Securing Face and Fingerprint Templates in Humanitarian Biometric Systems

Giuseppe Stragapede*[1], Sam Merrick*, Vedrana Krivokuća Hahn†, Justin Sukaitis‡, Vincent Graf Narbel‡
*Simprints, †Idiap Research Institute, ‡International Committee of the Red Cross (ICRC)

Abstract

In humanitarian and emergency scenarios, the use of biometrics can dramatically improve the efficiency of operations, but it poses risks for the data subjects, which are exacerbated in contexts of vulnerability. To address this, we present a mobile biometric system implementing a biometric template protection (BTP) scheme suitable for these scenarios. After rigorously formulating the functional, operational, and security and privacy requirements of these contexts, we perform a broad comparative analysis of the BTP landscape. PolyProtect, a method designed to operate on neural network face embeddings, is identified as the most suitable method due to its effectiveness, modularity, and lightweight computational burden. We evaluate PolyProtect in terms of verification and identification accuracy, irreversibility, and unlinkability, when this BTP method is applied to face embeddings extracted using EdgeFace, a novel state-of-the-art efficient feature extractor, on a real-world face dataset from a humanitarian field project in Ethiopia. Moreover, as PolyProtect promises to be modality-independent, we extend its evaluation to fingerprints. To the best of our knowledge, this is the first time that PolyProtect has been evaluated for the identification scenario and for fingerprint biometrics. Our experimental results are promising, and we plan to release our code[2].

1. Introduction

In humanitarian and emergency scenarios, biometric recognition can represent an efficient solution to verify or establish the identity of beneficiaries during the distribution of physical goods, such as medicines, food or blankets. In these contexts, IDs and access tokens are difficult to distribute, and passwords can be forgotten or shared with unentitled subjects. On the other hand, biometrics might introduce risks for data subjects, such as impersonation [25], inference of personal and sensitive information [30], linkage of individuals across application databases (cross-matching) [7], 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, which contained data about Afghan civilians The Taliban used them to determine who had a relationship with the US Army [3, 10]. For these reasons, the security standards upheld by humanitarian organizations are stricter, in line with the 'do-no-harm' principle [31].

Fortunately, 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 [12, 23]. BTP aims to develop methods that can be applied to biometric data to produce a protected version that can be safely stored, revealing little or no information about the data subjects. In particular, a BTP method should possess the following properties [11]: *(i)* recognition accuracy: the incorporation of the protection method into a biometric system should not degrade the system's recognition accuracy; *(ii)* irreversibility: it should be computationally infeasible to recover the original biometric data from its protected version; *(iii)* unlinkability and renewability: it should be possible to generate multiple distinct protected templates from the same subject's biometric data, so that the protected templates cannot be linked to each other. This would allow for the revocation and 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.

In this work, we propose a BTP-enhanced biometric system suitable for humanitarian use-cases. Firstly, we provide a rigorous formulation of the specific functional, operational, security and privacy requirements for the adopted BTP solution (Sec. 2). The characteristics of our use-case do not impair the generality of the gathered requirements, which could be applied to other security-critical mobile applications. Secondly, we perform a thorough analysis of the entire BTP landscape (Sec. 3), which results in identifying *PolyProtect*, a feature-transformation approach designed for mobile face verification [9], as the most suitable method, over well-known approaches such as homomorphic encryption (HE) or hashing-based solutions. Thirdly, we achieve

[1] giuseppe@simprints.com

[2] https://github.com/Simprints/Biometrics-SimPolyProtect

solid experimental results (Sect. 6), assessing PolyProtect from the perspective of verification (obtaining improved recognition rates in comparison with the unprotected system) and identification performance, irreversibility and unlinkability, in combination with *EdgeFace*, a novel state-of-the-art efficient feature extractor, on a face dataset from a field project in Ethiopia. Moreover, as PolyProtect promises to be modality-independent, we extend its evaluation to fingerprint biometrics, showing the true cross-modality potential of this BTP method. To the best of our knowledge, this is the first time that PolyProtect has been evaluated in the context of identification and for fingerprint biometrics, as well as on real-world datasets collected in humanitarian field projects.

2. BTP System Requirements

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 The designed BTP solution should be applicable to all modalities.

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.

Operational Requirements

O.1 – Computational Efficiency Lightweight BTP solutions should be preferred due to the mobile environment (*e.g.*, low-cost smartphones) 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.

Security and Privacy Requirements

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

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

Threat Model

To evaluate the security and privacy properties of a BTP method (irreversibility and unlinkability), it is first necessary to define a threat model, which characterises the type of attacker on which this 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 storage devices, which can be rooted to: obtain 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*.

3. 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 [19]: *(i)* method *type*, which can be *handcrafted* or *NN-based*; *(ii)* method *input*, which can be at *image-level* or at *feature-level*.

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* [12]. 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 [19].

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 include the fact that such protection methods are generally specific to the neural network on whose templates they are trained (conflicting with *F.3 – Feature Extractor-Independence*) [20, 21]. Moreover, assessing the scalability of these methods without retraining can be difficult (*F.5 – Easy New Enrolment*) [13], and the template renewability aspect also appears challenging (*F.6 – Template Revocability and Renewability*) [26]. Concerning the security and privacy analysis, in the adopted full-disclosure threat model, 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 information about the original embedding or image from different layers of the neural network (*S.1 – Irreversibility*). Such a thorough irreversibility analysis, which is out of the scope of this work, is still lacking in the literature for these kinds of methods.

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 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 focuses 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* module to 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 [19].

Feature-level Handcrafted Methods

We have ruled out NN-based and image-level methods due to their incompatibility with our project requirements. Nevertheless, the majority of BTP methods are both handcrafted and feature-level [19]. In this section, we narrow our focus to three approaches: homomorphic encryption (HE), hashing, and feature transformation approaches (Table 1).

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 (especially in resource-constrained humanitarian contexts). Consequently, most efforts towards HE-based BTP solutions have focused on reducing this computational complexity while simultaneously trying to minimise the resulting accuracy degradation [5,24]. In any case, encrypted templates remain secure only insofar as the corresponding decryption key remains secret, conflicting with the full-disclosure threat model adopted in our project. 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. In particular, *cryptographic* hash functions are designed to exaggerate small differences in the input. Consequently, the main challenge in their application to biometric templates is due to the intrinsic intra-class variability of biometric samples. For this reason, hash-based BTP methods tend to apply hashing to random, subject-specific codewords, which are bound to the biometric templates by some mathematical function, so that the matching operation takes place indirectly by reconstructing the codewords using probe samples. This is the case for fuzzy committment [15] or fuzzy vault-based [14] schemes. To reduce sensitivity to intra-class variations, *non-cryptographic* hashes [29] have been proposed: in this case, the same subject's templates are mapped to approximately the same code, allowing distance-based comparisons, without requiring an exact match as for cryptographic hash functions. However, we observed that all hashing methods proposed in the literature are likely to introduce some accuracy degradation [2,27,32] (*F.1 – Recognition Accuracy*). Moreover, their irreversibility and unlinkability have been demonstrated to be fragile in some cases (*S.1 – Irreversibility*, *S.2 – Unlinkability*) [16,17], or non-exhaustively evaluated [1,18].

Table 1. Having narrowed our focus to feature-level handcrafted BTP methods, we consider three approaches: homomorphic encryption, hashing, and PolyProtect [9]. Below, we roughly rate them against our project requirements and threat model. For the requirements, we consider four possible judgment values: satisfied ($\mathcal{S}$), possible ($\mathcal{P}$), challenging ($\mathcal{C}$), weak ($\mathcal{W}$).

Requirement		HE	Hashing	PolyProtect [9]
F.1	*Recognition accuracy*	$\mathcal{S}$	$\mathcal{C}$	$\mathcal{S}$
F.2	*Modality-Independence*	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{P}$
F.3	*Feature Extractor-Indep.*	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{S}$
F.4	*On-Device Recognition*	$\mathcal{C}$	$\mathcal{P}$	$\mathcal{S}$
F.5	*Easy New Enrolment*	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{S}$
F.6	*Template Revocability*	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{S}$
F.7	*Open-Source*	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{S}$
S.1	*Irreversibility*	$\mathcal{S}$*	$\mathcal{C}$	$\mathcal{S}$
S.2	*Unlinkability*	$\mathcal{S}$	$\mathcal{C}$	$\mathcal{S}$
O.1	*Computational Efficiency*	$\mathcal{W}$	$\mathcal{P}$	$\mathcal{S}$
O.2	*Time Effic.*	$\mathcal{W}$	$\mathcal{P}$	$\mathcal{S}$
O.2	*Offline Processing*	$\mathcal{W}$	$\mathcal{P}$	$\mathcal{S}$
Full-Disclosure Threat Model [11]		✗	*?*	✓

*Not under the full-disclosure threat model.

Feature transformation 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, entailing the use of a different comparison function to that adopted in the unprotected template domain (*e.g.*, Hamming distance). *PolyProtect*, a method designed for mobile face verification [9], falls into this category. The method was developed considering a full disclosure threat model, it shows no significant recognition accuracy degradation, and the code is available under a GPL-3.0 license[3], making the results easily reproducible.

4. 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.

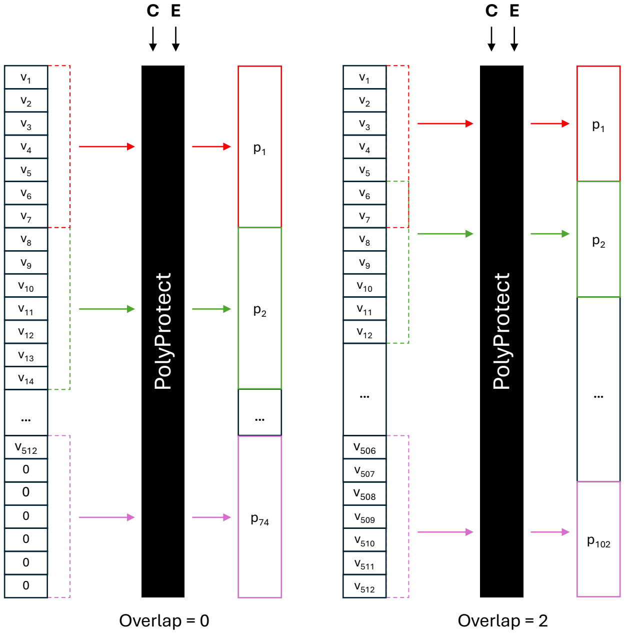

Figure 1. A graphical representation of PolyProtect is provided above. As an example, we consider overlap $o = \{0, 2\}$ for $m = 7$. The dimensionality of the protected template P varies according to both m and o.

5. Experimental Protocol

In this section, the experimental setup, methodology, and datasets adopted in our experiments are presented.

5.1. NN-based Feature Extractors

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 [6][4]. EdgeFace is based on a hybrid network architecture that leverages convolutional NN and Vision Transformer (ViT) capabilities [22]. 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

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

[4] We consider the *XS* version, which consists of only 1.77M parameters (6.83MB). Link: https://github.com/otroshi/edgeface

minutiae-based template representations, and their inherent two-dimensional nature, as each minutia point is typically described by its location coordinates and type; *(ii)* the lower level of abstraction of minutiae-based templates, in which each minutia 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[5] [28]. The model replicates the DeepPrint architecture [4], 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 encodes minutiae information.

5.2. Configuration of PolyProtect

Implementing PolyProtect requires setting four values: m (the number of unprotected template elements used to compute each element of the protected template P), o (the amount of overlap between consecutive sets of m elements), and the ranges of C (coefficients of the polynomial) and E (exponents of the polynomial).

The minimum value of m is set by the Abel-Ruffini theorem [9], which states that there is no closed form algebraic expression for solving polynomials of degree 5 or higher with arbitrary coefficients [8]. At the same time, 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, for $m = 5$ we would have 120 subjects, for $m = 6$ we would have 720, for $m = 7$ we would have 5040, etc. The upper limit of m is bound by the fact that the embeddings produced by the NN-based feature extractors consist of floating point values smaller than 1, which would cause large powers to effectively obliterate certain elements. In [9], the authors of PolyProtect set m to 5, since their baseline evaluation is carried out on a dataset consisting of fewer than 100 subjects. Accounting for bigger datasets, we set $m = 7$. 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.

5.3. Analysis of BTP Criteria

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 verification accuracy evaluation: *(i)* the *Normal* (N) scenario, in which the system should operate most of the time; and *(ii)* the *Stolen Coefficients and Exponents* (SCE) scenario, in which 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. Concerning the identification scenario, given the absence of the identity claim, 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. Then, a ranking is generated based on the distances, and the highest-scoring matches are returned.

In our irreversibility analysis, the full disclosure threat model described in Sec. 2 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. 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. Specifically, the embeddings to be recovered are 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. Following the attack defined in the original paper, a numerical solver[6] starts from 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. We also 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

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

[6] 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

either enrolled in a different application or used to replace a compromised PolyProtected template in the same application. This was simulated using the same numerical solver approach, 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. We invite the reader to consult [9] for more details about the definition of the system equations, which we omit for brevity.

The unlinkability of PolyProtect is evaluated using the framework proposed in [7]. 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 $D_{\leftrightarrow}^{sys}$, a global measure of the overall linkability of the underlying recognition system. Following the recommendation in [7], we compute 10 different PolyProtected templates per person. 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.

5.4. Datasets

For face biometrics we employ a dataset of face images obtained within the framework of a project currently being carried out 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. For fingerprint biometrics, we adopt an internal dataset collected in a field project in Ghana, consisting of 119 subjects with two samples each. Images were mostly acquired outdoors, from collaborative subjects, by trained data collectors. For face, low-end smartphones were used, with 8-12MP resolution. For fingerprints, dedicated scanners were used, with a resolution of 500DPI. Dirt and dust often accumulated on the scanner surface, making the capture challenging. The datasets cannot be made public due to participant privacy agreements, but the code will be made public so that interested researchers can perform evaluations on their own datasets of interest.

6. Experimental Results

This section presents our experimental results in terms of recognition accuracy, irreversibility, and unlinkability.

6.1. Recognition Accuracy

The verification performance is measured in terms of True Match Rate (TMR) at a given False Match Rate (FMR). We set thresholds corresponding to FMR=0.01% and 0.1%, as well as to the Equal Error Rate (EER). Table 2 shows 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). 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. On top of this, it is clear that by increasing the overlap value (down the rows), the performance improves. 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 2 contains the results in the SCE scenario. A reduction of the biometric performance is expected [9], as embeddings transformed with the same secret parameters are being compared. Indeed, in contrast to the N scenario, the baseline performance is, in almost

Table 2. Verification results. In bold, the results achieved by PolyProtect which improve the baseline performance.

o	TMR (%) @ FMR = 0.01% ↑	TMR (%) @ FMR = 0.1% ↑	EER (%) ↓	EER (%) ↓
	Face			Fingerprint
Bas.	94.27	95.97	1.35	40.21
N Scenario				
0	93.40	**97.26**	**0.83**	**7.13**
1	**94.67**	**97.24**	**0.72**	**6.82**
2	**94.78**	**98.07**	**0.57**	**6.69**
3	**94.97**	**97.71**	**0.65**	**6.51**
4	**95.65**	**98.51**	**0.62**	**6.19**
5	**95.97**	**98.54**	**0.55**	**6.30**
6	**96.39**	**98.98**	**0.45**	**6.19**
SCE Scenario				
0	87.54	92.80	2.25	40.83
1	89.38	93.95	2.21	40.68
2	90.74	93.99	1.74	41.73
3	90.28	94.59	1.91	40.5
4	91.95	95.05	1.51	40.64
5	93.36	95.76	1.49	**39.95**
6	94.20	**96.05**	**1.26**	40.34

all cases, the best one (except for TMR at FMR=0.1% and EER, with an overlap of 6). Moreover, similarly to the N scenario, the performance improves with an increase in the overlap value, thus limiting the accuracy degradation.

For fingerprint biometrics, we observe a baseline performance of 40.21% EER. As mentioned in Sec. 5, 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 to be much greater, reducing the EER to 7.13% 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 increasing the overlap value (down the rows) does not yield a clear improvement trend as in the face experiment.

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 performance discrepancy between the N and SCE scenarios 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.

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. All the protected query templates obtained are 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, without exploiting the discriminative power of incorporating subject-specific information. From this perspective, the comparisons are comparable to those carried out in the SCE verification scenario.

Table 3. Identification results in terms of True Positive Identification Rate-n (TPIR-n), which 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.

o	Face (TPIR-n (%), n)			Fingerprint (TPIR-n (%), n)		
	n = 1	n = 3	n = 10	n = 1	n = 3	n = 10
Bas.	98.09	98.51	99.36	11.67	23.33	41.67
0	93.74	96.16	97.92	8.00	19.83	37.83
1	95.10	96.62	98.03	8.50	18.67	38.33
2	95.78	97.37	98.66	7.50	18.33	39.17
3	95.78	97.37	98.41	8.33	18.83	41.00
4	96.47	97.86	98.88	9.50	18.17	39.33
5	97.39	97.96	98.85	8.67	21.33	41.17
6	97.41	98.47	99.19	10.17	19.67	41.00

Table 4. Irreversibility results in terms of Inversion Success Rate (ISR).

o	Face (ISR (%))		Fingerprint (ISR (%))	
	FMR = 0.01%	FMR = 0.1%	FMR = 1%	FMR = 10%
0	0	0	0	1.00
1	0	0	0	1.50
2	0	0	0	4.50
3	0	0	0	12.00
4	0	0.02	0	37.33
5	0.83	34.59	9.83	79.83
6	98.20	98.20	92.33	92.33

Table 3 shows that for face the baseline performance is better in all cases. However, by increasing the overlap value, and therefore the dimensionality of the output (protected) space, we observe an overall improvement in the identification rates. For TPIR-3 and TPIR-10, Table 3 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 would be considered acceptable. For the more stringent TPIR-1, the corresponding decrease would be greater (approximately 2.5% in Table 3). For fingerprints, the adopted dataset proves to be very challenging in the identification task as well, but the negative impact of PolyProtect seems limited in this case.

6.2. Irreversibility

Table 4 shows the results of the attempts to reconstruct an unprotected template from a single protected template. The Inversion Success Rate (ISR) is computed as the solution rate × match rate [9]. 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), since a stricter threshold would require a better approximation of the original input template. Another interesting observation is that, as the overlap value increases, the ISR increases, similarly to the recognition accuracy, revealing a trade-off between these two metrics. For the 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 [9].

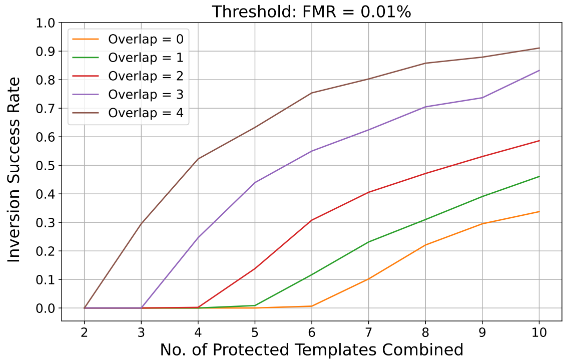

Figure 2. Attack via Record Multiplicity (ARM) for the face protected system.

Fig. 2 shows the results of the ARM experiment on the face dataset. Our ARM analysis encompasses a maximum overlap value of 4 since we observed non-zero ISR values for an overlap of 5 in the inversion of a single template. The used protected templates 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 $\times$ 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 that for a given overlap value, as the number of combined protected templates increases, the chances of a successful reconstruction are higher.

6.3. Unlinkability

Table 5 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 [7]. 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).

Compared to the naive (random) parameter selection, the strict one involves an additional comparison check [9]: 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. In contrast to the unprotected systems, the $D^{sys}_{\leftrightarrow}$ 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. The strict PolyProtect parameter selection shows that the $D^{sys}_{\leftrightarrow}$ values are further reduced, especially in the case of the fingerprint dataset.

Table 5. Unlinkability results, considering the naive (N) and strict (S) PolyProtect parameter (C and E) selection.

	$D^{sys}_{\leftrightarrow}$			
	Face		Fingerprint	
o	N	S	N	S
Bas.	0.757		0.277	
0	0.077	0.062	0.092	0.022
1	0.078	0.063	0.089	0.027
2	0.078	0.065	0.095	0.022
3	0.078	0.065	0.089	0.033
4	0.078	0.069	0.097	0.028
5	0.078	0.072	0.089	0.034
6	0.078	0.077	0.099	0.039

7. Conclusions

This article presented a BTP-enhanced mobile biometric system, suitable for humanitarian and emergency scenarios, which are characterised by particularly strict functional, operational, and security and privacy requirements (formulated in Sec. 2), as an unsafe use of biometrics can be particularly harmful for the data subjects. We provided a broad analysis of the existing BTP literature to identify a suitable method for our requirements (Sec. 3). We then selected PolyProtect [9], a feature-transformation approach, over approaches such as HE and hashing, mainly due to: 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); and its suitability for the full disclosure threat model [11].

We experimentally validated PolyProtect, observing results consistent with prior findings in recognition accuracy, irreversibility, and unlinkability. The novel insights provided in this work can be summarised as follows: *(i)* we configured our system to deal with a higher number of enrolled subjects (hence $m = 7$) and tested it on a challenging face dataset collected in a field project in Ethiopia, with very promising results, in combination with a more recent and efficient feature extractor, EdgeFace; *(ii)* we extended the evaluation of the recognition performance to the task of identification, showing that the performance degradation with respect to the unprotected system would probably be acceptable in practice; *(iii)* we tested PolyProtect on fingerprint embeddings from a dataset collected in Ghana, considering a fixed-length, freely available implementation of DeepPrint [4, 28], showing for the first time the true cross-modality potential of this BTP method.

References

[1] X. Dong, S. Kim, Z. Jin, J. Y. Hwang, S. Cho, and A. B. J. Teoh. Secure chaff-less fuzzy vault for face identification systems. *ACM Trans. on Multimidia Computing Communications and Applications*, 17(3):1–22, 2021.

[2] X. Dong, K. Wong, Z. Jin, and J.-l. Dugelay. A cancellable face template scheme based on nonlinear multi-dimension spectral hashing. In *Int. Workshop on Biometrics and Forensics*, 2019.

[3] K. EdalatNejad, W. Lueks, J. Sukaitis, V. G. 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.

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

[5] J. J. Engelsma, A. K. Jain, and V. N. Boddeti. HERS: Homomorphically Encrypted Representation Search. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):349–360, 2022.

[6] A. George, C. Ecabert, H. O. Shahreza, K. Kotwal, and S. Marcel. Edgeface: Efficient face recognition model for edge devices. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

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

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

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

[10] 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.

[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] 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.

[13] S. K. Jami, S. R. Chalamala, and A. K. Jindal. Biometric template protection through adversarial learning. In *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. IEEE, 2019.

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

[15] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *Proceedings of the 6th ACM conference on Computer and communications security*, page 28–36, 1999.

[16] D. Keller, M. Osadchy, and O. Dunkelman. Fuzzy commitments offer insufficient protection to biometric templates produced by deep learning. *arXiv preprint arXiv:2012.13293*, 2020.

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

[18] S. Kim, Y. Jeong, J. Kim, J. Kim, H. T. Lee, and J. H. Seo. Ironmask: Modular architecture for protecting deep face template. In *Proc. of the Conf. on Computer Vision and Pattern Recognition*, pages 16125–16134, 2021.

[19] V. Krivokuća Hahn and S. 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.

[20] A. Kumar Jindal, S. Chalamala, and S. 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.

[21] R. Kumar Pandey, Y. Zhou, B. Urala Kota, and V. 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.

[22] M. Maaz, A. Shaker, H. Cholakkal, S. Khan, S. W. Zamir, R. M. Anwer, and F. Shahbaz Khan. Edgenext: efficiently amalgamated cnn-transformer architecture for mobile vision applications. In *European conference on computer vision*, pages 3–20. Springer, 2022.

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

[24] V. 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.

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

[26] J. R. Pinto, M. V. Correia, and J. 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.

[27] C. Rathgeb, J. Merkle, J. Scholz, B. Tams, and V. Nesterowicz. Deep face fuzzy vault: Implementation and performance. *Computers & Security*, 113:102539, 2022.

[28] T. Rohwedder, D. Osorio-Roig, C. Rathgeb, and C. 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.

[29] A. Teoh, A. Goh, and D. 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.

[30] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper. On soft-biometric information stored in biometric face embeddings. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):519–534, 2021.

[31] 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.

[32] Y. Wang, B. Li, Y. Zhang, J. Wu, P. Yuan, and G. Liu. A biometric key generation mechanism for authentication based on face image. In *Int. Conf. on Signal and Image Processing*, 2020.