Face and Fingerprint Biometric Template Protection for Edge Devices in Humanitarian Action

Giuseppe Stragapede*, Sam Merrick*, Justin Sukaitis†, Vincent Graf Narbel†, Vedrana Krivokuća Hahn‡,
**Simprints Technology Ltd, Cambridge, United Kingdom*
†*International Committee of the Red Cross (ICRC), Geneva, Switzerland*
‡*Biometrics Security and Privacy Group, Idiap Research Institute, Martigny, Switzerland*

***Abstract*—While biometrics can dramatically increase the efficiency of operations in humanitarian and emergency scenarios, its usage might entail significant risks for the data subjects, which can be exacerbated in contexts of vulnerability. For this reason, the International Committee of the Red Cross (ICRC) and Simprints have partnered with the goal of building the first open-source mobile biometric product implementing a Biometric Template Protection (BTP) method. After laying out the specific functional, operational, and security and privacy project requirements, *PolyProtect*, a BTP method designed to operate directly on neural network face embeddings, is selected due to its effectiveness, modularity, and lightweight computational burden. We carry out an experimental evaluation of PolyProtect from the perspective of verification and identification performance, irreversibility, and unlinkability, in combination with *EdgeFace*, a novel state-of-the-art lightweight feature extractor, on a face dataset collected in a field project in Ethiopia. Moreover, as PolyProtect promises to be modality-independent, we evaluate its suitability for fingerprint biometrics, considering fixed-size neural network-based templates obtained from a dataset collected in a field project in Ghana. To the best of our knowledge, this is the first time that fingerprint biometrics are combined with PolyProtect. We observe that PolyProtect generally improves the verification performance for both modalities, showing the true cross-modality potential of this BTP method. Moreover, sound experimental results are also achieved from the perspective of irreversibility and unlinkability. For reproducibility purposes, we also include evaluations on publicly available synthetic datasets for both modalities. We plan to make our implementation open-source[1].**

***Index Terms*—Biometric Template Protection, Edge Devices, Face, Fingeprint, Humanitarian Action**

I. Introduction

One of the main operations of the International Committee of the Red Cross (ICRC) is the distribution of physical goods, such as food or blankets, in contexts of international conflict.

In these scenarios, biometric recognition can represent a convenient and efficient solution to verify the identity of beneficiaries, without requiring them to present physical objects (IDs, access tokens, or personal cards), which could be difficult to distribute and reissue in case of loss, nor knowledge of any specific information (passwords or PIN codes), which could be forgotten or shared with unentitled subjects.

However, the usage of biometrics might introduce risks for data subjects, such as impersonation [1], inference of personal and sensitive information [2], linkage of individuals across application databases (cross-matching) [3], etc. These risks can be exacerbated if the contexts in which beneficiaries live makes them especially vulnerable. For example, when the Taliban took control of Afghanistan, they gained access to biometric devices left behind by the US Army. These devices contained data about Afghan civilians, and the Taliban used them to determine who had a relation to the US Army [4], [5]. For these reasons, the security standards upheld by humanitarian organizations are stricter, in line with the 'do-no-harm' principle [6]. To this end, ICRC has adopted a dedicated biometrics policy designed to facilitate their responsible use and address the data protection challenges this poses [7].

Typically, biometric data are stored in the form of a *biometric template*, containing a compact representation produced by a feature extractor algorithm – also known as an *embedding*, when the template is generated by a neural-network (NN)-based feature extractor – such that it retains the most meaningful information to recognise individuals, discarding unnecessary information from the raw images. Additionally, templates are more convenient for matching operations and storage. However, it has been shown that templates, especially (face) embeddings, can be inverted to recover a close approximation of the underlying biometric signal (*e.g.*, face image) [8]. Fortunately, over the past two decades, the biometrics research community has invested significant efforts towards enhancing the security of biometric operations, with the establishment of the Biometric Template Protection (BTP) research field [9], [10]. BTP aims to develop methods that can be applied to biometric data to produce a protected version that can be safely stored. The main property of protected data is that, if they fall into the hands of an adversary, they reveal little or no information about the underlying biometric characteristic that was captured during the sample collection, thus protecting the data subjects. A BTP method should possess the following properties [11]:

1) Recognition accuracy: the incorporation of the protection method into a biometric recognition system should not result in a (significant) degradation of the system's recognition accuracy.
2) Irreversibility: it should be impossible (or computationally infeasible) to recover the original biometric data from its protected version.
3) Unlinkability and renewability: it should be possible

*Email: giuseppe@simprints.com

[1] https://github.com/Simprints/TBA/

to generate multiple distinct protected templates from the same subject's biometric data, such that the protected templates cannot be linked to each other. This would allow for the revocation and subsequent renewal of compromised templates, as well as the use of the same biometric characteristic across multiple databases, without the risk of cross-matching the data.

Simprints and ICRC have partnered in the context of the "Safe Biometrics for Humanitarian Aid" project to build the first open-source biometric product implementing a BTP solution that fulfils the strict security requirements for humanitarian use-cases. The designed solution will be integrated into the Simprints open-source mobile app, which is currently being deployed in several field projects.

The main contributions of this paper can be summarised as follows:

- A description of our use-case (Sec. II), which stems from the context of humanitarian action, with its specific functional, operational, security and privacy requirements that need to be fulfilled by the adopted BTP solution (Sec. III). The characteristics of our use-case do not impair the generality of the gathered requirements, which could be similar to a number of mobile security-critical applications.
- A thorough analysis of the entire BTP landscape (reported in Sec. IV), which results in identifying *PolyProtect*, a feature-transformation approach designed for mobile face verification [12], as the most suitable method, over well-known approaches such as homomorphic encryption (HE) or hashing-based solutions.
- A complete experimental evaluation of PolyProtect (Sec. V), based on a recent face recognition model designed for low computational resources, EdgeFace [13], on an internal face dataset collected in a field project in Ethiopia, and on the publicly available Vec2Face synthetic dataset [14]. Moreover, as PolyProtect promises to be modality-independent, we evaluate the suitability of this protection method for fingerprint biometrics, considering a neural network based feature extractor [15] on an internal dataset collected in Ghana, and on the publicly available SynFing synthetic dataset [16]. To the best of our knowledge, this is the first time that fingerprint biometrics are combined with PolyProtect. We adopt the methodology and experimental protocol proposed in [12] (Sec. VI). Our experimental results (Sec. VII) show that with respect to the unprotected system, PolyProtect can improve the system verification accuracy. Additionally, we benchmark the performance of the system in the more challenging identification task, noting a limited performance drop in some cases. Finally, sound experimental results are also achieved for both face and fingerprint modalities from the perspective of irreversibility and unlinkability, showing the true cross-modality potential of this BTP method.

II. Background Context

This section will provide the reader with some useful information concerning the premises under which the current work was developed. The system into which the designed BTP solution is meant to be integrated is briefly presented in Sec. II-A, followed by a discussion of the threat model considered in this project (Sec. II-B).

A. Simprints ID

Simprints operates through an in-house open-source Android mobile application (*SimprintsID*, *SID* for short) that enables community-based healthcare workers to enrol, identify or verify a beneficiary using biometrics. SID is a standalone application but can be easily integrated with any data collection platform[2], such as DHIS2, CommCare, SurveyCTO, etc. It can be used offline or online, enabling local staff to reach beneficiaries in some of the most remote field settings. SID can store data locally on the device without the need to connect to online servers, allowing the field worker to enrol new beneficiaries, as well as to verify and identify existing ones. When the field worker is within mobile data/Wi-Fi range, they can synchronise with the Simprints cloud database, to update records on their device and send data to the cloud database. SID currently supports two modalities: *(i)* fingerprint, enabled via a proprietary-design Bluetooth scanner, *Vero*, with a design focus on accuracy, mobility and robustness, for the extreme demands of last-mile health and humanitarian aid use-cases; *(ii)* face, using the camera hardware within the employed Android device.

B. Threat Model

While the evaluation of the recognition accuracy of a BTP method can be quite straightforward, as it would involve comparing the performance of the original (unprotected) biometric system with that of its protected counterpart, from the point of view of security it is first necessary to define a threat model, which characterises the type of attacker on which the security analysis is based. Several threat models are defined in the ISO/IEC 30316:2018 standard on performance testing of biometric template protection schemes [11]. In our use-case, a realistic scenario is represented by an attacker stealing one or more devices, which can be rooted to: obtain the full knowledge of the algorithms used for template extraction, template protection and comparison, as well as all the secrets; possibly execute all the submodules of the system that make use of the secrets. As per the mentioned standard, such conditions would correspond to the worst-case scenario, known as the *full disclosure model*.

III. BTP System Requirements

In the first phase of the project, the requirements that the BTP solution needs to satisfy were gathered through conversations with the ICRC's Data Protection Office and from ICRC documents on data protection and biometrics policy [6], [7]. The formulated requirements, grouped into three categories (functional, security and privacy, and operational), are reported below.

[2] https://simprints.gitbook.io/docs/development/simprints-for-developers/other-intergrations

Functional Requirements:

F.1 – Recognition Accuracy: There should be no degradation of the recognition performance of the protected biometric system with respect to its unprotected counterpart.

F.2 – Modality-Independence: As both Simprints and ICRC operations might involve different biometric modalities, the designed BTP solution should be applicable to all.

F.3 – Feature Extractor-Independence: It should be possible to combine the BTP solution with different biometric feature extractors.

F.4 – On-device Recognition: The enrolment, verification and identification of subjects should take place on the device without any internet connectivity.

F.5 – Easy New Enrolment: The enrolment of new subjects should not create any conflict with the running BTP solution nor with previously enrolled subjects.

F.6 – Template Revocability and Renewability: If the protected templates are compromised, *e.g.*, in the case of a reported missing device, it should be possible to revoke them and reissue new protected instances.

F.7 – Open-Source: The BTP solution should not use closed-source or commercial solutions, relying instead on technologies released under open-source-compatible license terms.

Security and Privacy Requirements:

S.1 – Irreversibility: The adherence of the BTP solution to the irreversibility principle should be demonstrated theoretically or empirically.

S.2 – Unlinkability: The adherence of the BTP solution to the unlinkability principle should be demonstrated theoretically or empirically.

Operational Requirements:

O.1 – Computational Efficiency: Lightweight BTP solutions should be preferred due to the mobile environment resource constraints.

O.2 – Time Efficiency: Fast BTP solutions should be preferred, *e.g.*, time should not represent an issue for identification against a large enrolment database stored locally.

O.2 – Offline Processing: The BTP solution should not rely on any remotely available resource.

IV. BTP Landscape Analysis

In light of the gathered requirements, an analysis of the BTP methods available in the literature was carried out to identify the most suitable solutions. To this end, a useful taxonomy of BTP methods is based on two independent aspects proposed in [17]: *(i)* method *type*, which can be *handcrafted* or *NN-based*; *(ii)* method *input*, which can be at *image-level* or at *feature-level*. Fig. 1 summarises the described BTP taxonomy. In the remainder of this section, we present the different BTP categories, and we rule out those which are in conflict with the outlined project requirements.

A. Method Type: Handcrafted vs. NN-based

In general terms, the sample(s) or features used as primary input to the BTP scheme are defined as *generative biometric data* [10]. Handcrafted approaches are explicitly defined algorithms applied to generative biometric data to produce protected instances of them. In contrast, NN-based approaches involve training a neural network to learn a suitable protection algorithm, to transform the generative biometric data to a protected template [17].

NN-based approaches are more recent than handcrafted ones, and they have received attention as they offer the possibility of avoiding explicitly defining the protection method. Nevertheless, the limitations of NN-based methods make them less preferable for our use-case. For example, such protection methods are generally specific to the neural network on whose templates they are trained (conflicting with *F.3 – Feature Extractor-Independence*) [18], [19]. Moreover, assessing the scalability of these methods without retraining can be difficult (*F.5 – Easy New Enrolment*) [20], and the template renewability aspect also appears challenging (*F.6 – Template Revocability and Renewability*) [21].

Concerning the security analysis, in the adopted full-disclosure threat model (Sec. II-B), the adversary has access to the trained model (*i.e.*, network architecture and all learned parameters). Consequently, the irreversibility analysis for NN-based BTP methods should consider how this knowledge could be used to extract additional information about the original embedding or image from different layers of the neural network (*S.1 – Irreversibility*). Such a thorough irreversibility analysis is still lacking in the literature for these kinds of methods.

B. Method Input: Image-level vs. Feature-level

Image-level methods are applied directly to the biometric sample (*e.g.*, a face image), following which a biometric feature extractor (such as a neural network) would be used to extract features from the protected image. In the case of feature-level methods, a biometric recognition system (most likely a neural network) would first be used to extract a set of features or a "template" (*i.e.*, an embedding, in the case of

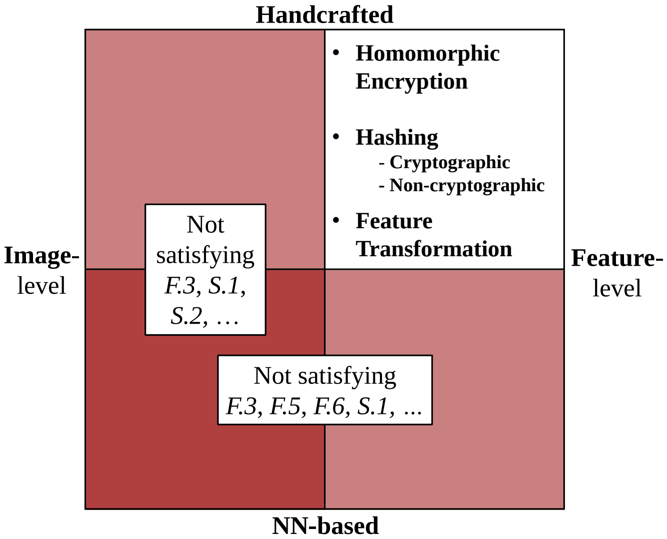


Fig. 1: Taxonomy of BTP methods proposed in [17]. We rule out the categories in red due to their incompatibility with several system requirements.

TABLE I: Having narrowed our focus to feature-level handcrafted BTP methods, we consider three approaches: homomorphic encryption, hashing, and PolyProtect [12]. Below, we roughly rate them against our project requirements and threat model. For the requirements, we consider four possible judgment values: satisfied, possible, challenging, weak.

Requirement		**HE**	**Hashing**	**PolyProtect [12]**
F.1	*Recognition accuracy*	Satisfied	Challenging	Satisfied
F.2	*Modality-Independence*	Satisfied	Possible**	Possible
F.3	*Feature Extractor-Independence*	Satisfied	Possible**	Satisfied
F.4	*On-Device Recognition*	Challenging	Possible	Satisfied
F.5	*Easy New Enrolment*	Satisfied	Possible	Satisfied
F.6	*Template Revocability*	Satisfied	Possible	Satisfied
F.7	*Open-Source*	Satisfied	Possible	Satisfied
S.1	*Irreversibility*	Satisfied*	Challenging	Satisfied
S.2	*Unlinkability*	Satisfied	Challenging	Satisfied
O.1	*Computational Efficiency*	Weak	Possible	Satisfied
O.2	*Time Efficiency*	Weak	Possible	Satisfied
O.2	*Offline Processing*	Weak	Possible	Satisfied
Full-Disclosure Threat Model [11]		Incompatible	Uncertain	Compatible

*Not under the full-disclosure threat model.

**Satisfying this requirement might entail using different hashing algorithms for different modalities or feature extractors.

a neural-network-based feature extractor) from the biometric sample, then the BTP algorithm would be applied to this template to generate the protected template.

The overwhelming majority of the works proposed in the literature focus on applying a BTP method at feature level. This preference for feature-level BTP may be attributed to the availability of several pre-trained NN biometric recognition models, which have been shown to be capable of extracting highly discriminative features from images. In this way, the BTP mechanism could be integrated as an *add-on* to pre-existing biometric systems, rather than having to additionally design a robust feature extractor for the protected images. To preserve the system modularity, requirement *F.3 – Feature-Extractor Independence* implies that the BTP method should not work directly on the generative data, allowing us to rule out this category of BTP methods. Additionally, from the perspective of the security and privacy analysis, there is scarcity in the scientific literature concerning the irreversibility and unlinkability properties of these systems (*S.1 – Irreversibility*, *S.2 – Unlinkability*), as well as what information about the unprotected template is leaked in different layers of NN-based BTP methods [17].

C. Feature-level Handcrafted Methods

As illustrated in Fig. 1, we have ruled out three of the four categories due to their incompatibility with our project requirements. Nevertheless, the majority of the methods proposed in the literature can be classified as both handcrafted and feature-level [17]. Consequently, a number of methods fall into this category. Within this section, we narrow our focus to three approaches: homomorphic encryption (HE), hashing, and feature transformation approaches.

1) Homomorphic Encryption (HE): HE enables us to perform operations on encrypted biometric data without having to first decrypt it, allowing us to maintain the same recognition accuracy, since the comparison score obtained in the encrypted domain is in principle equal to the unprotected system score. Nevertheless, the computational complexity of HE makes it challenging to apply HE as a BTP method. Consequently, most efforts towards HE-based BTP solutions have focused on reducing this computational complexity – for example via template quantisation or dimensionality reduction – while simultaneously trying to minimise the resulting accuracy degradation [22], [23]. In addition, encrypted templates remain secure only insofar as the corresponding decryption key remains secret, conflicting with the full-disclosure threat model adopted in this project.

2) Hashing: Hashing operations create a fixed-size, predictable output called a “hash”, such that it is mathematically impossible to recover the original input data from its hash. Cryptographic hash functions are purposely designed to exaggerate small differences in the input, which is the main challenge in their application to biometric templates due to the intrinsic intra-class variability of biometric measurements. Consequently, hash-based BTP methods tend to apply hashing to random, subject-specific codewords, which are bound to the biometric templates by some mathematical function, with the goal of performing the matching operation indirectly by reconstructing the codewords using probe samples. This is the case for fuzzy committment [24] or fuzzy vault-based [25] schemes. To reduce sensitivity to intra-class variations, non-cryptographic hashes [26] have been proposed: in this case, the same subject’s templates are mapped to approximately the same code, allowing distance-based comparisons between reference and probe hashes, in contrast to cryptographic hash functions, which would require an exact match.

We observed that both cryptographic and non-cryptographic hashing methods proposed in the literature are likely to introduce some accuracy degradation [27]–[29] (*F.1 – Recognition Accuracy*). Moreover, their irreversibility and unlinkability have been demonstrated to be fragile in some cases (*S.1 – Irreversibility*, *S.2 – Unlinkability*) [30], [31], or non-exhaustively evaluated [32], [33].

3) Feature Transformation: Alternative approaches are based on transforming templates with the help of subject-specific transformation functions. Although such feature transformations may resemble non-cryptographic hashing methods on the surface (*i.e.*, both are "transforms" in a sense), the main difference is that the protected templates generated using feature transformation BTP methods tend to lie in the same (or similar) domain as the original template (*e.g.*, floating-point values), so the same comparison function can often be applied, while hashes tend to be binary and thus usually necessitate the use of a different comparison function to that adopted in the unprotected template domain (*e.g.*, Hamming distance). Moreover, hashes have fixed length, while the size of the protected templates generated by feature transformations usually depends on factors like the original template size and certain transform parameters.

PolyProtect, a method designed for mobile face verification [12], falls into this category. The method was developed considering a full disclosure threat model, it shows no recognition accuracy degradation when parameters are properly tuned, and the code is publicly available[3] under a GPL-3.0 license, making the results easily reproducible. Due to its compatibility with our project requirements, we focused our attention on this method. Additionally, we report that most approaches proposed in the literature have not been comprehensively evaluated in terms of their ability to simultaneously satisfy all three properties of BTP methods, in contrast to PolyProtect, which has been evaluated *theoretically* and *empirically* for irreversibility, and *empirically* for unlinkability.

V. PolyProtect

In this section, the adopted feature-transformation BTP method, PolyProtect, is presented. Let $V = [v_1, v_2, ..., v_n]$ be an *n*-dimensional, real-number embedding extracted by a NN. The aim of PolyProtect is to map V to another real-number feature vector, $P = [p_1, p_2, ..., p_k]$ (where $k < n$), which is the protected version of V. This is achieved by mapping sets of m (where $m << n$) consecutive elements from V to single elements in P via multivariate polynomials defined by a set of m subject-specific, ordered, unique, non-zero integer coefficients, $C = [c_1, c_2, ..., c_m]$, and exponents, $E = [e_1, e_2, ..., e_m]$. From the perspective of security, the C and E parameters represent secret information. Consequently, they should be securely stored.

The first m consecutive elements of V (*i.e.*, $v_1, v_2, ..., v_m$) are mapped to the first element in P (*i.e.*, p_1) via Eq. 1:

$$p_1 = c_1 v_1^{e_1} + c_2 v_2^{e_2} + ... + c_m v_m^{e_m} \tag{1}$$

The elements of V used to generate p_2 depend on the value of overlap o between successive sets of elements. The minimum overlap is 0, in which case the elements of V in each set would be unique. The maximum overlap is $m-1$, in which case successive element sets would share $m-1$ elements. Eq. 2 defines the mapping from V to p_2 for overlap o:

$$p_2 = c_1 v_{m-o+1}^{e_1} + c_2 v_{m-o+2}^{e_2} + ... + c_m v_{m-o+m}^{e_m} \tag{2}$$

The remaining elements in P (*i.e.*, $p_3, ..., p_k$) are generated in a similar way, until all the elements in V have been used. If the last set in V is incomplete because the dimensionality of V is not divisible by the required number of element sets (defined by m and o), V is padded by a sufficient number of zeros to complete the last set.

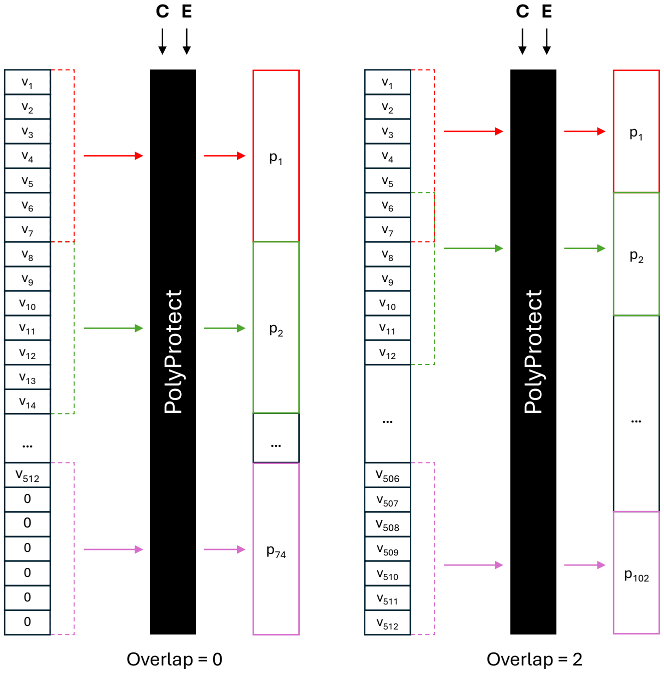


Fig. 2: A graphical representation of PolyProtect is provided above. As an example, we consider overlap $o = \{0, 2\}$ for $m = 7$. The dimensionality of the protected template P varies according to both m and o.

VI. Experimental Protocol

PolyProtect acts as an additional layer on top of a traditional NN-based biometric recognition system. Consequently, it is first necessary to establish the baseline performance of the corresponding unprotected system, to which the protected system will then be compared. To achieve this, we considered a state-of-the-art efficient face recognition model specifically designed for edge devices called EdgeFace [13]. The authors of EdgeFace made available pretrained versions of the model, which we adopt in our work without any fine-tuning. In particular, we consider the *XS* version, which consists of only 1.77M parameters (6.83MB). EdgeFace is based on a hybrid network architecture that leverages convolutional NN and Vision Transformer (ViT) capabilities [34]. It produces 512-dimensional output embeddings, which are compared to each other using the cosine distance.

For fingerprints, preliminary experiments showed the unsuitability of traditional minutiae-based templates for PolyProtect, essentially due to: *(i)* the variable length of minutiae-based template representations, and their inherent two-dimensional nature, as each minutia point typically consists of the coordinates on the screen and the feature type; *(ii)* the lower level of abstraction of minutiae-based templates, in which each minutia

[3] https://gitlab.idiap.ch/bob/bob.paper.polyprotect_2021/

point directly corresponds to a feature detected by the sensor (in contrast, NN-based embeddings are produced by a network trained to map identities to a high-dimensional space); *(iii)* the higher complexity of the minutiae-based matching algorithms in comparison with cosine distance (which is typically used for fixed-side NN-based templates). Consequently, despite the greater availability of minutiae-based fingerprint systems, we opted for a pretrained deep learning-based feature extractor[4] [15]. The model replicates the DeepPrint architecture [35], consisting of two main branches: one dedicated to the fingerprint texture representation, and the other designated to learning from the minutiae maps. The final 512-dimensional embedding consists of the concatenation of the outputs of the two branches, *i.e.*, the first half of the embedding encodes fingerprint texture information, and the second half of the embedding encodes minutiae information. The model was trained on a mixture of synthetic and real fingerprint images.

As described in Sec. V, implementing PolyProtect requires setting three values: m, the interval size to compute each element of the protected template P (*i.e.*, the overlap parameter, o), and the ranges of C (coefficients of the polynomial) and E (exponents of the polynomial). The rationale behind the selection of each of these values is explained below.

The choice of the value of m is guided by the Abel-Ruffini theorem, which states that there is no closed form algebraic expression for solving polynomials of degree 5 or higher with arbitrary coefficients [36]. At the same time, it is necessary to keep in mind that m corresponds to the length of the secret sequences of coefficients C and exponents E. In particular, the E values are selected as random permutations of integers from 1 to m. Therefore, to avoid having two enrolled subjects with the same set of exponents, an additional lower limit for the number of E values (and consequently for m) is set by the number of possible permutations without repetition with a sample size of m, which would correspond to the maximum number of enrolled subjects for that specific value of m. Consequently, for $m = 5$ we would have 120 subjects, for $m = 6$ we would have 720, for $m = 7$ we would have 5040, etc. Accounting for the number of subjects in typical projects, we selected $m = 7$. In turn, the upper limit of m is bound by the fact that the embeddings produced by the NN-based feature extractors used in this work consist of floating point values smaller than 1, which would cause large powers to effectively obliterate certain embedding elements during the PolyProtect mapping. In the original PolyProtect paper [12], the authors adopt the `mobile0-male` protocol of the Mobio face dataset [37], consisting of fewer than 100 subjects. Consequently, the authors used the minimum acceptable value of $m = 5$. Concerning the range of the coefficients C, the value adopted in the original paper was arbitrarily set to $[-50, 50]$. We extend this range to $[-100, 100]$ to compensate for the increase of m from 5 to 7. Given the fixed-size embeddings of dimensionality 512 adopted in this work, Table II shows the sizes of the protected embeddings P, which depend on the overlap value.

TABLE II: Given the dimensionality of the unprotected embedding V, the dimensionality of its protected counterparts P depends on the overlap value.

Overlap	V	P
0	R_{512}	R_{74}
1	R_{512}	R_{86}
2	R_{512}	R_{102}
3	R_{512}	R_{128}
4	R_{512}	R_{170}
5	R_{512}	R_{254}
6	R_{512}	R_{506}

A. Recognition Accuracy

1) Verification: The aim of this analysis is to determine whether the incorporation of PolyProtect into a deep-neural-network-based verification system would degrade its recognition accuracy. To conduct this analysis, we start from a self-collected dataset of 942 subjects (more details in Sec. VI-D). To reproduce the original experiments, the subjects are split into *dev* and *eval* sets (in equal proportions), each subject having a single reference and query sample image. Following the original paper, we consider two scenarios for the recognition accuracy evaluation: *(i)* the *Normal* (N) scenario; and *(ii)* the *Stolen Coefficients and Exponents* (SCE) scenario. The PolyProtected system should operate in the Normal scenario most of the time. Here, we assume that each enrolled user dutifully employs their own C and E parameters in the generation of their PolyProtected templates, as envisioned by the design of the PolyProtect scheme. In the Stolen Coefficients and Exponents scenario, a subject attempts to authenticate as a different one by stealing the target's C and E parameters, and applying them to their own embedding to generate their PolyProtected template. While the SCE scenario should be uncommon in practice, it is still important to consider it as the worst-case scenario.

2) Identification: Due to the nature of the aid distribution operations carried out in field projects, in which a field worker uses the Simprints ID app to attest the identities of multiple beneficiaries, it is often not possible for the beneficiaries to make the identity claim needed for verification, *i.e.* for retrieving a specific reference (enrolled) template to be compared with a query one. Among other reasons, this might happen because of the lack of physical ID documents. Consequently, identification is often the preferred operational mode by design. From the perspective of PolyProtect, we know that it is necessary to use subject-specific secret parameters to transform the input templates. However, given the absence of the identity claim in the identification scenario, it would be impossible to know which subject-specific parameters to use for the transformation. Therefore, the query template must be transformed by PolyProtect considering all sets of transformation parameters (C and E) already registered in the database. That is, if M subjects are enrolled in the system, then M protected templates would be generated from the query one. Each of the M protected query templates is then compared to the corresponding protected reference template, *i.e.*, the one transformed with the same set of C and E parameters

[4] https://github.com/tim-rohwedder/fixed-length-fingerprint-extractors

during enrolment. Finally, a ranking is generated based on the distances, and the highest-scoring matches are returned by the system. Although identification might entail M times the number of transformations required in the verification scenario, given its fast and lightweight nature PolyProtect does not have a significant impact on the total computation time.

B. Irreversibility

In the context of PolyProtect, the full disclosure threat model described in Sec. II-B is reflected as follows: knowledge of the algorithm including the number of embedding elements (m) used to generate each PolyProtected element, the overlap value (o), as well as the subject-specific C and E parameters that define the PolyProtect polynomials. Moreover, we assume that the adversary has access to one or more PolyProtected templates, P, corresponding to a particular embedding, V, as well as knowledge of the distribution of unprotected templates, which is representative of the embeddings used to create the PolyProtected templates to which the adversary has access. The adversary's goal, therefore, is to use all this information to attempt to recover a subject's original embedding, V, from one or more of their PolyProtected templates, P. Following the original approach in [12], our goal is to recover the evaluation set reference embeddings, which the adversary does not have access to. In contrast, we assume that the adversary has access to the development set, which is used to estimate the distribution of each one of the 512 values in the evaluation set reference embeddings as well as the match threshold.

The irreversibility of a BTP method can be demonstrated theoretically or empirically. The attempt to reconstruct the unprotected template from a single protected template would correspond to solving an *underdetermined* system of k equations in 512 variables (unknowns), leading to $512 - k$ degrees of freedom, which in turn leads to the conclusion that PolyProtect is *theoretically* irreversible. However, a numerical solver[5] can be used to converge to an approximate solution for V from a set of initial guesses. Following the attack defined in the original paper, the employed numerical solver starts from the guesses obtained from the development set to estimate a solution for each of the 512 elements in each V in the evaluation set, from the corresponding P. As soon as the solver indicates that a solution for V has been found, the process is stopped. Since the solver operations can be quite time-consuming, a time-out of 20 minutes for a single template inversion is set, typically going off in fewer than 5% of the cases.

The irreversibility analysis can be extended to consider an Attack via Record Multiplicity (ARM): the adversary has access to multiple PolyProtected templates from the same V, which they attempt to combine to recover an approximation of V. This type of attack could occur in the scenario where the same embedding is used to generate different PolyProtected templates (using different C and E parameters), then each PolyProtected template is either enrolled in a different application or used to replace a compromised PolyProtected template in the same application. In our ARM analysis, we consider the worst-case scenario where the same V is used to generate all of a subject's P templates. Therefore, V is associated with 10 different Ps (each generated using different C and E parameters), attempting to recover an approximation of V using 2 to 10 of its corresponding Ps. This was simulated using the numerical solver approach explained above, but considering $k \times p$ equations (where k is the dimensionality of the P templates, and p is the number of Ps that the adversary is assumed to have access to), instead of only k equations, making the system of equations not underdetermined but often overdetermined. We invite the reader to consult [12] for more details about the definition of the system equations, which we omit for brevity.

C. Unlinkability

Starting from the same unprotected template V, a BTP method is said to benefit from the property of unlinkability if it is possible to produce sufficiently different protected template instances P and P' such that they are not linkable to each other. In practice, this property allows us to renew a compromised protected template, as well as to generate different protected templates for different applications without the risk of cross-matching the underlying identity. In the case of PolyProtect, this implies using different C and E parameters in the $V \rightarrow P$ mappings.

The unlinkability of PolyProtect is evaluated using the framework proposed in [3]. This framework is based on mated and non-mated score distributions, which represent the comparison scores between different protected templates from the same subject and between different protected templates from different subjects, respectively. The unlinkability is measured in terms of two metrics: $D_{\leftrightarrow}(s)$, a local score-wise measure of the degree of linkability based on the likelihood ratio between mated and non-mated scores, and $D_{\leftrightarrow}^{sys}$, a global measure of the overall linkability of the underlying recognition system.

Following the recommendation in [3], we compute 10 different PolyProtected templates per person[6]. Then, each PolyProtected template is compared to every other PolyProtected template from the same subject to generate a set of mated comparison scores, and to all PolyProtected templates from every other subject to generate a set of non-mated comparison scores. We also calculate the unlinkability of the corresponding unprotected embeddings in the same way. This process is repeated for 10 trials. The resulting 10 sets of mated and non-mated comparison scores are then concatenated (separately), and the concatenated scores are used to evaluate the unlinkability of the PolyProtected templates.

D. Datasets

The datasets considered in the experimental part of this work have either been collected by Simprints in field projects

[5] The numerical solver used is Python's *scipy.optimize.root* function with the *lm* method. Link: https://docs.scipy.org/doc/scipy/reference/optimize.root-lm.html

[6] In contrast to the original PolyProtect paper, due to restrictions in the used datasets, we derived the 10 different PolyProtected templates from exactly the same image, not 10 different images belonging to the same subject.

or are publicly available synthetic datasets with a commercially non-restrictive license.

For face biometrics, we employ a dataset of face images obtained within the framework of a project currently being carried out by Simprints in Ethiopia. It consists of 942 subjects with 2 captures per subject: 57% of the subjects are females, and 43% are males, while the mean age is approximately 24.5 years ($\sigma = 16.5$). All subjects have East African origins. We also include an evaluation on the Vec2Face[7] synthetic dataset [14]. The selected subset consists of the first 1000 subjects. Each subject's data consists of two samples, randomly selected from each dataset class folder.

For fingerprint biometrics, we adopt an internal dataset collected in a field project in Ghana, consisting of 119 subjects with two samples each. In addition, we consider a subset of the SynFing dataset[8], based on the StyleGan2 architecture [16]. As in the case of synthetic face datasets, we considered the first 1000 subjects, whose data consists of two samples, one used for enrolment and one for query.

VII. Experimental Results

A. Recognition Performance

1) Verification:

a) Face biometrics: The verification performance is measured in terms of True Match Rate (TMR) at a certain False Match Rate (FMR), for which we consider thresholds corresponding to FMR=0.01% and 0.1%, as well as the Equal Error Rate (EER). Table III includes the results for face biometrics on the internal dataset (on the left) and on the Vec2Face dataset [14] (on the right). The first row of the table features the baseline performance of the unprotected systems. Then, the remaining rows are split in two halves, respectively for the N and SCE scenarios. Finally, in each individual row, a different overlap value is considered.

TABLE III: Face verification results. In bold, the results achieved by PolyProtect which improve the baseline performance.

o	TMR (%) @ FMR = 0.01%	TMR (%) @ FMR = 0.1%	EER (%)	TMR (%) @ FMR = 0.1%	EER (%)
	Internal Dataset			Vec2Face [14]	
Bas.	94.27	95.97	1.35	71.00	6.41
N Scenario					
0	93.40	**97.26**	**0.83**	**77.28**	8.76
1	**94.67**	**97.24**	**0.72**	**80.80**	7.88
2	**94.78**	**98.07**	**0.57**	**83.30**	6.86
3	**94.97**	**97.71**	**0.65**	**86.80**	**6.10**
4	**95.65**	**98.51**	**0.62**	**90.14**	**4.88**
5	**95.97**	**98.54**	**0.55**	**92.32**	**3.68**
6	**96.39**	**98.98**	**0.45**	**94.90**	**2.26**
SCE Scenario					
0	87.54	92.80	2.25	64.26	11.54
1	89.38	93.95	2.21	68.06	11.08
2	90.74	93.99	1.74	70.08	10.12
3	90.28	94.59	1.91	**73.32**	9.42
4	91.95	95.05	1.51	**77.42**	8.38
5	93.36	95.76	1.49	**79.80**	7.80
6	94.20	**96.05**	**1.26**	**83.28**	6.90

Firstly, we can observe that for almost any overlap value, PolyProtect improves the verification performance with respect to the baseline performance (except for TMR at FMR=0.01%, with an overlap of 0 for the internal dataset, and with overlaps of 0, 1, 2 for Vec2Face [14]). This might be due to the fact that by combining subject-specific information (C and E parameters), in the protected space embeddings belonging to the same subject are pushed together, while embeddings belonging to different subjects are moved away from each other. Consequently, due to the transformation performed by PolyProtect, the *intra-user* variability is reduced, and the *inter-user* distances increase. On top of this, it is clear that by increasing the overlap value (down the rows), the performance improves. As noted in Table II, PolyProtect always reduces the dimensionality of the input template. In particular, the higher the overlap value, the higher the number of dimensions of the protected space (from 512 values of the input unprotected template, for an overlap of 6 the output template will have a dimensionality of 506, while for an overlap of 0 the dimensionality of the output template will be 74). So, it makes sense that the use of larger overlap values, which generate PolyProtected templates of higher dimensionality, will result in higher recognition accuracy.

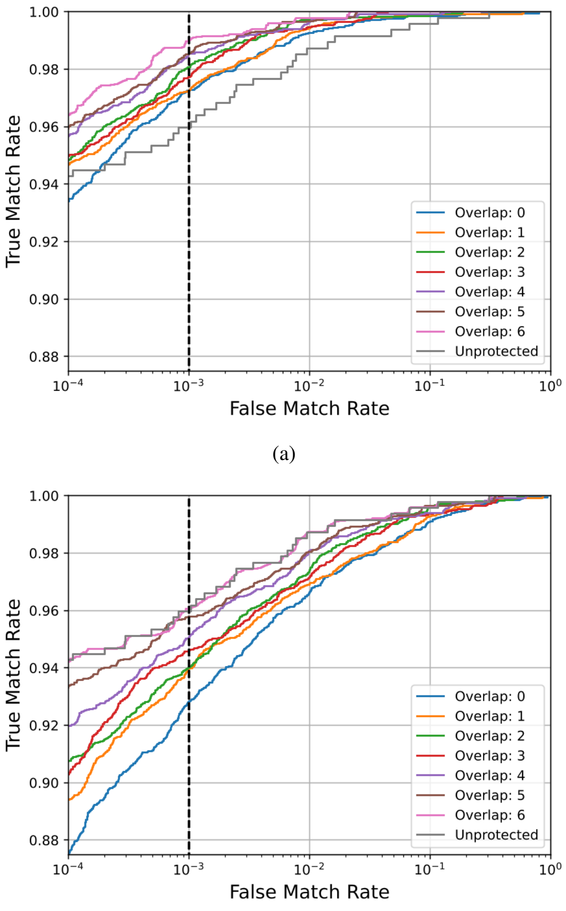
The bottom part of Table III contains the results in the SCE scenario. A reduction of the biometric performance is expected [12], as embeddings transformed with the same secret parameters are being compared. Indeed, in contrast to the N scenario, the baseline performance is, in almost all cases, the best one (except for TMR at FMR=0.1% and EER, with an overlap of 6 for the internal dataset, and for TMR at FMR = 1% with an overlap of 3, 4, 5, 6 for Vec2Face [14]). Moreover, similarly to the N scenario, the performance improves with an increase in the overlap value, thus limiting the accuracy degradation.

Fig. 3 shows the Receiver Operating Characteristic (ROC) curves for face biometrics (the N scenario is on the top, the SCE on the bottom). The black dashed line corresponds to the operating point at 0.1% FMR, *i.e.*, the central column for each dataset in Table III. In Fig. 3(a), we can observe how the grey curve, corresponding to the baseline performance, is below the other curves, which correspond to the systems implementing PolyProtect. In contrast, in Fig. 3(b), the same grey curve appears to be overlapping several times with the pink curve, corresponding to the highest of the PolyProtect curves (overlap equal to 6).

b) Fingerprint biometrics: Table IV contains the verification results for fingerprint biometrics on the internal dataset (on the left) and on the SynFing dataset [16] (on the right). In this case, we observe a baseline performance of respectively 40.21% and 18.81% EER. As mentioned in Sec. VI, we employ pretrained models without any fine tuning. Due to the worse model performance (compared to the face recognition system), in this case we omit the results at stricter operating points such as FMR = 0.1%. With a lower baseline recognition accuracy, in the N scenario the impact of PolyProtect seems

[7] https://github.com/HaiyuWu/Vec2Face

[8] https://github.com/rafaelbou/fingerprint-generator



(a)

(b)

Fig. 3: ROC curves for face biometrics (internal dataset): (a) N scenario, (b) SCE scenario.

to be much greater, reducing the EER to 7.13% (internal dataset) and 3.19% (SynFing [16]) for an overlap of 0. In field operations, where, due to harsh operational conditions, baseline performance levels typically tend to be lower than on datasets assembled in laboratory settings, this kind of contribution could be of great added value. Moreover, interestingly, we notice that for both datasets increasing the overlap value (down the rows) does not yield a clear improvement trend as in the face experiment. This might be due to the fact that the biometric information retained in the protected template is not as discriminative as in the case of face. Consequently, the more compact protected templates obtained with lower overlap values might not lose as much meaningful information as a result of the PolyProtect transformation. An additional possible explanation could be that in the fingerprint embeddings the texture and minutiae information occupy distinct parts of the embedding (*i.e.*, they are computed separately by the two branches of the feature extractor, and are then concatenated [15]). As a result, these features are less “spread out” in the embedding space compared to the face features, which are produced by a single-branched network, EdgeFace. Therefore, increasing the dimensionality of the protected template does not necessarily have a beneficial impact on inter-class distances.

TABLE IV: Fingerprint verification results. In bold, the results achieved by PolyProtect which improve the baseline performance.

	EER (%)			
	Internal Dataset		SynFing [16]	
o	N Scenario	SCE Scenario	N Scenario	SCE Scenario
Bas.	40.21		18.81	
0	**7.13**	40.83	**3.19**	20.80
1	**6.82**	40.68	**3.75**	21.57
2	**6.69**	41.73	**3.23**	20.43
3	**6.51**	40.35	**3.62**	19.63
4	**6.19**	40.64	**3.42**	20.61
5	**6.30**	**39.95**	**3.39**	19.80
6	**6.19**	40.34	**3.39**	19.38

Additionally, we observe that, as in the case of face biometrics, the EER values obtained in the SCE scenario are almost always higher than in the baseline system (the only exception is represented by the case of an overlap of 5 on the internal dataset). The performance discrepancy between the two PolyProtect scenarios (*i.e.*, N and SCE) seems to confirm the role played by the subject-specific secret information (C and E parameters) towards producing a more discriminative mapping in the protected domain.

2) Identification: As detailed in Sec. VI-A2, in the identification scenario, given the absence of an identity claim, the query template must be transformed by PolyProtect considering all sets of transformation parameters (C and E) registered in the database. Each of the M protected query templates is then compared to the corresponding protected reference template, *i.e.*, the one transformed with the same set of C and E parameters during enrolment. Given this aspect, identification is inevitably more difficult than verification. In fact, the comparisons always take place between pairs of protected templates that undergo exactly the same transformation. In other words, it is not possible to exploit the discriminative power of incorporating different, subject-specific information. From this perspective, the comparisons are comparable to those carried out in the SCE verification scenario.

Such considerations are reflected in the results obtained in terms of True Positive Identification Rate (TPIR), as presented in Table V. Identification systems are typically configured to return a fixed number n of the most similar candidates. Consequently, the TPIR-n represents the percentage of identification attempts where the query subject is included in the ranked list of the n most similar candidates returned after searching a biometric reference database (we consider $n = 1, 3, 10$). The table is split into two parts: the top one containing the results on the face datasets, the bottom one containing the results on the fingerprint datasets. The results are presented for different overlap values (down the rows) and rank values (across the columns).

For face, we can see that in all cases the baseline performance is better. However, by increasing the overlap value, and therefore the dimensionality of the output (protected) space, we observe an overall improvement in the identification rates. In field projects, it is generally preferred to use TPIR-3 and TPIR-10 over TPIR-1, so that the local staff can manually scrutinise the ranked 3- or 10-item shortlist returned by the biometric system to select the final best match. In the case of the internal face dataset, for TPIR-3 and TPIR-10, Table V shows that for intermediate overlap values such as 2 or 3 the decrease in identification accuracy when PolyProtect is employed, is approximately 1% in absolute terms, which might be considered acceptable. For the more stringent TPIR-1, the corresponding decrease would be greater (approximately 2.5%). That is, the recognition performance degradation is milder in the most frequently used project field settings. Concerning the Vec2Face dataset [14], we can observe that the identification rates (especially the TPIR-1) are more significantly affected when setting low overlap values, which correspond to lower dimensionality of the protected templates. For fingerprint, the internal dataset proves to be very challenging in the identification task as well, but the negative impact of PolyProtect on the performance seems limited in this case. As for SynFing [16], starting from an overlap value of 4, we observe higher TPIR values than for the baseline system. As observed in the verification experiment, the better results achieved for the PolyProtected fingerprint system in the case of identification might be due to the worse baseline performance.

B. Irreversibility

The results of the attempts to reconstruct an unprotected template from a single protected template are presented in Table VI. The resulting inversion success rates are reported for both face and fingerprint modalities, for different overlap and threshold values. In terms of metrics, the match rate is computed at the two thresholds established on the baseline systems' development set of embeddings ($FMR = 0.01\%, 0.1\%$). Then, the Inversion Success Rate (ISR) is computed as the solution rate × match rate, like in [12]. The experiments carried out on the face dataset are on the left, while the ones on the fingerprint dataset are on the right. Due to the different performance achieved in terms of recognition accuracy, for each of the systems we consider different thresholds, *i.e.*, FMR = 0.01%, 0.1% for face, and FMR = 1%, 10% for fingerprint.

TABLE V: Identification results obtained for both modalities. In bold, the results achieved by PolyProtect which improve the baseline performance.

	TPIR-*n* (%), *n*					
o	1	3	10	1	3	10
Face						
	Internal Dataset			Vec2Face [14]		
Bas.	98.09	98.51	99.36	72.20	81.20	88.40
0	93.74	96.16	97.92	45.00	57.38	69.20
1	95.10	96.62	98.03	48.48	60.58	72.84
2	95.78	97.37	98.66	52.00	63.50	75.88
3	95.78	97.37	98.41	55.30	66.84	77.84
4	96.47	97.86	98.88	59.74	70.96	81.60
5	97.39	97.96	98.85	63.92	74.44	83.90
6	97.41	98.47	99.19	69.02	77.92	86.12
Fingerprint						
	Internal Dataset			SynFing [16]		
Bas.	11.67	23.33	41.67	31.80	46.00	61.00
0	8.00	19.83	37.83	24.42	37.52	51.98
1	8.50	18.67	38.33	27.10	41.20	57.32
2	7.50	18.33	39.17	29.84	43.18	58.80
3	8.33	18.83	41.00	31.22	45.66	60.96
4	9.50	18.17	39.33	31.38	45.78	**62.40**
5	8.67	21.33	41.17	**34.86**	**50.70**	**65.96**
6	10.17	19.67	41.00	**36.48**	**53.92**	**67.82**

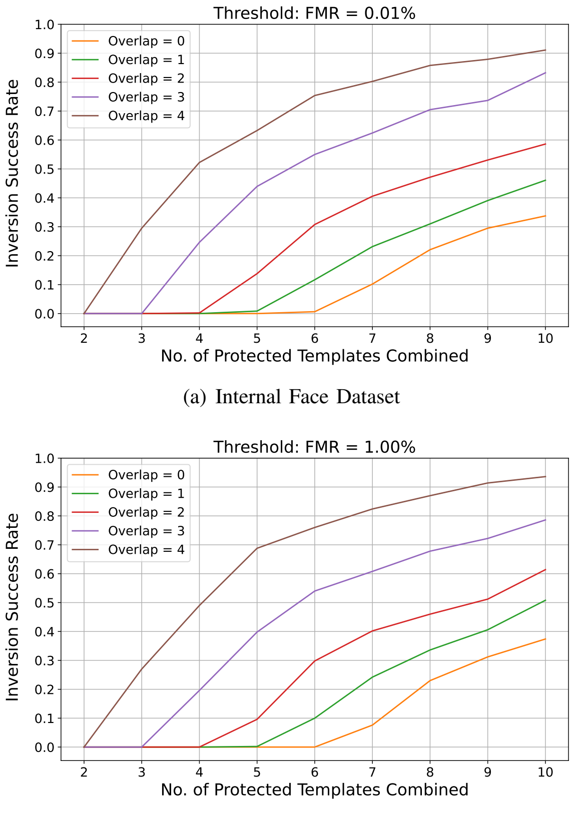
TABLE VI: Irreversibility results in terms of Inversion Success Rate (ISR) for both modalities.

	Face			Fingerprint		
o	Thres.	ISR (%)		Thres.	ISR (%)	
		Internal Dataset	Vec2Face [14]		Internal Dataset	SynFing [16]
0	FMR = 0.01%	0	0	FMR = 1%	0	0
1		0	0.14		0	0
2		0	0.64		0	0
3		0	0.70		0	0
4		0	1.80		0	0
5		0.83	3.26		9.83	8.60
6		98.20	6.08		92.33	90.60
0	FMR = 0.1%	0	0	FMR = 10%	1.00	1.00
1		0	0.44		1.50	0.80
2		0	1.14		4.50	2.80
3		0	1.44		12.00	7.80
4		0.02	2.66		37.33	24.00
5		34.59	3.94		79.83	72.40
6		98.20	6.12		92.33	90.60

It is evident that the inversion success rate is, in general, lower when the baseline systems operate at a stricter match threshold (at a lower FMR). This can be observed in all cases (both modalities, and both thresholds). This is because a stricter threshold would require better approximation of the original input template. In other words, the stricter the threshold in practice, the less likely the inversion attack would be to succeed. This is especially noticeable for the fingerprint system operating with a threshold corresponding to the 10% FMR point (bottom right corner). As we can see, even with an overlap of 0, we observe a successful reconstruction in 1.0% of the cases, whereas for the threshold at 1% FMR a non-zero ISR is observed for a minimum overlap of 5.

Another interesting observation is that, as the overlap value increases, the ISR increases. For the internal face dataset, it takes at least an overlap of 5 at the stricter threshold to observe 0.83% of successful reconstructions. Then, the ISR reaches almost 100% for an overlap of 6. With the more lenient threshold, we observe a 34.59% ISR with an overlap of 5. This trend (*i.e.*, increasing ISR as the amount of overlap increases) is due to the number of equations in the underdetermined system of equations assembled for attempting the reconstruction starting from a single template, *i.e.*, the greater the overlap, the greater the number of equations, and the more constrained the system becomes, so it becomes easier to solve for V. (Sec. VI-B). The number of equations for each overlap value is reported in the 'P' column of Table II, whereas the number of unknowns is always 512. It is interesting to observe that in the case of the Vec2Face dataset [14], although slightly increasing together with the overlap value, the ISR reaches 6.12% as its maximal value for the the more lenient threshold, showing that inversion attacks are less successful on this dataset. Additionally, for three of out the four datasets considered, the maximum number of successful reconstructions reaches the same maximum value at the different thresholds: 98.20% for the internal face dataset, 92.33% for the internal fingerprint datasets, and 90.60% for SynFing [16]. This might be due to the fact that some protected templates are harder for the numerical solver to invert (*i.e.*, solve for V) regardless of the threshold, or the 20-min timer set for each template reconstruction did not allow the numerical solver to converge.

Fig. 4 graphically displays the results of the ARM experiment on the internal dataset for face and on the SynFing dataset [16] for fingerprints, which represent the two datasets that yield the best recognition performance for each modality. In this case, attempts were made to reconstruct an unprotected template by combining a variable number of protected templates. Once again, the inversion success rates reported for face (Fig. 4(a)) and fingerprint (Fig. 4(b)) systems are computed using different thresholds ($FMR = 0.01\%$ for face, and $FMR = 1\%$ for fingerprint). The limitation of our ARM analysis to an overlap of 4 is due to the fact that we observe non-zero ISR values for an overlap of 5 considering the inversion of a single template.



(a) Internal Face Dataset

(b) SynFing Fingerprint Dataset [16]

Fig. 4: Attack via Record Multiplicity (ARM) on the (a) face and (b) fingerprint protected systems. For each modality, we considered the dataset on which the best recognition performance was achieved.

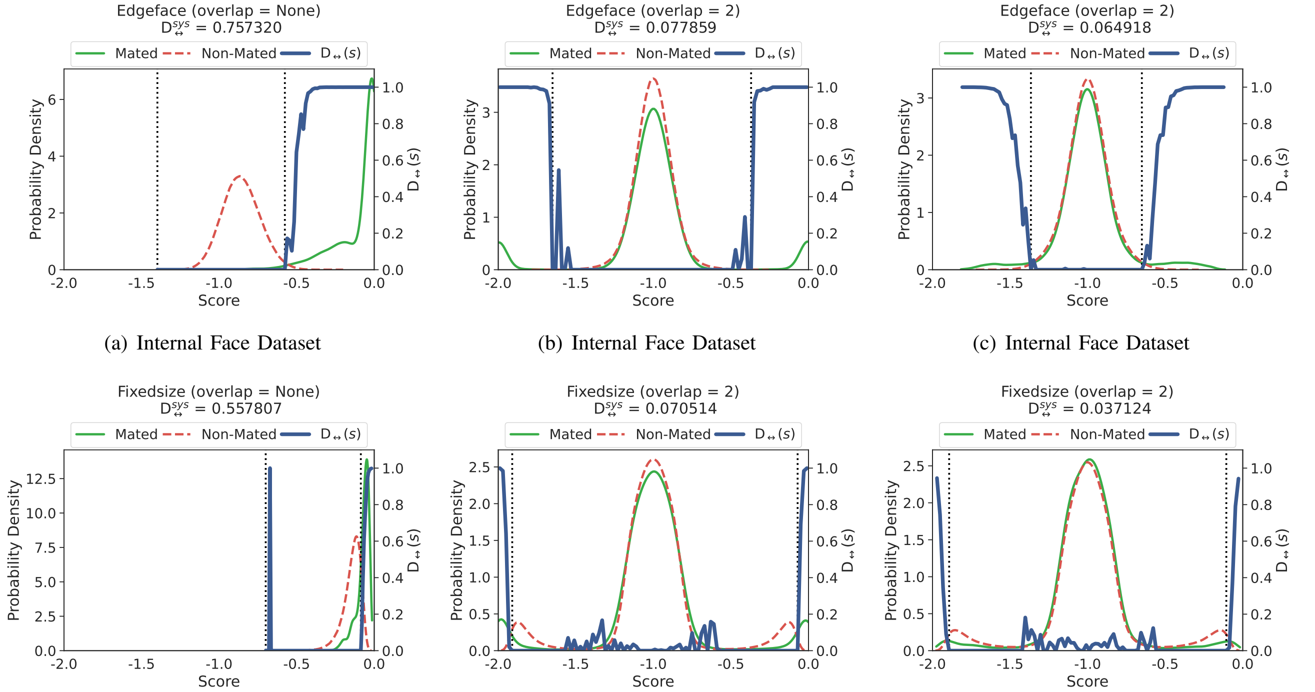
The protected templates that were used, were all generated from the same embedding but using different C and E parameters. The number of equations in the system to be solved by the numerical solver consists of the number of protected templates that are combined × the dimensionality of each protected template. In turn, the number of unknowns would be equal to the dimensionality of each unprotected template. Overall, we can observe very similar trends across the two graphs: for a given overlap value (each curve individually), as the number of combined protected templates increases, the chances of a successful reconstruction are higher. In turn, for higher overlap values, we observe higher inversion success rates when fewer protected templates are combined. For instance, for an overlap of 0, it takes a combination of 7 protected templates to start noticing a non-zero inversion success rate, whereas for an overlap of 4 it takes a combination of 3 protected templates.

C. Unlinkability

The unlinkability property is evaluated using the framework proposed in [3]. This method measures unlinkability in the context of the mated and non-mated score distributions, which represent the comparison scores between different protected templates from the same subject and between different protected templates from different subjects, respectively. The unlinkability is measured in terms of two metrics: $D_{\leftrightarrow}(s)$, a local score-wise measure of the degree of linkability based on the likelihood ratio between mated and non-mated scores, and $D_{\leftrightarrow}^{sys}$, a global measure of the overall linkability of the underlying recognition system.

Fig. 5 shows the unlinkability plots, considering for each modality the dataset on which we observed the best recognition performance, *i.e.*, the internal dataset for face (top row) and SynFing [16] for fingerprint (bottom row). The graphs on the left represent the baseline (unprotected) scenario, the central and right ones respectively correspond to the naive and strict parameter selection in the PolyProtected scenario (as per the analysis in [12]). Due to space restrictions, for the PolyProtect systems we show only the plots for overlap = 2, since this appears to be one of the optimal values for achieving an effective recognition performance *vs.* irreversibility trade-off (together with an overlap of 3). Each graph contains three curves: the mated (green, solid) and non-mated (red, dashed) distributions, and $D_{\leftrightarrow}(s)$ curves. Ideally, for full unlinkability, the mated and non-mated distributions should overlap as much as possible.

Due to the good recognition performance of the employed FR system, in the face baseline performance graph (top left corner) the overlap of the mated and non-mated distributions is limited, causing a relatively high $D_{\leftrightarrow}^{sys}$ value (0.757). Moreover, past a score value of approximately -0.5 (on the horizontal axis), we can observe that $D_{\leftrightarrow}(s)$ rapidly reaches a degree of linkability close to 1. On the other hand, considering the fingerprint baseline performance (bottom left corner), we see that the overlap between the mated and non-mated distributions is greater. This implies that the unprotected



(a) Internal Face Dataset

(b) Internal Face Dataset

(c) Internal Face Dataset

(d) SynFing Fingerprint Dataset [16]

(e) SynFing Fingerprint Dataset [16]

(f) SynFing Fingerprint Dataset [16]

Fig. 5: Unlinkability plots. Face: (a) baseline system; (b) PolyProtect with naive parameter selection; (c) PolyProtect with strict parameter selection. Fingerprint: (d) baseline system; (e) PolyProtect with naive parameter selection; (f) PolyProtect with strict parameter selection. For each modality, we considered the dataset on which the best recognition performance was achieved.

fingerprint templates are less linkable compared to the unprotected face templates, which is also evidenced by the lower $D^{sys}_{\leftrightarrow}$ value of the baseline fingerprint recognition system. In contrast, in the central column (naive PolyProtect parameter selection) for both modalities the overlap of the mated and non-mated distributions is much more significant, showing that the protected templates benefit from unlinkability. Moreover, the $D_{\leftrightarrow}(s)$ curve shows that for the central part of the score range, the degree of linkability is very close to zero for face (top graph), and quite limited for fingerprint (bottom graph).

Compared to the naive parameter selection, the strict one (right column) involves an additional comparison check: sets of C and E parameters are selected only if they are capable of producing a protected template P such that the comparison scores with all the other protected templates originating from the same unprotected template V, obtained with the other assigned sets of C and E parameters, are within a required score range. Otherwise, a new set of parameters is randomly generated until the aforementioned condition is satisfied. The idea behind this strict process of selecting the C and E parameters is to ensure that different protected templates generated from the same face embedding would be unlinkable, *i.e.*, comparison of these templates would generate scores in the “unlinkable” score range.

Table VII summarises the global $D^{sys}_{\leftrightarrow}$ measures for all overlaps. Concerning the metrics adopted, $D^{sys}_{\leftrightarrow}$ measures the overall system linkability, where a value of 0 would indicate that the system is fully unlinkable, whereas a value of 1 would indicate that the system is fully linkable [3]. We observe that $D^{sys}_{\leftrightarrow}$ for our baseline systems, which use unprotected face embeddings, is closer to 1 (or at least significantly further from 0 compared to the protected systems), *i.e.*, 0.757 for the internal face dataset, 0.911 on Vec2Face [14], 0.277 for the internal fingerprint dataset, and 0.558 for SynFing [16] (Table VII), indicating that unprotected face embeddings from the same subject (*e.g.*, used across different applications) are linkable to a considerable extent. In particular, the higher $D^{sys}_{\leftrightarrow}$ value for the face recognition system might be due to the fact that the underlying feature extractor model performs better than the fingerprint model. Moreover, in the case of the internal

TABLE VII: Unlinkability results obtained for both modalities, considering the naive (N) and strict (S) PolyProtect parameter (C and E) selection.

	$D^{sys}_{\leftrightarrow}$							
	Face				Fingerprint			
	Internal		Vec2Face [14]		Internal		SynFing [16]	
o	N	S	N	S	N	S	N	S
Bas.	0.757		0.911		0.277		0.558	
0	0.077	0.062	0.079	0.047	0.092	0.022	0.067	0.032
1	0.078	0.063	0.079	0.045	0.089	0.027	0.072	0.037
2	0.078	0.065	0.079	0.048	0.095	0.022	0.071	0.037
3	0.078	0.065	0.079	0.048	0.089	0.033	0.071	0.037
4	0.078	0.069	0.079	0.050	0.097	0.028	0.067	0.037
5	0.078	0.072	0.079	0.052	0.089	0.034	0.074	0.036
6	0.078	0.077	0.080	0.060	0.099	0.039	0.075	0.037

fingerprint dataset, the low $D_{\leftrightarrow}^{sys}$ value can be explained by the worse recognition performance achieved by the employed fingerprint model on this challenging dataset.

In contrast, the $D_{\leftrightarrow}^{sys}$ values for our protected systems are reduced by a factor of 10, and are close to 0, with the naive parameter selection, suggesting that different protected templates generated from the same subject's face or fingerprint embedding are almost fully unlinkable. Although the differences between the different overlap values do not appear to be significant, we observe that for an overlap of 0, we obtain in all cases the least linkable protected templates.

If we consider the strict PolyProtect parameter selection, we can see that the $D_{\leftrightarrow}^{sys}$ values are further reduced (graphically represented in Fig. 5). In particular, the reduction is more significant in the case of the fingerprint datasets. As pointed out in [12], this phenomenon is related to the two bumps in the mated score distributions of the protected systems' score range when the C and E parameters are naively generated. The authors suggest that this may be due to the relationship between the signs of the corresponding elements among each pair of protected templates that is compared through the cosine distance to generate those scores. Therefore, the system does not achieve equal unlinkability across the entire score range. However, by selecting the C and E parameters in a stricter way such that the mated scores would be forced to lie within the "unlinkable" score range, it seems possible to remove the mated score distribution bumps at the extreme ends of the score range.

VIII. Conclusions

This article illustrated the main outcomes of our analysis of existing Biometric Template Protection (BTP) solutions, which may be suitable for the project "Safe Biometrics for Humanitarian Aid", a partnership between Simprints and ICRC. The goal of the project is to build the first biometric open-source product implementing a BTP solution that fulfils the strict security requirements for humanitarian use-cases.

After motivating the adoption of the full disclosure threat model (as specified in the ISO/IEC 30316:2018 standard on performance testing of biometric template protection schemes [10]), we laid out our system requirements, which were gathered throughout several conversations with the ICRC's Data Protection Office, and from the ICRC's documents on data protection and biometrics policy [6], [7].

Then, we surveyed the existing BTP literature in order to identify a suitable solution for our requirements. To achieve this, we considered the taxonomy proposed in [17], and ruled out categories that did not meet our needs: image-level BTP methods and NN-based methods. Consequently, our focus narrowed to handcrafted, feature-level BTP methods, which can be further divided into three groups: homomorphic encryption (HE)-based approaches, hashing, and feature-transformation approaches. In light of our requirements, we identified the latter as the most suitable method category. In particular, PolyProtect [12], a BTP method proposed for mobile face verification, was selected for the continuation of the project. A detailed explanation of the reasons that motivated our choice can be found in Sec. III. In summary, the main advantages offered by PolyProtect include its extremely lightweight computational burden; its property of using the same mathematical operation for matching in the unprotected and protected space (*i.e.*, cosine distance for both unprotected and protected templates), which allows it to operate transparently as an *add-on* security layer on top of existing feature extractor models; its comprehensive and publicly available experimental evaluation[9], and the GPL-3.0 license under which it is made available.

We then proceeded to evaluate PolyProtect according to the experimental protocol proposed in [12]. Overall, the trends observed in our experimental results are comparable to those in the original paper in terms of recognition accuracy, irreversibility and unlinkability. From this perspective, the main novel aspects of this work can be summarised as follows. Firstly, in comparison with the original paper [12], we considered a more recent feature extractor, EdgeFace, which was designed for efficient face recognition, satisfying the typical mobile environment constraints (the *XS* version considered requires only 6.83MB for storage) [13]. Similarly to the latest generation face feature extractors, EdgeFace produces output templates with dimensionality of 512, in contrast to the 128-value templates of the two feature extractors adopted in the original study. Secondly, we configured our system to deal with a higher number of enrolled subjects (hence $m = 7$) and tested it on a challenging internal dataset collected in a field project in Ethiopia and on the Vec2Face [14] synthetic face dataset, with very promising results. Thirdly, we extended the evaluation of the recognition performance to the task of identification, showing that this is more challenging for PolyProtect given the necessity of combining subject-specific transformation parameters with the unprotected biometric template. Nevertheless, the performance degradation with respect to the unprotected system is limited and would be acceptable in practice. Fourthly, as PolyProtect promises to be modality independent, we tested it on fingerprint embeddings, considering a fixed-length, freely available implementation of DeepPrint[10] [15], [35], showing the true cross-modality potential of this BTP method.

Future avenues for research involve the experimental evaluation of several aspects to gain further insights. From the perspective of optimising recognition performance, parameters such as the degree of the polynomial (directly related to the range of the E parameters), as well as the range of the C parameters, might offer margins for improvement. Counting on the availability of a higher number of protected templates, attacks involving the training of neural network-based classifiers can be designed, to investigate what information about the data subjects is retained in the protected representations. Finally, we plan to integrate PolyProtect into the Simprints open-source mobile app[11], which is currently being deployed in several field projects.

[9] https://gitlab.idiap.ch/bob/bob.paper.polyprotect_2021

[10] https://github.com/tim-rohwedder/fixed-length-fingerprint-extractors

[11] https://github.com/Simprints/TBA/

Acknowledgment

This work was funded by the "Safe Biometrics for Aid Distribution" project, a partnership between the ICRC and Simprints.

References

[1] Hatef Otroshi Shahreza and Sébastien Marcel. Face reconstruction from facial templates by learning latent space of a generator network. *Advances in Neural Information Processing Systems*, 36, 2024.

[2] Philipp Terhörst, Daniel Fährmann, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. On soft-biometric information stored in biometric face embeddings. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):519–534, 2021.

[3] Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2018.

[4] K. EdalatNejad, W. Lueks, J. Sukaitis, V. Graf Narbel, M. Marelli, and C. Troncoso. Janus: Safe Biometric Deduplication for Humanitarian Aid Distribution. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 116–116, 2024.

[5] Human Right Watch. New Evidence that Biometric Data Systems Imperil Afghans. https://www.hrw.org/news/2022/03/30/new-evidence-biometric-data-systems-imperil-afghans, 2022. Online; accessed 10 July 2024.

[6] The International Committee of Red Cross. Handbook on Data Protection in Humanitarian Action. https://www.icrc.org/en/publication/430501-handbook-dataprotection-humanitarian-action-second-edition, 2020. ICRC.

[7] The International Committee of Red Cross. Policy on the Processing of Biometric Data. https://www.icrc.org/en/document/icrc-biometrics-policy, 2019. ICRC.

[8] Hatef Otroshi Shahreza and Sébastien Marcel. Blackbox Face Reconstruction from Deep Facial Embeddings Using A Different Face Recognition Model. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 2435–2439, 2023.

[9] Karthik Nandakumar and Anil K Jain. Biometric template protection: Bridging the performance gap between theory and practice. *IEEE Signal Processing Magazine*, 32(5):88–100, 2015.

[10] International Standard Organization (ISO). ISO/IEC 24745:2022 – Biometric Information Protection. https://www.iso.org/standard/75302.html, 2022. Information Security, Cybersecurity and Privacy Protection.

[11] International Standard Organization (ISO). ISO/IEC 30136:2018 – Performance testing of biometric template protection schemes. https://www.iso.org/standard/53256.html, 2018. Information technology.

[12] Vedrana Krivokuća Hahn and Sébastien Marcel. Towards Protecting Face Embeddings in Mobile Face Verification Scenarios. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1):117–134, 2022.

[13] Anjith George, Christophe Ecabert, Hatef Otroshi Shahreza, Ketan Kotwal, and Sébastien Marcel. Edgeface: Efficient face recognition model for edge devices. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[14] Haiyu Wu, Jaskirat Singh, Sicong Tian, Liang Zheng, and Kevin W Bowyer. Vec2face: Scaling face dataset generation with loosely constrained vectors. *arXiv:2409.02979*, 2024.

[15] Tim Rohwedder, Dailé Osorio-Roig, Christian Rathgeb, and Christoph Busch. Benchmarking fixed-length fingerprint representations across different embedding sizes and sensor types. In *2023 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2023.

[16] Rafael Bouzaglo and Yosi Keller. Synthesis and reconstruction of fingerprints using generative adversarial networks. *arXiv:2201.06164*, 2022.

[17] Vedrana Krivokuća Hahn and Sébastien Marcel. Biometric Template Protection for Neural-Network-Based Face Recognition Systems: A Survey of Methods and Evaluation Techniques. *IEEE Transactions on Information Forensics and Security*, 18:639–666, 2023.

[18] Rohit Kumar Pandey, Yingbo Zhou, Bhargava Urala Kota, and Venu Govindaraju. Deep secure encoding for face template protection. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRw)*, pages 9–15, 2016.

[19] Arun Kumar Jindal, Srinivas Chalamala, and Santosh Kumar Jami. Face template protection using deep convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRw)*, pages 462–470, 2018.

[20] Santosh Kumar Jami, Srinivasa Rao Chalamala, and Arun Kumar Jindal. Biometric template protection through adversarial learning. In *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. IEEE, 2019.

[21] João Ribeiro Pinto, Miguel V Correia, and Jaime S Cardoso. Secure triplet loss: Achieving cancelability and non-linkability in end-to-end deep biometrics. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2):180–189, 2020.

[22] Vishnu Naresh Boddeti. Secure Face Matching Using Fully Homomorphic Encryption. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10, 2018.

[23] Joshua J. Engelsma, Anil K. Jain, and Vishnu Naresh Boddeti. HERS: Homomorphically Encrypted Representation Search. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):349–360, 2022.

[24] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. page 28–36, 1999.

[25] Ari Juels and Madhu Sudan. A fuzzy vault scheme. *Designs, Codes and Cryptography*, 38:237–257, 2006.

[26] 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 Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1892–1901, 2006.

[27] Christian Rathgeb, Johannes Merkle, Johanna Scholz, Benjamin Tams, and Vanessa Nesterowicz. Deep face fuzzy vault: Implementation and performance. *Computers & Security*, 113:102539, 2022.

[28] Xingbo Dong, KokSheik Wong, Zhe Jin, and Jean-luc Dugelay. A cancellable face template scheme based on nonlinear multi-dimension spectral hashing. In *Int. Workshop on Biometrics and Forensics*, 2019.

[29] Yazhou Wang, Bing Li, Yan Zhang, Jiaxin Wu, Pengwei Yuan, and Guimiao Liu. A biometric key generation mechanism for authentication based on face image. In *Int. Conf. on Signal and Image Processing*, 2020.

[30] Danny Keller, Margarita Osadchy, and Orr Dunkelman. Fuzzy commitments offer insufficient protection to biometric templates produced by deep learning. *arXiv preprint arXiv:2012.13293*, 2020.

[31] Danny Keller, Margarita Osadchy, and Orr Dunkelman. Inverting binarizations of facial templates produced by deep learning (and its implications). *IEEE Transactions on Information Forensics and Security*, 16:4184–4196, 2021.

[32] Sunpill Kim, Yunseong Jeong, Jinsu Kim, Jungkon Kim, Hyung Tae Lee, and Jae Hong Seo. Ironmask: Modular architecture for protecting deep face template. In *Proc. of the Conf. on Computer Vision and Pattern Recognition*, pages 16125–16134, 2021.

[33] Xingbo Dong, Soohyong Kim, Zhe Jin, Jung Yeon Hwang, Sangrae Cho, and Andrew Beng Jin Teoh. Secure chaff-less fuzzy vault for face identification systems. *ACM Trans. on Multimidia Computing Communications and Applications*, 17(3):1–22, 2021.

[34] Muhammad Maaz, Abdelrahman Shaker, Hisham Cholakkal, Salman Khan, Syed Waqas Zamir, Rao Muhammad Anwer, and Fahad Shahbaz Khan. Edgenext: efficiently amalgamated cnn-transformer architecture for mobile vision applications. In *European conference on computer vision*, pages 3–20. Springer, 2022.

[35] Joshua J Engelsma, Kai Cao, and Anil K Jain. Learning a fixed-length fingerprint representation. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):1981–1997, 2019.

[36] Jeremy Gray and Jeremy Gray. The unsolvability of the quintic. *A History of Abstract Algebra: From Algebraic Equations to Modern Algebra*, pages 97–114, 2018.

[37] Christopher McCool, Roy Wallace, Mitchell McLaren, Laurent El Shafey, and Sébastien Marcel. Session variability modelling for face authentication. *IET Biometrics*, 2(3):117–129, 2013.