General Framework to Evaluate Unlinkability in Biometric Template Protection Systems

Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch

***Abstract*— The wide deployment of biometric recognition systems in the last two decades has raised privacy concerns regarding the storage and use of biometric data. As a consequence, the ISO/IEC 24745 international standard on biometric information protection has established two main requirements for protecting biometric templates: irreversibility and unlinkability. Numerous efforts have been directed to the development and analysis of irreversible templates. However, there is still no systematic quantitative manner to analyze the unlinkability of such templates. In this paper, we address this shortcoming by proposing a new general framework for the evaluation of biometric templates' unlinkability. To illustrate the potential of the approach, it is applied to assess the unlinkability of the four state-of-the-art techniques for biometric template protection: biometric salting, bloom filters, homomorphic encryption, and block remapping. For the last technique, the proposed framework is compared with other existing metrics to show its advantages.**

***Index Terms*— Unlinkability, privacy, template protection, biometrics, performance testing.**

I. Introduction

BIOMETRICS refers to automated recognition of individuals based on their biological or behavioural characteristics, such as iris or signature [1]. Its advantages over traditional authentication methods (e.g., no need to carry tokens or remember passwords, harder to circumvent or stronger link between the subject and the action or event), have allowed a wide deployment of biometric systems in the last decade, including large-scale national and international initiatives such as the Unique ID program of the Indian government [2] or the Smart Borders project of the European Comission [3]. However, unprotected storage of biometric reference templates poses severe privacy threats, e.g. identity theft, cross-matching or limited renewability. In fact, biometric data are defined as sensitive data within the European Union (EU) General Data Protection Regulation 2016/679 [4], which means that the use of these data is subjected to the right of privacy preservation.

Considering those privacy issues, biometric template protection schemes have been developed in the last two decades [5], [6], and several standardization efforts [7]–[9] have been directed to this topic. Biometric template protection schemes are commonly categorized as *biometric cryptosystems*, also referred to as helper data schemes, and *cancelable biometrics*, also referred to as feature transformation approaches. Biometric cryptosystems are designed to securely bind a digital key to a biometric characteristic or generate a digital key from a biometric signal [10]. Cancelable biometrics consist of intentional, repeatable distortions of biometric signals based on transforms that provide a comparison of biometric templates in the transformed domain [6], [11].

As defined in the ISO/IEC International Standard 24745 on biometric information protection [7], in order to protect the privacy of the individuals, "knowledge of the transformed biometric reference cannot be used to determine any information about the generating biometric sample(s) or features", which makes clear reference to the necessity of storing *irreversible* biometric templates. But not only that, the ISO/IEC standard continues by stating "[. . . and] the stored biometric references should not be linkable across applications or databases". That is, protected templates are not only required to be *irreversible*, but also *unlinkable*. Only by fulfilling both requirements can we grant the privacy to which subjects are entitled.

As stated in [9], a standardised benchmark protocol for biometric template protection schemes, in terms of recognition accuracy, security and privacy, is necessary for a further deployment of biometric verification systems. However, whereas the irreversibility of protected templates or the accuracy degradation with respect to unprotected systems have been thoroughly analysed in the literature, little attention has been paid to the objective evaluation of their unlinkability [12]. In fact, there is still no general metric, protocol or framework to assess, in an objective way, the unlinkability of biometric templates in order to be able to establish a fair benchmark for the performance of different protection algorithms. Due to the limitations of the current proposals to evaluate this property (for more details, see Sect. II), no standardised metric has been included in the current ISO/IEC 30136 project on performance testing of biometric template protection schemes [8].

To tackle the aforementioned issue, in the present article we propose a general framework that addresses this need. The novel methodology is partially inspired by the initial ideas presented by Ferrara *et al.* [13], where the authors consider three sets of scores distributions resulting from the

Manuscript received May 9, 2017; revised November 7, 2017 and December 22, 2017; accepted December 22, 2017. Date of publication December 29, 2017; date of current version February 7, 2018. This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP (www.crisp-da.de). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Adams W. K. Kong. *(Corresponding author: Marta Gomez-Barrero.)*

M. Gomez-Barrero, C. Rathgeb, and C. Busch are with the da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany (e-mail: marta.gomez-barrero@h-da.de; christian.rathgeb@h-da.de; christoph.busch@h-da.de).

J. Galbally is with the European Commission, DG Joint Research Centre, 21027 Ispra, Italy (e-mail: javier.galbally@ec.europa.eu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.2788000

comparison of templates enrolled in different applications using different application-specific parameters. However, probably due to the lack of an appropriate measure to benchmark them, no objective numerical analysis was carried out in their work, only a visual analysis between the score distributions. In this context, we provide two different metrics that enable the quantitative assessment of templates' unlinkability and the objective benchmark among systems. The two metrics are: i) a local score-wise measure, $D_{\leftrightarrow}(s)$, based on the likelihood ratio between the score distributions; and ii) a global measure, $D_{\leftrightarrow}^{sys}$, independent of the score domain, and thereby allowing a fairer benchmark of the overall systems' unlinkability. Both measures yield values in a closed range and build upon the solid theory behind likelihood ratios for calibration of scores or interpretation of evidence in forensic environments [14]–[18]. Furthermore, they are defined for the complete domain of scores in order to allow a more straightforward benchmark of different schemes, useful for instance for evaluating systems in competitions or for commercial purposes.

The rest of the paper is structured as follows. Previous works on unlinkability analysis of particular biometric template protection systems are summarised in Sect. II. General concepts on biometric template protection and unlinkability are included in Sect. III. The new framework for the systematic analysis of the unlinkability of biometric template protection systems is described in Sect. IV. A protocol for the linkability analysis of biometric templates is provided in Sect. V. Then, the framework is applied to evaluate the unlinkability of four different state-of-the-art template protection schemes in Sect. VI, and conclusions are drawn in Sect. VII

II. Related Works

As mentioned in Sect. I, few works related to biometric template protection (BTP) have included some type of analysis of templates' unlinkability. These are summarised in Table I. For instance, Linnartz and Tuyls analyse in [19] the information leaked by the stored reference templates and the danger posed by replay attacks. On the other hand, Dodis *et al.* show that public data can disclose information about the original biometric sample, and measure this leak in terms of the average min entropy [20]. In addition, Buhan *et al.* generalise the concept of fuzzy extractors to continuous distributions in [21], establishing as well a relationship between the False Match Rate (FMR) of the system and the min entropy of the templates.

From a more general perspective, a framework for the security and privacy analysis of biometric systems is developed in [22]. In that work, a general system, involving four logical entities (i.e., sensor, server, database and comparator), is defined, and the security of the system is analysed assuming different adversary models where some of the aforementioned entities may be malicious. While the information flow and the ability to reconstruct templates is analysed, no metric or protocol is provided to measure the unlinkability of a given system.

Other works have proposed specific cross-matching attacks (which can also be considered as a *linkage function*, as will be seen in Sect. III) in order to link templates produced by several template protection schemes, and analysed the threat posed in terms of their success chances. For instance, in [23], the distributions of dissimilarity scores for normal and attacking comparisons are depicted, but no quantitative measure of the linkability is provided. In [24], a cross-matching attack against fuzzy vault schemes is proposed. However, in this case, due to the time consuming nature of the attack, only ten vaults are analysed, reporting a success rate of the attack of 40% (which could be understood as a very attack-specific linkability metric).

For the particular case of biometric cryptosystems, the unlinkability property has been frequently referred to as *indistinguishability* [25]–[27], being its definition identical to the unlinkability definition included in [7 and 8, Sec. III]. In their work [25], Simoens *et al.*, following a cryptographic perspective, analyse the unlinkability of biometric sketches playing the so-called *indistinguishability game*. Assuming the attacker or adversary has obtained a set of sketches, his objective is to identify related sketches. In order to measure his success chances, they estimate his advantage over a random guess. This is done by setting bounds or computing probabilities which rely on the Error Correcting Codes (ECC) theory, which is fundamental to the development of fuzzy sketches. Therefore, such an analysis can not be extrapolated to systems based, for instance, on cancelable biometrics approaches. In addition, one of the main limitations of the theoretical derivation is the assumption of handling uniform data, which does not hold for biometric data.

A similar approach is followed in [26] for the analysis of Quantization Index Modulation (QIM) biometric cryptosystems. Since "the concept describes the advantage of an attacker with respect to a perfect indistinguishable system", which is hard to achieve in practice due to the inherent correlation of biometric data, the authors propose an alternative practical evaluation in which they benchmark Equal Error Rates (EER) obtained for regular comparisons of templates protected with a single key (i.e., recognition accuracy analysis) against comparisons of templates protected with different keys (i.e., unlinkability analysis). For such comparisons, they use the same distinguisher function used for the indistinguishability game. The main drawback of this practical approach is that, even if an increase of the EER implies some degree of unlinkability of the system (i.e., the system loses discriminative power when different keys are used to protect the templates), such unlinkability increase is not quantified.

In [27], Buhan *et al.* consider for the first time a *continuous* value, instead of binary, for the linkability of two given templates for a fuzzy scheme. In other words, unlinkability is regarded as a continuous property, being possible to assign different *degrees of linkability* to the templates. To that end, they introduce "a *classification function* to decide whether the query sketch and the target sketch are related". This function is hence used to link templates. Then, the advantage of an eventual attacker over a random guess is theoretically modelled, and a relationship is established with the False Match Rate (FMR) and False Non-Match Rate (FNMR) of the system. It should be noted that similar unrealistic assumptions to those of [25] are made in the theoretical derivation. In addition, the FMR and FNMR are estimated in terms of the verification function

TABLE I

SUMMARY OF THE MOST RELEVANT METHODOLOGIES FOR UNLINKABILITY ASSESSMENT AND THEIR MAIN PROPERTIES: *i*) WHETHER THEY ARE GENERAL ENOUGH TO BE APPLIED TO DIFFERENT BTP SCHEMES, *ii*) WHETHER THEY REGARD LINKABILITY AS A BINARY OR CONTINUOUS CHARACTERISTIC, *iii*) WHAT KIND OF METRIC IS PROPOSED, *iv*) WHETHER A QUANTITATIVE METRIC IS PROPOSED AND *v*) WHAT ASSUMPTIONS ARE MADE. DRAWBACKS ARE HIGHLIGHTED IN BOLD LETTERS

Ref.	General?	Binary / Continuous	Metric	Quantitative?	Assumptions
[23]	Yes	-	Score distributions	**No**	-
[24]	Yes	**Binary**	Success rate	Yes	-
[25]	**No**: Biometric sketches	**Binary**	Adversary's advantage	Yes	**Uniform data**
[26]	**No**: QIM	**Binary**	Adversary's advantage	Yes	**Uniform data**
[26]	Yes	**Binary**	**Accuracy oriented curves (ROC)**	**No**	-
[27]	**No**: Fuzzy scheme	Continuous	Adversary's advantage	Yes	**Uniform data**
[28]	Yes	**Binary**	**Accuracy oriented curves (ROC)**	**No**	-
[29]	Yes	**Binary**	**Accuracy oriented curves (DET)**	**No**	-
[30]	Yes	**Binary**	**Accuracy oriented curves (DET)**	**No**	-
[31]	Yes	Continuous	**Accuracy oriented curves (CMC)**	Yes	-
[13]	Yes	-	Score distributions	**No**	-
[32]	Yes	-	Score distributions	**No**	-

of the system, which yields less discriminative information in terms of unlinkability than the classification function used by the attacker.

Kelkboom *et al.* present a practical evaluation of the robustness of a fuzzy commitment scheme to several cross-matching attacks in [28]. To that end, and following the approach presented in [26] for the EERs, the Receiving Operating Characteristic (ROC) curves of the system for the accuracy and unlinkability evaluations are compared: if the accuracy shown by the ROC curves decreases, then the system is assumed to be unlinkable. Again, a relationship of the vulnerability to the linkability attack and the FMR and FNMR is established, assuming the extracted bits to be independent with equal bit-error probability, a fact that does not always hold in biometric data due to its inherent correlation. In a similar manner, Nagar *et al.* analyse the unlinkability of a biometric template protection scheme in [29], where the error rates for the unlinkability analysis are referred to as Cross Match Rate (CMR) and False Cross Match Rate (FCMR).

Following this same concept of using accuracy curves, Piciucco *et al.* define in [30] the Renewable Template Matching Rate (RTMR) as the percentage of correctly linked templates. For the evaluation, a Detection Error Trade-off (DET)-like curve is depicted, where FNMR is shown for each RTMR for a subset of the database used in the experimental evaluation. In this case, mated scores for the FNMR are computed comparing templates protected with the *same key* and extracted from the same instance, whereas non-mated scores for the RTMR are obtained from the comparison of templates protected with *different keys* and extracted from different instances. The authors point out that the FNMR *vs*. RTMR curve should be similar to the common DET curve (FNMR *vs*. FMR) to indicate that matching templates protected with different keys is at least as hard as achieving a false match with templates obtained from different subjects. Therefore, if the curves are visually similar, unlinkability is obtained.

In [31], a linkage function based on Principal Components Regression is defined for fuzzy commitment systems using Universal Background Models (UBM). Building upon this linkage function, the eventual attacker decides which is the identity hidden by both templates. In order to measure linkability, the authors focus on the probability that the attacker chooses the correct identity in a top-N list, and how this probability deviates from a random guess (full unlinkability). In other words, the authors analyse the variation of the Cumulative Match Curves (CMC) used to evaluate the accuracy of identification biometric systems. While this is a clever approach to solve the problem with some highly desirable properties (i.e., it is general, continuous and provides a quantitative measure), it still presents some limitations such as: *i*) it does not give the linkability level for a given score (i.e., what is the probability that, given score s_0, the two templates stem from the same subject?); and *ii*) it does not provide one unique general linkability measure for the whole system, but a different value for each size N of the identities list.

In contrast to this DET, CMC or ROC approach, which can hide how linkable templates are for specific subsets of scores (see Sect. VI-E), [13], [32] have addressed the problem of unlinkability evaluation directly analysing the similarity scores. In Ferrara *et al.* [13], consider three sets of scores distributions resulting from the comparison of templates enrolled in different applications using different secret keys. More specifically, templates were extracted from *i*) the same sample of a given instance, *ii*) different samples of the same instance or *iii*) samples of different instances. The analysis in [32] focuses only on the last two distributions, which are the most likely to occur in a real-world scenario. In both works, only a visual benchmark between the score distributions is carried out, presumably due to a lack of an appropriate metric. Traditionally, such difference between probability densities has been estimated in terms of the Kullback-Leibler (*KL*) divergence [33] between two discrete distributions, P and Q, which is defined as:

$$D_{KL}(P||Q) = \sum_{s} P(s) \ln \left(\frac{P(s)}{Q(s)} \right) \tag{1}$$

where $D_{KL} \geq 0$, and $D_{KL} = 0$ holds iff $P \simeq Q$, i.e. the smaller D_{KL}, the higher the similarity between distributions. However, this measure presents several limitations for the unlinkability

analysis due to three main reasons: i) it gives only an overall measure of the unlinkability of the system, not being possible to measure the level of unlinkability for different domains of the linkage scores, ii) it is not bounded, thus making it difficult to benchmark the unlinkability of different systems, and iii) it is not defined for $Q(s) = 0$ if $P(s) \neq 0$, hence not taking into account important ranges of scores, or not being at all defined for fully separable distributions.

In summary, although being all valuable contributions, the aforementioned articles share some common shortcomings, as highlighted in red in Table I:

- Unrealistic assumptions on uniformity of biometric data [25]–[27].
- Non general approaches: many of the methods proposed are developed for a specific system or template protection technique but cannot be applied to others. As a consequence, it is difficult or simply not possible to use them to benchmark different systems [25]–[27].
- Linkability is regarded as a binary decision, either templates are fully linkable, or fully unlinkable, and no *degree* of linkability is considered [24]–[26], [28]–[30].
- Some approaches only suggest how linkability could be measured, but do not give any quantitative measure, just a subjective analysis [13], [23], [32].
- Other methods recommend the use of metrics employed for verification accuracy evaluations, such as DET or ROC curves, not suitable for the linkability evaluation, as it will be shown in [26–31, Sec.VI-E].

The proposed framework tackles such limitations and offers the following advantages:

- No assumptions are made on the data, neither on bits' independence (as understood in entropy studies) nor on uniformity.
- Only a classification function, named as "linkage function", is assumed to exist, in order to assess the non-binary nature of the unlinkability property [27]. In addition, to make the framework as general as possible, the function can take arguments in both the original (i.e., unprotected) or the protected domain, as in [29].
- The proposed metrics evaluate linkability based on score distributions obtained from the linkage function, independently of what the linkage function is. This allows for a general metric, since it can be computed for any Lebesgue integrable linkage function.
- Not only a global measure for the unlinkability of the templates is provided, but also a local measure for each linkage score, in order to allow a more thorough evaluation.
- Using always the same metric for different linkage functions has the advantage of allowing to monitor the changes in a system's linkability when different functions are used to compare the templates.

III. Concepts on Biometric Template Protection and Unlinkability

Throughout the article we will use the Harmonized Biometric Vocabulary (HBV) defined in the ISO/IEC 2382-37 [34]. For any clarification on the concepts used we refer the reader to the mentioned standard.[1] Given that they are often used throughout the article, for the sake of clarity, we will only include here the next definitions:

- *Biometric characteristic*: "biological and behavioural characteristic of an individual from which distinguishing, repeatable biometric features can be extracted for the purpose of biometric recognition". For example, a fingerprint or an iris are two different biometric characteristics.
- *Biometric instance*: for some characteristics, an individual possesses several instances. For example, the right index fingerprint is a different instance from the left thumb, even if they serve to identify the same data subject.
- *Mated samples*: "paired biometric probe and biometric reference that are from the same biometric characteristic of the same biometric data subject". For example, two fingerprint samples from the same right index finger.
- *Non-mated samples*: "paired biometric probe and biometric reference that are not from the same biometric instance". For example, two fingerprint samples from different fingers.

Within ISO/IEC IS 24745 [7], unlinkability is defined as "a property of two or more biometric references that they cannot be linked to each other or to the subject(s) from which they were derived". The challenge is hence to determine whether two protected templates, $\mathbf{T}_1$ and $\mathbf{T}_2$, enrolled in two different applications, conceal the same biometric instance (i.e., they represent different samples of biometric data extracted from the same biometric instance - e.g., the same left index finger).

From an analytic perspective, and taking into account the works described in Sect. II, the unlinkability definition given in the ISO/IEC IS 24745 presented above can be reformulated as a gradual property of the templates:

> **Definition of linkability**: two templates are *fully linkable* if there exists some method to decide that they were extracted, with all certainty, from the same biometric instance. Two templates are *linkable to a certain degree* if there exists some method to decide that it is more likely that they were extracted from the same instance than from different instances.

Given the aforementioned definition of linkability, it follows that this property is fully related to the *method* (i.e., linkage function) used to decide if two templates stem from the same instance. In fact, the different cross-matching attacks developed in the works described in Sect. II are particular examples of linkage functions.

The scenario to be analysed is depicted in more detail in Fig. 1. Given two applications, the biometric reference sample $\mathbf{M}$ is presented to the each of them during enrolment. Features are extracted and the *Pseudonymous Identifier Encoder* (*PIE*), taking as input both the sample $\mathbf{M}$ and a secret key $\mathbf{K}_i$, with $i = \{1, 2\}$, computes the corresponding reference template, $\mathbf{T}_i = PIE\,(\mathbf{M}, \mathbf{K}_i)$.

Following the aforementioned definition and the approach presented in [27], where a classification function is used to evaluate unlinkability in a continuous manner, a *linkage*

[1] Available at http://standards.iso.org/ittf/PubliclyAvailableStandards/c055194_ISOIEC_2382-37_2012.zip or http://www.christoph-busch.de/standards.html

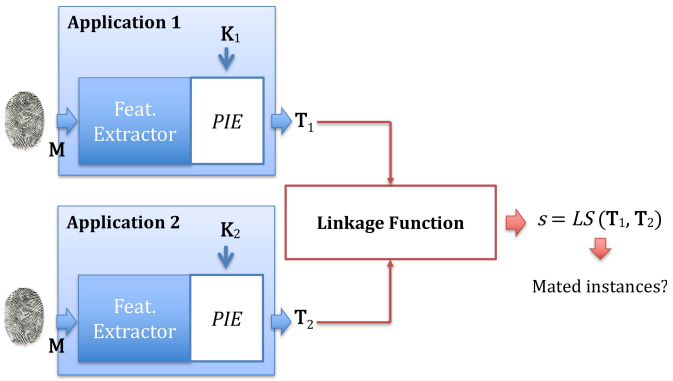


Fig. 1. General diagram for linking templates. For two different applications, at enrollment the *PIE* takes as input the reference biometric sample $\mathbf{M}$ and the corresponding secret key $\mathbf{K}_i$ to generate the protected templates $\mathbf{T}_i = PIE(\mathbf{M}, \mathbf{K}_i)$. Then, taking as input $\mathbf{T}_1$ and $\mathbf{T}_2$, a linkage score between them, $s = LS(\mathbf{T}_1, \mathbf{T}_2)$, can be generated and used to decide whether both templates conceal mated instances.

function will provide a linkage score between the analysed templates: $s = LS(\mathbf{T}_1, \mathbf{T}_2)$. Such function might be the similarity score computed by the BTP scheme or any other function exploiting a vulnerability of the system. This score will be used to determine whether both templates conceal the same instance. In other words, a BTP scheme is considered linkable *to some degree* if, given a score $s = LS(\mathbf{T}_1, \mathbf{T}_2)$, the likelihood that both templates $\mathbf{T}_1$ and $\mathbf{T}_2$ conceal the same instance is larger than the likelihood that they conceal different instances.

In summary, the proposed framework to evaluate unlinkability assumes the following:

- The existence of a given linkage function *LS* that, given two templates $\mathbf{T}_1$ and $\mathbf{T}_2$, produces a score $s = LS(\mathbf{T}_1, \mathbf{T}_2)$. This linkage function can be the *PIC* (Pseudonymous Identifier Comparator) of the original system or some other function thought to exploit some vulnerability of the system. The only requisite is that it produces continuous normalizable scores.
- Access to the linkage score s between templates $\mathbf{T}_1$ and $\mathbf{T}_2$, $s = LS(\mathbf{T}_1, \mathbf{T}_2)$.

Now, to extend formality to the problem being addressed, some mathematical notations are introduced in this section. Let us define the following hypothesis:

$$H_m = \{\text{both templates belong to mated instances}\}$$
$$H_{nm} = \{\text{both templates belong to non-mated instances}\}$$

Based on those hypothesis, we can define two types of score distributions, where $s = LS(\mathbf{T}_1, \mathbf{T}_2)$ is the linkage score between two templates, as defined in Fig. 1:

- *Mated samples* distribution: scores computed from templates extracted from samples of a single instance of the same subject and enrolled in different applications, using different keys:

$$\mathbf{T}_1 = PIE(\mathbf{M}_1, \mathbf{K}_1), \quad \mathbf{T}_2 = PIE(\mathbf{M}_1, \mathbf{K}_2) \tag{2}$$

  It represents the conditional probability of obtaining a score s knowing that two templates come from mated instances, that is, $p(s|H_m)$.

- *Non-mated samples* distribution: scores yielded by templates generated from samples of different instances and enrolled in different applications, using different keys:

$$\mathbf{T}_1 = PIE(\mathbf{M}_1, \mathbf{K}_1), \quad \mathbf{T}_2 = PIE(\mathbf{M}_2, \mathbf{K}_2) \tag{3}$$

  It represents the conditional probability of obtaining a score s knowing that two templates come from non-mated instances, that is, $p(s|H_{nm})$.

In contrast to [13], where three score distributions are considered, here only two distributions will be analysed as in [32]. The main reason is that the first distribution in [13], comprising linkage scores of exactly the same sample enrolled in different systems, is included here as part of the *Mated samples* distribution (i.e., the mated samples distribution will represent both cases i) and ii)). Furthermore, the score distribution related to case i) in [13] is a very unusual case in real scenarios as, in most cases, different enrolled samples will be acquired by each system. Nevertheless, the proposed framework is general and can also be directly applied to the case of having three score distributions.

IV. Measuring Linkability

Based on the definition of linkability given in Sect. III, two templates are linkable to some degree if there is some method that allows us to determine that it is more likely that they come from the same instance than from different instances. Therefore, according to that definition, if a method allows us to determine that two templates are more likely to come from *different* instances, those two templates are *unlinkable*.

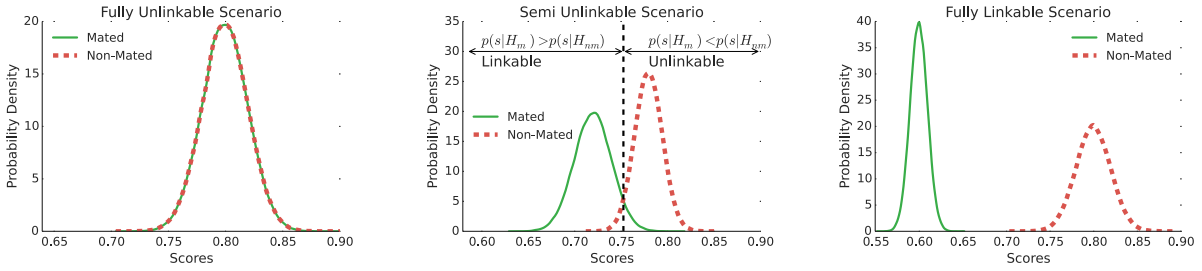
That is, if for a given score s of the linkage function, $p(H_m|s) > p(H_{nm}|s)$, then it is more likely that $\mathbf{T}_1$ and $\mathbf{T}_2$ belong to the same instance, and, therefore, the templates are linkable (to some degree). If, on the contrary, $p(H_{nm}|s) \geq p(H_m|s)$, then it is more likely that $\mathbf{T}_1$ and $\mathbf{T}_2$ belong to different instances. That is, for that score s, the linkage function would fail to link both templates. Therefore, it follows that, to have a fully unlinkable system for the whole domain of scores s of the linkage function, there has to be a complete overlap between the aforementioned *mated* and *non-mated* score distributions.

Following the discussion presented to this point, we define two different measures for the linkability of biometric templates:

- Local measure $D_{\leftrightarrow}(s)$: it evaluates the linkability of the templates in a score-wise basis.
- Global measure $D_{\leftrightarrow}^{sys}$: it gives an overall measure of the linkability of the whole system, independent of the score domain of the system at hand, thereby allowing a benchmark among different systems.

To better illustrate the relationship between the linkage scores and the proposed metrics, Fig. 2 represents generic *Mated samples* distributions (in solid green) and *Non-mated samples* distributions (in dashed red) for: (a) fully unlinkable, (b) semi-linkable and (b) fully linkable templates. As we observe,

- Under a *fully unlinkable* scenario (Fig. 2a), both distributions are identical. In this case, the probability that, for a given score, the templates protect the same



(a) (b) (c)

Fig. 2. Examples of *Mated samples* (solid green) and *Non-mated samples* (dashed red) distibutions yielded by (a) fully unlinkable, (b) semi linkable and (c) fully linkable templates.

instances or different instances is the same and, therefore, the templates cannot be linked.

- Under a *fully linkable* scenario (Fig. 2c), the *Mated samples* and *Non-mated samples* distributions are fully separable. Therefore, given the linkage score between two templates, we can make a decision with almost all certainty for all the score domain on whether the templates that produced any of the scores stem from the same instance (i.e., $s \in [0.55, 0.65]$).
- Under a *semi linkable* scenario (Fig. 2b), we observe that templates can be linked only for a subset of the scores: $s < 0.75$. For that subset, it is more likely that both templates stem from mated instances. For $s > 0.75$, templates are more likely to conceal different instances, and are not linkable.

In the following subsections, we describe how both the local and the global measures are computed and we discuss their main properties.

A. Local Measure $\mathrm{D}_{\leftrightarrow}(s)$*: System Score-Wise Linkability*

$\mathrm{D}_{\leftrightarrow}(s) \in [0, 1]$ evaluates the linkability of a system for each *specific linkage score* s. As such, this metric is appropriate to analyse within one system in which parts of the score domain it fails to provide unlinkability. If for a specific score s_1, a system yields $\mathrm{D}_{\leftrightarrow}(s_1) = 1$, it means that, *in case* the linkage function produced s_1, we would be able to link both templates $\mathbf{T}_1$ and $\mathbf{T}_2$ to the same instance with almost all certainty. On the other hand, $\mathrm{D}_{\leftrightarrow}(s_0) = 0$ should be interpreted as full unlinkability for that particular score s_0. In other words, *if* s_0 were produced by the linkage function, it would be more likely that both templates stemmed from different instances, hence failing to link them to a single data subject. All intermediate values of $\mathrm{D}_{\leftrightarrow}(s)$ between 0 and 1 report an increasing degree of linkability.

As highlighted in Sect. III and Fig. 2b, the key on the success of linking to templates lies on determining whether, given a score s, it is more likely that two templates stem from mated samples than from non-mated samples. Or, in other words, on whether $p(H_m|s) > p(H_{nm}|s)$ for a given score s. Therefore, such linkability can be accounted for in terms of the difference of conditional probabilities of each hypothesis H_m and H_{nm} for a given score s:

$$\mathrm{D}_{\leftrightarrow}(s) = p(H_m|s) - p(H_{nm}|s) \tag{4}$$

However, these two conditional probabilities are unknown. As defined in Sect. III, what can be computed a priori, and is known for each biometric template protection system and linkage function, are the *Mated* and *Non-mated samples* distributions (i.e., $p(s|H_m)$ and $p(s|H_{mn})$), that is, the probability of observing s knowing that two templates, protected with different keys, belong to mated samples or to non-mated samples.

In the following we explain how $\mathrm{D}_{\leftrightarrow}(s)$ is computed based on the known distributions $p(s|H_m)$ and $p(s|H_{mn})$. To do so, we start the computation from the likelihood ratio:

$$LR(s) = \frac{p(s|H_m)}{p(s|H_{nm})} \tag{5}$$

Bayes theorem can be applied to Eq. 5 to obtain

$$LR(s) = \frac{\frac{p(H_m|s)\cdot p(s)}{p(H_m)}}{\frac{p(H_{nm}|s)\cdot p(s)}{p(H_{nm})}} = \frac{p(H_m|s)}{p(H_{nm}|s)} \cdot \frac{p(H_{nm})}{p(H_m)} \tag{6}$$

Therefore, the probabilities shown in Eq. 4 are related as follows

$$\frac{p(H_m|s)}{p(H_{nm}|s)} = LR(s) \cdot \omega \tag{7}$$

where $\omega = p(H_m)/p(H_{nm})$ denotes the ratio between the unknown prior probabilities of the *Mated samples* and *Non-mated samples* distributions. From this last equation, we have

$$p(H_m|s) = LR(s) \cdot p(H_{nm}|s) \cdot \omega \tag{8}$$

Even if the values of those probability functions are unknown, s must belong to any of those two distributions. As a consequence,

$$p(H_{nm}|s) = 1 - p(H_m|s) \tag{9}$$

We can thus re-write Eq. 8 as

$$p(H_m|s) = LR(s) \cdot (1 - p(H_m|s)) \cdot \omega \tag{10}$$

$$\Leftrightarrow p(H_m|s) = \frac{LR(s) \cdot \omega}{1 + LR(s) \cdot \omega} \tag{11}$$

Now, as mentioned in Eq. 4, in order to define the local linkability measure $\mathrm{D}_{\leftrightarrow}(s)$ we are interested in the difference $p(H_m|s) - p(H_{nm}|s)$, which applying Eq. 9 yields:

$$\mathrm{D}_{\leftrightarrow}(s) = p(H_m|s) - p(H_{nm}|s) = 2 \cdot p(H_m|s) - 1 \tag{12}$$

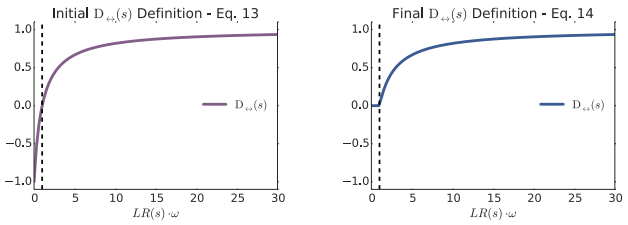


Fig. 3. (a) Initial definition of $D_{\leftrightarrow}(s)$, according to Eqs. 4 and 13, and (b) final definition of $D_{\leftrightarrow}(s)$, according to Eq. 14, where $D_{\leftrightarrow}(s) = 0$ for $LR(s) \cdot \omega < 1$. The dashed black line represents $LR(s) \cdot \omega = 1$, and thus $D_{\leftrightarrow}(s) = 0$.

Combining Eqs. 11 and 12, we can re-write the local linkability measure as

$$D_{\leftrightarrow}(s) = 2\frac{LR(s)\cdot\omega}{1+LR(s)\cdot\omega} - 1 \tag{13}$$

This measure is depicted in Fig. 3a. As we may observe, this leads to two different scenarios, depending on the value of $LR(s) \cdot \omega$, separated by a dashed vertical line:

- If $LR(s) \cdot \omega \leq 1$ (i.e., left of the vertical line, where $p(H_m|s) \leq p(H_{nm}|s)$), we can deduce, with some certainty, that both templates do *not* belong to the same instance. As a consequence, we cannot link both templates to the same data subject. In that case, $D_{\leftrightarrow}(s)$ yields a negative value, and hence the system is not linkable for that score. However, in order to have a single measure for unlinkable templates, a single value for this range of scores would be desired: $D_{\leftrightarrow}(s) = 0 \ \ \forall s \ \mid \ LR(s) \cdot \omega \leq 1$.
- If $LR(s) \cdot \omega > 1$ (i.e., right of the vertical line, where $p(H_m|s) > p(H_{nm}|s)$), we can state that it is more likely that both templates belong to mated instances, thereby making the templates somewhat linkable for those score values. In fact, in this case, the higher $LR(s)$, the more linkable the templates are. As a consequence, $D_{\leftrightarrow}(s)$ yields an increasing value in (0, 1], with higher values for more linkable templates (i.e., the higher $LR(s)$, the closer $D_{\leftrightarrow}(s)$ is to 1).

Keeping those remarks in mind, we define $D_{\leftrightarrow}(s)$ as a two-part function of s in Fig. 3b, depending on the value of the corresponding $LR(s) \cdot \omega$, as follows:

- $D_{\leftrightarrow}(s) = 0$ for s such that $LR(s) \cdot \omega \leq 1$ (i.e., unlinkable score values). See Fig. 3b for values $LR(s) \cdot \omega \leq 1$ (i.e., values to the left of the vertical dashed line).
- For the linkable score values (i.e., $LR(s) \cdot \omega > 1$), we define $D_{\leftrightarrow}(s)$ as in Eq. 13, hence defining a continuous function with $D_{\leftrightarrow}(s) = 0$ for $LR(s) \cdot \omega = 1$.

Therefore, we finally have

$$D_{\leftrightarrow}(s) = \begin{cases} 0 & \text{if } LR(s)\cdot\omega \leq 1 \\ 2\dfrac{LR(s)\cdot\omega}{1+LR(s)\cdot\omega} - 1 & \text{if } LR(s)\cdot\omega > 1 \end{cases} \tag{14}$$

This function $D_{\leftrightarrow}(s)$ presents some very interesting and desirable properties that make it specially well-behaved to be used as linkability metric (see Appendix A for the complete mathematical proof of these properties):

- Domain: $D_{\leftrightarrow}(s)$ is defined over the whole score domain.
- Continuity: $D_{\leftrightarrow}(s)$ is continuous in the whole domain.
- Range: $D_{\leftrightarrow}(s)$ is bounded in [0, 1].
- Monotonicity: $D_{\leftrightarrow}(s)$ is a monotonically increasing function.

It should be noted that in the definition we have a variable s (i.e., the linkage score) and a constant value ω (i.e., the quotient between the prior probabilities). For a given system, if the prior probabilities are available, those statistics should be used to compute ω. Otherwise, and in order not to bias the analysis towards one type of comparison, we can assume that both mated and non-mated comparison attempts are equally probable (i.e., worst-case scenario in the unlinkability analysis, see below). In other words, we assume that $p(H_m) = p(H_{nm})$, and thus set $\omega = 1$.

Regarding those prior probabilities, it should be highlighted that they are different from the mated and non-mated probabilities derived from normal verification attempts, where we can assume $p(H_m) \gg p(H_{nm})$. For the particular case of the unlinkability analysis and linkage scores, comparisons are carried out between a given template, enrolled in application A, and all templates enrolled in application B. Therefore, assuming that N different subjects are enrolled in database B, the probability of obtaining a mated and a non-mated score are, respectively:

$$p(H_m) = 1/N \tag{15}$$

$$p(H_{nm}) = (N-1)/N \tag{16}$$

As a consequence, for all $N \geq 2$, $p(H_m) \leq p(H_{nm})$, being equal only for $N = 2$. That leads to values $\omega \leq 1$, being $\omega = 1$ the worst-case scenario for the unlinkability evaluation, where only two subjects are enrolled in system B. For a more in depth analysis of this parameter, a specific experiment is performed in Sect. VI-D.

B. Global Measure $D_{\leftrightarrow}^{sys}$: System Overall Linkability

As described previously, it is also useful to have an estimation of the *unlinkability of the whole system* (and not just on a score-wise basis as is the case of $D_{\leftrightarrow}(s)$), which may allow a fairer benchmark of the unlinkability level of two or more systems.

For this purpose, we introduce the global metric $D_{\leftrightarrow}^{sys} \in [0, 1]$, which gives an estimation of the global linkability of a system, *independently* of the score. This way, if a system has $D_{\leftrightarrow}^{sys} = 1$ (i.e., case in which both the *Mated samples* and *Non-mated samples* distributions have no overlap, as shown in Fig. 2c), it means that it is fully linkable for all the scores of the *Mated samples* distribution domain (i.e., where $D_{\leftrightarrow}(s) = 1$). That is, if we evaluate the linkage function on two protected templates $\mathbf{T}_1$ and $\mathbf{T}_2$, we can decide (with almost all certainty) whether they conceal or not the same instance. Similarly, $D_{\leftrightarrow}^{sys} = 0$ (i.e., Fig. 2a, where both score distributions totally overlap) means that the system is fully unlinkable for the whole score domain. That is, independently of the score produced by the linkage function, it is equally probable that the two templates stem from the same instance (H_m) than from different instances (H_{nm}). All intermediate values of $D_{\leftrightarrow}^{sys}$ between 0 and 1 report a decreasing degree of unlinkability (i.e., increasing degree of linkability).

Therefore, we are interested on measuring how likely it is to get a score stemming from the *Mated samples* distribution. This can be achieved computing the difference $p(H_m \cap s) - p(H_{nm} \cap s)$ and integrating it over the whole score domain:

$$\begin{aligned}\int & p(H_m \cap s) - p(H_{nm} \cap s)\,\mathrm{d}s \\ &= \int p(s) \cdot (p(H_m|s) - p(H_{nm}|s))\,\mathrm{d}s \\ &= p(H_m) \int p(s|H_m) \cdot (p(H_m|s) - p(H_{nm}|s))\,\mathrm{d}s \\ &\quad + p(H_{nm}) \int p(s|H_{nm}) \cdot (p(H_m|s) - p(H_{nm}|s))\,\mathrm{d}s \end{aligned} \tag{17}$$

The second equality holds since $p(s) = p(s|H_m) \cdot p(H_m) + p(s|H_{nm}) \cdot p(H_{nm})$.

Regarding the success on linking templates, we are only interested in the first addend, that is, in the probabilities stemming from the *Mated samples* distribution. In addition, as in the definition of the local linkability measure $\mathrm{D}_{\leftrightarrow}(s)$, two templates can be linked only if $p(H_m|s) > p(H_{nm}|s)$. Therefore, we define $\mathrm{D}_{\leftrightarrow}^{sys}$ as

$$\mathrm{D}_{\leftrightarrow}^{sys} = \int\limits_{\substack{p(H_m|s) \\ > p(H_{nm}|s)}} p(s|H_m) \cdot (p(H_m|s) - p(H_{nm}|s))\,\mathrm{d}s \tag{18}$$

where the first addend in Eq. 17 has been normalised by $p(H_m)$ in order to obtain a measure independent of the prior probability.

Keeping that in mind, we can finally re-write Eq. 18 in terms of $\mathrm{D}_{\leftrightarrow}(s)$ and $p(s|H_m)$, hence defining $\mathrm{D}_{\leftrightarrow}^{sys}$ as:

$$\mathrm{D}_{\leftrightarrow}^{sys} = \int p(s|H_m) \cdot \mathrm{D}_{\leftrightarrow}(s)\,\mathrm{d}s \tag{19}$$

This way, the final value of $\mathrm{D}_{\leftrightarrow}^{sys}$ depends on: *i*) the domain of scores where the system is linkable (determined by $\mathrm{D}_{\leftrightarrow}(s)$); *ii*) how linkable the system is in that domain of scores (given by $\mathrm{D}_{\leftrightarrow}(s)$); and *iii*) how probable it is that such scores are produced (given by $p(s|H_m)$). Therefore, this new global measure assigns different levels of linkability to intermediate scenarios, not fully unlinkable or fully linkable.

Derived from the properties of the well-behaved $\mathrm{D}_{\leftrightarrow}(s)$ function, the global metric $\mathrm{D}_{\leftrightarrow}^{sys}$ also presents some desirable properties (see Appendix B for the complete mathematical proof):

- $\mathrm{D}_{\leftrightarrow}^{sys}$ is properly defined.
- Range: $\mathrm{D}_{\leftrightarrow}^{sys}$ is bounded in [0, 1].

C. Local and Global Metrics: Discussion

Since the mated and non-mated score distributions are statistical distributions, their integral between $-\infty$ and $+\infty$ is exactly one. As a consequence, if there is a range of scores $\{s_{\text{linkable}}\}$ where $p(s|H_m) > p(s|H_{nm})$, there must be a range of scores $\{s_{\text{unlinkable}}\}$ where $p(s|H_{nm}) > p(s|H_m)$. Therefore, given the definition of unlinkability presented in Section III, if two templates $\mathbf{T}_1$ and $\mathbf{T}_2$ are compared using a linkage function and they produce a score s_0 that falls in $s_0 \in \{s_{\text{unlinkable}}\}$, the two templates are unlinkable. In other words, for that score s_0 the system is unlinkable, and hence $\mathrm{D}_{\leftrightarrow}(s_0) = 0$. On the other hand, by obtaining that score s_0 we have learned something about the system: there must be some other score $s_1 \in \{s_{\text{linkable}}\}$ where the system is linkable, because, as mentioned above, we are dealing with statistical distributions.

In summary, if we obtain s_0 where $p(s_0|H_{nm}) > p(s_0|H_m)$, we know that: *i*) the system is unlinkable for s_0, $\mathrm{D}_{\leftrightarrow}(s_0) = 0$; *ii*) the system must be linkable for some s_1, whose value we don't know and for which $\mathrm{D}_{\leftrightarrow}(s_1) \neq 0$; and *iii*) as a consequence, the global measure $\mathrm{D}_{\leftrightarrow}^{sys} \neq 0$. Therefore, by obtaining a score in the unlinkable range of scores s_0, we actually learn that the system, as a whole, is linkable at some point, and this will be reflected by $\mathrm{D}_{\leftrightarrow}^{sys}$. However, this does not mean that $\mathrm{D}_{\leftrightarrow}(s_0)$ should be different from 0, because, given the unlinkability definition, for that specific score, the system is not linkable (i.e., we cannot link the two templates to the same instance, on the contrary, we can link them to different data subjects).

V. Proposed Linkability Evaluation Protocol

After presenting and showing the complementarity of both unlinkability measures (i.e., local and global), we propose here a general protocol for the evaluation of the unlinkability of a particular biometric template protection system.

First, it should be noted that, in practice, linkability is defined as the ability to link templates across different applications (i.e., stored in databases used by different applications). Templates stored in different databases are protected using different keys. As a consequence, in order to evaluate the linkability of templates protected with a specific algorithm, several different databases need to be created, containing templates extracted from the same biometric samples and using a different key for each database. With this in mind, the proposed protocol to evaluate the linkability of templates would run as follows:

1) Generate K databases of protected templates each of them using a different key. It is recommended that $K > 5$.
2) Compute the *Mated samples* score distribution for the selected linkage function, *across the K databases* generated in step 1. As in any recognition accuracy evaluation of a traditional biometric system, the larger the database, the more statistically significant the results of the evaluation are. Please see the end of this section for some relevant examples on how to conduct a biometric accuracy evaluation.
3) Compute the *Non-Mated samples* score distribution for the selected linkage function, *across the K databases* generated in step 1. Analogous considerations as those described for the *Mated samples* score distribution should be taken into account.
4) If the prior probabilities $p(H_m)$ and $p(H_{nm})$ are available, use them to compute ω. Otherwise, we can assume that $p(H_m) = p(H_{nm})$, and thus set $\omega = 1$.
5) Compute $\mathrm{D}_{\leftrightarrow}(s)$.
6) Compute $\mathrm{D}_{\leftrightarrow}^{sys}$.
7) Report $\mathrm{D}_{\leftrightarrow}(s)$ plots, together with the *Mated samples* and *Non-Mated samples* distributions, and the

corresponding global linkability values $D_{\leftrightarrow}^{sys}$, as in the examples analysed Sect. VI (Figs. 4 and 5).

8) Analyse the plots and $D_{\leftrightarrow}^{sys}$ values: where does $D_{\leftrightarrow}(s)$ reach a maximum? Is it close to one? Where is $LR(s) \cdot \omega = 1$? What is the global value $D_{\leftrightarrow}^{sys}$?

It should be noted that the aforementioned steps should be repeated for each linkage function considered, always following the same score computation protocol. This way, the more functions considered, the more significant and thorough the evaluation is. In particular, if F different functions are used in parallel to link the templates, F scores $s_1 = LS_1(\mathbf{T}_1, \mathbf{T}_2), \ldots, s_F = LS_F(\mathbf{T}_1, \mathbf{T}_2)$ will be obtained for a single pair of templates. Those scores will lead to the corresponding global linkability values $D_{\leftrightarrow,1}^{sys}, \ldots, D_{\leftrightarrow,F}^{sys}$. As a consequence, the system will be at least as vulnerable as the most challenging function considered: $D_{\leftrightarrow}^{sys} = \max_f \{D_{\leftrightarrow,f}^{sys}\}$.

In addition, and in order to make the evaluation reproducible, the following information should be reported for each linkage function:

- The database and particular protocol followed for computing the *Mated samples* and *Non-Mated samples* distributions. If two or more systems are compared, the same database and protocol should be followed in order to allow for a fair benchmark. A good example of how to define such protocol may be found in [35] and [36].
- Details on how the linkage function is computed (e.g.., input, knowledge required).
- Define the type of information the attacker has access to. As in any other security evaluation, it is important to set the adversary model considered in the evaluation. It should be noticed that, depending on the security model considered, the level of linkability of a system may vary.

In summary, it is important to notice that the level of linkability of a system depends on: i) the type of linkage function considered, and ii) the security model assumed for the evaluation.

The linkability protocol defined above is general in the sense that it can be applied to any *Lebesgue integrable* linkage function and any security model, therefore providing a general framework to benchmark linkability across systems (or for one single system under different linkage functions and / or adversary models). We believe that this assumption on the nature of the scores is not very restrictive, since a majority of the usual linkage functions used in biometric recognition produce continuous normalizable scores (e.g., distance-based functions). Nonetheless, the metrics could eventually be applied to other non-continuous linkage functions by mapping them to a distance-based function.

Finally, while the computation of mated and non-mated sets of scores is part of the linkability protocol described in this section, the description of how to perform such accuracy evaluation falls out of the scope of the present article. However, the interested reader can find plenty of works in the literature addressing this problem, being two good examples [35], [36].

In order to facilitate the use of the present framework and allow reproducibility of the article, a Python implementation of the metrics will be made available through the da/sec website[2] and the da/sec Github account.[3]

VI. Analysing the Unlinkability of BTP Systems

In this section, following the protocol proposed in Sect. V, we apply the metrics described in Sect. IV to evaluate the unlinkability of four previously proposed BTP schemes. In addition to that analysis, the impact of using different values for ω is studied in Sect. VI-D and a comparison with other metrics proposed in the literature is carried out in Sect. VI-E.

A. Experimental Setup

To show the generality of the proposed framework, four systems based on different biometric characteristics, using different features and comparators, and protected with different BTP algorithms, will be considered.

Iris verification + random XOR protection: as unprotected system we use a particular implementation within the publicly available University of Salzburg Iris Toolkit v1.0[4] [37] of the Log-Gabor based algorithm proposed by Masek [38]. Dissimiliarity scores are computed in terms of the Hamming Distance. Iris-codes are then protected by XORing them with random binary strings [39].

On-line signature verification + Homomorphic Encryption protection: as unprotected system, a state-of-the-art approach based on global features has been chosen [40]. From a set of 100 global features extracted from the x and y coordinates, the best 40 normalized features according to [41] are selected to form the final template. Dissimilarity scores are computed based on the Euclidean distance. Then, templates are protected with Homomorphic Encryption [42].

Face verification + Bloom filter protection: in the unprotected domain, the Log-Gabor Binary Pattern Histograms Sequences algorithm proposed in [43] is used. In particular, experiments are run on a publicly available implementation within the FaceRecLib[5] [44] and the Bob Toolbox. Dissimilarity scores are computed based on histograms intersections. Templates are then protected using Bloom filters[6] [45].

Fingervein verification + block re-mapping protection: as unprotected system we use the maximum curvature method presented in [46] to extract the connected vein pattern.[7] Similarity scores are computed in terms of the cross-correlation of rotated templates. Finally, block re-mapping [47] is applied to protect the templates.

Following the recommendations given in the linkability evaluation protocol presented in Sect. V, in the experiment we have considered $K = 10$ different keys to protect the templates. This simulates a case in the real world where the same subjects are enrolled in ten different applications and an attacker tries to link the templates in the ten databases to each other. Subsequently, the *Mated samples* and *Non-Mated*

[2] https://dasec.h-da.de/research/biometrics/unlinkability/

[3] https://github.com/dasec/unlinkability-metric

[4] http://www.wavelab.at/sources/

[5] https://pypi.python.org/pypi/facereclib

[6] https://github.com/dasec/face-bf-btp

[7] https://de.mathworks.com/matlabcentral/fileexchange/35716-miura-et-al-vein-extraction-methods

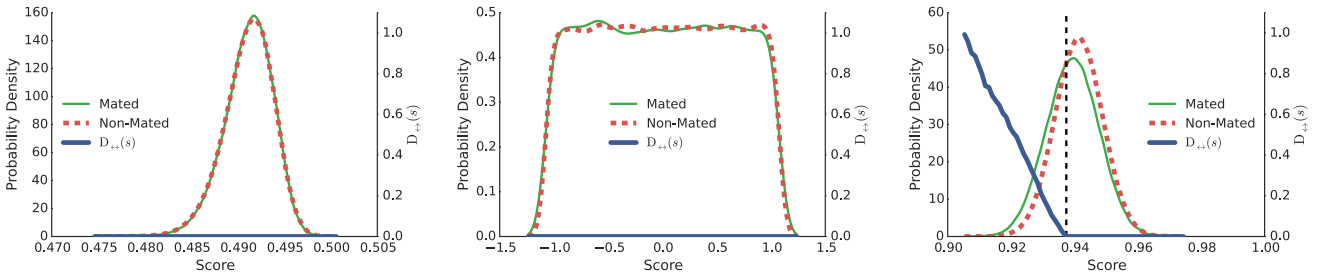


Fig. 4. Examples of unlinkability analysis for three different BTP systems. The dashed vertical lines represent $LR(s) = 1$. (a) Iris System + XOR BTP, $D_{\leftrightarrow}^{sys} = 0.0$. (b) Sign. System + HE BTP, $D_{\leftrightarrow}^{sys} = 0.0$. (c) Face System + Bloom Filter BTP, $D_{\leftrightarrow}^{sys} = 0.7$.

samples distributions are computed from scores stemming from templates stored in different databases (i.e., protected with different keys).

Keeping those remarks in mind, the iris, signature and face systems have been analysed over the corresponding sub-corpora of the multimodal BioSecure database [48], which comprises data of 210 different subjects. The iris and face sub-corpora include four samples of each eye and face, respectively (in the present study only the left eye sample is considered). On the other hand, the on-line signature sub-corpus comprises sixteen samples of each subject. In particular, for the *Mated samples* distribution, the following scores are computed: i) for the iris- and face-based systems, all possible mated comparisons where the reference sample is protected with a different key than the probe sample (i.e., 56,700 scores), and ii) for the signature-based systems, the first four signatures are used for enrolment and the twelve remaining ones are used to compute the scores (i.e., 113,400 scores). Then, for the *Non-Mated samples* distribution, in both cases the scores are obtained comparing one sample of the iris, face or the signature stored reference, with the first sample of the remaining subjects, protected with different keys (i.e., 987,525 scores all both characteristics).

The fingervein system used for the comparison with the state-of-the-art in unlinkability analysis is evaluated on the UTFVP database[8] [49]. The database comprises data from 60 different subjects, from whom the vascular pattern of the index, ring and middle finger of both hands was collected twice at each of the two acquisition sessions ($60 \times 6 \times 4 = 1,440$ fingervein samples). In the experiments, for the *Mated samples* distribution all possible scores are computed (i.e., 2,160 scores). Then, for the *Non-Mated samples* distribution, the scores are obtained comparing the enrolled fingervein with the first sample of the remaining instances, protected with different keys (i.e., 64,520 scores).

Regarding implementation details, LRs are computed in a point-wise fashion. In addition, since we have no prior evidence about the unknown prior probabilities of the *Mated samples* and *Non-mated samples* distributions to estimate the most appropriate value for ω, we will assume that $p(H_m) = p(H_{nm})$, and hence $\omega = 1$.

Finally, the experimental evaluation comprises four stages: i) the iris, signature and face systems are analysed in Sect. VI-B, utilising the dissimilarity score of the original *PIC* (Pseudonymous Identifier Comparator) as linkage function, ii) then further linkage functions trying to exploit some extra-knowledge or vulnerability of the original *PIC* are studied in Sect. VI-C for the face-based system, iii) the impact of different values of ω on the metrics is subsequently analysed in Sect. VI-D for both the facial and iris based templates, and iv) the fingervein system is evaluated with both the newly proposed metric and other general metrics described in Sect. II in order to show the advantages of the presented framework.

[8]http://www.sas.el.utwente.nl/home/datasets

B. First Linkage Function: Systems' PIC Scores

The first set of linkage functions analysed are the original *PIC*s described for each of the four protected systems considered [11], [39], [42], [45]. As such, in this case we only assume knowledge of the dissimilarity score computed by each scheme: $s = LS(\mathbf{T}_1, \mathbf{T}_2) = PIC(\mathbf{T}_1, \mathbf{T}_2)$. No further knowledge is assumed. Therefore, this is the trivial "zero-effort" linkage score, where only *PIC* based comparisons between protected templates are considered.

The corresponding *Mated samples* (solid green) and *Non-mated samples* (dashed red) distributions are depicted in Fig. 4: iriscodes protected with an XOR with a random string in Fig. 4a, on-line signature protected with Homomorphic Encryption in Fig. 4b, and face protected with Bloom filters in Fig. 4c. In all cases, the proposed score-wise linkability measure $D_{\leftrightarrow}(s)$ is depicted in blue, and the global measure $D_{\leftrightarrow}^{sys}$ is shown in the corresponding figure title.

We may observe in Figs. 4a and 4b that both types of protected templates are robust to linkage functions based on the dissimilarity score of the *PIC*. As it may be seen, the *Mated samples* and *Non-mated samples* distributions completely overlap, leading to $D_{\leftrightarrow}(s) = 0$ for the whole domain of scores. Accordingly, $D_{\leftrightarrow}^{sys} = 0$, as it corresponds to fully unlinkable templates. On the other hand, face protected templates (Fig. 4c) are still slightly linkable, showing a global linkability measure of $D_{\leftrightarrow}^{sys} = 0.07$. This is due to the fact that, for dissimilarity scores $s < 0.94$, it is more likely that templates stem from mated instances. However, since the probability of obtaining such scores is very low ($p(s|H_m) < 0.005$), the system is almost fully unlinkable - hence the low value for $D_{\leftrightarrow}^{sys}$.

Therefore, in these template protection systems, the proposed metrics allow us to conclude, based on objective values,

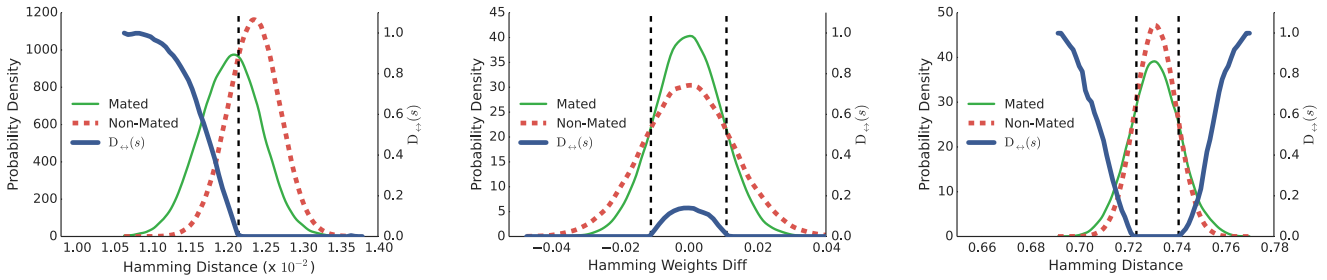


Fig. 5. Unlinkability analysis of the facial BTP scheme under three linkage functions different from the original *PIC*. (a) Reconstruction Function, $D_{\leftrightarrow}^{sys} = 0.25$. (b) Hamming Weights Function, $D_{\leftrightarrow}^{sys} = 0.08$. (c) (b) XOR Function, $D_{\leftrightarrow}^{sys} = 0.06$.

that the BTP approaches implemented to secure the templates serve their purpose and provide full or a very high degree of unlinkability.

C. Further Linkage Functions Different From the PIC

As it was highlighted in Sect. V, a complete unlinkability analysis should also include the evaluation of the robustness of the system to specifically designed linkage functions which exploit particular vulnerabilities of BTP approaches [50]. To that end, the aforementioned distributions should be estimated not only for the *PIC* score of the system, but also for other more sophisticated functions and their corresponding linkage scores. Accordingly, we now analyse the vulnerabilities of the protected face verification system described above to three additional linkage functions. In all cases, the same protocol for the score computation described for the face case study in Sect. VI-A has been followed, and analogous distributions for the corresponding linkage scores are depicted in Fig. 5. As it may be oserved, the score distributions are different from those obtained for the first linkage function based on the *PIC* score (see Fig. 4c), since different information is exploited to discriminate between mated and non-mated comparisons. A brief description of the linkage functions provided below, and for more details on the linkage functions, the reader is referred to [23], [45], and [51].

Reconstruction function (Fig. 5a): a methodology for the reconstruction of iriscodes given their corresponding Bloom filter based templates is proposed in [23], which can be also used to link templates. In this case, the linkage function will operate in a two-stage manner: i) reconstruct each unprotected template ($\mathbf{T}_1^{\text{rec}}$, $\mathbf{T}_2^{\text{rec}}$) from the protected counterpart ($\mathbf{T}_1$, $\mathbf{T}_2$), and ii) compute the Hamming distance between the reconstructed templates: $s = LS(\mathbf{T}_1, \mathbf{T}_2) = HD(\mathbf{T}_1^{\text{rec}}, \mathbf{T}_2^{\text{rec}})$. Therefore, for the reconstruction step, knowledge of the secret key of the system is assumed.

As we may observe in Fig. 5a, for half of the scores ($s < 1.2 \times 10^{-2}$), it is more likely that the templates which yielded s conceal the same instance. Which, in turn, leads to a successful linkage of the templates (i.e., $D_{\leftrightarrow}(s) > 0$). Moreover, for a wide part of those scores ($s < 1.15 \times 10^{-2}$), we can assume that with almost all certainty the two compared templates belong to the same instance, since $p(s|H_{nm}) = 0$ and $p(s|H_m) > 0$. Consequently, $D_{\leftrightarrow}(s) = 1$ in that range. The vulnerabilities to this linkage function are also reflected in the global linkability measure, $D_{\leftrightarrow}^{sys} = 0.25$, which is a value higher than that of the *PIC* score ($D_{\leftrightarrow}^{sys} = 0.07$). Being able to link templates is hence more likely using this reconstruction function than utilizing the scores output by the *PIC*.

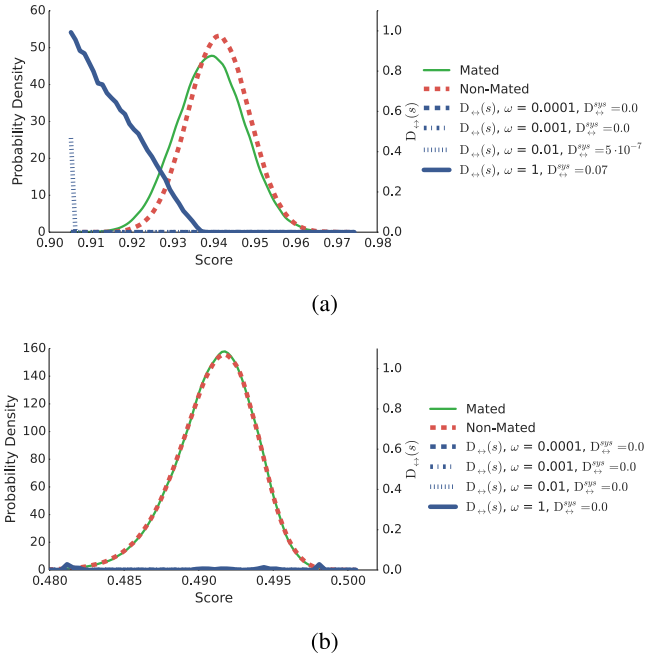


Fig. 6. Unlinkability analysis of the facial and iris based BTP systems for different values of ω. (a) Face System + Bloom Filter BTP - Analysis of ω. (b) Iris System + XOR BTP - Analysis of ω.

Hamming weights function (Fig. 5b): a function to link templates is proposed in [51], where templates protected with different keys and belonging to the same instance are shown to have similar Hamming Weights. Therefore, this linkage function will evaluate the Hamming Weight difference between the given protected templates: $s = LS(\mathbf{T}_1, \mathbf{T}_2) = |HW(\mathbf{T}_1) - HW(\mathbf{T}_2)|$. In this case, only knowledge of the templates is required.

For this function, the global linkability level is the very similar to the level achieved for the *PIC* score: $D_{\leftrightarrow}^{sys} = 0.07$. This means that this linkage function cannot extract any further information about whether two templates $\mathbf{T}_1$ and $\mathbf{T}_2$ conceal the same instance. More in detail, only for a small subset of the scores ($s \in [-0.01, 0.01]$) it is slightly more likely that both templates conceal the same instance, and therefore $D_{\leftrightarrow}(s) \in (0, 0.2]$. Since those scores are the most probable

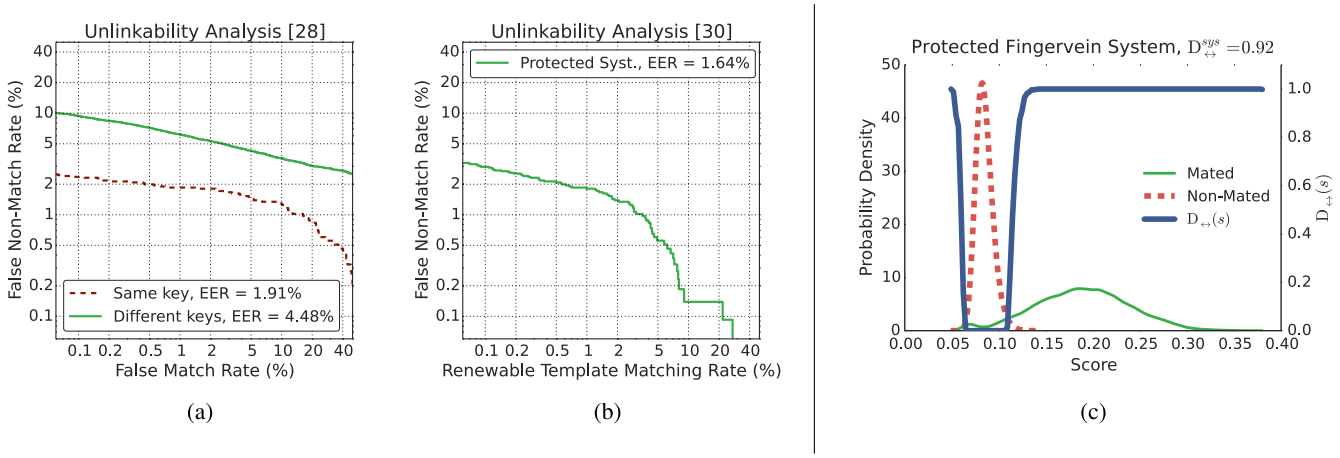


Fig. 7. Unlinkability evaluation of the fingervein system using: i) previously proposed metrics: left of the vertical line, figures (a) and (b), and ii) the framework proposed in this paper, right of the vertical line, figure (c).

(i.e., they yield the highest values in $p(s|H_m)$), the global linkability of the system raises to a value of $D_{\leftrightarrow}^{sys} = 0.08$.

XOR function (Fig. 5c): in the original concept of Bloom filter-based template protection [52], unlinkability was provided with an XOR operation. However, due to its linearity, templates protected with different keys can be linked exploiting such linearity by i) permuting one of the templates ($\mathbf{T}_2^{\text{perm}}$) and ii) finding the Hamming Distance with respect to the first template: $s = LS(\mathbf{T}_1, \mathbf{T}_2) = HD(\mathbf{T}_1, \mathbf{T}_2^{\text{perm}})$. In this case, only knowledge of the templates is required.

For this function, the global linkability level is again the very similar to the level achieved for the *PIC* score. In fact, an even smaller portion of scores on the tails of the distributions allow a successful linkage of the templates with almost all certainty, since $p(s|H_{nm}) = 0$ and $p(s|H_m) > 0$. Accordingly, $D_{\leftrightarrow}(s) = 1$ for those scores. Since those scores yield very low values for $p(s|H_m)$, the chances of obtaining those scores are low, hence showing a global linkability for the system of only $D_{\leftrightarrow}^{sys} = 0.06$.

Finally, following the evaluation guidelines proposed in Sect. V and taking into account the four linkage functions analysed (the *PIC* score in Sect. VI-B and the three functions evaluated in the present section), we can conclude that the global linkability value of the system is $D_{\leftrightarrow}^{sys} = \max\{0.07, 0.25, 0.08, 0.06\} = 0.25$.

D. Analysis of Parameter ω

In addition to evaluating different linkage functions, the present framework allows the analysis of different scenarios in terms of the a priori probabilities $p(H_m)$ and $p(H_{nm})$. This can be done varying its ratio, ω (see Sect. IV-A for the definition of this parameter). To that end, four different values $\omega = \{0.0001, 0.001, 0.01, 1\}$ have been used in Fig. 6 to analyse the *PIC* linkage scores of the facial and iris based BTP schemes.

As it may be observed, the higher ω is, the higher the value of $D_{\leftrightarrow}(s)$ for each score s with $p(s|H_m) > 0$ is. This is due to the fact that the ratio $(LR(s) \cdot \omega) / (1 + LR(s) \cdot \omega)$ increases with ω. Regarding the unlinkability property, this also reflects the fact that, for a higher ω, the attacker knows that the probability of having a mated comparison is higher (a higher ω implies a smaller number of enrolled subjects), and therefore the probabilities of linking the subjects increases.

Finally, such increase in $D_{\leftrightarrow}(s)$ also has an impact on the global linkability of the system: for $\omega = 1$, $D_{\leftrightarrow}^{sys}$ reaches its maximum value.

E. Advantages Over Previously Proposed Metrics

In this last set of experiments, we show how the proposed framework is able to determine that a system is linkable even when other existing metrics fail to unveil this vulnerability. To that end, we analyse the fingervein system with two of the general unlinkability metrics that have been proposed in the state-of-the-art [28], [30] and with the framework proposed in this article, for the *PIC* linkage function. It should be highlighted that, although the approaches presented in [28] and [30] are based on the analysis of the DET curves, in the end unlinkability is considered as a binary value: systems are either linkable or unlinkable. The results are depicted in Fig. 7.

In the first approach proposed in [28] (see Fig. 7a), regular DET curves are depicted in two scenarios: i) the recognition accuracy evaluation (dashed red), where both mated and non-mated scores are computed on templates protected with a single key, and ii) the unlinkability analysis (solid green), where mated scores stem from templates extracted from the *same* instance and protected with *different* keys, and non-mated scores from templates extracted from *different* instances and protected with *different* keys. It should be noted that in [29], the FMR of the second scenario is denoted Cross-Match Rate (CMR) and the FNMR is denoted False Cross Match Rate (FCMR). As it may be observed, the error rates increase for the unlinkability analysis with respect to the accuracy analysis. In particular, the EER increases from 1.91% to 4.48%. Therefore, the task of discriminating templates extracted from mated and non-mated instances is harder when the templates are protected with different keys. According to the analysis carried out in [26], [28], and [29], we should hence conclude that the protected system is unlinkable.

In the second approach proposed in [30], the DET-like curve comparing the FNMR of templates protected with a single key

and the RTMR of templates protected with different keys is depicted in Fig. 7b. In this case, the curve shows a similar behaviour to the accuracy curve shown in Fig. 7a in dashed red. Therefore, as the authors suggest in [30], the task of discriminating templates protected with different keys is as hard as achieving a false match with a random template, thereby making the protected system unlinkable.

Finally, the protected system using different keys is analysed in Fig. 7c with the metrics proposed in the present article. As it may be observed, $\mathrm{D}_{\leftrightarrow}^{sys}$ yields a value of 0.92, very close to 1, thereby indicating that the templates are almost fully linkable. This is corroborated by the mated and non-mated distributions, which are almost fully separable. As a consequence, $\mathrm{D}_{\leftrightarrow}(s) = 1$ for almost the entire domain of scores for both systems. That means that the simple linkage function that uses the similarity scores of the *PIC* is enough to link protected templates, hence contradicting the results of the evaluations carried out following [28], [30].

VII. Conclusions

We have proposed in the present article two new quantitative measures for the unlinkability analysis of biometric templates, which can be applied to any biometric template protection scheme, regardless of its overall strategy (i.e., cancelable biometrics, cryptobiometrics or encrypted system). On the one hand, $D_{\leftrightarrow}(s)$ provides a score-wise analysis of the linkability of the templates, in order to carry out a thorough and detailed evaluation of the templates' unlinkability. On the other hand, $D_{\leftrightarrow}^{sys}$ evaluates the system as a whole, thereby allowing a benchmark of the linkability of different systems. Furthermore, the necessary steps towards a complete unlinkability evaluation have been proposed in order to develop a full security benchmark for biometric template protection schemes, key for the further deployment of biometric systems, as stated in [9]. We therefore believe that the proposed framework will contribute to the advancement of biometric technologies in the future. To that end, and with the aim to make the article reproducible, an implementation of the metrics has been made public through the da/sec website and the da/sec Github account.

As with any other security oriented property, linkability is not an intrinsic property of the system. Instead, it is directly related to particular threats in the form of i) linkage functions and ii) security models assumed, which must be analysed on a one by one basis. This is therefore the approach followed in the proposed framework. Both metrics evaluate in a quantitative manner the degree of linkability of protected biometric templates with respect to a given linkage function, which may exploit any particular weakness of the system at hand.

For the evaluation, only access to the scores obtained from the corresponding linkage functions is needed, without assumptions on the input data or the nature of the templates and systems. We deem that such requirement is straightforward for both system developers and evaluators, as no access to inner modules or communication channels of the system is required.

A direct consequence of this sole requirement of the framework is the general applicability of the metrics. Contrary to some previously proposed unlinkability metrics, where assumptions on the underlying algorithms of the system or the data were made, there is no restriction on the type of systems which can be analysed, and the number of linkage functions for each system. The only requirement is the use of Lebesgue integrable linkage functions. We believe this is a minor limitation, since the majority of usual linkage functions meet that condition, and other non-continuous functions can be mapped to distance-based functions for the analysis. This has been shown by evaluating four previously proposed biometric template protection systems based on different characteristics and protection strategies, as well as several linkage functions.

In addition, and with respect to the previously proposed approaches for unlinkability assessment, it has been shown that the proposed metrics can reveal linkability vulnerabilities concealed to other metrics. Furthermore, our metrics do so by providing not only local information ($\mathrm{D}_{\leftrightarrow}(s)$) but also a global measure ($\mathrm{D}_{\leftrightarrow}^{sys}$), both taking into account the continuous nature of linkability (i.e., instead of a binary decision, different degrees of unlinkability can be achieved). As a consequence, the framework provides a fair benchmark of several systems and linkage functions as long as the same evaluation protocol is followed (e.g., same data are used).

Finally, it should be noted that the proposed framework can be applied not only to the evaluation of biometric recognition schemes, but also to other fields where privacy protection is key. For instance, whereas user profiling in social media can bring benefits such as content [53] or travel recommendation [54], sensitive information can also be recovered out of seemingly anonymous data [55]. As stated in [55], a direct consequence of this fact is that "individuals may be confronted with social exclusion, prejudice and discrimination risks both in their workplace and in their social environment". The proposed framework can be therefore used to analyse how vulnerable the information stored in the cloud is to such profiling (i.e., linking) activities, and further help to prevent these undesired practices.

Appendix
Mathematical Proofs

We present here the proofs of the properties of both linkability measures.

A. Local Measure $\mathrm{D}_{\leftrightarrow}(s)$ Properties

1) Domain: Since $LR(s)$ is defined over the whole domain of s scores, $\mathrm{D}_{\leftrightarrow}(s)$ is defined for any two $p(s|H_m)$ and $p(s|H_{nm})$ distributions.

2) Continuity: Additionally, $\mathrm{D}_{\leftrightarrow}(s)$ is continuous since it is piecewise continuous and it is also continuous at $LR(s) \cdot \omega = 1$:

$$\lim_{LR(s)\cdot\omega\to 1^-} \mathrm{D}_{\leftrightarrow}(s) = \lim_{LR(s)\cdot\omega\to 1^-} 0 = 0 \tag{20}$$

$$\lim_{LR(s)\cdot\omega\to 1^+} \mathrm{D}_{\leftrightarrow}(s) = \lim_{LR(s)\cdot\omega\to 1^+} 2\frac{LR(s)\cdot\omega}{1+LR(s)\cdot\omega} - 1 = 0 \tag{21}$$

3) Range: Furthermore, $\mathrm{D}_{\leftrightarrow}(s)$ is bounded in $[0, 1]$:

$$\lim_{LR(s)\cdot\omega\to 0} \mathrm{D}_{\leftrightarrow}(s) = \lim_{LR(s)\cdot\omega\to 0} 0 = 0 \tag{22}$$

$$\lim_{LR(s)\cdot\omega\to +\infty} \mathrm{D}_{\leftrightarrow}(s) = \lim_{LR(s)\cdot\omega\to +\infty} 2\frac{LR(s)\cdot\omega}{1+LR(s)\cdot\omega} - 1 \underset{\text{L'Hôpital}}{=} \lim_{LR(s)\cdot\omega\to +\infty} 2\frac{1}{1} - 1 = 1 \tag{23}$$

4) Monotonicity: $\mathrm{D}_{\leftrightarrow}(s)$ is a monotonically increasing function, since its derivative is always non-negative:

$$\mathrm{D}'_{\leftrightarrow}(s) = \begin{cases} 0 & \text{if } LR(s)\cdot\omega \le 1 \\ 2\dfrac{\omega}{(1+LR(s)\cdot\omega)^2} > 0 & \text{if } LR(s)\cdot\omega > 1 \end{cases} \tag{24}$$

being the second term positive since all figures involved are positive numbers. As a consequence of its monotonicity, $\mathrm{D}_{\leftrightarrow}(s)$ provides appropriate increasing values for the evaluation of a monotonic property such as templates linkability.

B. Global Measure $\mathrm{D}^{sys}_{\leftrightarrow}$ Properties

1) Properly Defined: Since $\mathrm{D}_{\leftrightarrow}(s)$ and $p(s|H_m)$ are continuous functions, their product is also continuous. As a consequence, by the Riemann-Lebesgue theorem [56], the product is integrable, hence being $\mathrm{D}^{sys}_{\leftrightarrow}$ properly defined. This property also holds for discrete functions, since they are also Lebesgue integrable.

2) Range: Let us now prove that $\mathrm{D}^{sys}_{\leftrightarrow} \in [0, 1]$. On the one hand, since $p(s|H_m)$ is a probability density, which, integrated over $[s_{min}, s_{max}]$ yields an area of one, we have

$$\mathrm{D}^{sys}_{\leftrightarrow} = \int \mathrm{D}_{\leftrightarrow}(s)\cdot p(s|H_m)\mathrm{d}s \le \int 1\cdot p(s|H_m)\mathrm{d}s = 1 \tag{25}$$

Similarly,

$$\mathrm{D}^{sys}_{\leftrightarrow} = \int \mathrm{D}_{\leftrightarrow}(s)\cdot p(s|H_m)\mathrm{d}s \ge \int 0\cdot p(s|H_m)\mathrm{d}s = 0 \tag{26}$$

REFERENCES

[1] A. K. Jain, "Technology: Biometric recognition," *Nature*, vol. 449, pp. 38–49, Sep. 2007.

[2] Government of India. (2012). *Unique Identification Authority of India*. [Online]. Available: https://uidai.gov.in/

[3] European Comission. (2013). *Smart Borders*. [Online]. Available: http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/index_en.htm

[4] *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)*, document Regulation (EU) 2016/679, European Council, 2016.

[5] P. Campisi, Ed. *Security and Privacy in Biometrics*. London, U.K.: Springer, 2013.

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

[7] *Information Technology—Security techniques—Biometric Information Protection*, document ISO/IEC 24745:2011, ISO/IEC JTC1 SC27 Security Techniques, ISO, 2011.

[8] *Information Technology—Performance Testing of Biometric Template Protection Schemes*, ISO/IEC FDIS 30136, ISO/IEC JTC1 SC37 Biometrics, International Organization for Standardization, 2017.

[9] S. Rane, "Standardization of biometric template protection," *IEEE MultimediaMag.*, vol. 21, no. 4, pp. 94–99, Oct. 2014.

[10] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: Issues and challenges," *Proc. IEEE*, vol. 92, no. 6, pp. 948–960, Jun. 2004.

[11] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Syst. J.*, vol. 40, no. 3, pp. 614–634, Apr. 2001.

[12] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP J. Inf. Secur.*, vol. 2011, no. 1, p. 3, 2011.

[13] M. Ferrara, D. Maltoni, and R. Cappelli, "A two-factor protection scheme for MCC fingerprint templates," in *Proc. BIOSIG*, 2014, pp. 1–8.

[14] C. Champod and D. Meuwly, "The inference of identity in forensic speaker recognition," *Speech Commun.*, vol. 31, no. 2, pp. 193–203, 2000.

[15] A. M. Bazen and R. N. J. Veldhuis, "Likelihood-ratio-based biometric verification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 86–94, Jan. 2004.

[16] J. Gonzalez-Rodriguez, A. Drygajlo, D. Ramos-Castro, M. Garcia-Gomar, and J. Ortega-Garcia, "Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition," *Comput. Speech Lang.*, vol. 20, nos. 2–3, pp. 331–355, 2006.

[17] T. Ali, L. Spreeuwers, R. Veldhuis, and D. Meuwly, "Effect of calibration data on forensic likelihood ratio from a face recognition system," in *Proc. Int. Conf. Biometrics, Theory, Appl., Syst. (BTAS)*, 2013, pp. 1–8.

[18] D. Ramos, R. Haraksim, and D. Meuwly, "Likelihood ratio data to report the validation of a forensic fingerprint evaluation method," *Data Brief*, vol. 10, pp. 75–92, Feb. 2017.

[19] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Proc. Int. Conf. Audio-Video-Based Biometric Person Authentication (AVBPA)*, 2003, pp. 393–402.

[20] Y. Dodis, L. Reyzin, and A. Smith, "How to generate strong keys from biometrics and other noisy data," in *Proc. Eurocrypt*, 2004, pp. 523–540.

[21] I. Buhan, J. Doumen, P. Hartel, and R. Veldhuis, "Fuzzy extractors for continuous distributions," in *Proc. ACM Int. Conf. Comput. Sci. (ICCS)*, 2007, pp. 353–355.

[22] K. Simoens, J. Bringer, H. Chabanne, and S. Seys, "A framework for analyzing template security and privacy in biometric authentication systems," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 833–841, Apr. 2012.

[23] J. Bringer, C. Morel, and C. Rathgeb, "Security analysis of bloom filter-based iris biometric template protection," in *Proc. Int. Conf. Biometrics (ICB)*, 2015, pp. 527–534.

[24] A. Kholmatov and B. Yanikoglu, "Realization of correlation attack against the fuzzy vault scheme," *Proc. SPIE*, vol. 6819, p. 68190O, Mar. 2008.

[25] K. Simoens, P. Tuyls, and B. Preneel, "Privacy weaknesses in biometric sketches," in *Proc. IEEE Signal Process.*, May 2009, pp. 188–203.

[26] 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 *Proc. Int. Conf. Data Privacy Manage. Auto. Spontaneous Secur. (DPM/SETOP)*, 2009, pp. 78–92.

[27] I. Buhan, J. Merchan, and E. Kelkboom, "Efficient strategies to play the indistinguishability game for fuzzy sketches," in *Proc. Int. Workshop Inf. Forensics Secur. (WIFS)*, 2010, pp. 1–6.

[28] E. J. C. Kelkboom, J. Breebaart, T. A. M. Kevenaar, I. Buhan, and R. N. 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.

[29] A. Nagar, K. Nandakumar, and A. K. Jain, "Biometric template transformation: A security analysis," *Proc. SPIE*, p. 75410O, Jan. 2010.

[30] E. Piciucco, E. Maiorana, C. Kauba, A. Uhl, and P. Campisi, "Cancelable biometrics for finger vein recognition," in *Proc. SPLINE*, Jul. 2016, pp. 1–5.

[31] 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.

[32] 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, 2014.

[33] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[34] *Information Technology—Vocabulary—Part 37: Biometrics*, document ISO/IEC 2382-37:2012, ISO and IEC, ISO/IEC TC JTC1 SC37 Biometrics, 2012.
[35] M. Ferrara, D. Maltoni, and R. Cappelli, "Non-invertible minutia cylinder-code representation," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1727–1737, Dec. 2012.
[36] A. Mansfield and J. Wayman, "Best practices in testing and reporting performance of biometric devices," Nat. Phys. Lab., Teddington, U.K., NPL Rep. CMSC 14/02, Aug. 2002, [Online]. Available: http://www.cesg.gov.uk/
[37] A. Uhl and P. Wild, "Weighted adaptive hough and ellipsopolar transforms for real-time iris segmentation," in *Proc. Int. Cong. Biometrics (ICB)*, 2012, pp. 1–8.
[38] L. Masek and P. Kovesi, "MATLAB source code for a biometric identification system based on iris patterns," M.S. thesis, School Comput. Sci. Softw. Eng., Univ. Western Australia, Crawley, WA, Australia, 2003.
[39] J. Zuo, N. K. Ratha, and J. H. Connell, "Cancelable iris biometric," in *Proc. ICPR*, 2008, pp. 1–4.
[40] M. Martinez-Diaz, J. Fierrez, R. P. Krish, and J. Galbally, "Mobile signature verification: Feature robustness and performance comparison," *IET Biometrics*, vol. 3, no. 4, pp. 267–277, 2014.
[41] J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Performance and robustness: A trade-off in dynamic signature verification," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 1697–1700.
[42] M. Gomez-Barrero, J. Fierrez, J. Galbally, E. Maiorana, and P. Campisi, "Implementation of fixed-length template protection based on homomorphic encryption with application to signature biometrics," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2016, pp. 191–198.
[43] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1. Oct. 2005, pp. 786–791.
[44] M. Günther, R. Wallace, and S. Marcel, "An open source framework for standardized comparisons of face recognition algorithms," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 547–556.
[45] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, and J. Fierrez, "Unlinkable and irreversible biometric template protection based on Bloom filters," *Inf. Sci.*, vols. 370–371, pp. 18–32, Nov. 2016.
[46] N. Miura, A. Nagasaka, and T. Muyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *IEICE Trans. Inf. Syst.*, vol. 90, no. 8, pp. 1185–1194, 2007.
[47] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Syst. J.*, vol. 40, pp. 614–634, 2001.
[48] J. Ortega-Garcia *et al.*, "The multiscenario multienvironment biosecure multimodal database (BMDB)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1097–1111, Jun. 2010.
[49] 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)*, 2013, pp. 1–5.
[50] K. Simoens *et al.*, "Criteria towards metrics for benchmarking template protection algorithms," in *Proc. ICB*, 2012, pp. 498–505.
[51] J. Hermans, B. Mennink, and R. Peeters, "When a bloom filter is a doom filter: Security assessment of a novel iris biometric template protection system," in *Proc. BIOSIG*, 2014, pp. 1–6.
[52] C. Rathgeb, F. Breitinger, and C. Busch, "Alignment-free cancelable iris biometric templates based on adaptive bloom filters," in *Proc. Int. Conf. Biometrics, ICB*, 2013, pp. 1–8.
[53] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," in *Proc. ACM Int. Conf. Recommender Syst.*, 2010, pp. 199–206.
[54] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, and G. Chen, "Travel recommendation using geo-tagged photos in social media for tourist," *Wireless Pers. Commun.*, vol. 80, no. 4, pp. 1347–1362, 2015.
[55] L. Mitrou, M. Kandias, V. Stavrou, and D. Gritzalis, "Social media profiling: A panopticon or Omniopticon tool?" in *Proc. Int. Conf. Surveill. Stud. Netw. (CSSN)*, 2014, pp. 1–15.
[56] T. M. Apostol, *Mathematical Analysis*. Pearson, 1974, pp. 169–172.

Marta Gomez-Barrero received the M.Sc. degree in computer science and mathematics and the Ph.D. degree in electrical engineering from the Universidad Autonoma de Madrid, in 2011 and 2016, respectively. Since 2016, she has been a Post-Doctoral Researcher with the Center for Research in Security and Privacy, Germany. Her current research focuses on the development of privacy-enhancing biometric technologies as well as presentation attack detection methods, within the wider fields of pattern recognition and machine learning. She was a recipient of a number of distinctions, including the EAB European Biometric Industry Award 2015, the Siew-Sngiem Best Paper Award at ICB 2015, the Archimedes Award for young researches from Spanish Ministry of Education in 2013, and the Best Poster Award at ICB 2013.

Javier Galbally received the M.Sc. degree in electrical engineering from the Universidad de Cantabria, Spain, in 2005, and the Ph.D. degree in electrical engineering from the Universidad Autnoma de Madrid, Spain, in 2009. He was an Assistant Professor with the Universidad Autnoma de Madrid until 2012. In 2013, he joined the European Commission in the DG Joint Research Centre, where he is currently a Post-Doctoral Researcher. His research interests are mainly focused on pattern and biometric recognition, including biometric systems security and vulnerabilities, biometric template protection, and inverse biometrics. He was a recipient of a number of distinctions, including the IBM Best Student Paper Award at ICPR 2008, finalist of the EBF European Biometric Research Award 2009, the Best Ph.D. Thesis Award 2010 by the UAM, the Best Poster Award at IJCB 2013 and 2017, and the Best Paper Award at ICB 2015.

Christian Rathgeb is currently a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt, Germany, and also a Principal Investigator with the Center for Research in Security and Privacy. His research includes pattern recognition, iris biometrics, and privacy enhancing technologies for biometric systems. He served on various program committees and conferences, journals, and magazines as a reviewer. He is currently a member of the European Association for Biometrics and a Program Chair of the International Conference of the Biometrics Special Interest Group.

Christoph Busch received the Diploma degree from the Technical University of Darmstadt (TUD), Darmstadt, Germany, and the Ph.D. degree in computer graphics from TUD, in 1997. He joined the Fraunhofer Institute for Computer Graphics, Darmstadt, in 1997. He has been a Lecturer in biometric systems with DTU, Copenhagen, since 2007. He is currently a member with the Faculty of Computer Science and Media Technology, Norwegian University of Science and Technology, Norway, and holds a joint appointment with the Faculty of Computer Science, Hochschule Darmstadt. His research includes pattern recognition, multimodal and mobile biometrics, and privacy enhancing technologies for biometric systems. He has co-authored over 400 technical papers and has been a speaker at international conferences. He is currently a Co-Founder of the European Association for Biometrics and a Convener of WG3 in ISO/IEC JTC1 SC37 on Biometrics.