Efficient Known-Sample Attack for Distance-Preserving Hashing Biometric Template Protection Schemes

Original
Efficient Known-Sample Attack for Distance-Preserving Hashing Biometric Template Protection Schemes / Lai, Y.; Jin, Z.; Wong, K.; Tistarelli, M.. - In: IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. - ISSN 1556-6013. - 16:(2021), pp. 3170-3185. [10.1109/TIFS.2021.3073802]

Efficient Known-Sample Attack for Distance-Preserving Hashing Biometric Template Protection Schemes

Yenlung Lai, Zhe Jin, *Member, IEEE*, KokSheik Wong, and Massimo Tistarelli, *Senior Member, IEEE*

***Abstract*— The rapid deployment of biometric authentication systems raises concern over user privacy and security. A biometric template protection scheme emerges as a solution to protect individual biometric templates stored in a database. Among all available protection schemes, a template protection scheme that relies on distance-preserving hashing has received much attention due to its simplicity and efficiency in offering privacy protection while archiving decent authentication performance. In this work, we introduce an efficient attack called known sample attack and demonstrate that most state-of-art template protection schemes that utilize distance-preserving hashing can be compromised in practice (within few seconds), especially when the output is significantly smaller than the original input sample size. These findings further motivated our subsequent work in proposing a secure authentication mechanism to resist such an attack with proper study over the distribution of the input samples. Furthermore, we conducted revocability, unlinkability analysis to demonstrate the satisfactory of general biometric template protection requirements; and showed the resistance of various security and privacy attacks, i.e., false acceptance attack, and attack via record multiplicity.**

***Index Terms*— Biometric, Known-Sample attack, secure authentication.**

Below is the revised funding acknowledgement

I. INTRODUCTION

BIOMETRICS refers to the automatic verification or identification process using the physiological or behavioral characteristics of humans. Some typical biometric traits include fingerprint, face, and iris, which are inherently and permanently associated with individuals. Due to its attractive features such as token/ID card-free and ease of use (e.g., no need to remember the complex password), the biometric authentication system is widely deployed in many applications that demand identity management [1]. However, since biometric is permanently associated with individuals, direct exposure of personal biometric data to a third party may lead to security and privacy issues. Specifically, once the database that utilized to store individual biometric data (i.e., template) is compromised, the attacker could transform the stored template to its original form, which leads to severe privacy invasion and permanent identity loss for individual users.

Therefore, the security and privacy issues of biometric template storage are of great concern. As a remedy, biometric template protection (BTP), which is a protection scheme, is put forward by researchers to address the concerns mentioned above [2].

Briefly, BTP is designed with the primary goal of transforming an unprotected biometric template into a protected biometric template using a parameterized function. An effective biometric template protection scheme should satisfy the following four requirements: *non-invertibility, recovability, unlinkability,* and *performance preservation* [3].

In this paper, we focus on distance-preserving hashing BTP: a BTP scheme that utilizes a heuristic hash function $f : \mathbb{R}^k \to \{0, 1\}^n$, with distance-preserving property, to generate a hashed template over the hashed domain $\{0, 1\}^n$. Most conventional distance-preserving hashing BTPs fulfill the properties of non-invertibility, revocability, and unlinkability. However, to realize performance preservation, such BTP must preserve the relative distance between different biometric templates after the BTP applied. Such a goal is necessary to ensure that similar templates render high similarity scores for better recognition utility. Nonetheless, the distance preserving property induces information leakage and jeopardizes the system security. We introduce an *efficient* security attack for existing distance-preserving hashing BTP. We show that without proper designation of the hash function, such distant preservation property could lead to a severe security breach, hence leaving the security of the system in doubt for practical use. Subsequently, we propose a countermeasure to resist such an attack while preserving the original authentication performance.

The rest of this paper is organized as follows: a literature survey on existing research on BTP is covered in Section II. Our motivations and contributions of this paper are highlighted in Section III. An efficient attack, which is robust against the current state-of-the-art distance preserving hashing BTP schemes, is put forward in Section IV. A proposal of solution

---

Manuscript received January 4, 2020; revised September 20, 2020 and January 13, 2021; accepted April 8, 2021. This work was supported in part by the grants from the MSCA-Rise European Project IDENTITY and the Italian Ministry for Research Project SPADA and in part by the Ministry of Higher Education (MOHE) Malaysia through Fundamental Research Grant Scheme under Grant FRGS/1/2018/ICT02/MUSM/03/3. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shantanu D. Rane. *(Corresponding author: Zhe Jin.)*

Yenlung Lai, Zhe Jin, and KokSheik Wong are with the School of Information Technology, Monash University Malaysia, Subang Jaya 47500, Malaysia (e-mail: yenlung.lai@monash.edu; jin.zhe@monash.edu; wong.koksheik@monash.edu).

Massimo Tistarelli is with the Computer Vision Laboratory, Universita di Sassari, 07100 Sassari, Italy (e-mail: tista@uniss.it).

Digital Object Identifier 10.1109/TIFS.2021.3073802

Below is the revised funding acknowledgement.
This research was partially supported by the EU H2020 Marie Curie project IDENTITY and by grants by the Italian Ministry for Research (PRIN 2015 "Cosmos") and the Ministry for Economic Development project "SPADA" and the Ministry of Higher Education (MOHE) Malaysia through Fundamental Research Grant Scheme under Grant FRGS/1/2018/ICT02/MUSM/03/3.

to resist against the attack is given in V. Our experiments and evaluations are covered in Section VI. A concluding remark is given in Section VII.

II. RELATED WORK

A. Distance-Preserving Hashing BTP

One of the representative approaches to construct a BTP scheme is by 'hashing'. In different with the conventional cryptographic one-way hashing, e.g., SHA-512, the term 'hashing' in BTP context relies on heuristic distance-preserving hash function $f : \mathbb{R}^k \rightarrow \{0, 1\}^n$ to generate a hashed template $y = f(x) \in \{0, 1\}^n$ from its original biometric template $x \in \mathbb{R}^k$. The hashed template obtains its non-invertible characteristic through information loss via dimensional reduction, where $n < k$. Follow the studies in [4], [5], reconstructing the input x from y is equivalent to solving an under-determined linear system, which is computationally hard if $n \ll k$.

In general, there are two main categories where distance-preserving hashing is utilized for BTP in the literature, namely, Bio-hashing and Locality Sensitive Hashing (LSH).

1) Bio-Hashing: The earliest attempt of applying such technique to biometric is Bio-hashing [6] for protecting human fingerprint template. In Bio-hashing, the hashing operation is performed by using randomly generated orthogonal matrices which are implemented as dimension-reducing mapping to project the original fingerprint template to a random string of lower dimension. The projection supports distance-preserving property, where the pairwise distance of the fingerprint templates is preserved in the hashed domain. There are some similar lines of work on Bio-hashing applied to different biometric modalities, including palm [7], iris [8], and human speech [9].

2) Locality Sensitive Hashing (LSH): On the other hand, recent approaches (including the state-of-the-art works) [10]–[12] use LSH for BTP. Briefly, LSH refers to the use of multiple hash functions h_i over a LSH hashing family $H = \{h_i : \mathbb{R}^k \rightarrow U\}_{i=1}^{n}$, where individual hash function h_i is designed to hash the input $x, y \in \mathbb{R}^k$ to an output hash space $U = \{0, 1\}$ where $n < k$. LSH ensures the input pair x, y with small distance (viz., high similarity) renders a higher probability of collision in the hashed domain and vice versa. There are few reported ways for designing the hash function $h_i \in H$. For instances, Lai *et al.* [11] construct a LSH family $H = \{h_i : \{0, 1\}^k \rightarrow \{0, \ldots, q - 1\}\}_{i=1}^{n}$ for iris template protection, while Jin *et al.* [10] construct a LSH family $H = \{h_i : \mathbb{R}^k \rightarrow \{0, \ldots, q - 1\}\}_{i=1}^{n}$ for fingerprint template protection. Both constructions utilize a set of randomly generated projection matrices of dimension $q \times k$ for h_i to project the input features into q dimensional subspace, where the index of maximum value selected over $\{0, \ldots, q - 1\}$ is returned.

Formal speaking, Bio-hashing and LSH exploit the random projection process to preserve the original inputs' distance in the hashed domain.

Specifically, the random projection could be viewed as a multiplicative data perturbation such that $y = A \cdot x$ for a random matrix A and input x.

Theorem 1 [13]: Given two vectors $x, x' \in \mathbb{R}^k$. Let $A \in \mathbb{R}^{n\times k}$ be a $n \times k$ random matrix whose elements A_{ji} (where $j = 1, \cdots, n$ and $i = 1, \cdots, k$) are i.i.d. drawn from some distributions with $\mathbb{E}\left[A_{ji}\right] = \mu$ and $Var(A_{ji}) = 1$. Recall that random projection computes $y = \frac{1}{n}A \cdot x$ and $y' = \frac{1}{n}A \cdot x'$. Then for x and x' such that $\|x\| \leq 1$ and $\|x'\| \leq 1$, it follows that

$$(1 + \mu^2)\,\mathbb{E}\left[\left\|y - y'\right\|^2\right] - \left\|x - x'\right\|^2 \leq 2\mu^2 k. \qquad (1)$$

Eq (1) of the above theorem implies that regardless of which type of random matrix, the corresponding output distance, i.e., $\left\|y - y'\right\|^2$, would inevitably increase with the increment of the input distance $\left\|x - x'\right\|^2$, and vice versa. This demonstrates that the Bio-hashing and LSH distance exhibit distance-preserving property, and they can be generally named as distance-preserving hashing, categorized under the distance-preserving transformation (DPT).

B. Related Works in Privacy Preservation Using DPT

The studies of distance-preserving property for privacy-preserving data mining in a broad sense have been an area of research since 1991 [14]. The main goal is to protect user data privacy from a database via DPT. Some notable literature refer to the works by Kim and Winkler [15], Tendick [16], and Evans *et al.* [17]. However, the question of how well is x being hidden in y remains unclear, which deserves a careful study. Potential attackers without any prior knowledge can only do very little (if any) in recovering the original sample x. However, it is unrealistic for such zero prior knowledge to happen in many practical situations. Motivated by such reasoning, a lot of works have been done by considering the vulnerability of distance-preserving transformation. We briefly highlight some notable literature as follow. For a more general survey, we direct the interested reader to [18].

First, Liu *et al.* [19] reported that the attacker could exploit the distance-preserving property in reverting the original sample x. They realized the principal component analysis (PCA) could be a useful tool for a reasonable estimation of the original and transformed covariance matrices, which later leads to the recovery of the original data. Their work has inspired Turgay *et al.* [20] to recover the original data values with very high confidence for PCA based attack. A more robust type of attack extended from PCA based attack is proposed by Guo *et al.* [21] to show security breach in projection-based transformation (isometric). They applied traditional independent component analysis (ICA) over a set of known samples and perturbed samples. Information leakage allows the derivation of a transformation matrix that could lead to a close approximation of the original sample. Chen *et al.* [22] have also pointed out the security concern over distance preserving transformation such as geometric data perturbation, including random rotation perturbation, random translation perturbation, and noise addition. Subsequently, Wong *et al.* [23] have shown that the original input data is uniquely recoverable when one can solve the K-independent linear equation systems.

For the biometric line of research, recent works by Galbally *et al.* [24] reported the utilization of some heuristic algorithms to revert the original biometric template (human iris) through the exploitation of the preserved distances among hashed template. Specifically, they utilized a genetic algorithm, which aimed to minimize the 'fitness' function corresponding to the pairwise hashed template distance. The minimization process performs iteratively with several guesses defined by a population of synthetic iris data. Apart from using a genetic algorithm, another attack carried out by Feng *et al.* [25] incorporated multiple layers of perceptron learning to minimize the hashed distance over the set of synthetic real-value templates. Their results demonstrated the close reconstruction of human face images from the synthetic real-value template by utilizing the hill-climbing technique [26]. Recent work by Kaplan *et al.* [27] have shown that any distance-preserving transformation is also relation-preserving. Even under the scenario where only the relative order of the distance or similarity is preserved, the data breach is still inevitable due to similarity information leakage.

The schemes mentioned above have identified severe security threats to distance-preserving transformation, leading to doubts and curiosity about non-invertibility.

C. Optimal Distance-Preserving Hashing

Upon closer look, all the attacks mentioned in the previous sub-section rely on the information leakage due to DPT.

Follow the works in [13], to resist this kind of attacks, one must reduce such information leakage, which can be described as the mutual information, denotes as $I(D_I|D_o)$, where D_o is the distribution of the original interclass distance, and D_I is the distribution of the interclass distance d_I after hashing with $f(.)$. In our context, we refer the term *dissimilarity* when D_o and D_I are normalized to $[0, 1]$. Because optimizing the mutual information $I(D_I|D_o)$ is NP-complete, it is more practical for one to optimize the upper bound of $I(D_I|D_o)$. Specifically, let $H(W)$ denote the entropy of distribution W. It follows that $I(D_I|D_o) \leq H(D_I) \leq H(D_o)$. Then, by assuming that D_I asymptotically follows unimodal distribution, where small distribution's variance presumably leads to small entropy [28], the term $H(D_I)$ can be replaced with the variance of D_I, denoted as $\mathrm{Var}(D_I)$. Optimizing $I(D_I \mid D_o)$ can be done via minimizing $\mathrm{Var}(D_I)$. In other words, the inter-class distance over the hashed domain shall be made as small as possible with equidistance to reduce the mutual information leakage. Doing this is sufficient to show the resistance against the aforementioned attacks over DPT.

On the other hand, for good recognition utility, the system must ensure the samples sourced from similar subjects can match successfully with high probability. Given this, it is desirable that after hashing with $f(.)$, there is a large gap between the distribution of the interclass's distance (denoted as D_I) and the distribution of the intraclass's distance (denoted as D_g). Ideally, we wish that $\mathbb{E}[D_I] \to 1$ while $\mathbb{E}\big[D_g\big] \to 0$, where the similarity scores ($1 - D_I$ and $1 - D_g$) obtained between the hashed templates shall remain insensitive up to a threshold $\tau > 0$ with respect to their original similarity scores for both interclass and intraclass. The relation between the hashed similarity scores (after hashing with $f(.)$) and the original similarity scores is depicted in Fig. 1. Note that an S-curve characterizes such a non-linear relationship, where the hashed similarity score is shown to remain negligibly small given the original similarity score is less than the acceptance threshold, i.e., $\tau = 0.59$ as depicted in Fig. 1. Such a non-linear relationship is sufficient to ensure negligible similarity information leakage over the hashed templates while offering correctness (i.e., authenticity) for genuine users who can present another biometric template w and show a similarity score $s(.,.)$ of at least τ with the enrolled template w', i.e., $s(w, w') \geq \tau$.

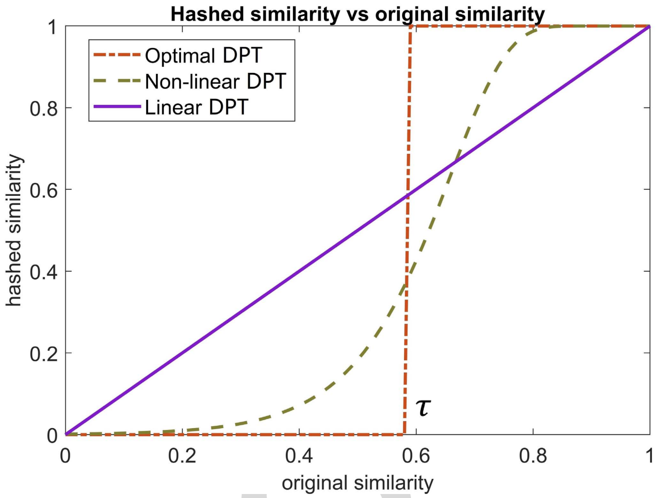


Fig. 1. Overview diagram illustrating the different relationships between the hashed similarity scores and original similarity scores.

III. Motivations and Contributions

Our motivations and contributions of this paper can be summarized as follow:

A. DPT-Based Attacks on BTP Schemes

Liu *et al.* [19] have generalized two attacks over DPT, namely, *known samples attack* (KSA), and *known input-output attack* (KIOA). The former attack assumes the potential attacker has acquired a small subset of the database samples, while the latter assumes that the attacker has the exact knowledge over several pairs of input and their corresponding output. The known input-output attack asserts a significantly stronger assumption upon the attackers' power in acquiring the input-output pairs. Our work emphasizes on KSA due to its popularity for DPT analysis [19]–[21], [27]. Here, we propose an attack for the BTP scheme relies on distance-preserving hashing, specifically for face biometric features. Comparatively, our attack only requires one known sample, which is indeed more realistic and practical than the 5% of the whole dataset requirement imposed by [19] and 4 ∼ 6 samples mentioned in [27]. Moreover, the attack carried out in [20] assumed that the attacker has a distance matrix of the private data, which directly implies that the attacker knows the original sample's distribution. In our case, the attacker is considered much 'weaker' in that he/she has no prior knowledge of the original sample's distribution. However, we allow the attacker to model and change the input distribution iteratively by introducing additional perturbation (i.e., adding noise). We shall show later that even under such 'weaker' attacking environment, our results of attacking the BTP scheme that adopted distance-preserving hashing is prominence, where the number of iterations required to produce a valid authentication result can always describe using some polynomial ($\mathsf{poly}(n)$) in the hashed template's length (n). These results lead to significantly less time required to launch a successful attack over a short template.

B. Realization of Non-Linear DPT

To resist against KSA, one should minimize the similarity information leakage between the original similarity scores and the hashed similarity scores [13]. To achieve this, an optimal, or at least non-linear, DPT is desired. As a metric for evaluation, the degree of minimization of leakage can be directly visualised from the gradient of the S-curve as depicted in Fig. 1. Clearly, steeper gradient over the S-curve indicates higher degree of minimization, hence better security and privacy protection over the DPT transformed data. To realize such non-linear DPT, we therefore reformulate the design of the distance-preserving hashing for BTP. Specifically, we construct a new hashing family from the conventional locality sensitive hashing family to realize such non-linear relationship over the input and hashed domains. Besides, the new hashing family inherits the good properties of the conventional LSH hashing family such as *efficiency* and *simplicity*, and resistance against KSA as a secure distance-preserving hashing BTP scheme. Most importantly, we also conducted revocability, unlinkability analysis; and show the resistance of various security and privacy attacks, i.e. false acceptance attack, attack via record multiplicity.

IV. OUR ATTACK

A. New KSA Attack Formalization

We begin with the intuition of the proposed KSA attack that incorporated the structure information of the input biometric.

Attack Intuition: Let $f : \mathbb{R}^k \to \{0,1\}^n$ be a convention distance-preserving hash function. Let $w \in \mathbb{R}^k$ be the enrolled biometric template. Given a targeted hashed template $f(w) \in \{0,1\}^n$ and a dissimilarity score $\varepsilon' \in [0,1]$, our goal is to find a sample $w^* \in \mathbb{R}^k$ s.t. $d(f(w), f(w^*)) \leq \varepsilon'$. Due to the distance-preserving property of $f(\cdot)$, if the dissimilarity score (after transformation) is $d(f(w), f(w^*)) \leq \varepsilon'$, then the original dissimilarity score (before transformation) $d(w, w^*) \leq \varepsilon$ should hold for arbitrary $\varepsilon > 0$. Using the naive brute-force search for w^* is a practically infeasible or at least inefficient approach due to the field size $|\mathbb{R}^k|$, which increases exponentially with k. However, if we are able to look for a noise distribution $\mathcal{D}$ and some noise samples $w_{e,i} \in \mathcal{D}$ (for $i \in \{1,2,\ldots,N\}$) s.t. $d(f(w), f(w_{e,i})) \leq \varepsilon'$, then the searching can be reduced to look for any $w_{e,i} \in \mathcal{D}$ where $d(f(w), f(w_{e,i})) \leq \varepsilon'$ and $d(w, w_{e,i}) \leq \varepsilon$ hold. This allows us to reduce our search space for all $w_{e,i} \in \mathcal{D}$ (rather than deal with $w^* \in \mathbb{R}$) over a smaller subspace parametrized by $|\mathcal{D}|$, which is relatively easier to be modelled compared to $|\mathbb{R}^k|$.

Algorithm 1 Proposed KSA

```
1:  function ATTACK_f($w^*$, $f(w)$, $N$, $\varepsilon'$, $\varepsilon$, $\lambda$, $\mathcal{S}$)
2:      $\chi \leftarrow_\$ \{0,1\}^k$
3:      $\sigma \leftarrow \mathcal{S}$          ▷ select $\sigma$ from $\mathcal{S}$ without repetition
4:      Set $M = \sigma \cdot U$          ▷ $U \in 1^k$ is vector of one
5:      for $i = 1 : N$ do
6:          $e_i \leftarrow_\$ \chi$     ▷ select $e_i \in \chi$ without repetition, where $\forall e_i \in \chi$, $\|e_i\| = \lfloor k\varepsilon \rfloor$
7:          $w_{e,i} = M \circ e_i + w^*$          ▷ where $\circ$ denotes the Hadamard product of $M$ and $e_i$
8:          Compute $d_i = d(f(w), f(w_{e,i}))$ and output $(d_1, d_2 \ldots, d_N)$     ▷ we refer $d(x,y) = \frac{1}{\pi}\arccos(x \cdot y)$.
9:      end for
10:     Set $\varepsilon_0 = \min(d_1, d_2 \ldots, d_N)$
11:     if $\varepsilon_0 > \lambda/\varepsilon'$ then
12:         Back to Step 3
13:     else
14:         Output $w_{e,i}$ corresponds to $\min(s_1, s_2 \ldots, s_N)$
15:     end if
16: end function
```

Algorithmically, to look for such $w_{e,i} \in \mathcal{D}$, we have to first initialize a random distribution $\mathcal{D}$ over $\mathbb{R}^k$. This can be achieved by knowing at least one sample $w^* \in \mathbb{R}^k$. More specific, we make use on the known sample's distribution to construct a smaller subset $\mathcal{S}$, which later is used to realize $\mathcal{D}$. The noisy sample can be generated by perturbing the input sample w^* using a randomly selected real values $\sigma \in \mathcal{S} \in \mathcal{D}$. Meanwhile, we also denote a distribution $\chi \in \{0,1\}^k$ s.t. for all random sampled $e \in \chi$, the weight $\|e\| = \lfloor k\varepsilon \rfloor$ is parameterized by the original dissimilarity score $\varepsilon > 0$ s.t. $d(w, w_{e,i}) \leq \varepsilon$. The sampled e will be used to determine the position of w^*, over $0, \ldots, k-1$, to be perturbed using the randomly selected σ to model $\mathcal{D}$ precisely. More detailed discussion on how we construct $\mathcal{S}$ and sample e are covered in the next sub-section.

Let $M \in \mathbb{R}^{k \times k}$ be a perturbation matrix. Given some reference hashed dissimilarity $d(f(w), f(w^*)) = \varepsilon'$, the goal of looking for $w_{e,i} \in \mathcal{D}$ can be achieved by minimizing the dissimilarity score ε' using N number of noise samples $w_{e,1}, w_{e,2}, \ldots, w_{e,N}$ until one yields a dissimilarity score $\varepsilon_0 \leq \lambda\varepsilon'$ with a ratio $\lambda > 0$. Clearly, $\lambda > 0$ means the minimized dissimilarity score ε_0 is desired to be lower than the reference score ε' for meaningful minimization result.

Our attack algorithm with input $N, \varepsilon', \varepsilon, \lambda, f(w)$ and w^* depicted in Algorithm 1. The output of Algorithm 1 is a noisy sample $w_{e,i}$ that corresponds to the minimized dissimilarity score ε_0.

B. Attack Complexity and Efficiency

Note that the runtime complexity of Algorithm 1 is bounded by $\mathcal{O}\big(|\mathcal{S}|Nk^2\big)$. To look for $|\mathcal{S}|$, we make use of the possibility of self-enrollment of a potential attacker. Hence at least one sample $w^* \in \mathbb{R}^k$ can be obtained in reality. More precisely, the distribution of w^* can be identified by looking at the minimum and maximum value of w^*, i.e., $w^* \in [\min(w^*),\ \max(w^*)]$. We can define $|\mathcal{S}|$ to be the size of the subset $\mathcal{S}$ where $\mathcal{S} \in [\min(w^*),\ \max(w^*)]$. Doing this will narrow our focus to a smaller subset $\mathcal{S}$, which is very much more manageable compared to $\mathbb{R}$. For any value $\sigma \in \mathcal{S}$ (chosen uniformly at random from $\mathcal{S}$), it should be noted that our attack is efficiently bounded over a subspace of size $|2\sigma|^k$. Therefore, we should have the desired distribution $\mathcal{D} \in [2(\min(w^*)),\ 2(\max(w^*))]^k$ for all $\sigma \in \mathcal{S}$, and $\mathcal{S} \in [\min(w^*),\ \max(w^*)]$ should follow.

Formally, the dissimilarity score minimization can be conceived as a process of searching for a similar point $w_{e,i} \in \mathcal{D}$ s.t. $d(w, w_{e,i}) \leq \varepsilon$ given $d(f(w), f(w_{e,i})) \leq \varepsilon'$ holds. For each iteration, a sample from $\mathcal{D}$ will be selected as $w_{e,i}$ to minimize $d(f(w), f(w_{e,i}))$. Trivially, for any input sample of size k over $\mathbb{R}$, there are at most $\mathbb{R}^k$ different samples over the input space. If a minimization solution exists, at most $\mathbb{R}^k$ random guesses are required. However, one needs to consider the exponentially large number of possibilities (of combinations) when k is increasing, e.g., long input length. Nonetheless, we will show that attack complexity can be relieved to $\mathcal{O}\left(nk^2\right)$ parametrized by an integer $m > 0$, the input length k, and the original dissimilarity score $\varepsilon > 0$ as shown in below.

For $\max(w^*) - \min(w^*) \leq 1$, using a parameter (integer) $m > 0$, we could construct a subset $\mathcal{S} \in [\min(w^*),\ \max(w^*)]$ of size

$$|\mathcal{S}| = \frac{\max(w^*) - \min(w^*)}{2^{-m}} \leq 2^m \tag{2}$$

For instance, given $m = 2$, $\max(w^*) = 2$ and $\min(w^*) = 1$, a subset $\mathcal{S}$ can be constructed as $\mathcal{S} = \{1, 1.25, 1.5, 2\}$ with $|S| \leq 2^2 = 4$.

Recall for any $\varepsilon > 0$, any random sampled e_i should have weight equal to $\|e_i\| = \lfloor k\varepsilon \rfloor$. Hence, Step 7 of Algorithm 1 is equivalent to perturbing exactly $\lfloor k\varepsilon \rfloor$ locations of w^* with M and e_i. For an input w^* of size k, it follows that by *Stirling's approximation*, we can always set (for $\varepsilon \in (1/k, 1/2)$):

$$N = 2^{\lfloor kh_2(\varepsilon) \rfloor} \leq \binom{k}{k\varepsilon}, \tag{3}$$

where $h_2(\varepsilon) = -\varepsilon \log(\varepsilon) - (1 - \varepsilon) \log(1 - \varepsilon)$ is the binary entropy function.

For any hashed template $f(w) \in \mathcal{D}_f$ in some random distribution $\mathcal{D}_f$ over $\in \{0, 1\}^n$, let the total number of points over $\mathcal{D}_f \in \{0, 1\}^n$ be $n = 2^{\lfloor kh_2(\varepsilon)+m \rfloor} - 1$. We therefore have the intermediate results as follow $2^{\lfloor kh_2(\varepsilon) \rfloor + m} = n + 1 \leq 2^{kh_2(\varepsilon)+m}$, which leads us to the inequality below

$$kh_2(\varepsilon) + m \geq \log(n + 1). \tag{4}$$

Follow Eq (4) above, to look for $f(w_{e,i})$ (viewed as a point over $\mathcal{D}_f \in \{0, 1\}^n$) and check whether $d(f(w), f(w_{e,i})) = \varepsilon_0 \leq \lambda\varepsilon'$ using Algorithm 1, the logarithm of the number of point can be found over $\mathcal{D}_f \in \{0, 1\}^n$ must be bounded at most $kh_2(\varepsilon) + m$. In other words, the overall attack complexity is asymptotically (for large m) described as $\mathcal{O}\left(nk^2\right) = \mathsf{poly}(n)$, which is polynomial time. Then, we have the following claim for our attack efficiency.

Claim 1: Given $\varepsilon \in (1/k, 1/2)$ and the subset $\mathcal{S}$ of size $|\mathcal{S}| = 2^m$ with an integer $m > 0$, for any targeted BTP transformation function $f : \mathbb{R}^k \to \{0, 1\}^n$ with output template over a random distribution $\mathcal{D}_f \in \{0, 1\}^n$ that consists of $2^{\lfloor kh_2(\varepsilon)+m \rfloor} - 1 = n$ number of points, the Algorithm 1 will halt in $\mathcal{O}\left(nk^2\right)$ with $N = 2^{\lfloor kh_2(\varepsilon) \rfloor}$.

C. Acquiring the Pre-Images

Here, we discuss the capability of our proposed attack in getting large number of similar points that are, contributed by any noisy sample $w_{e,i} \in \mathcal{D}$, close to the targeted sample w, i.e., $d(w, w_{e,i}) \leq \varepsilon$.

Given the information of $\mathcal{S}$ and N, the distribution $\mathcal{D}$ can be revealed and the number of points over $\mathcal{D}$ can be known precisely. More specific, note that the perturbation value σ is chosen uniformly at random from the subset $\mathcal{S} \in [\min(w^*), \max(w^*)]$, and the random string e_i is chosen uniformly at random follows distribution χ of weight $\lfloor k\varepsilon \rfloor$. Every iteration in running Algorithm 1 will output a random noisy sample $w_{e,i} \in \mathcal{D}$ (see Step 7 of Algorithm 1) corresponding to the selected values of $\sigma \in \mathcal{S}$ and e_i. Follows Eq (2) and Eq (3), the number of possible values for $w_{e,i}$ can be expressed as $n+1 = N|\mathcal{S}| \leq 2^{\lfloor kh_2(\varepsilon) \rfloor + m}$. Note that a point in $\mathcal{D}$ can be revealed as $w_{e,i} \in \mathcal{D}$. Given the distribution $\mathcal{D}$ with number of points not greater than $2^{\lfloor kh_2(\varepsilon) \rfloor + m} - 1$, at most n iterations would suffice to try all the noisy samples over $\mathcal{D}$ using Algorithm 1. In view of this, the proposed KSA attack implicitly constructed a known distributions $\mathcal{D}$ of at most $2^{\lfloor kh_2(\varepsilon) \rfloor + m} - 1$ number of points where each point, a.k.a. the noisy sample $w_{e,i}$, should distribute randomly and uniformly over $\mathcal{D}$. Therefore, it is appropriate to treat the matching in every single iteration to be independent and identically distributed. To be specific, we define

$$X = \sum_{i=1}^{n+1} X_i, \tag{5}$$

where X_i denotes the independent variable s.t. $X_i = 1$ if $d(f(w), f(w_{e,i})) \leq \varepsilon_0 \leq \lambda\varepsilon'$. Hence, X follows binomial distribution with $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$ for some probability $p_i \in [0, 1]$.

Conceivably, X can be interpreted as the number of successful minimization result over the known distribution $\mathcal{D}$ that yields $d(f(w), f(w_{e,i})) = \varepsilon_0 \leq \lambda\varepsilon'$. Without loss of generality, since $\lambda\varepsilon' \leq \varepsilon'$, hence, a successful minimization result would mean that a similar point $(w_{e,i})$ can be found over $\mathcal{D}$ where $d(f(w), f(w_{e,i})) \leq \varepsilon'$ holds, which implies $d(w, w_{e,i}) \leq \varepsilon$.

Arguing that the number of similar points can be found is different given different input samples w^*, the exact value of X must be different as well. Hence, it is reasonable to bound the number of similar points as a variable based on their original similarity score $1 - \varepsilon$. Then by *Chernoff bound* (for $\varepsilon \in (1/k, 1/2)$):

$$\Pr[X \leq n(1 - \varepsilon)] \leq \exp\left(-\tfrac{n\varepsilon^2}{2}\right). \tag{6}$$

Based on the Eq (6) above, we can conclude that given an arbitrary random $\varepsilon \in (1/k, 1/2)$, for sufficiently large n, the number of similar points s.t. $d(w, w_{e,i}) \le \varepsilon$ that can be found over $\mathcal{D}$ is unlikely to be smaller than $n(1-\varepsilon)$. From this point of view, after n number of iterations, the probability for Algorithm 1 in getting $n(1-\varepsilon)$ similar points is at least $1 - \exp(-n\varepsilon^2/2)$, which is close to one if n is set to be large enough. These obtained similar points are also known as the *pre-images* of w where both $d(w, w^*) \le \varepsilon$ and $d(w, w_{e,i}) \le \varepsilon$ should follow. The above results also indicate that any computational unbounded attacker in running Algorithm 1 must be able to obtain at least $n(1-\varepsilon)$ number of pre-images with high probability when n is large enough.

Given n is large, it follows that m must be set sufficiently large as well s.t. $n \le 2^{kh_2(\varepsilon)+m} - 1$ in order to support the efficiency argument of Algorithm 1 follows Claim 1, especially under the case when ε is small. In other words, a larger m is necessary to provide more information, hence, more points can be found over $\mathcal{D}$. Doing so is in our favor of looking for a similar point over larger distribution $\mathcal{D}$, which contains more points within $\mathcal{O}(nk^2)$ operations.

Indeed, Nagar *et al.* have demonstrated that acquiring the pre-images of the enrolled sample w is sufficient to compromise the BTP schemes, i.e., Bio-hashing. (see [29], Section 5). In their works, for any targeted sample $w \in \mathbb{R}^k$, they proposed to use $t > 0$ number of known biometric samples $w_i^* \in \mathbb{R}^k$ (for $i = 1, \ldots, t$), collected from a database s.t. $d_i = d(f(w), f(w_i^*)) \le \delta$, to estimate $x \in \mathbb{R}^k$ by minimizing the 2-norm distance follows argmin $\left\|x - w_i^*\right\|_2 \le \varepsilon$. The minimization is done by using the Matlab built-in isqlin function to obtain a series of estimated results $x_i, \ldots x_t$ equivalent to the similar points that are ε-close to w. Then, the pre-image of w, denoted as $\hat{x}$, is computed among $x_i, \ldots x_t$ follows $\hat{x} = \frac{\sum_{i=1}^{t} x_i/d_i}{\sum_{i=1}^{t} 1/d_i}$.

Our proposed KSA attack improved Nagar *et al.*'s approach in two perspectives. Firstly, the proposed KSA required only a single known sample of w^*, which can be trivially obtained through self-enrolment. Secondly, the proposed KSA has incorporated the input structure of the biometric distribution. Specifically, the number of pre-images obtained is described as a function of the points of distribution ($\mathcal{D}$). The incorporated structural information of the biometric distribution offers a better attack efficiency guarantee in looking for the similar points that are ε-close to the original biometric sample w.

V. COUNTERMEASURE FOR DPT-BASED ATTACKS

In this section, first, we present a few definitions and briefly walk through the randomized strategy, which is a crucial background study in the following subsection. Then, we put forward a countermeasure for DPT based attack, particularly to resist against KSA. We reformulate the conventional LSH hashing adopted by the BTP scheme as proposed in [10]–[12] to construct a non-linear DPT for our goal.

A. Definitions

Definition 1 (Locality Sensitive Hashing): Let $d_1 < d_2$ be two distances of some distance measure $d(\cdot,\cdot)$. A family $\mathcal{H}$ of functions is said to be (d_1, d_2, p_1, p_2)-sensitive if $\forall h \in \mathcal{H}$ then the following hold true:

$$\Pr[h(x) = h(y)] \ge p_1, \ \text{if } d(x, y) \ge p_1, \text{and}$$

$$\Pr[h(x) = h(y)] \le p_2, \ \text{if } d(x, y) \le p_2.$$

Given an (d_1, d_2, p_1, p_2)-sensitive family $\mathcal{H}$, one can construct another family $\mathcal{H}'$ where each member of $\mathcal{H}'$ consists of exactly k members from $\mathcal{H}$. We called such new family to be (d_1, d_2, p_1^k, p_2^k)-sensitive, which is defined below:

Definition 2: Given an (d_1, d_2, p_1, p_2)-sensitive family $\mathcal{H}$, we say another family $\mathcal{H}'$ is (d_1, d_2, p_1^k, p_2^k)-sensitive if it consists of members of a set $\{h_1, \ldots, h_k\}$ from $\mathcal{H}$, where $h(x) = h(y)$ (over $\mathcal{H}$) if and only if $h_i(x) = h_i(y)$ for $i = 1, \ldots, k$ (over $\mathcal{H}'$).

B. Randomized Strategy for LSH Family

One typical way to construct an LSH family of (d_1, d_2, p_1, p_2)-sensitive is by random projection.

Random projection has been used by Gormans *et al.* [30] in solving the relaxed version of maximum cut problem. In particular, given a graph $G(V, E)$ and nonnegative weight $z_{ij} = z_{ji}$ on the edges $(i, j) \in E$, the max-cut problem is a computational problem that aims to find the set of vertices $S \subset V$ follows a cut $(S, \bar{S})$ where the weight of the edges with one endpoint in S and the other in $\bar{S}$ (the complement of S) is maximized [31]. The relaxed version of max-cut problem is to maximize the objective function described as $\frac{1}{2}\sum_{i<j} z_{ij}(1 - w_i \cdot w_j)$ where w_i and w_j are two vectors over $\mathbb{R}^k$. Gormans *et al.* used a random vector r (uniformly distributed on a unit sphere) to partition the set of vertices S and its complement $\bar{S}$ into those vectors $w \in \mathbb{R}^k$ that lie above the hyperplane (i.e., the inner product $r \cdot w$ is positive) and below the hyperplane (i.e., the inner product $r \cdot w$ is negative) while maximizing the objective function $\frac{1}{2}\sum_{i<j} z_{ij}(1 - w_i \cdot w_j)$.

The Lemma below characterizes the above randomized strategy that renders a locality sensitive hashing family which is (d_1, d_2, p_1, p_2)-sensitive with the distance measure referring to the cosine distance (i.e., angle between w and w') described as $d(w, w') = \frac{1}{\pi}\arccos(w \cdot w')$.

Lemma 1 [30]:

$$\Pr\left[\mathsf{sgn}(r_i \cdot w) \ne \mathsf{sgn}(r_i \cdot w')\right] = \frac{1}{\pi}\arccos(w \cdot w').$$

In our case, we apply random projection to project the input biometric template (a vector) $w \in \mathbb{R}^k$ using multiple random Gaussian vectors with mean zero and variance one, and a signum function $\mathsf{sgn}(r \cdot w) \in \{0, 1\}$, yielding an output vector $v \in \{0, 1\}^n$ described as follow $v = [\mathsf{sgn}(r_1 \cdot w), \ldots, \mathsf{sgn}(r_n \cdot w)]$, where $\mathsf{sgn}(r_i \cdot w) = 0$ if $r_i \cdot w \ge 0$ and $\mathsf{sgn}(r_i \cdot w) = 1$ if $r_i \cdot w < 0$. The output vector v is a core element to be used in our proposed countermeasure for DPT based attack, and to construct a new LSH family, which are discussed in details in the next sub-section.

C. Formalization of the Proposed Technique

The formalization of our proposed countermeasure for DPT based attack adopts the LSH family constructed via

Algorithm 2 Proposed Transformation

```
function TRANS$_{f \in \mathcal{H}_r}(w, r, s, u, b)$
    $n = s \times u \times b$
    for $i = 1 : n$ do
        $v_i = f_i(w, r_i)$
    end for
    Set $v = (v_1, \dots, v_n)$
    Reshape $v \to v \in \{0, 1\}^{s \times ub}$
    Convert every $b$ bits into a unit of integer in range $\{0, \dots, 2^b - 1\}$
    Output $v \in \{0, \dots, 2^b - 1\}^{s \times u}$
end function
```

Algorithm 3 Proposed Authentication

```
function AUTH$(v, w', r, s, u, b, \tau)$
    $v' \leftarrow$ TRANS$_{f \in \mathcal{H}_r}(w', r, s, u, b)$
    Initialize score $X = 0$;
    for $i = 1, \dots, s$ do
        if Each row of $v'$ and $v$ collided in at least $\tau$ positions of units then
            Set $X = X + 1$
        end if
    end for
    Output $X/s$
end function
```

randomized strategy. We follow Definition 1 and 2 to construct a new LSH family derived from the randomized strategy, which offers non-linearity for our security goal.

Notation: Suppose we are given an input sample $w \in \mathbb{R}^k$ (for enrolment). Let $f \in \mathcal{H}_r$ denote the hashing function over the LSH family of randomized strategy $\mathcal{H}_r$, where $f : \mathbb{R}^k \to \{0, 1\}^n$. In particular, we have $f_i(r_i, w) = \mathsf{sgn}(r_i \cdot w)$ for $i = 1, \dots, n$ with random Gaussian vector $r_i \in \mathcal{N}(0, 1)$ and signum function $\mathsf{sgn}(.)$. We set $n = s \times b \times u$, and use s, b and u to denote stripe, bit and unit, respectively.

Main Idea: Our core idea is to reformulate the LSH function to generate a fixed number of points that can be directly expressed using the number of stripes over the hashed domain.

For high recognition utility, our formulation must ensure that similar points, that are ε-close together, i.e., $d(w, w') \leq \varepsilon$, can be found with overwhelming probability given their hashed similarity is large (i.e., the matching score is high, close to one, after hashing). On the other hand, it should exhibit negligible probability to look for the similar points when the hashed similarity is small (i.e., the matching score is negligible small, close to zero, after hashing). To achieve this, we define a radius of τ for each stripe (point) over the hashed domain. Such radius could be quantified by the number of units in a single stripe, which consists of b number of bits. With an adequately selected τ, we can tolerate the errors in the similar input samples to ensure authenticity with overwhelming probability. It follows that a highly non-linear relationship between the original similarity scores and output hashed similarity scores can be obtained, hence establishing resistance against KSA while keeping high recognition utility.

Overview Procedure (Transformation): Our procedure to generate the hashed template is quite simple and can be summarized as follow. First, the input template $w \in \mathbb{R}^k$ is being hashed by $f_1, \dots, f_n$ with $r_1, \dots, r_n$ to output a binary vector v of size n. Next, v will be reshaped into a 2-D matrix of size $s \times ub$. We called the individual row of the resulting matrix - a stripe. Precisely, a stripe consists of u number of units, and every unit is represented by b binary symbols (bit). Each unit can be conveniently viewed as an integer over the set of $\{0, \dots, 2^b - 1\}$. Let $r = (r_1, \dots, r_n)$ be the collection of all random Gaussian vectors. The transformation takes (w, r, s, u, b) as input, and its pseudocode is presented as Algorithm 2.

Overview Procedure (Authentication): Given another input template $w' \in \mathbb{R}^k$, using the same published parameters (r, s, u, b), the same transformation (Algorithm 2) is utilized to generate its corresponding hashed vector $v' \in \{0, 1\}^{s \times ub}$. Authentication can then be viewed as a score counting process as follow: For each stripe ($i = 1, \dots, s$) in v and v', a score count X_i is recorded if there is at least τ number of colliding units. The total score count is simply $X = \sum_{i=1}^{s} X_i$. Then, X is normalized and outputted as the similarity score, i.e., $X/s \in [0, 1]$. The authentication mechanism, which takes (w', r, s, u, b, τ) as the input, is presented as Algorithm 3.

D. Non-Linearlity Derivation

Here, we derive the non-linearity property of our proposed algorithm pair (TRANS, AUTH).

Let $d(w, w') = \frac{\arccos(w \cdot w')}{\pi}$ be the *dissimilarity* between w and w', which corresponds to their distance measured by the angle between them. Therefore, $p = 1 - d(w, w')$ refers to the *similarity* measure. By Lemma 1, we have the colliding probability of single bit over a single stripe to be:

$$\Pr\big[f_i(r_i \cdot w) = f_i(r_i \cdot w')\big] = 1 - d(w, w') = p.$$

Recall that each unit consists of exactly b number of bits. We shall see that for each single unit, it should come from a (d_1, d_2, p_1^b, p_2^b)-sensitive family $\mathcal{H}'$ (see Definition 2). It follows that the colliding probability for one single unit is equivalent to colliding exactly b number of bits. This can be expressed as:

$$\Pr\big[f_i(r_i \cdot w) = f_i(r_i \cdot w') \;\big|\; i = 1, \dots, b\big] = p^b.$$

The probability of no unit colliding is $1 - p^b$. Let z be number of colliding units. Clearly, z follows a binomial distribution and we denote p_c the probability of *at least* τ number of units colliding. Therefore,

$$p_c(u, b, \tau, p) = \Pr[z \geq \tau] = \sum_{i=\tau}^{u} \binom{u}{i} (p^b)^i (1 - p^b)^i. \quad (7)$$

By Eq (7) and Definition 2, each stripe is considered as an $(d_1, d_2, p_{c1}, p_{c2})$-sensitive LSH family $\mathcal{H}''$ constructed from an (d_1, d_2, p_1^b, p_2^b)-sensitive LSH family $\mathcal{H}'$.

The derived p_c has direct effect on the final computed authentication score $X = \sum_{i=1}^{s} X_i$, where $X_i = 1$ if the $i-$th stripe has at least τ number of colliding units. Given all stripes are independent, then X should follows i.i.d with $\Pr[X_i = 1] = p_c$ and $\Pr[X_i = 0] = 1 - p_c$. Therefore we shall have the expected score count expressed as $\mathbb{E}[X] = sp_c$ and variance $\mathrm{Var}(X) = s(p_c)(1 - p_c)$. Follows Eq (5), one shall notice that our proposed transformation offers well-defined number of stripe s which can be interpreted as the number of points over the hashed domain $\{0, 1\}^n$.

Note that the score count X is highly non-linear with respect to the original dissimilarity $d(w, w')$ measurement (see the functionality of p_c in Eq (7)). Fig. 2 depicts the non-linear relationship between the derived p_c and the input dissimilarity score $d(w, w')$. Observe that a larger number of bits b and τ would lead to a greater degree of non-linearity, where the gradient of the S-curve becomes steeper. Besides, a larger number of u promotes more colliding units. Therefore, the input templates with small dissimilarity $d(w, w')$ can easily attain overwhelming value for p_c (i.e., close to one). The argument above gives rise to our *correctness* claim for the genuine user with a higher value of τ.

VI. EXPERIMENTS AND EVALUATION

Experiments Set-up and Protocol: For input biometric templates, we adopt a pre-trained convolution neural network dedicated to face recognition, namely InsightFace [32]. InsightFace employs a loss function named additive angular margin loss for learning. With InsightFace that is pre-trained with MS-Celeb-1M, a face vector with a size of $k = 256$ can be obtained. Besides, we adopt the Labelled face in the wild (LFW) dataset [33], which consists of 7,701 images of 4,281 subjects. We follow the protocol outlined in [33], where a total number of 6,000 face pairs are divided into ten disjoint subsets for cross-validation. Each subset contains 3000 genuine pairs and 3000 impostor pairs, resulting in a total number of 3000 genuine matching scores and 3000 imposter matching scores. All the while, we only consider single set of random Gaussian vector $(r_1, \ldots, r_n)$ for random projection used in (TRANS, AUTH). Equal error rate (EER) is considered as the performance metric, which is the error rate when the false acceptance rate (FAR) and false rejection rate (FRR) are equal.

For attacks using Algorithm 1, for each imposter matching, we can conveniently set the distance between the hashed templates as $\varepsilon_0 = 1$. If such distance is at most $\lambda\varepsilon'$, Algorithm 1 will halt and stop in Step 4. Otherwise, Algorithm 1 will continue to minimize ε_0. The minimization process intending to achieve $\varepsilon_0 \le \lambda\varepsilon'$ for all imposter matching, yielding a total number of 3000 minimized dissimilarity score ε_0, namely the KSA attack scores, for performance evaluation of the proposed KSA attack. All experiments are conducted by using PC with processor core i5-2.50 GHz with 8GB RAM, graphic card GTX 1050 Ti, and with MATLAB Ver. R2018a.

A. Evaluation of Proposed Attack on Bio-Hashing and LSH

We evaluate the proposed known sample attack in this section. Our attack focuses on Bio-hashing and conventional LSH. In particular, for LSH, we refer to the randomized strategy in generating the hashed vector v. Note that the randomized strategy can be viewed as a special case of the recently proposed hashing scheme [10] when the output is in binary, i.e., $q = 2$.

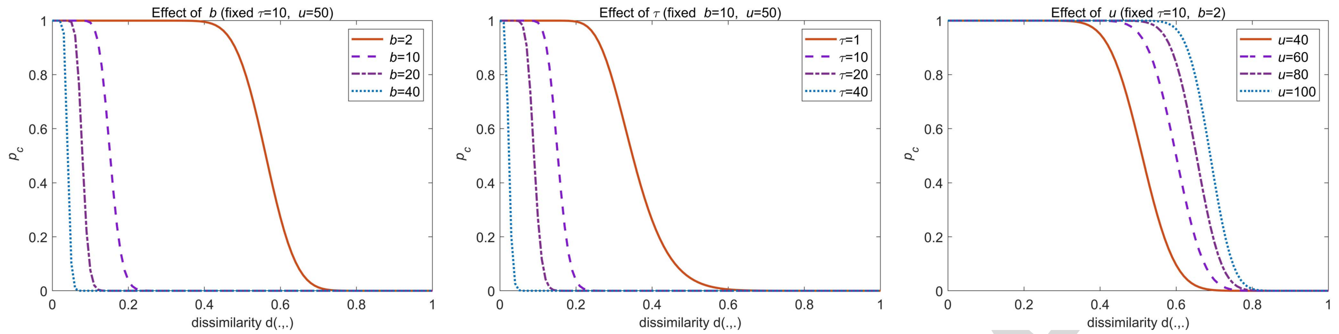
Parameters Control: Among all the necessary inputs $(w^*, f(w), N, \varepsilon', \varepsilon, \lambda)$, there are only four parameters, namely, N, ε', λ, and m, to be adjusted. Here, we set $\varepsilon = 10/256$, $\lambda = 1/4$, and limiting $N = 800$. The value of $m = 7$ is chosen by computing $|\mathcal{S}| = |0.2178 - (-0.1978)|/2^{-8} = 106.4 \le 2^7$ to get a set of values for $\mathcal{S}$ over the range of $[-0.1978, 0.2178]$. Follow Eq (4), $|\mathcal{S}| \le 2^7 < 2^8$. Considering the attack efficiency (see Claim 1, Eq (4)), the logarithm of the number of point can be found over the hashed domain's distribution $\mathcal{D}_f \in \{0, 1\}^n$ must be bounded at most $kh_2(\varepsilon) + m$. However, without proper designation of the transformation function, one could not assure the number of points over $\mathcal{D}_f \in \{0, 1\}^n$ will be at most $2^{\lfloor kh_2(\varepsilon)+m \rfloor} - 1$. This means if Eq (4) does not hold, then the derived KSA attack efficiency is obsoleted and no guarantee on $n(1 - \varepsilon)$ (follows Eq (6)) number of similar points can be found over $\mathcal{D}$ by using Algorithm 1. Nevertheless, a straightforward way to ensure efficiency of Algorithm 1 is to reduce the hashed output length n. In light of this, our proposed attack is highly efficient for the conventional Bio-hashing and LSH with security relying on dimensional reduction, i.e., $n < k$. Thus, our evaluation only focuses on small n.

Fig. 3 and 4 depict the results for Bio-hashing and LSH for output length of $n = 40, 60, 80$, and 100. The average time taken for obtaining a single dissimilarity attack score is observed to be 1.942851 secs and 2.04068 for Bio-hashing and LSH, respectively. Our results show that for fixed parameter m and ε', smaller output length would lead to better attacking result in the sense that the mean of the KSA attack scores follows closer to the mean of the genuine score distribution.

B. Performance Evaluation of Proposed Transformation and Authentication

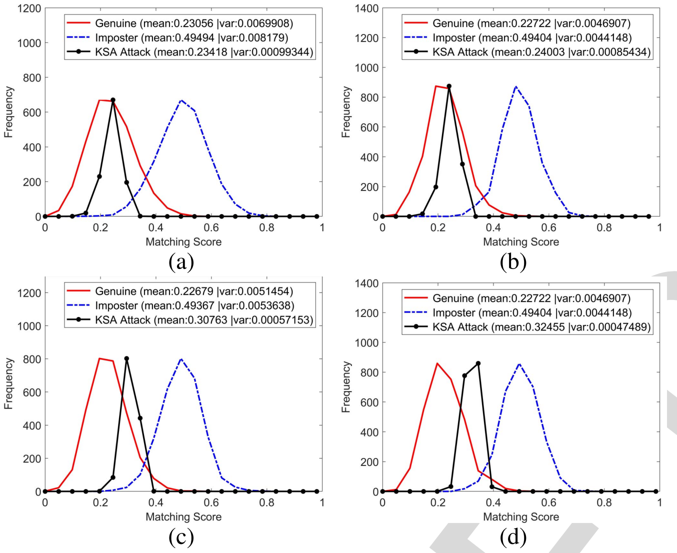
This section presents the performance evaluation of (TRANS, AUTH). With reference to (TRANS, AUTH), there are four parameters to be considered, namely s, u, b and τ. Noting the proposed authentication algorithm AUTH records the number of stripes that have at least τ number of colliding units. The authentication procedure can be perceived as a similar point searching process, where the adversary is required to look for an arbitrary number of similar points, close to w s.t. $d(w, w_{e,i}) \le \varepsilon$, where $\varepsilon > 0$ corresponding the minimum number of colliding units between different stripes (τ). It should be noted that the generated stripes are independent of each other. Thus a larger value of s will offer higher confidence to the final score count X, where $X \to sp_c$ should converge by *law of large number*.

Parameters Control: Recall that, $p_c(u, b, \tau, p)$ is parametrized by u, b, and τ where $p = 1 - d(w, w') = 1 - \frac{\arccos(w \cdot w')}{\pi}$ (follows Lemma 1). Let $\varepsilon = \frac{\arccos(w \cdot w')}{\pi}$. It is convenient to define the original dissimilarity follows $d(w, w') = \varepsilon$, which means $p_c(u, b, \tau, 1 - \varepsilon)$ is now a function of ε. The value of ε



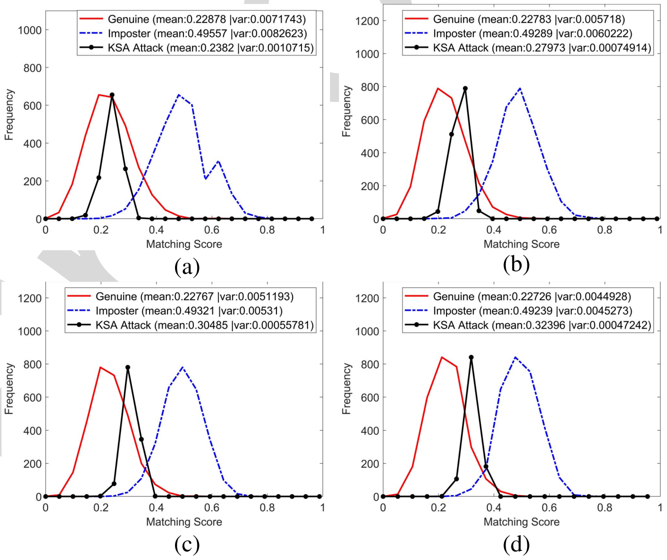
(a) (b) (c)

Fig. 2. Non-linearity relation parametrized by (a) $b = 5, 10, 15, 20$ (fixed $\tau = 5, u = 50$), and (b) $\tau = 5, 10, 15, 20$ (fixed $b = 10, u = 50$), and (c) $u = 40, 60, 80, 100$ (fixed $\tau = 10, b = 2$).



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

Fig. 3. Proposed KSA on Bio-hashing (a) $n = 40$, (b) $n = 60$, (c) $n = 80$, (d) $n = 100$.



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

Fig. 4. Proposed KSA on LSH (randomized strategy) (a) $n = 40$, (b) $n = 60$, (c) $n = 80$, (d) $n = 100$.

follows various distributions according to the input biometric template or input types, which is hard to predict. Therefore, choosing a set of values u, b, τ with optimal authentication performance and security for arbitrary value of $\varepsilon > 0$ will be our main interest. Based on Fig. 2, we know that the increment of b and τ would yield the same non-linearity effect by shifting the S-curve to the left with steeper gradient. On the other hand, the increment of u would shift the S-curve to the right with a steeper gradient. Therefore, we can choose b to be a constant to adjust the shifting of the S-curve to the left or right by increasing τ or u, respectively. Doing so would allow us to examine the non-linearity effect over the authentication performance and select the optimal parameter set corresponding to the original input distribution. All the while, we set $s = 50$ as the constant with different combinations for u, b, and τ. We set $u = 40, 60, 80, 100$, $\tau = 10, 12, 14, \ldots, 30$, and repeat each setting with $b = 1, 2, 3$, and 4.

The authentication performance (in term of EER) for various settings of u, b and τ is recorded in Table I. The original performance of the input sample (without transformation) is recorded to be 0.73% of EER. The best authentication performance we could obtain after applying our proposal is 0.75% of EER. Clearly, this authentication performance is closely preserved by referring to its original one. Given $s = 50, b = 2$ and $u = 50$, the output score distributions for genuine and imposter authentication with different value of τ is shown in Fig. 5. In general, given (u, s, b), a right choice of value τ could lead to large separation between the genuine and imposter score distributions. This scenario is mainly due to the non-linearity effect derived in Section V-D.

C. Security Evaluation of Proposed Transformation and Authentication

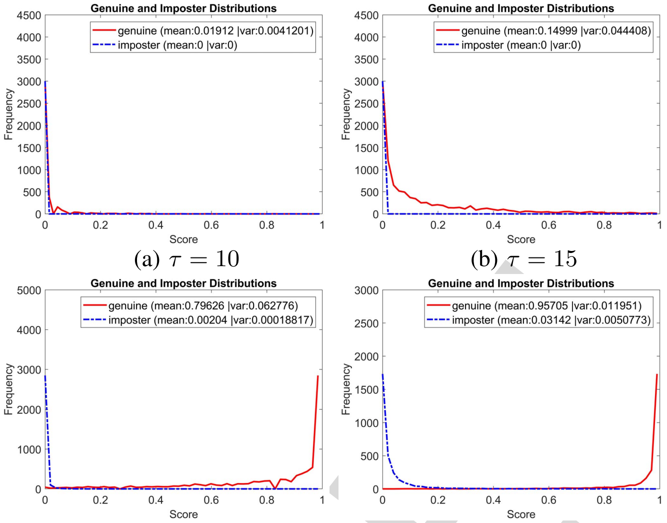
Here, we show how our proposal can resist against KSA. We adopt the newly proposed KSA (Algorithm 1) for our security evaluation of (TRANS, AUTH).

Recall that the matching (similarity) score outputted by AUTH (Algorithm 3) can be interpreted as the number of similar points or the amount of pair of stripes that have at least τ unit colliding. The applied KSA would have to *maximize* such a similarity score to compromise the system. In this sense, we have to reverse the stopping criteria of Algorithm 1

TABLE I

EER (%) RECORDED FOR DIFFERENT VALUE OF u, b, AND τ

***b*=1**	$\tau = 10$	$\tau = 12$	$\tau = 14$	$\tau = 16$	$\tau = 18$	$\tau = 20$	$\tau = 22$	$\tau = 24$	$\tau = 26$	$\tau = 28$	$\tau = 30$
$u = 40$	1.45	1.18	1.03	0.93	**0.92**	1.08	1.82	3.93	14.53	31.85	44.60
$u = 60$	8.55	3.47	1.55	1.32	1.03	0.88	0.88	**0.83**	0.87	0.85	1.08
$u = 80$	28.03	17.42	9.88	5.07	2.47	1.42	1.08	0.93	**0.80**	0.88	0.87
$u = 100$	40.92	33.78	25.08	17.22	10.73	6.20	3.35	1.82	1.22	1.02	**0.98**
***b*=2**	$\tau = 10$	$\tau = 12$	$\tau = 14$	$\tau = 16$	$\tau = 18$	$\tau = 20$	$\tau = 22$	$\tau = 24$	$\tau = 26$	$\tau = 28$	$\tau = 30$
$u = 40$	16.87	8.33	3.38	1.47	1.15	0.93	**0.75**	0.88	0.95	0.97	1.45
$u = 60$	43.42	37.38	29.77	21.32	13.75	8.23	4.57	2.28	1.22	1.00	**0.88**
$u = 80$	-	47.48	45.13	41.27	36.35	30.72	24.57	18.22	13.18	9.07	**5.67**
$u = 100$	-	-	48.83	47.82	46.23	43.72	40.03	36.27	31.70	26.53	**21.80**
***b*=3**	$\tau = 10$	$\tau = 12$	$\tau = 14$	$\tau = 16$	$\tau = 18$	$\tau = 20$	$\tau = 22$	$\tau = 24$	$\tau = 26$	$\tau = 28$	$\tau = 30$
$u = 40$	41.67	33.85	23.97	14.68	7.93	3.40	1.63	1.22	0.98	0.87	**0.83**
$u = 60$	49.47	48.50	46.80	44.10	39.97	34.47	27.93	21.67	15.40	10.02	**6.08**
$u = 80$	-	-	49.58	49.15	48.37	47.12	45.37	42.52	39.30	34.88	**30.42**
$u = 100$	-	-	-	49.87	49.67	49.50	49.02	48.45	47.45	45.95	**44.23**
***b*=4**	$\tau = 10$	$\tau = 12$	$\tau = 14$	$\tau = 16$	$\tau = 18$	$\tau = 20$	$\tau = 22$	$\tau = 24$	$\tau = 26$	$\tau = 28$	$\tau = 30$
$u = 40$	-	-	-	-	25.87	17.37	9.87	4.78	1.93	1.30	**0.88**
$u = 60$	-	-	-	-	47.78	46.08	43.58	39.30	34.77	29.12	**23.18**
$u = 80$	-	-	-	-	49.78	49.65	49.27	48.52	47.68	46.63	**44.47**
$u = 100$	-	-	-	-	-	-	-	-	-	-	-



(a) $\tau = 10$ (b) $\tau = 15$

(c) $\tau = 25$ (d) $\tau = 30$

Fig. 5. Genuine and Imposter score distributions of proposed scheme with fixed $s = 50, b = 2$.

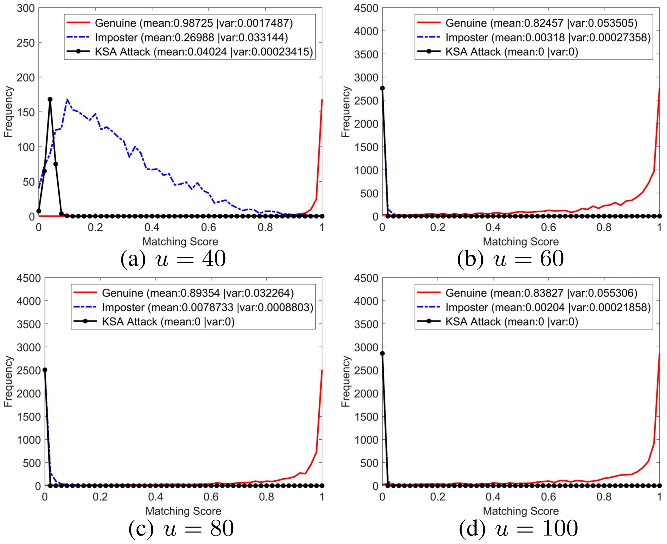
Algorithm 4 KSA for Proposed Scheme

```
1: function ATTACK_{TRANS,AUTH}($w^*, v, N, \varepsilon', \varepsilon, \lambda, s, u, b, \tau, \mathcal{S}$)
2:     $\chi \leftarrow_{\$} \{0, 1\}^k$
3:     $\sigma \leftarrow \mathcal{S}$            ▷ select $\sigma$ from $\mathcal{S}$ without repetition
4:     Set $M = \sigma \cdot U$              ▷ $U \in 1^k$ is vector of one
5:     for $i = 1 : N$ do
6:         $e_i \leftarrow \chi$      ▷ select $e_i \in \chi$ without repetition, where $\forall e_i \in \chi$, $\|e_i\| = \lfloor k\varepsilon \rfloor$
7:         $w_{e,i} = M \circ e_i + w^*$      ▷ where $\circ$ denotes the Hadamard product of $M$ and $V_i$
8:         Compute $s_i$= AUTH($v, w_{e,i}, r, s, u, b, \tau$) and output $(s_1, s_2 \ldots, s_N)$
9:     end for
10:    Set $\varepsilon_0 = \max(s_1, s_2 \ldots, s_N)$
11:    if $\varepsilon_0 < \lambda/\varepsilon'$ then
12:        Back to Step 3
13:    else
14:        Output $w_{e,i}$ corresponds to $\max(s_1, s_2 \ldots, s_N)$
15:    end if
16: end function
```

(i.e., line 11) and change it to $\varepsilon_0 < \varepsilon'/\lambda$. Doing that ensures the Algorithm 1 will output meaningful maximization result for all reference similarity scores ε_i (for $i = 1, \ldots, N$) obtained by using AUTH.

Let $v = \text{TRANS}_{f \in \mathcal{H}_r}(w^*, r, s, u, b)$ be the transformed known sample. Our proposed KSA (Algorithm 1) can be adopted in an reverse manner with (TRANS, AUTH) described in Algorithm 4. Remark that incorporating (TRANS, AUTH) into Algorithm 4 explicitly allows the attacker to have complete knowledge over the designed system as follow the *Kerckhoffs's principle*.

Parameters Control: For evaluation, we use the same KSA setup for Bio-hashing and LSH with $N = 800$, $\varepsilon = 10/256$, $\lambda = 1/4$ and $m = 7$. The parameters considered for (TRANS, AUTH) are $s = 50$, and $b = 1$. The above setup is used for different output stripe size of $u = 40$, 60, 80, and 100, and $\tau = 18, 24, 26$, and 30 with respect to their best authentication performance for $b = 1$ (see Table I). Fig. 6 depicted the result of KSA for (TRANS, AUTH) as described in Algorithm 4. As expected, the non-linearity property of (TRANS, AUTH) offers a strict constraint in looking for a similar point over the hashed domain. This can be explained with the function of $p_c(u, b, \tau, 1 - \varepsilon)$ (see Eq (7)) where only input with small cosine dissimilarity $\varepsilon = \frac{\arccos(w \cdot w')}{\pi}$ can show at least τ colliding units in the hashed domain with overwhelming probability. Therefore, the observed KSA attack scores follow the imposter score's distribution with small variance.



(a) $u = 40$ (b) $u = 60$ (c) $u = 80$ (d) $u = 100$

Fig. 6. Proposed KSA on algorithm pair (TRANS, AUTH).

D. Potential Security Attacks

In this subsection, we review the potential security attacks on BTP and demonstrate how our scheme resists against these attacks.

False Acceptance Attack: One of the major issues for BTP schemes relies on the false acceptance attack [34], [35]: a biometric system with high false acceptance rate is deemed as a low performing and insecure system and the transformed template stored in this system thus cannot be considered secure. False acceptance attack has been rigorously investigated in the recent works [10], [36], [37] to design a secure BTP scheme.

To be specific, let m_0 be the (minimum) entropy of a biometric source in a random distribution W. For two random k bits samples $(w, w^*) \in W$ derived from the same subject (genuine case), where w is the enrolled sample, and w^* is the query sample. To show meaningful security, the matching mechanism must only accept w^* given the hamming distance $\|w \oplus w^*\| \leq m_0$, which means that the original dissimilarity score $d(w, w^*) \leq \varepsilon \leq m_0/k$ must hold for any $\varepsilon \leq m_0/k$, where m_0/k denotes the (minimum) entropy rate of W. Otherwise, the system would accept another query sample $w' \in W'$ over a random distribution W' with dissimilarity score $d(w, w') = \varepsilon > m_0/k$, which leads to a false acceptance. It should be noted that a false acceptance would imply that the biometric source has lost all its entropy and shows no security, i.e., $m_0/k - \varepsilon < 0$.

A typical way to evaluate the false acceptance attack security is by measuring FAR. However, such measurement is crude in the sense that it does not consider the input distribution of the biometric source and cannot show meaningful security to any source that consists of a large number of errors. To further explain this, note that the FAR is depend upon the dissimilarity score of ε. Any source with a large number of errors will introduce a high dissimilarity score, which means the matcher that accepts w^* s.t. $d(w, w^*) \leq \varepsilon$ must set large ε to reduce FRR which subsequently increases FAR. Biometric traits typically demonstrate "more error than entropy", for instance, the human iris [38]. The human iris is believed to offer high entropy, i.e., $m_0 = 249$ bits. However the k bits binary template, namely Iriscode, generated from human iris usually contains error ($k\varepsilon$) that is more than 249 bits, i.e., $k\varepsilon > m_0$.[1] Given above discussion, it is inevitable that a false acceptance is expected given any two Iriscodes (w, w') derived from different subjects with dissimilarity score $d(w, w') = \varepsilon > m_0/k$. Moreover, it is imprudent to believe that the distribution of the biometric source W can be modeled precisely, especially for high entropy source. The attacker might have higher computation power to model W and lead to lower attack complexity, i.e., a lower value of m_0. Nevertheless, we can conveniently bound the entropy rate of distribution W follows $m_0/k \geq \varepsilon$ for all $w' \in W'$ that comes with a maximum dissimilarity score equal to ε. Since $m_0 \geq k\varepsilon$ is necessary to prevent a false acceptance given any sample $w' \in W'$, it follows that the false acceptance security can be claimed given the system knows the value of ε.

Based on the above reasoning, to show meaningful false acceptance security for larger class of biometric sources (including more error than entropy sources), it is desirable to design a BTP transformation as a function of the input distribution where the knowledge on the original dissimilarity score ε is perceived as a necessity. In fact, it is easy to verify that the proposed transformation and authentication algorithm pair (TRANS, AUTH) enjoys such property with the denoted p_c known as the probability of *at least* τ number of units colliding expressed as $p_c(u, b, \tau, p)$ where $p = 1 - d(w, w')$, and $d(w, w') = \varepsilon = \frac{\arccos(w \cdot w')}{\pi}$ corresponds to the original dissimilarity score (cosine dissimilarity) of the input samples (w, w'). Moreover, because the generated stripes (after transforming using TRANS) are independence to each other The output score X/s should asymptotically converge to p_c by law of large number (for value of $s \gg 1$). In other words, the relation in between p_c and $d(w, w')$ shown in Fig. 2 is asymptotically good for false acceptance security evaluation of (TRANS, AUTH).

Generally, by using *Bayes's theorem*, the relationship of the probability $\Pr[z \geq \tau] = p_c$ given the input dissimilarity score $d(w, w') \leq \varepsilon$ can be described as:

$$\Pr\Big[z \geq \tau \;\Big|\, d(w, w') \leq \varepsilon\Big] = \frac{\Pr[z \geq \tau]\Pr\Big[d(w, w') \leq \varepsilon \;\Big|\, \Pr[z \geq \tau]\Big]}{\Pr[d(w, w') \leq \varepsilon]}.$$

The term $\Pr\Big[d(w, w') \leq \varepsilon \;\Big|\, \Pr[z \geq \tau]\Big]$ is the acceptance rate, i.e., a person is identified as a valid user. In reality, the person in performing the authentication should be random (either genuine user imposter), therefore we shall let $\Pr\Big[d(w, w') \leq \varepsilon\Big] = 0.5$ and $\Pr\Big[d(w, w') \leq \varepsilon \;\Big|\, \Pr[z \geq \tau]\Big] = 0.5$, yielding

$$\Pr\Big[z \geq \tau \;\Big|\, d(w, w') \leq \varepsilon\Big] = \Pr[z \geq \tau] = p_c. \qquad (8)$$

It should be noted that Eq (8) reduces the worst-case scenario, with referring to the maximum value of ε, to the

---

[1] We direct the interested reader to refer to [39], [40] [41] for more details regarding the issues on "more error than entropy" biometric sources.

average-case false acceptance security of (TRANS, AUTH) depending on the average selection of parameter u, τ, b with arbitrary value of $\varepsilon > 0$. Recall that we can bound the (minimum) entropy m_0 of the biometric sources of distribution W follows $m_0 \geq k\varepsilon$. In such a case, it is convenient to define $\lfloor k\varepsilon \rfloor = -\lfloor \log(1/p_c) \rfloor$, as the false acceptance complexity, which leads us to the following claim to show meaningful false acceptance security for large classes of biometric sources with (minimum) entropy at least equal to the false acceptance complexity.

Claim 2: [2] *Given any attacker is able to sample $w' \in W'$ over some random distribution $W' \in \mathbb{R}^k$ s.t. the original dissimilarity $d(w, w')$ is at most ε, where $w \in \mathbb{R}^k$ is the targeted attack biometric template. The average-case false acceptance security of* (TRANS, AUTH) *is $p_c(u, \tau, b, 1 - \varepsilon)$ for any $\varepsilon > 0$. In particular, the input distribution $W \in \mathbb{R}^k$ for all $w \in W$ must possess (minimum) entropy equal to $m_0 \geq k\varepsilon \geq \lfloor k\varepsilon \rfloor = -\lfloor \log(1/p_c) \rfloor$.*

Attack via Record Multiplicity (ARM): ARM refers to a privacy attack, which utilized multiple compromised hashed templates with and without the associated information, i.e., helper data, parameters, etc. to reconstruct the original biometric template [42], [43]. For a biometric recognition system to be useful, it should allow the user to enroll in multiple applications. These enrolled templates shall store in different data storage, which can be easily compromised and make available to the third party. Because of this, ARM is conceived as a highly practical attack given a large deployment of biometric recognition systems.

In reality, to get access to the biometric system, potential attackers need not invert the hashed template completely; instead, only a close approximation of the original biometric template is necessary and sufficient [29]. Hence, it is desirable to analyze the ARM security in terms of the attack complexity to reconstruct a fraction of the original template, which is sufficient to get access to the system by using an arbitrary number of the hashed templates.

To show that the proposed algorithm pair (TRANS, AUTH) resist against the ARM, we can reduce ARM to false acceptance attack: for any random sample $w' \in W'$ efficiently reconstructed via ARM, i.e., within polynomial time, that is ε-close to the enrolled template w, the attacker can get access into the system by a false acceptance in polynomial time. The above statement clearly described that if the attacker can launch a successful false acceptance attack, then he/she can also launch a successful ARM attack efficiently if the reconstruction of the sample w' can be done efficiently, i.e., in polynomial time.

In fact, given the proposed KSA attack, we have demonstrated that the sampling process for the noisy sample $w_{e,i} \in \mathcal{D}$ where $d(w, w_{e,i}) \leq \varepsilon$ can be done in polynomial time (see Section IV) by only using one known sample w^* that is trivially obtained through self enrolment. Therefore, the proposed KSA attack can be considered as a more robust notion of ARM attack without the need for the attacker to compromise multiple template storages. Argued in this way, to show resistance against ARM, a non-linear DPT is desirable, which can be accomplished using the proposed algorithm pair (TRANS, AUTH) for transformation and authentication.

Non-Linear to Liner Mapping on the DPT Curve: Here we also explore the possibility of any attacker could perform a mapping from the non-linear DPT curve to a more linear one (see Fig. 1), which leads to the dispute against a system that exhibits a non-linear DPT curve looking close to the optimal DPT need not be necessarily better in security as compared to the linear case.

To support the justification that a non-linear DPT offers better security guaranty, we first note that the knowledge of the non-liner DPT curve need not to be kept in secret. We also note that the proposed transformation and authentication (TRANS,AUTH) functions are only useful when the value of τ is known, means a proper value of τ must be selected to show meaningful non-linear property in such a way that the gap between the genuine and imposter distribution is maximized. Therefore, any attacker and system provider must know the DPT curve, i.e., the parameter set (u, τ, b) while designing the biometric system.

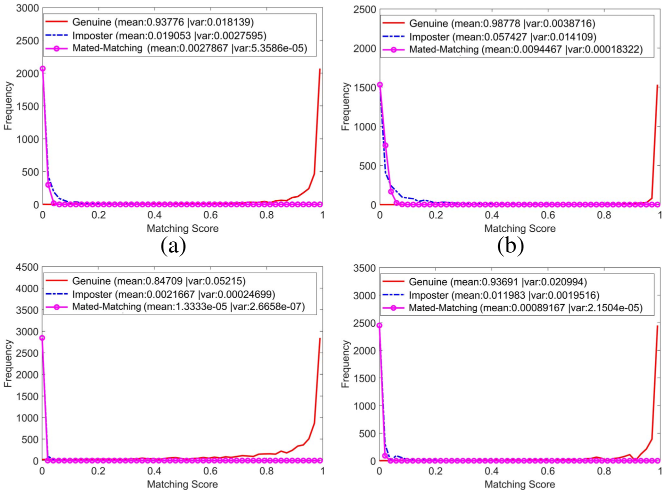
Since the mapping from a non-liner DPT curve to more linear one implies the changes in the S-curve and its gradient, which is parameterized by the parameters (u, τ, b). In such a case, mapping from non-linear DPT curve to linear is possible if there are multiple systems, say q number, where a targeted user has generated his/her biometric samples $(w_1, \ldots, w_q) \in W$ (e.g., generated from the user's face biometric) over a random distribution W, and enrolled w_i into the i-th system. Clearly, a non-linear mapping would succeed if one of the available systems (among q) behaves a linear DPT curve. On the contrary, such mapping can be avoided if all the systems have a proper choice of (u, τ, b) that renders a non-linear DPT curves. Doing this is necessary to ensure the security of the biometric samples $(w_1, \ldots, w_q) \in W$ to be enrolled into different systems for personal authentication.

E. Revocability and Unlinkability

Revocability Evaluation: To evaluate the revocability of the algorithm pair (TRANS, AUTH), we follow the same protocol mentioned in Section VI (first paragraph) to generate 3000 mated-matching scores, which are the matching scores between different hashed templates, generated using different set of random Gaussian vectors $(r_1, \ldots, r_n)$, over the *same* subject. We evaluate the revocability of the algorithm pair (TRANS, AUTH) under different values of $b = 1, 2, 3, 4$ with respect to different parameter settings of (u, τ, b) that render the lowest EER as tabulated in Table I. The genuine and imposter scores' distributions (both involved in only single set of random Gaussian vectors) are plotted together with the mated-scores' distribution (involved 3000 different sets of random Gaussian vectors) in a graph. Fig. 7 depicted four different graphs of different parameter settings with constant

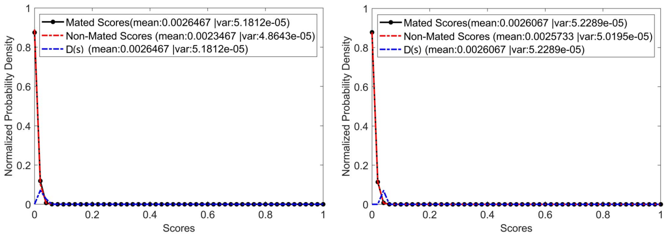
---

[2] Note that the derived false acceptance complexity does not assert any computational assumption over the attacker site. In other words, we allow the attacker to have unlimited computation power to model the biometric input distribution W and assume he/she is able to sample a $w' \in W'$ from W' where $d(w, w') \leq \varepsilon$ holds under such information-theoretical (computationally unbounded) setting.

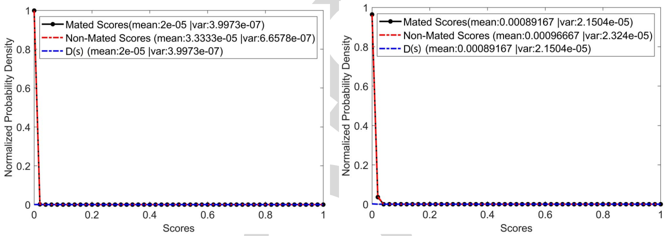


(a) (b) (c) (d)

Fig. 7. Revocability evaluation: the graphs of the genuine, imposter, and mated-matching scores' distributions.



(a) $D_{sys} = 0.06980$ (b) $D_{sys} = 0.09770$

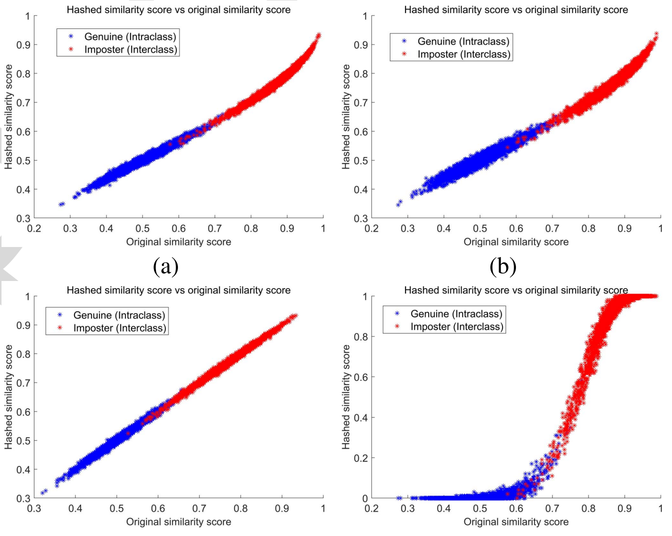


(c) $D_{sys} = 0.00003$ (d) $D_{sys} = 0.00160$

Fig. 8. Unlinkability evaluation: the graphs of the mated-matching and non-mated matching scores' distributions.

$s = 50$: (a) $u = 80, \tau = 26, b = 1$, (b) $u = 40, \tau = 22, b = 2$, (c) $u = 40, \tau = 30, b = 3$, and (d) $u = 40, \tau = 30, b = 4$ respectively. Note that a large degree of overlapping occurs between the imposter and mated-matching scores' distributions are observed. This result implies that the refreshed templates are sufficiently distinctive, albeit they are generated from the same subject. Indeed, the new transformed sample generated with a different set of random Gaussian vectors acts as an 'imposter' to the old one since they are uncorrelated. This verifies the revocability of (TRANS, AUTH) in generating new templates to replace the old one with a different set of random Gaussian vectors.

Unlinkability Evaluation: To evaluate the unlinkability of the algorithm pair (TRANS, AUTH), we dopted the framework proposed by Gomez *et al.* [44]. Let $\Pr[s \mid M_s]$ be the probability densities of a given similarity score $s \in [0, 1]$ that belongs mated-matching group. On contrary, let $\Pr\left[s \mid M_s'\right]$ denote the probability densities of score s belongs to the non-mated group M_s': the matching scores generated with (TRANS, AUTH) over different hashed templates generated using different set of random Gaussian vector $(r_1, \ldots, r_n)$ under the *different* subjects. The unlinkability property can be characterized by the local linkability defined as $D(s) = 2\frac{\omega LR(s)}{1+\omega LR(s)} - 1$ given $\omega LR(s) = \Pr[s \mid M_s] / \Pr\left[s \mid M_s'\right] > 1$, where $LR(s)$ is the likelihood ratio and $\omega = \Pr[M_s] / \Pr\left[M_s'\right]$ which can be conveniently set equal to one. The system's linkability is then defined as $D_{sys} = \int D(s) \Pr[s \mid M_s]\, ds$. Specifically, $D_{sys} \in [0, 1]$ and the system is completely linkable given $D_{sys} = 1$. Therefore, to attain unlinkability of a BTP scheme, it is desirable to show that D_{sys} is negligible small. Referring to the same parameter settings in revocability evaluation, Fig. 8 depicted four different graphs, each contains 3000 mated-matching scores and 3000 non-mated matching scores. The results show that the mated and non-mated scores' distributions are significant overlapping (for all four graphs) with small value of D_{sys}. Therefore we assert that the algorithm pair (TRANS, AUTH) supports unlinkability.



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

Fig. 9. Comparison of the non-linearlity DPT curve of proposed technique to (a) Bio-hashing, (b) IOM-hashing, (c) IFO-hashing, (d) proposed.

F. Comparison With Existing Approaches

Non-Linearlity: We compare our proposal (best performance setting $s = 50, u = 40, b = 2, \tau = 22$) with the best performance setting for Bio-hashing [4] (0.73% EER) and other notable LSH scheme such as Index of Max hashing (IOM) [10] (0.75% EER) and Indexing First One hashing (IFO) [11] (1.38% EER). Fig. 9 depicted the comparison results. Our proposal yields a highly non-linear relationship between the original similarity scores versus the hashed similarity scores in comparison to others.

Decision Environment: The degree that one can confidently decide whether the observed sample belongs to the genuine (red) or imposter distribution (blue) is as shown in Fig. 5.

Note that the error rate is proportional to the overlapped region between the genuine and imposter distributions.

The decision environment for dual distributions reveals the extent to which the genuine and imposter distribution can be separated, thus determining how reliable the decision can be made for individual authentication. Following the works by Daugman [38], for two-choice decision task such as biometric decision making, we can measure the separation of these two distributions by their decidability d' defined in Eq (9), where (μ_1, μ_2) and (σ_1, σ_2) refer to the two means and standard deviation, respectively, of two different distributions.

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}}. \tag{9}$$

The measure of d' is independent w.r.t. any acceptance threshold. Instead, it reflects the cost for the system in reducing the FAR via increasing FRR, or vice versa. Therefore, one can succinctly use d' to calibrate the performance of every biometric technology.

Based on the studies in the previous Section II-C, to optimize the mutual information leakage, the best is to hope for achieving $\mathbb{E}[D_I] \rightarrow 1$ and $\mathbb{E}\big[D_g\big] \rightarrow 0$ with $\mathrm{Var}(D_I)$ is minimized. Therefore the gap between the distribution of the interclass's distance and the distribution of the intraclass's distance must be large enough.

Note that a large gap between the distribution of the interclass's distance D_I (imposter distribution) and the distribution of the intraclass's distance D_g (genuine distribution) implies high decidability. More precisely, the decidability can be described in term of $\mathbb{E}[D_I]$, $\mathbb{E}\big[D_g\big]$, $\mathrm{Var}(D_I)$, and $\mathrm{Var}(D_g)$ as $d' = \frac{|\mathbb{E}[D_I]-\mathbb{E}[D_g]|}{\sqrt{(\mathrm{Var}(D_I)+\mathrm{Var}(D_g))/2}}$. Since our optimization goal is to minimize $\mathrm{Var}(D_I)$ while keeping $\mathbb{E}[D_I] \rightarrow 1$ and $\mathbb{E}\big[D_g\big] \rightarrow 0$. Therefore, such goal can be achieved by maximizing d', which suggests a steeper gradient of the S-curve (highly non-linearity) depicted in Fig. 1.

Follow Fig. 5 (d), the computed decidability in our proposal is 10.03. Besides, in our experiment, the measured d' for Bio-hashing, IOM-hashing and IFO-hashing are 4.92, 5.32, and 2.52 respectively. The comparison on the recorded d' with the recent proposed state-of-the-art BTP schemes [37], [45], [36], [46], [47] is tabulated in Table II. Such comparison is performed under the scenario when the user and attacker have complete knowledge on the transformation function and parameters used.

Observe that our proposal can achieve a higher d' value among most of the state-of-the-art BTP schemes. The achievable $d' = 10.03$ is higher as compared to a non-ideal (crossed platform) iris recognition system, which is 7.3 as reported in [38]. It is also worth highlighting that the non-linearity between the original similarity scores and the hashed similarity scores can be strengthened by increasing the parameter u with a proper selection of b and τ, which promotes the maximization of the system's decidability d'. This is in our favor of reducing the mutual information leakage (i.e., minimizing $\mathrm{Var}(D_I)$) to show resistance against the DPT based attacks, while maintaining a good recognition utility (keeping $\mathbb{E}\big[D_g\big] \rightarrow 0$) as discussed in Section II.

Last but not least, we examine the performance in terms of FRR against FAR for various distance preserving hashing BTP schemes using the detection error trade-off (DET) curve, as shown in Fig. 10. As it can be observed, the proposed scheme achieved superior performance with the lowest FRR (4.4%) at zero FAR (0 %). On average, this implies only about 4 rejections (i.e. 4.4% FRR) over 100 trials of a genuine user to be authenticated, while no unauthorized persons is accepted incorrectly (i.e. zero FAR). This result suggests that the proposed scheme is feasible in real application scenarios.

TABLE II

COMPARISON OF SYSTEM'S DECIDABILITY WITH OTHER EXISTING BTP SCHEMES BASED ON THEIR RECORDED (HIGHEST) d'

BTP Schemes	**Decidability**, d'
Bio-hashing [6] (for fingerprint)	4.92
IOM-hashing [10] (for fingerprint)	5.34
IFO-hashing [11] (for human iris)	5.67
Kaur et al. [37] (for fingervein, palmvein, and face)	9.74
Sadhya at al. [45] (for human iris)	2.39
Qiu et al. [46] (for palm print)	9.20
Walia et al. [47] (for human iris and pericular feature)	13.47
Walia et al. [36] (for human iris, fingerprint, and face)	5.38
Proposed (for face)	10.03

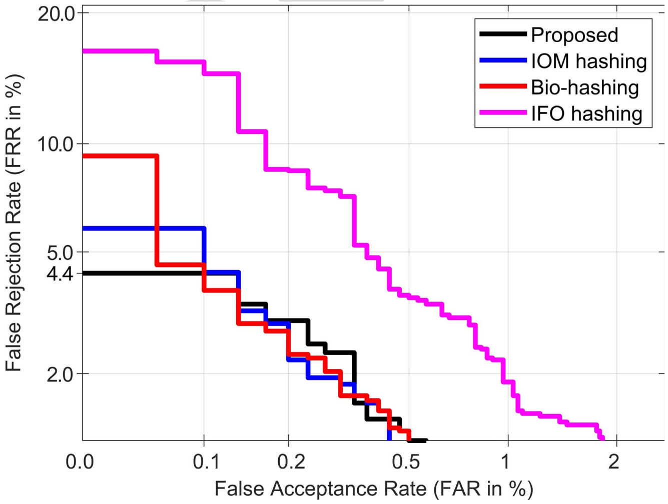


Fig. 10. DET curves for various distance preserving hashing BTP schemes.

VII. CONCLUSION

In this work, we explore the vulnerability in the existing distance-preserving hashing BTP scheme. We demonstrate an efficient security attack, i.e., KSA, for distance-preserving hashing BTP. Our results show that the potential attacker can model the input samples' distribution and obtain the pre-images of the enrolled biometric sample. This scenario is worse when the hash function's output length is set to very small, that is preferred by most distance-preserving hashing BTP schemes for irreversibility purpose via dimension reduction. We also provide some discussions over the mutual information leakage due to the published distance-preserving hashing BTP. Noticing the non-linearity relationship between the input distance and hashed distance is crucial to provide authenticity for similar subjects while avoiding false matching for distinct subjects. The above reasons motivated our work on a pair of transformation and authentication algorithm (Trans, Auth) to give a highly non-linear relationship between the input and hashed domains. The algorithm pair (Trans, Auth) offers efficiency and simplicity for fast and secure authentication with a biometric template (we used face vector in our experiment). Most importantly, it showed resistance against KSA for polynomial-time bounded attackers under known distribution $\mathcal{D}$ scenario and satisfied the four criteria to be used as a secure BTP scheme.

References

[1] S. Nanavati, M. Thieme, and R. Nanavati, *Biometrics, Identity Verification in a Networked World*. Hoboken, NJ, USA: Wiley, 2002.

[2] S. Prabhakar, S. Pankanti, and A. K. Jain, “Biometric recognition: Security and privacy concerns,” *IEEE Secur. Privacy*, vol. 1, no. 2, pp. 33–42, Mar. 2003.

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

[4] A. B. J. Teoh, Y. W. Kuan, and S. Lee, “Cancellable biometrics and annotations on BioHash,” *Pattern Recognit.*, vol. 41, no. 6, pp. 2034–2044, Jun. 2008.

[5] A. B. J. Teoh, A. Goh, and D. C. L. Ngo, “Random multispace quantization as an analytic mechanism for BioHashing of biometric and random identity inputs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1892–1901, Dec. 2006.

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

[7] T. Connie, A. Teoh, M. Goh, and D. Ngo, “PalmHashing: A novel approach for dual-factor authentication,” *Pattern Anal. Appl.*, vol. 7, no. 3, pp. 255–268, Dec. 2004.

[8] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, “Sectored random projections for cancelable iris biometrics,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 1838–1841.

[9] J. Portelo, A. Abad, B. Raj, and I. Trancoso, “Secure binary embeddings of front-end factor analysis for privacy preserving speaker verification,” in *Proc. INTERSPEECH*, 2013, pp. 2494–2498.

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

[11] Y.-L. Lai *et al.*, “Cancellable iris template generation based on indexing-first-one hashing,” *Pattern Recognit.*, vol. 64, pp. 105–117, Apr. 2017.

[12] K.-Y. Chee *et al.*, “Cancellable speech template via random binary orthogonal matrices projection hashing,” *Pattern Recognit.*, vol. 76, pp. 273–287, Apr. 2018.

[13] Y. Chen, Y. Wo, R. Xie, C. Wu, and G. Han, “Deep secure quantization: On secure biometric hashing against similarity-based attacks,” *Signal Process.*, vol. 154, pp. 314–323, Jan. 2019.

[14] B. Kumar Pandya, U. Kumar Singh, K. Dixit, and K. Bunkar, “Effectiveness of multiplicative data perturbation for privacy preserving data mining,” *Int. J. Adv. Res. Comput. Sci.*, vol. 5, no. 6, pp. 112–115, 2014.

[15] J. Kim and W. Winkler, “Multiplicative noise for masking continuous data,” *Statistics*, vol. 1, p. 9, Apr. 2003.

[16] P. Tendick, “Optimal noise addition for preserving confidentiality in multivariate data,” *J. Stat. Planning Inference*, vol. 27, no. 3, pp. 341–353, Mar. 1991.

[17] T. Evans, L. Zayatz, and J. Slanta, “Using noise for disclosure limitation of establishment tabular data,” in *Proc. Annu. Res. Conf., US Bur. Census*, Washington, DC, USA, vol. 20233, 1996, pp. 65–86.

[18] B. D. Okkalioglu, M. Okkalioglu, M. Koc, and H. Polat, “A survey: Deriving private information from perturbed data,” *Artif. Intell. Rev.*, vol. 44, no. 4, pp. 547–569, Dec. 2015.

[19] K. Liu, C. Giannella, and H. Kargupta, “An attacker’s view of distance preserving maps for privacy preserving data mining,” in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery*. Berlin, Germany: Springer, 2006, pp. 297–308.

[20] E. O. Turgay, T. B. Pedersen, Y. Saygın, E. Savaş, and A. Levi, “Disclosure risks of distance preserving data transformations,” in *Proc. Int. Conf. Sci. Stat. Database Manage.* Berlin, Germany: Springer, 2008, pp. 79–94.

[21] S. Guo and X. Wu, “Deriving private information from arbitrarily projected data,” in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2007, pp. 84–95.

[22] K. Chen, G. Sun, and L. Liu, “Towards attack-resilient geometric data perturbation,” in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, Apr. 2007, pp. 78–89.

[23] W. K. Wong, D. W.-L. Cheung, B. Kao, and N. Mamoulis, “Secure kNN computation on encrypted databases,” in *Proc. 35th SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 2009, pp. 139–152.

[24] J. Galbally, A. Ross, M. Gomez-Barrero, J. Fierrez, and J. Ortega-Garcia, “Iris image reconstruction from binary templates: An efficient probabilistic approach based on genetic algorithms,” *Comput. Vis. Image Understand.*, vol. 117, no. 10, pp. 1512–1525, Oct. 2013.

[25] Y. C. Feng, M.-H. Lim, and P. C. Yuen, “Masquerade attack on transform-based binary-template protection based on perceptron learning,” *Pattern Recognit.*, vol. 47, no. 9, pp. 3019–3033, Sep. 2014.

[26] A. Adler, “Sample images can be independently restored from face recognition templates,” in *Proc. Can. Conf. Electr. Comput. Eng. Toward Caring Humane Technol. (CCECE)*, vol. 2, 2003, pp. 1163–1166.

[27] E. Kaplan, M. E. Gursoy, M. E. Nergiz, and Y. Saygin, “Known sample attacks on relation preserving data transformations,” *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 2, pp. 443–450, Mar. 2020.

[28] H. W. Chung, B. M. Sadler, and A. O. Hero, “Bounds on variance for unimodal distributions,” *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 6936–6949, Nov. 2017.

[29] A. Nagar, K. Nandakumar, and A. K. Jain, “Biometric template transformation: A security analysis,” in *Proc. 2nd Media Forensics Secur.*, vol. 7541. San Jose, CA, USA: International Society for Optics and Photonics, Feb. 2010, p. 75410.

[30] M. X. Goemans and D. P. Williamson, “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming,” *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[31] S. Poljak and Z. Tuza, “Maximum cuts and large bipartite subgraphs,” in *DIMACS Series*, vol. 20. New Brunswick, NJ, USA: AMS, 1995, pp. 181–244.

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

[33] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,” in *Proc. Workshop Faces Real-Life Images, Detection, Alignment, Recognit.*, 2008.

[34] C. Vielhauer and R. Steinmetz, “Handwriting: Feature correlation analysis for biometric hashes,” *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 4, 2004, Art. no. 389304.

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

[36] G. S. Walia, G. Jain, N. Bansal, and K. Singh, “Adaptive weighted graph approach to generate multimodal cancelable biometric templates,” *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1945–1958, 2020.

[37] H. Kaur and P. Khanna, “Random distance method for generating unimodal and multimodal cancelable biometric features,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 709–719, Mar. 2019.

[38] J. Daugman, “How iris recognition works,” in *The Essential Guide to Image Processing*. Amsterdam, The Netherlands: Elsevier, 2009, pp. 715–739.

[39] S. Simhadri, J. Steel, and B. Fuller, “Cryptographic authentication from the iris,” in *Proc. Int. Conf. Inf. Secur.* New York, NY, USA: Springer, 2019, pp. 465–485.

[40] G. Itkis, V. Chandar, B. W. Fuller, J. P. Campbell, and R. K. Cunningham, “Iris biometric security challenges and possible solutions: For your eyes only? Using the iris as a key,” *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 42–53, Sep. 2015.

[41] Y.-L. Lai and Z. Jin, “Input-dependent error sketching model enabled information theoretical secure sketch,” *IEEE Access*, vol. 8, pp. 134681–134694, 2020.

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

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

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

[45] D. Sadhya and B. Raman, "Generation of cancelable iris templates via randomized bit sampling," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 11, pp. 2972–2986, Nov. 2019.

[46] J. Qiu, H. Li, and C. Zhao, "Cancelable palmprint templates based on random measurement and noise data for security and privacy-preserving authentication," *Comput. Secur.*, vol. 82, pp. 1–14, May 2019.

[47] G. S. Walia, K. Aggarwal, K. Singh, and K. Singh, "Design and analysis of adaptive graph based cancelable multi-biometrics approach," *IEEE Trans. Dependable Secure Comput.*, early access, May 25, 2020, doi: 10.1109/TDSC.2020.2997558.

Yenlung Lai received the B.Sc. degree in physics from Universiti Tunku Abdul Rahman (UTAR), Malaysia, in 2015. He is currently pursuing the Ph.D. degree with Monash University Malaysia. His research interests include information security and biometrics.

Zhe Jin (Member, IEEE) received the B.I.T. degree (Hons.) in software engineering and the M.Sc. (I.T.) degree from Multimedia University, Malaysia, in 2007 and 2011, respectively, and the Ph.D. degree in engineering from Universiti Tunku Abdul Rahman Malaysia in 2016. He is currently a Senior Lecturer with the School of Information Technology, Monash University Malaysia. He has published more than 40 refereed journals, conference articles, including IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE DSC, and *Pattern Recognition*. His research interests include biometric security, computer vision, and machine learning. He was awarded the Marie Skłodowska-Curie Research Exchange Fellowship and visited the University of Salzburg, Austria, and the University of Sassari, Italy, as a Visiting Scholar under the EU Project IDENTITY.

KokSheik Wong received the B.S. and M.S. degrees in computer science and mathematics from Utah State University, USA, in 2002 and 2005, respectively, and the D.Eng. degree from Shinshu University, Japan, under the scholarship of Monbukagakusho.

He was with Multimedia University from 2009 to 2010, and the University of Malaya from 2010 to 2016. He is currently an Associate Professor with the School of Information Technology, Monash University Malaysia. His research interests include multimedia signal processing, data hiding, multimedia perceptual encryption, joint encryption, and data-hiding method, as well as their applications. He is a member of the APSIPA. In 2015, his student's thesis received the Best Ph.D. Thesis Award from the IEEE Signal Processing Society, Malaysia Section. He currently serves as an Associate Editor for the *Journal of Information Security and Applications* (JISA), *Malaysian Journal of Computer Science*, and *IIEEJ Transactions on Image Electronics and Visual Computing*. He also serves as the Vice-Editor-in-Chief for *APSIPA Newsletter* and the EiC for *APSIPA Newsletter* in 2020.

Massimo Tistarelli (Senior Member, IEEE) received the Ph.D. degree in computer science and robotics from the University of Genoa.

He is currently a Full Professor (with tenure) in computer science and the Director of the Computer Vision Laboratory, University of Sassari, Italy. Since 1986, he has been involved as a project coordinator and task manager in several projects on computer vision and biometrics funded by the European Community. Since 1994, he has been the Director of the Computer Vision Laboratory, Department of Communication, Computer and Systems Science, University of Genoa, and currently at the University of Sassari, leading several National and European projects on computer vision applications and image-based biometrics. He is a Founding Member of the Biosecure Foundation, which includes all major European research centers working in biometrics. He is one of the world-recognized leading researchers in the area of biometrics, especially in the field of face recognition and multimodal fusion. He is coauthor of more than 150 scientific papers in peer reviewed books, conferences, and international journals. He is the Principal Editor of the Springer books "*Handbook of Remote Biometrics*" and "*Handbook of Biometrics for Forensic Science*." His main research interests include biological and artificial vision (particularly in the area of recognition, three-dimensional reconstruction, and dynamic scene analysis), pattern recognition, biometrics, visual sensors, robotic navigation, and visuo-motor coordination.

Prof. Tistarelli is a fellow of the IAPR and the Vice President of the IEEE Biometrics Council. He has organized and chaired several world-recognized several scientific events and conferences in the area of computer vision and biometrics, and he has been an Associate Editor for several scientific journals, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *IET Biometrics*, *Image and Vision Computing*, and *Pattern Recognition Letters*. Since 2003, he has been the Founding Director for the International Summer School on Biometrics (16th Edition).