: DIFFERENT FEATURE EXTRACTIONS

We have biometric samples (e.g., two face images) C_1 and C_2, and extract features with two different feature extractors ($\mathsf{e}_1(\cdot)$ and $\mathsf{e}_2(\cdot)$), yielding $U_{1,1} = \mathsf{e}_1(C_1)$ and $U_{1,2} = \mathsf{e}_2(C_1)$ as well as $U_{2,1} = \mathsf{e}_1(C_2)$ and $U_{2,2} = \mathsf{e}_2(C_2)$. We protect each extracted feature and have four protected templates $T_{1,1} = \mathsf{p}(U_{1,1}, k_{1,1})$, $T_{1,2} = \mathsf{p}(U_{1,2}, k_{1,2})$, $T_{2,1} = \mathsf{p}(U_{2,1}, k_{2,1})$, and $T_{2,2} = \mathsf{p}(U_{2,2}, k_{2,2})$. We use a scoring function $\mathsf{s}(\cdot,\cdot)$ and have $S_1 = \mathsf{s}(T_{1,1}, T_{1,2})$ and $S_2 = \mathsf{s}(T_{2,1}, T_{2,2})$. We would like to determine whether an adversary can find if C_1 and C_2 are for the same person or not, given S_1 and S_2?

SC. 3: DIFFERENT TEMPLATE PROTECTION SCHEMES

We have two unprotected templates, U_1 and U_2, which are protected with two different BTP schemes $\mathsf{p}_1(\cdot,\cdot)$ and $\mathsf{p}_2(\cdot,\cdot)$, yielding $T_{1,1} = \mathsf{p}_1(U_1, k_{1,1})$ and $T_{1,2} = \mathsf{p}_2(U_1, k_{1,2})$ as well as $T_{2,1} = \mathsf{p}_1(U_2, k_{2,1})$ and $T_{2,2} = \mathsf{p}_2(U_2, k_{2,2})$. We use two scoring functions, $\mathsf{s}_1(\cdot,\cdot)$ and $\mathsf{s}_2(\cdot,\cdot)$ correspond to BTP scheme $\mathsf{p}_1(\cdot,\cdot)$ and $\mathsf{p}_2(\cdot,\cdot)$, respectively, and have $S_1 = \mathsf{s}_1(T_{1,1}, T_{1_2})$ and $S_2 = \mathsf{s}_2(T_{2,1}, T_{2,2})$. We aim to determine whether an adversary can find if U_1 and U_2 are for the same person or not, given S_1 and S_2?

[8] We can also consider a bi-modal bometric recognition system which uses face and voice data for recognition, and thus extracts separate templates from face and voice samples.

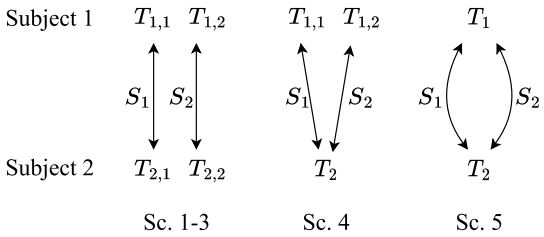


FIGURE 2. Different scenarios where we have two scores from different leaked templates.

SC. 4: DIFFERENT PROTECTED TEMPLATES

We have three protected templates $T_{1,1}$, $T_{1,2}$, and T_2 and also a scoring function $\mathsf{s}(\cdot,\cdot)$. We know that $T_{1,1} = \mathsf{p}(U_1, k_{1,1})$, $T_{1,2} = \mathsf{p}(U_1, k_{1,2})$ are for the same person (U_1) and aim to determine whether having $S_1 = \mathsf{s}(T_{1,1}, T_2)$ and $S_2 = \mathsf{s}(T_{1,2}, T_2)$, an adversary can find if $T_2 = \mathsf{p}(U_2, k_2)$ is also for the same person or not? (i.e., subjects 1 and 2 are the same person or not?)

SC. 5: DIFFERENT SCORING FUNCTIONS

We have two protected templates $T_1 = \mathsf{p}(U_1, k_1)$ and $T_2 = \mathsf{p}(U_2, k_2)$ and also two scoring functions $\mathsf{s}_1(\cdot,\cdot)$ and $\mathsf{s}_2(\cdot,\cdot)$. We would like to determine whether having $S_1 = \mathsf{s}_1(T_1, T_2)$ and $S_2 = \mathsf{s}_2(T_1, T_2)$, an adversary can find if T_1 and T_2 are for the same person (mated) or for different persons (non-mated)? Note that in contrast to scenarios 1-4, in this scenario the adversary does not have any additional knowledge about any potential links between leaked templates, but uses two scoring functions to facilitate the hypothesis testing task.

Fig. 2 illustrates different scenarios where we have two scores from different leaked templates. It is worth mentioning that each of these scenarios can be extended to any number of templates/scores or combined with other scenarios. For example, by combining Sc. 4 and Sc. 5, with two scoring functions we can have four similarity scores: $s_1 = S_1(T_{1,1}, T_2)$, $s_2 = S_1(T_{1,2}, T_2)$, $s_3 = S_2(T_{1,1}, T_2)$, and $s_4 = S_2(T_{1,2}, T_2)$. For simplicity, we do not discuss such combinations in this paper, however, the proposed method can be extended for such scenarios too.

III. MEASURING LINKABILITY USING MULTIPLE SIMILARITY SCORES

In this section, we overview the theoretical properties of maximal linkability which are relevant to the present setting of multiple templates and similarity scores. First, we review the definition of maximal linkability introduced in [22]. Then, we discuss the properties of maximal linkability which follow from well known properties of maximal leakage and composition across multiple views. Finally, we address the behavior of maximal linkability with respect to the five composition scenarios outlined in Section II.

A. MAXIMAL LINKABILITY

Maximal linkability as introduced in [22] is based on an information-theoretic measure called maximal leakage [31, Theorem 1]. Maximal linkability measures the amount of information revealed by two templates about the two possible hypotheses: the templates are mated, and the templates are not mated. That is, 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 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. Maximal linkability of two systems producing templates (T_1, T_2) is then given as

$$\mathrm{M}^{sys}_{\leftrightarrow} = \mathcal{L}(H \to (T_1, T_2)) \tag{1}$$
$$= \log \sum_{(t_1,t_2)\in\mathcal{T}_1\times\mathcal{T}_2} \max\{p(t_1, t_2|h_m), p(t_1, t_2|h_{nm})\} \tag{2}$$
$$.$$

The biometric systems 1 and 2 could be two different instances of the same simple system; this is the scenario studied in [22]. They could also be different systems like the ones in Section II.

In general, estimating $\mathrm{M}^{sys}_{\leftrightarrow}$ can be an intractable problem since the space $\mathcal{T}_1 \times \mathcal{T}_2$ is very large. One way to overcome this is to estimate linkability through a similarity score. 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 in [22] as maximal leakage between two hypothesis:

$$\mathrm{M}^{\mathsf{s}}_{\leftrightarrow} = \mathcal{L}(H \to S). \tag{3}$$

That is, for discrete S,

$$\mathrm{M}^{\mathsf{s}}_{\leftrightarrow} = \log \sum_{s\in\mathcal{S}} \max\{p(s|h_m), p(s|h_{nm})\}, \tag{4}$$

and for continuous S,

$$\mathrm{M}^{\mathsf{s}}_{\leftrightarrow} = \log \int_{\mathcal{S}} \max\{p(s|h_m), p(s|h_{nm})\}\, \mathsf{ds}. \tag{5}$$

Maximal linkabilities $\mathrm{M}^{sys}_{\leftrightarrow}$ and $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ have strong interpretations in terms of bounds on the False Match Rate (FMR) and False Non-match Rate (FNMR) in the hypothesis test defined above. We refer the reader to [22] for more in-depth discussion of this interpretation.

Finally, we remark that maximal linkability $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$ is well defined even when the similarity scores is a vector. For example, if the similarity score $s = (s_1, s_2)$ is a tuple, then Eq. (4) becomes

$$\log \sum_{(s_1,s_2)\in\mathcal{S}_1\times\mathcal{S}_2} \max\{p(s_1, s_2|h_m), p(s_1, s_2)|h_{nm})\}, \tag{6}$$

and Eq. (5) becomes

$$\log \int_{\mathcal{S}_1\times\mathcal{S}_2} \max\{p(s_1, s_2|h_m), p(s_1, s_2|h_{nm})\}\, \mathsf{ds}. \tag{7}$$

An example for a two-dimensional synthetic distribution is provided in Fig. 3. Likewise, the same hypothesis testing interpretation from [22] holds for the example in Fig. 3.

B. PROPERTIES OF MAXIMAL LEAKAGE

In addition to the operational interpretation in terms of hypothesis testing, maximal linkability inherits theoretical properties from maximal leakage. For example, we know that

$$0 \le \mathrm{M}^{\mathsf{s}}_{\leftrightarrow} \le \mathrm{M}^{sys}_{\leftrightarrow} \le 1, \tag{8}$$

see [22, Lemma 1]. Specifically, two important groups of such theoretical properties are *data processing inequalities* and *composition theorems*. These properties generally hold for most of the well-known measures of information (e.g. mutual information), as well as privacy (e.g. differential privacy). Such properties are useful in analyzing the behavior of $\mathrm{M}^{sys}_{\leftrightarrow}$ and $\mathrm{M}^{\mathsf{s}}_{\leftrightarrow}$, as we discuss next.

1) DATA PROCESSING INEQUALITY

Intuitively, the data processing inequality says that new information cannot be gained by processing an observation. It is well known that maximal leakage satisfies the data processing inequality, see for example [31]. That is, if $X \leftrightarrow Y \leftrightarrow Z$ are random variables which form a Markov chain, then

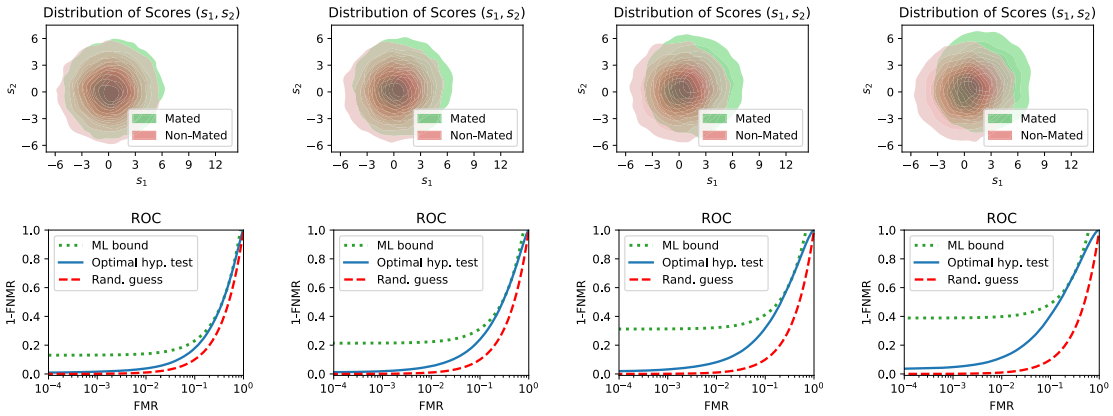
$$\mathcal{L}(X \to Z) \le \mathcal{L}(X \to Y), \tag{9}$$

and

$$\mathcal{L}(X \to Z) \le \mathcal{L}(Y \to Z). \tag{10}$$

In terms of biometric systems, (9) could be alternatively stated as

$$\mathcal{L}(H \to S) \le \mathcal{L}(H \to (T_1, T_2)). \tag{11}$$



	(a)	(b)	(c)	(d)
Mated	$\mathcal{N}\left(\begin{bmatrix}0.50\\0.25\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}1.0\\0.50\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}1.5\\0.75\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}2.0\\1.0\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$
Non-mated	$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}5 & 0\\0 & 5\end{bmatrix}\right)$
$\mathrm{M}_{\leftrightarrow}^{\mathsf{s}_1}, \mathrm{M}_{\leftrightarrow}^{\mathsf{s}_2}$	0.1158, 0.0683	0.2372, 0.1353	0.3339, 0.1868	0.4235, 0.2355
$\mathrm{M}_{\leftrightarrow}^{(\mathsf{s}_1,\mathsf{s}_2)}$	0.1766	0.2793	0.3918	0.4735

FIGURE 3. 2D histograms of synthetic distributions of mated and non-mated scores (first row) and their corresponding ROC plots (second row). The distribution of mated and non-mated scores are indicated in the third and fourth rows, respectively. In each case, the value of maximal linkability for each individual score (i.e., $\mathrm{M}_{\leftrightarrow}^{\mathsf{s}_1}$ and $\mathrm{M}_{\leftrightarrow}^{\mathsf{s}_2}$) and for the tuple of $s = (s_1, s_2)$, (i.e., $\mathrm{M}_{\leftrightarrow}^{(\mathsf{s}_1,\mathsf{s}_2)}$) is also indicated in the fifth and sixth rows, respectively. The linkability $\mathrm{M}_{\leftrightarrow}^{(\mathsf{s}_1,\mathsf{s}_2)}$ for the joint similarity score is always higher than the linkability for individual scores since in the joint case more information is available to the adversary. In the ROC plots, the green dotted curves indicate the maximal likability bound for the adversary hypothesis test (from Lemma 1), the solid blue curves show the optimal possible hypothesis test by the adversary, and the dashed red curves depict the random guess accuracy.

This is because (T_1, T_2) is processed to obtain a scoring function output S; thus, the amount of information in S about H cannot be greater than in (T_1, T_2).

Given any similarity function s on $\mathcal{T}_1 \times \mathcal{T}_2$, the second inequality in Eq. 8 follows from the data processing inquality and in general $\mathrm{M}_{\leftrightarrow}^{\mathsf{s}}$ underestimates the true linkability $\mathrm{M}_{\leftrightarrow}^{sys}$. Now, suppose that we have two scoring functions. That is $S_1 = \mathsf{s}_1(T_1, T_2)$ and $S_2 = \mathsf{s}_2(T_1, T_2)$. Then

$$H \leftrightarrow (T_1, T_2) \leftrightarrow (S_1, S_2) \leftrightarrow S_1 \tag{12}$$

and

$$H \leftrightarrow (T_1, T_2) \leftrightarrow (S_1, S_2) \leftrightarrow S_2. \tag{13}$$

From this we have that

$$\max\left(\mathrm{M}_{\leftrightarrow}^{\mathsf{s}_1}, \mathrm{M}_{\leftrightarrow}^{\mathsf{s}_2}\right) \leq \mathrm{M}_{\leftrightarrow}^{(\mathsf{s}_1,\mathsf{s}_2)} \leq \mathrm{M}_{\leftrightarrow}^{sys}. \tag{14}$$

In other words, considering two scoring functions gives us a better estimate of the true linkability than individual scores.

Similar to [22, Lemma 3], we can bound the adversary hypothesis test using two similarity scores as follows:

Lemma 1: Suppose $\hat{H}$ is a decision rule for the hypothesis H based on observing $S_1 = \mathsf{s}_1(T_1, T_2)$ and $S_2 = \mathsf{s}_2(T_1, T_2)$ and taking values on $\{h_m, h_{nm}\}$. In other words, $H \leftrightarrow (S_1, S_2) \leftrightarrow \hat{H}$. Let

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

be the False Match and False Non-match Rates for this decision rule. Let $\mathrm{M}_{\leftrightarrow}^{(\mathsf{s}_1,\mathsf{s}_2)}$ be the maximal s-linkability score of the system. Then

$$(1 - \mathrm{FMR}) + (1 - \mathrm{FNMR}) \leq 2^{\mathrm{M}_{\leftrightarrow}^{(\mathsf{s}_1,\mathsf{s}_2)}}. \tag{15}$$

Fig. 3 illustrates examples for two-dimensional synthetic distributions of scores and their corresponding maximal linkability.

2) COMPOSITION THEOREMS

Composition theorems track how much an information or privacy measure changes if multiple noisy views of the event of interest are available. For example, a composition theorem for maximal leakage [31, Lemma 6] says that if $Z_1 \leftrightarrow X \leftrightarrow Z_2$ form a Markov chain, then

$$\mathcal{L}(X \to (Z_1, Z_2)) \leq \mathcal{L}(X \to Z_1) + \mathcal{L}(X \to Z_2). \tag{16}$$

On the other hand, if X, Z_1, Z_2 do not satisfy the Markov chain condition, we can still say that [31, Corollary 2]

$$\mathcal{L}(X \to (Z_1, Z_2)) \leq \mathcal{L}(X \to Z_1) + \mathcal{L}(X \to Z_2 | Z_1), \tag{17}$$

where $\mathcal{L}(X \rightarrow Z_2|Z_1)$ denotes the so-called *conditional maximal leakage*.

In terms of biometric systems, if we have that

$$S_1 \leftrightarrow (\tilde{T}_1, \tilde{T}_2) \leftrightarrow H \leftrightarrow (T_1, T_2) \leftrightarrow S_2, \tag{18}$$

then from (16)

$$\mathrm{M}_{\leftrightarrow}^{(\mathsf{S}_1,\mathsf{S}_2)} \leq \mathrm{M}_{\leftrightarrow}^{\mathsf{S}_1} + \mathrm{M}_{\leftrightarrow}^{\mathsf{S}_2}. \tag{19}$$

On the other hand, if we just have that

$$H \leftrightarrow (T_1, T_2) \leftrightarrow S_1 \text{ and } H \leftrightarrow (\tilde{T}_1, \tilde{T}_2) \leftrightarrow S_2, \tag{20}$$

then the Markov chain condition $(\tilde{T}_1, \tilde{T}_2) \leftrightarrow H \leftrightarrow (T_1, T_2)$ may not hold. We need to use the bound in (17) to obtain

$$\mathrm{M}_{\leftrightarrow}^{(\mathsf{S}_1,\mathsf{S}_2)} \leq \mathrm{M}_{\leftrightarrow}^{\mathsf{S}_1} + \mathcal{L}(H \rightarrow S_2|S_1), \tag{21}$$

where $\mathcal{L}(H \rightarrow S_2|S_1)$ is

$$\log \max_{s_1 \in \mathcal{S}_1} \sum_{s_2 \in \mathcal{S}_2} \max \{p(s_2|s_1, h_m), p(s_2|s_1, h_{nm})\}. \tag{22}$$

Therefore, by having access to two similarity scores, the adversary cannot learn more than the true linkability of the system. However, they may learn more than simply the sum of these two linkability.

C. MAXIMAL LINKABILITY FOR MULTIPLE SIMILARITY SCORES

In this section we overview how the properties of maximal linkability could be applied to the five different scenarios in Section II.

DIFFERENT SCORING FUNCTIONS

In Sc. 5 described in Section II, the adversary observes two scores, S_1 and S_2 from template pair (T_1, T_2). By applying the data processing inequality we can thus obtain that this scenario satisfies Eq. 14. While it is not possible to get an upper bound on true linkability $\mathrm{M}_{\leftrightarrow}^{sys}$ by observing S_1 and S_2, the joint linkability $\mathrm{M}_{\leftrightarrow}^{(\mathsf{S}_1,\mathsf{S}_2)}$ gives a better estimate of $\mathrm{M}_{\leftrightarrow}^{sys}$ than $\max\left(\mathrm{M}_{\leftrightarrow}^{\mathsf{S}_1}, \mathrm{M}_{\leftrightarrow}^{\mathsf{S}_2}\right)$. Moreover, if S_1 or S_2 really capture the relevant information about the linkability of (T_1, T_2), one would expect that considering them jointly would not lead to a significant increase in adversary's ability to link that two templates.

DIFFERENT PROTECTED TEMPLATES

Sc. 4 described in Section II behaves similarly to Sc. 5. The difference is in Sc. 4 two protected templates are available for the first user instead of one template in Sc. 5. That is, the adversary has $(T_{1,1}, T_{1,2})$ for the first user, and T_2 for the second user. Thus,

$$H \leftrightarrow \left((T_{1,1}, T_{1,2}), T_2\right) \leftrightarrow (T_{1,i}, T_2), \quad i \in \{1, 2\}. \tag{23}$$

From the data processing inequality we see that the linkability of such a system will be generally higher than $\mathrm{M}_{\leftrightarrow}^{sys}$ since, by definition,

$$\mathrm{M}_{\leftrightarrow}^{sys} = \mathcal{L}(H \rightarrow (T_{1,1}, T_2)) = \mathcal{L}(H \rightarrow (T_{1,2}, T_2)). \tag{24}$$

It is also no longer possible to determine how the joint linkability $\mathrm{M}_{\leftrightarrow}^{(\mathsf{S}_1,\mathsf{S}_2)}$ will relate to $\mathrm{M}_{\leftrightarrow}^{sys}$. On the one hand, it could be lower since some information is lost by considering the scoring functions instead of the templates. On the other hand, it could be higher since the adversary learns something about two different linked templates $(T_{1,1}, T_{1,2})$.

With respect to the data processing inequality, Sc. 2 (i.e., different feature extractors) and Sc. 3 (i.e., different template protection schemes) described in Section II also resemble Sc. 4. That is, the adversary has access to two templates from each user and thus access to more information. In general, the linkability of this overall composite systems will be higher than $\mathrm{M}_{\leftrightarrow}^{sys}$, where

$$\mathrm{M}_{\leftrightarrow}^{sys} = \mathcal{L}(H \rightarrow (T_{1,1}, T_{2,1}) = \mathcal{L}(H \rightarrow (T_{1,2}, T_{2,2})). \tag{25}$$

It is also not clear how the joint linkability $\mathrm{M}_{\leftrightarrow}^{(\mathsf{S}_1,\mathsf{S}_2)}$ will relate to $\mathrm{M}_{\leftrightarrow}^{sys}$. Therefore, performing robustness analysis to see how the linkability changes with multiple views could be very interesting.

DIFFERENT BIOMETRIC MODALITIES

In Sc. 1 described in Section II, we have that the same two individuals are compared with scoring functions derived from voice and face templates. Let $\mathrm{M}_{\leftrightarrow}^{v}$ denote the linkability of the voice system by itself, and $\mathrm{M}_{\leftrightarrow}^{f}$ denote the linkability of the image system by itself. Let $\mathrm{M}_{\leftrightarrow}^{sys}$ denote the linkability of the overall bi-modal system and observe that $\max\left(\mathrm{M}_{\leftrightarrow}^{v}, \mathrm{M}_{\leftrightarrow}^{f}\right) \leq \mathrm{M}_{\leftrightarrow}^{sys}$. However, since the voice and image templates are already linked at an individual level, we cannot find an upper bound for $\mathrm{M}_{\leftrightarrow}^{sys}$ with the sum of individual linkabilities $\mathrm{M}_{\leftrightarrow}^{v}$ and $\mathrm{M}_{\leftrightarrow}^{f}$. Thus, it is important to analyze the whole bi-modal system as a single system and not individual parts.

From the perspective of the scoring functions, we have

$$H \leftrightarrow \left((T_{1,1}, T_{1,2}), (T_{2,1}, T_{2,2})\right) \leftrightarrow (S_1, S_2). \tag{26}$$

By applying the data processing inequality we can thus obtain that this scenario satisfies Eq. 14. That is, the adversary cannot learn more than the true linkability $\mathrm{M}_{\leftrightarrow}^{sys}$. Once again, the joint linkability $\mathrm{M}_{\leftrightarrow}^{(\mathsf{S}_1,\mathsf{S}_2)}$ gives a better estimate of $\mathrm{M}_{\leftrightarrow}^{sys}$ than $\max\left(\mathrm{M}_{\leftrightarrow}^{\mathsf{S}_1}, \mathrm{M}_{\leftrightarrow}^{\mathsf{S}_2}\right)$.

COMPOSITION FOR ALL SCENARIOS

We emphasize that, from the composition perspective, Eq. 19 is not guaranteed to hold for any of the five scenarios. That is, even a sum of two individual scores is not an upper bound on the true score. Instead, we need Eq. 22.

Note, however, we can modify the setting in the following way. Suppose two sets of face images (C_1, C_2) and $(\tilde{C}_1, \tilde{C}_2)$ are known to be both mated or both non-mated (but, the adversary does not know which). However, the images themselves come from two (if mated) or four (in non-mated) different people. Then

$$S_1 \leftrightarrow (T_{1,1}, T_{2,1}) \leftrightarrow H \leftrightarrow (T_{1,2}, T_{2,2}) \leftrightarrow S_2. \tag{27}$$

In this case, Eq. 19 is a valid bound on the overall linkability.

TABLE 1. Summary of BTP schemes used in our experiments.

BTP scheme	output	score function
BioHashing [32]	binary	Hamming distance*
MLP-Hash [33]	binary	Hamming distance*
IoM-GRP [34]	integer	number of collisions
HE [35]	ciphertext	Euclidean distance* (in ciphertext)

*To have similarity values, distance functions are multiplied by -1.

IV. EXPERIMENTS

In this section, we present the experimental results for evaluating the linkability of *multiple* protected biometric templates based on our measure explained in Section III. First, in Section IV-A, we describe our experimental setup for the used biometric systems and implementation details. In Section IV-B, we analyze the numerical results for different scenarios (defined in Section II) in biometric systems where multiple pieces of information are available (more specifically, two similarity scores) for linkability measurement of protected biometric systems. In Section IV-C, we discuss the extension of the scenarios studied in Section IV-B to the situation where the adversary can find three similarity scores to perform hypothesis tests. Finally, in Section IV-D, we further discuss our experimental findings of measuring the linkability of the biometric systems.

A. EXPERIMENTAL SETUP

We consider different BTP schemes, different modalities (face and voice), and SOTA DNN-based feature extractors, to evaluate the linkability of protected templates with multiple similarity scores.

1) BTP SCHEMES

In our experiments, we use different BTP schemes, including BioHashing [32], Multi-Layer Perceptron (MLP) Hashing [33], Index-of-Maximum (IoM) Hashing [34] (i.e., Gaussian random projection-based hashing, shortly GRP), and Homomorphic Encryption (HE) based on Brakerski/Fan-Vercauteren (BFV) [35] algorithm. Table 1 summarizes BTP schemes we use in our experiments, and compares their outputs and corresponding scoring functions.

2) BIOMETRIC MODALITIES

In our experiments, we use two different biometric modalities, including face and voice.[9] We build different biometric recognition systems for the these biometric modalities using state-of-the-art feature extractor models:

Face Recognition For face recognition, we use ArcFace-InsightFace [36], ElasticsFace [37], and FaceNet [38] models as three different SOTA feature extractors. We consider MOBIO [39] dataset, which is a bi-modal dataset including face and voice data taken with mobile devices from 150 individuals and collected in 12 sessions (6)-11 samples in each session) for each subject. For mated pairs, we consider all possible combinations of samples for different subjects. For non-mated pairs, we consider only the first 10 samples for each subject, and then we use all possible pairs of samples from different subjects.

Speaker (voice) Recognition For speaker (voice) recognition, we use ECAPA-TDNN model [40] as the feature extractor, and use voice data in MOBIO [39] dataset to gnerate protected voice templates. To generate mated and non-mated pairs, we use the same protocol as described for the face recognition system.

Table 2 summarises different biometric recognition systems used in our experiments. This table also reports the dimension of unprotected templates extracted by each feature extractor as well as the recognition performance of each model in terms of Equal Error Rate (EER). Note that in our experiments in sections IV-B2-IV-B5, we use face templates, and in the experiment in section IV-B1 we use pairs of face and voice templates of same subjects in the MOBIO dataset.

3) IMPLEMENTATION DETAILS AND SOURCE-CODE

In our experiments, we use the Bob[10] toolbox [41], [42] to build the biometric recognition systems. We also use the open-source implementations (in Bob) of the BioHashing, MLP-Hash, IoM-GRP, and HE schemes [33], [43], [44], [45]. For HE, we use its implementation using the SEAL-Python[11] wrapper on Python 3.8 for the C++ SEAL library [46]. The source code of all our experiments is publicly available to help researchers reproduce our results and build upon our work.[12]

B. ANALYSIS OF DIFFERENT SCENARIOS IN BIOMETRIC SYSTEMS

In this section, we consider different scenarios in Section II and for each scenario we describe a case based on biometric systems based on biometric modalities and BTP schemes explained in Section IV-A. In all cases, we assume that the adversary could find two similarity scores to perform hypothesis tests, and we evaluate the linkability of multiple protected templates with different biometric modalities (in Section IV-B1), different feature extractors (in Section IV-B2), different BTP schemes (in Section IV-B3), different keys (in Section IV-B4), and different scoring functions (in Section IV-B5). In our experiments in Sections IV-B1, IV-B2, and IV-B5, we consider BioHash-protected templates since BioHashing is the simplest BTP scheme in Table 1. Similarly, we use face templates in our experiments in Sections IV-B2-IV-B5 since the face is one of the most popular biometric characteristics. However, we should note that similar experiments with other BTP schemes and other

[9] Maximal linkability has been also used to evaluate the linkability of other biometic modalities, such as finger vein [22], for single biometric template, and it can be similarly applied to multiple leaked templates for different types of biometric modalities.

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

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

[12] https://gitlab.idiap.ch/bob/bob.paper.access2024_linkability_multiple

TABLE 2. Summary of biometric recognition systems used in our experiments.

Modality	Dataset	# Mated	# Non-mated	Feat. Extractor	Feat. Length	EER(%)
Face	MOBIO (face)	1,516,300	2,235,000	ArcFace	512	0.02
				ElasticFace	512	0.02
				FaceNet	128	0.60
Voice	MOBIO (voice)	1,516,300	2,235,000	ECAPA-TDNN	192	1.40

TABLE 3. Linkability of BioHash-protected templates for features from different biometric modalities.

mod. #1 (b_f)	**mod. #2** (b_v)	$M_{\leftrightarrow}^{s_f}$	$M_{\leftrightarrow}^{s_v}$	$M_{\leftrightarrow}^{(s_f,s_v)}$
Face (ArcFace)	Voice (ECAPA)	0.0169	0.0162	0.0232
Face (ElasticFace)	Voice (ECAPA)	0.0143	0.0162	0.0212
Face (FaceNet)	Voice (ECAPA)	0.0302	0.0162	0.0344

TABLE 4. Linkability of BioHash-protected templates for different face feature extractors.

Feat. Ext. #1 (e_1)	**Feat. Ext. #2** (e_2)	$M_{\leftrightarrow}^{s_1}$	$M_{\leftrightarrow}^{s_2}$	$M_{\leftrightarrow}^{(s_1,s_2)}$
ArcFace	ElasticFace	0.0156	0.0149	0.0227
ArcFace	FaceNet	0.0135	0.0149	0.0209
ElasticFace	FaceNet	0.0295	0.0149	0.0337

TABLE 5. Linkability of different BTP schemes for ArcFace templates.

BTP #1 (p_1)	**BTP #2** (p_2)	$M_{\leftrightarrow}^{s_1}$	$M_{\leftrightarrow}^{s_2}$	$M_{\leftrightarrow}^{(s_1,s_2)}$
BioHashing	MLP-Hashing	0.0156	0.0096	0.0188
BioHashing	IoM-GRP	0.0156	0.0024	0.0171
BioHashing	HE	0.0156	0.0042	0.0178
MLP-Hashing	IoM-GRP	0.0096	0.0024	0.0107
MLP-Hashing	HE	0.0096	0.0042	0.0118
IoM-GRP	HE	0.0024	0.0042	0.0088*

* $M_{\leftrightarrow}^{(s_1,s_2)}$ is greater than $M_{\leftrightarrow}^{s_1} + M_{\leftrightarrow}^{s_2}$.

biometric characteristics can be implemented using our open-source paper package.

1) LINKABILITY OF PROTECTED TEMPLATES WITH DIFFERENT BIOMETRIC MODALITIES (SC. 1)

In a multi-modal biometric recognition system the protected templates of each biometric modalities can be stored in the database of the system. In this experiment, we use voice signals and corresponding face images from the MOBIO dataset. We consider voice features extracted by ECAPA-TDNN and face features extracted by different models (ArcFace, ElasticFace, and FaceNet) and protect features of each modality separately using BioHashing. Table 3 reports the linkability of multiple protected biometric templates from voice and face modalities (i.e., Sc. 1 defined in Section II) denoted with $M_{\leftrightarrow}^{(s_f,s_v)}$ and compares it with linkability of face templates ($M_{\leftrightarrow}^{s_f}$) and voice templates ($M_{\leftrightarrow}^{s_v}$). As the results in this table show, pairs of voice and face protected templates have more linkability than their individual protected templates. This means the pair of face and voice templates provide more information to the adversary.

2) LINKABILITY OF PROTECTED TEMPLATES WITH DIFFERENT FEATURE EXTRACTORS (SC. 2)

In some biometric systems, different feature extractors may be used and the final decision is made by fusing the scores from different templates available for each subject. In this experiment, we consider face images from the MOBIO dataset and extract features using different feature extractors, including ArcFace, ElasticFace, and FaceNet. We protect features extracted by each model separately using BioHashing. Table 4 reports the linkability of multiple protected biometric templates extracted from different face feature extractor models (i.e., Sc. 2 defined in Section II) denoted with $M_{\leftrightarrow}^{(s_1,s_2)}$ and compares it with linkability of individual templates extracted by each model (i.e., $M_{\leftrightarrow}^{s_1}$ and $M_{\leftrightarrow}^{s_2}$). As the results in this table show, protected pairs of features extracted from different models reveal more linkability than individual protected templates. Therefore, the adversary can have more information if templates from different feature extractors are available.

3) LINKABILITY OF PROTECTED TEMPLATES WITH DIFFERENT TEMPLATE PROTECTION SCHEMES (SC. 3)

Considering the required level of security, different BTP schemes may be used in different biometric systems. In a particular case, the same user can be enrolled in two systems and the adversary may get access to templates of the same users in both systems. In another case, different BTP schemes may be used in the same system. In this experiment, we extract ArcFace from face images in the MOBIO dataset and use different BTP schemes including BioHashing, MLP-Hashing, IoM-Hashing, and HE. The results of the linkability evaluation of multiple protected templates with different BTP schemes (i.e., Sc. 3 defined in Section II) are reported in Table 5, and show that multiple templates leak more information. In particular, in the case of IoM-GRP and HE, we observe that the linkability of multiple protected templates is greater than the summation of the linkability of protected templates with individual BTP schemes.

4) LINKABILITY OF PROTECTED TEMPLATES WITH DIFFERENT KEYS (SC. 4)

In a single biometric system, the same user may be registered with different keys at different times. This can happen

TABLE 6. Linkability of protected templates with different keys for ArcFace templates.

BTP	$M^{s_1}_{\leftrightarrow}$	$M^{s_2}_{\leftrightarrow}$	$M^{(s_1,s_2)}_{\leftrightarrow}$
BioHashing	0.0156	0.0156	0.0156
MLP-Hashing	0.0096	0.0096	0.0096
IoM-GRP	0.0024	0.0024	0.0024
HE	0.0042	0.0031	0.0100*

* $M^{(s_1,s_2)}_{\leftrightarrow}$ is greater than $M^{s_1}_{\leftrightarrow} + M^{s_2}_{\leftrightarrow}$.

because of the particular application or even in a typical system if the user is removed and registered again into the system. If the adversary gains access to both protected templates with different keys, the linkability of the pairs of protected templates is different than single protected templates. In this experiment, we use ArcFace features extracted from face images of the MOBIO dataset, and generate protected templates using different BTP schemes including BioHashing, MLP-Hashing, IoM-Hashing, and HE. For each BTP scheme, we generated two sets of protected templates using different keys. Table 6 compares the linkability of protected biometric templates if the adversary has access to single templates or multiple templates with different keys (i.e., Sc.4 defined in Section II). The results in this table show that multiple protected templates with different keys have more linkability than single protected templates. Particularly, for HE the results in this table show that the linkability of multiple protected templates with different keys is greater than the summation of the linkability of single protected templates.

5) LINKABILITY OF PROTECTED TEMPLATES USING DIFFERENT SCORING FUNCTIONS (SC. 5)

In every protected biometric system, if the protected templates are leaked, the adversary can use different scoring functions to perform a hypothesis test to identify if the two templates are mated or non-mated. In this experiment, we consider ArcFace features extracted from face images of the MOBIO dataset and protected with BioHashing. We apply different scoring functions[13] (including Hamming distance, Euclidean distance, Cosine distance, Kulsinski distance, Russell-Rao distance, Sokal-Michener distance, and Correlation distance) for BioHash-protected templates and consider the scores available from all pairs of score functions. Table 7 reports the linkability of biometric templates when using two scoring functions (i.e., Sc.5 defined in Section II). The results in this table show that in such a hypothesis test, the linkability of protected templates is higher than using each scoring function separately and the adversary gains a better hypothesis test. However, theoretical properties of maximal linkability discussed in Section III tell us that it is still less than the true linkability of the system. As a matter of fact, our theoretical predictions in Section III are also consistent

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

TABLE 7. Linkability of BioHash-protected templates of AcrFace with different scoring functions.

Func. #1 (s_1)	Func. #2 (s_2)	$M^{s_1}_{\leftrightarrow}$	$M^{s_2}_{\leftrightarrow}$	$M^{(s_1,s_2)}_{\leftrightarrow}$
Hamming	Euclidean	0.0156	0.0162	0.0162
Hamming	Cosine	0.0156	0.0245	0.0272
Hamming	Correlation	0.0156	0.0158	0.0159
Hamming	Kulsinski	0.0156	0.0270	0.0287
Hamming	Russell-Rao	0.0156	0.0266	0.0277
Hamming	Sokal-Michener	0.0156	0.0166	0.0166
Euclidean	Cosine	0.0162	0.0245	0.0274
Euclidean	Correlation	0.0162	0.0158	0.0163
Euclidean	Kulsinski	0.0162	0.0270	0.0282
Euclidean	Russell-Rao	0.0162	0.0266	0.0276
Euclidean	Sokal-Michener	0.0162	0.0166	0.0166
Cosine	Correlation	0.0245	0.0158	0.0268
Cosine	Kulsinski	0.0245	0.0270	0.0270
Cosine	Russell-Rao	0.0245	0.0266	0.0268
Cosine	Sokal-Michener	0.0245	0.0166	0.0275
Correlation	Kulsinski	0.0158	0.0270	0.0288
Correlation	Russell-Rao	0.0158	0.0266	0.0276
Correlation	Sokal-Michener	0.0158	0.0166	0.0166
Kulsinski	Russell-Rao	0.0270	0.0266	0.0271
Kulsinski	Sokal-Michener	0.0270	0.0166	0.0283
Russell-Rao	Sokal-Michener	0.0266	0.0166	0.0276

with the experimental results in Table 7 where we can see for different score functions we have $M^{(s_1,s_2)}_{\leftrightarrow} \leq M^{s_1}_{\leftrightarrow} + M^{s_2}_{\leftrightarrow}$.

C. EXTENDING STUDIED SCENARIOS TO THREE SIMILARITY SCORES

In our experiments in Sections IV-B1-IV-B5, we considered the scenarios where the adversary could eventually find two similarity scores to perform hypothesis tests. For each of the scenarios analyzed in Section IV-B, we can extend our linkability measurement to the situation where there are more than two (e.g., three) scores available for the adversary's hypothesis tests. For example, let us consider the scenario discussed in Section IV-B5 (i.e., Sc. 5) and extend to the situation where the adversary applies three scoring functions to find mated and non-mated protected templates. In such a case, we can use our method to find linkability based on three similarity scores. Fig. 4 illustrates the linkability of BioHash-protected of ArcFace templates if the adversary tries three different scoring functions in their hypothesis test. This figure also depicts the linkability of protected templates based on one and two similarity scores available for the adversary's hypothesis test. As we can observe by increasing the number of scoring functions, the adversary achieves a higher linkability. However, the value of maximal linkability is still slightly higher than the maximum of linkability based on a single similarity score in each case. Other scenarios (Sc. 1-4) can be similarly extended to the situations where the adversary can have three or more similarity scores between multiple templates.

D. DISCUSSION

Our experiments in Sections IV-B and IV-C evaluate the linkability of biometric systems when the adversary can find multiple similarity scores from different stages (based on defined scenarios) of a protected biometric system.

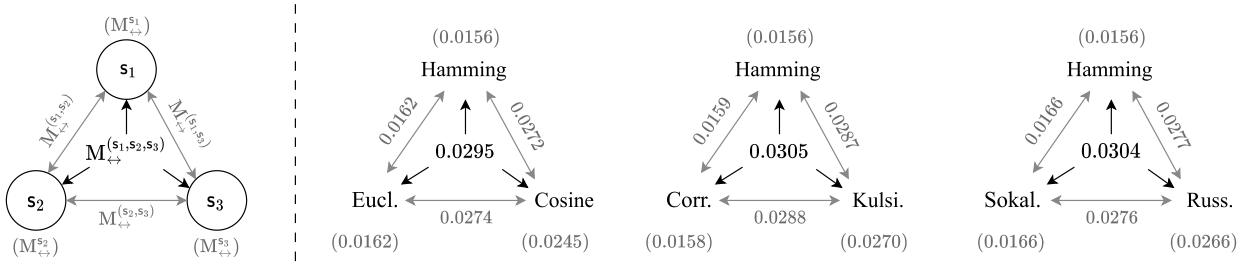


FIGURE 4. Linkability of protected templates using multiple scoring functions for BioHash-protected templates of ArcFace.

With more available information, it is natural to assume that there is more linkability between protected templates, and thus an adversary can achieve better accuracy when performing hypothesis tests to distinguish mated and non-mated templates. Our experiments confirm that within biometric systems, multiple available similarity scores facilitate the linkability of templates. However, the unlinkability of the protected biometric systems degrades gradually with more available similarity scores. In most cases studied in Sections IV-B and IV-C, we observe that linkability often does not exceed the summation of linkability of each similarity score,[14] and in many cases, it is slightly greater than the maximum of linkability based on a single similarity score.

In general, we are interested in estimating $M^{sys}_{\leftrightarrow}$ for each biometric system to find the true linkability of the system. However, as discussed in Section III, it is not computationally possible to calculate $M^{sys}_{\leftrightarrow}$ for real biometric systems in practice. An alternative approach is to use scoring functions to compute a proxy for $M^{sys}_{\leftrightarrow}$ and using several scoring functions at once gives a better estimate of $M^{sys}_{\leftrightarrow}$. In particular, in our experiments in Sections IV-B5 and IV-C, we showed that we can use two and three scoring functions, respectively, to better estimate the linkability of system with different scoring functions. Different scoring functions (as in Sc. 5) can be also applied to other scenarios (i.e., Sc. 1-4) defined in Section II to improve our estimation of the linkability of multiple biometric templates.

For a number of cases in Sc. 3 and Sc. 4 in our experiments in Section IV-B, we observe that maximal linkability using multiple similarity scores can be higher than the summation of maximal linkability of each individual score, which was expected from our theoretical analysis in Section III. Therefore, as mentioned in Section III, it is important to perform robustness analysis and evaluate the linkability of protected templates based on multiple similarity scores available for an adversary.

V. CONCLUSION

In this paper, we focused on measuring the linkability of protected biometric templates when an adversary can access multiple protected templates from different biometric systems, a single multi-modal biometric system, or even a single unimodal biometric system. We defined maximal linkability for the case where an adversary can find multiple similarity scores from the available protected templates. We considered five different scenarios where the adversary gains access to multiple biometric templates from different biometric modalities, different feature extractors, different template protection schemes, or with different keys, and also two protected templates with different scoring functions. In each of these scenarios, the adversary can find multiple similarity scores and can perform a hypothesis test to determine mated and non-mated biometric templates. In our experiments, we focused on the situation where the adversary can find two similarity scores for their hypothesis test and measured the linkability of protected templates. However, our approach can be extended to more than two similarity scores. In particular, we showcased measuring linkability of protected templates where the adversary can find three similarity scores for their hypothesis test. To our knowledge, the linkability of multiple protected biometric templates has not been studied in the literature, and thus this paper paves the way for more comprehensive linkability studies of protected biometric templates. Our proposed measure can particularly be used to evaluate the linkability of protected templates at different stages within the same biometric system, across different biometric systems, and within multi-modal biometric systems.

[14]Even if as discussed in Section III, there is no theoretical guarantee not to exceed the summation for scenarios 1-4 defined in Section II.

REFERENCES

[1] *Information Technology—Biometric Multimodal and Other Multibiometric Fusion*, Standard ISO/IEC TR 24722:2015, Int. Org. for Standardization Int. Electrotechnical Committee, 2015, p. 2015.

[2] S. Vhaduri and C. Poellabauer, "Multi-modal biometric-based implicit authentication of wearable device users," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3116–3125, Dec. 2019.

[3] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in *Proc. 12th Eur. Signal Process. Conf.*, Sep. 2004, pp. 1221–1224.

[4] M. Khamis, M. Hassib, E. V. Zezschwitz, A. Bulling, and F. Alt, "GazeTouchPIN: Protecting sensitive data on mobile devices using secure multimodal authentication," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, Nov. 2017, pp. 446–450.

[5] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2019.

[6] K. Su, G. Yang, B. Wu, L. Yang, D. Li, P. Su, and Y. Yin, "Human identification using finger vein and ECG signals," *Neurocomputing*, vol. 332, pp. 111–118, Mar. 2019.

[7] L. M. Dinca and G. P. Hancke, "The fall of one, the rise of many: A survey on multi-biometric fusion methods," *IEEE Access*, vol. 5, pp. 6247–6289, 2017.

[8] *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)*, Apr. 2016.

[9] *740 ILCS 14/biometric Information Privacy Act*, 2008.

[10] *IEEE Standard for Biometric Privacy*, 2021, pp. 1–37.

[11] 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.

[12] 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.

[13] 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.

[14] P. Kaur, N. Kumar, and M. Singh, "Biometric cryptosystems: A comprehensive survey," *Multimedia Tools Appl.*, vol. 82, no. 11, pp. 16635–16690, May 2023.

[15] *Information Technology—Security Techniques—Biometric Information Protection*, Standard ISO/IEC JTC1 SC27, 2022.

[16] *Information Technology—-Biometric Performance Testing and Reporting—Part 1: Principles and Framework*, Standard ISO/IEC JTC1 SC37, Int. Org. for Standardization Int. Electrotechnical Committee, 2021.

[17] *Information Technology—Security Techniques—Performance Testing of Biometric Template Protection Schemes*, Standard ISO/IEC 30136, 2018.

[18] H. O. Shahreza and S. Marcel, "Breaking template protection: Reconstruction of face images from protected facial templates," in *Proc. IEEE 18th Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2024, pp. 1–25.

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

[20] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proc. Biometrics Symp.*, Sep. 2007, pp. 1–6.

[21] C. Li and J. Hu, "Attacks via record multiplicity on cancelable biometrics templates," *Concurrency Comput. Pract. Exper.*, vol. 26, no. 8, pp. 1593–1605, Jun. 2014.

[22] H. Otroshi Shahreza, Y. Y. Shkel, and S. Marcel, "Measuring linkability of protected biometric templates using maximal leakage," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2262–2275, 2023.

[23] I. Buhan, J. Breebaart, J. Guajardo, K. de 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*. Cham, Switzerland: Springer, 2010, pp. 78–92.

[24] E. J. C. Kelkboom, J. Breebaart, Tom. 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.

[25] A. Nagar, K. Nandakumar, and A. K. Jain, "Biometric template transformation: A security analysis," in *Media Forensics and Security*, vol. 7541. Bellingham, WA, USA: SPIE, 2010, pp. 237–251.

[26] 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.

[27] E. A. Rua, E. Maiorana, J. L. A. 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, Feb. 2012.

[28] 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)*, Sep. 2014, pp. 1–8.

[29] 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.

[30] 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.

[31] 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.

[32] 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, Nov. 2004.

[33] H. O. Shahreza, V. K. Hahn, and S. Marcel, "MLP-Hash: Protecting face templates via hashing of randomized multi-layer perceptron," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2023, pp. 605–609.

[34] 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.

[35] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic encryption," *Cryptol. ePrint Arch.*, vol. 1, no. 1, pp. 1–18, Jul. 2012.

[36] 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. 4685–4694.

[37] 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. 1577–1586.

[38] 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.

[39] 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.

[40] B. Desplanques, J. Thienpondt, and K. Demuynck, "Ecapa-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification," 2020, *arXiv:2005.07143*.

[41] 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. 1–17.

[42] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, "Continuously reproducing toolchains in pattern recognition and machine learning experiments," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Aug. 2017, pp. 1–10.

[43] 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.

[44] 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 Security (WIFS)*, Dec. 2021, pp. 1–6.

[45] 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.

[46] (2020). *Microsoft SEAL (Release 3.6)*. [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 the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, and is a Research Assistant with the Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland, where he received H2020 Marie Skłodowska-Curie Fellowship (TReSPAsS-ETN) for his doctoral program. During his Ph.D., Hatef also experienced 6 months as a visiting scholar with the Biometrics and Internet Security Research Group at Hochschule Darmstadt, Germany. He is also the winner of the European Association for Biometrics (EAB) Research Award 2023. His research interests include machine learning, deep learning, computer vision, biometrics, and biometric template protection.

YANINA Y. SHKEL (Member, IEEE) is an Assistant Professor in the School of Computer and Communication Sciences at École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. She received the B.S. degrees 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. Before joining the Graduate School, she was a Database Developer with Morningstar, Inc., and in 2013, she spent her time with 3M Corporate Research, as an Intern. From 2014 to 2019, she was a Postdoctoral Researcher with Princeton University, Princeton, NJ, USA, and with the University of Illinois at Urbana–Champaign, Champaign, IL, USA. From 2019 to 2023 she was a Scientist at EPFL, Switzerland. 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 is a recipient of the NSF Center for Science of Information (CSoI) Postdoctoral Fellowship and the Swiss National Science Foundation Starting Grant.

SÉBASTIEN MARCEL (Senior Member, IEEE) heads the Biometrics Security and Privacy group at Idiap Research Institute (Switzerland) and conducts research on face recognition, speaker recognition, vein recognition, attack detection (presentation attacks, morphing attacks, deepfakes) and template protection. He received his Ph.D. degree in signal processing from Université de Rennes I in France (2000) at CNET, the research center of France Telecom (now Orange Labs). He is Professor at the University of Lausanne (School of Criminal Justice) and a lecturer at the 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.