Criteria Towards Metrics for Benchmarking Template Protection Algorithms*

Koen Simoens[1], Bian Yang[2], Xuebing Zhou[3], Filipe Beato[1],
Christoph Busch[2,3], Elaine M. Newton[4], and Bart Preneel[1]

[1] K.U.Leuven – COSIC and IBBT, Heverlee, Belgium
[2] Gjøvik University College – NISLab, Gjøvik, Norway
[3] Hochschule Darmstadt – CASED, Darmstadt, Germany
[4] National Institute of Standards and Technology, Gaithersburg, MD, USA

Abstract

Traditional criteria used in biometric performance evaluation do not cover all the performance aspects of biometric template protection (BTP) and the lack of well-defined metrics inhibits the proper evaluation of such methods. Previous work in the literature focuses, in general, on a limited set of criteria and methods. This paper provides the first holistic approach to the evaluation of biometric template protection that is able to cover a whole range of methods. We present a selection of well-defined criteria and some metrics that are compliant with the reference architecture for template protection as defined in the recently adopted standard ISO/IEC 24745 (2011), which is applicable to nearly all known BTP methods. The criteria have been grouped in three categories of performance: technical, protection, and operational.

1. Introduction

Biometrics provide an alternative to passwords and other token-based authentication because they do not require users to memorize or carry a credential, and they are more tightly-bound form factors for identification than identity documents. However, privacy issues, which are a direct consequence of the properties that are desired from biometric characteristics, *e.g.*, uniqueness or permanence, have been raised repeatedly [25, 27]. Biometric data may reveal sensitive, *e.g.*, medical, information and they uniquely identify an individual. This implies that a biometric sample or template can be used as a unique identifier to link information across different applications. Moreover, once biometric data have been compromised they can be used for spoofing by constructing artificial samples. Furthermore, biometric characteristics are limited in number and cannot be renewed. Because of these issues biometric data should be protected, i.e., made uninterpretable and unlinkable without authorization, but without losing the capability to identify a person or to verify a person's identity. These are the main objectives of biometric template protection (BTP) methods. Cavoukian and Stoianov [12] summarize this in their white paper on biometric encryption by stating that this is a positive-sum technology: both privacy and security can be assured without giving in on one or the other.

Despite the variety[1] of template protection schemes that have been proposed in the literature [5, 13, 18, 19, 21, 24, 26, 32, 35] there is still a lack of well-established metrics for evaluating BTP methods. This lack makes it impossible to perform a proper evaluation or direct comparison of BTP methods. Previous works in the literature focus mainly on the security and privacy aspects or only on a particular type of algorithm. Scheirer and Boult [29] proposed several attacks on the fuzzy vault scheme [17] and biometric encryption [32]. Simoens *et al*. evaluated the irreversibility and unlinkability of schemes based on error-correcting codes, Nagar, Nandakumar, and Jain [22] analysed similar properties for cancelable fingerprint transformations[26] and the biohashing method [34]. In addition, some initiatives [1] focus only on the technical performance aspect, i.e., the accuracy and efficiency of BTP algorithms. Finally, a number of frameworks [10, 13, 21, 27] have been proposed to model template protection. However, so far none of these has been able to cover all known BTP methods.

1.1. Objectives and contributions

The primary goal of this work is to progress towards ranking and independent benchmarking of different BTP algorithms. We identify and select criteria that are relevant

*This work was supported by grant number 60NANB10D217 awarded by NIST. F. Beato is supported by FCT Grant SFRH/BD/70311/2010.

[1] For a survey on BTP methods the reader is referred to the work of Jain, Nandakumar and Nagar [16] and the work of Rathgeb and Uhl [28].

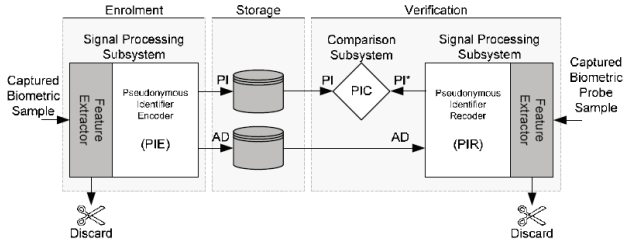

Figure 1. Reference architecture for template protection [2].

for BTP performance assessment. However, criteria do not tell us how well an algorithm actually performs. Tools are needed that measure the performance on these criteria and that produce numbers or values that allow comparison of algorithms on a particular criterion. This paper proposes a set of relevant criteria, consistently defined in a generic architecture, and its main contribution is a consolidation in the evaluation of BTP methods. In addition, metrics are presented for some of the criteria, based on the literature or new ideas.

1.2. Organization

The criteria and metrics are defined in the ISO reference architecture for template protection (ISO/IEC 24745 [2]). The architecture and corresponding terminology are briefly summarized in Section 2. The criteria and metrics have further been grouped in three categories. Section 3 discusses technical performance. Section 4 is dedicated to how well BTP protects biometric data. Section 5 discusses operational performance, which relates to modality independence, interoperability and quality of performance. Section 6 provides a summary of conclusions.

2. Preliminaries

Before presenting the performance metrics, we describe the framework in which the metrics have been defined. Figure 1 shows the reference architecture for the protection of biometric information that has been standardized in ISO/IEC 24745 [2]. In this generic architecture, which applies to nearly all known methods[2], a biometric sample is transformed during enrolment into a *renewable biometric reference*, which is defined in [2] as a *"Revocable or renewable identifier that represents an individual or data subject within a certain domain by means of a protected binary identity (re)constructed from the captured biometric sample."* The main idea behind a renewable reference is that an individual can be enrolled in different applications with the same characteristic, *e.g.*, face, but through different secure references that cannot be linked to each other. The term *protected template (PT)* is used as a synonym for renewable biometric reference in this paper.

[2]The standard [2] includes a mapping of the BTP methods in [36, 18, 32, 17, 21, 13, 6, 9, 26, 33, 34, 14, 5, 20, 37] to the reference architecture.

2.1. Pseudonymous identifier (PI) and auxiliary data (AD)

A protected template, as defined in [2], consists of a *pseudonymous identifier (PI)* and possibly *auxiliary data (AD)*. The PI represents the individual in an application context and is used as a reference for verification, but does not allow the retrieval of the enrolment data. Multiple unlinkable PIs can be derived from the same characteristic. The AD is the part of the PT that helps to reconstruct the PI during verification. It mostly depends on the enrolment data and may contain elements that allow diversification, i.e., the creation of multiple PIs. The AD is not necessarily stored along with the PI but both are needed during verification.

In general it is assumed that the AD is public because it is part of the PT. It is discouraged to have secret AD, however, it is not strictly required by the standard. It is tolerated that some schemes require secrecy of the AD, *e.g.*, if plain encryption is used the encryption key is considered AD. As a consequence, the secrecy of the AD cannot be considered as an evaluation criterion on its own. However, it can be the conclusion of an evaluation that some property, *e.g.*, irreversibility, cannot be satisfied unless the AD is secret.

2.2. Functional components

During enrolment features are extracted from a captured biometric sample and fed into a *pseudonymous identifier encoder (PIE)*. The PIE is a system, process or algorithm that produces a renewable biometric reference, which consists of a PI and, possibly, AD. During verification a new PI* is recoded by the *pseudonymous identifier recoder (PIR)* from a freshly captured sample and auxiliary data, which was generated during enrolment. The recoded PI* is compared with a reference PI by the *pseudonymous identifier comparator (PIC)*, which outputs a comparison score.

3. Technical performance

In biometric systems widely-deployed today, *e.g.* by law enforcement and border control, the technical performance is of primary interest, and is widely tested operationally. Technical performance includes the following aspects: accuracy of the recognition algorithm (error rates), throughput, and storage requirements. Besides these common aspects, BTP has some unique technical performance aspects: performance degradation (compared to unprotected algorithms), diversity, and the error rate of failing to generate a PT. We give concise definitions for these criteria and discuss them in detail below.

3.1. Accuracy

Definition 1 (Accuracy)**.** *Statistical reflection of trustworthiness of the decisions (match and non-match) made by a biometric system, represented by standardized error rates.*

The common and standardized metrics for measuring the accuracy of biometric recognition algorithms are defined in [3]. These metrics apply to both unprotected and protected template algorithms. Obviously, to be able to compare different algorithms, both protected and unprotected, the same database and testing protocols should be used. The most well known accuracy metrics are the false-match-rate (FMR) and false-non-match-rate (FNMR), which reflect the accuracy of the comparison algorithm, and the false-acceptance-rate (FAR) and the false-rejection-rate (FRR), which reflect the accuracy at system level. The difference between the two levels is determined by the failure-to-acquire-rate (FTA). An additional measure for BTP is the failure to encode a PI, *e.g.*, due to low entropy in a sample.

3.2. Accuracy degradation

Definition 2 (Accuracy degradation)**.** *The accuracy performance decrease caused by BTP algorithms.*

Suppose we observe the accuracy results from a biometric system in two cases - with and without template protection - with the rest of the testing context being the same. In most existing BTP algorithms some accuracy degradation will occur. If we observe an error rate E (*e.g.*, FMR, FNMR, EER, etc.) from an accuracy performance test over the unprotected templates, and observe the same error rate in a different value E_p from a test over the protected templates, then we can define two accuracy degradation representations: the absolute accuracy degradation rate $(E_p - E)$ and the relative accuracy degradation rate $(E_p - E)/E$.

3.3. Throughput

Definition 3 (Throughput)**.** *The number of biometric transactions processed continuously by an individual biometric processing unit (*e.g.*, feature extractor, feature comparator, PIE, PIR, and PIC) in a defined time interval.*

These processing units are the BTP algorithm components (PIE, PIR, PIC) that can be implemented in the same hardware and software development environment for comparison. In terms of time consumed per transaction, both the creation (encoding / recoding) and the comparison time for PTs are required for evaluation.

Besides a BTP algorithm's efficiency, throughput is also related to both the biometric system's efficiency (data processing, communication time and system stability) and human factors, i.e., whether subjects (and system operators, if any) are well-trained at the human-machine interface or not, whether they are in a hurry, or nervous or taking time, etc. A fair evaluation in throughput of different BTP algorithms requires approximately the same implementation and testing environment [3]. Unlike the accuracy degradation defined in Section 3.2 throughput is not necessarily influenced negatively by BTP in an interoperable biometric system.

3.4. Storage requirements

Definition 4 (Storage requirements)**.** *Requirements imposed by biometric systems in different applications on the size of PTs and the implementation of BTP algorithms.*

The storage requirements are highly dependent on the applications. While PCs and central databases can provide enough storage capacity, embedded systems or small personal tokens such as smart cards or RFID chip are very limited in storage resources. In the latter case, the size of the implementation of the PIE, the PIR and the PIC has to be taken into account. Obviously, the code size (footprint) of the implementation of the BTP algorithm can only be compared when it is evaluated on the same target platform. Because BTP algorithms have not been standardized yet, they may have template sizes that are distinctly different from unprotected templates. In some of the BTP algorithms, the length of protected templates can even be adjusted to achieve expected effects in accuracy, security / privacy, and other performance aspects.

3.5. Diversity

Definition 5 (Diversification capacity)**.** *Maximum number of independent protected templates that can be generated from the same biometric feature by a BTP algorithm.*

The renewability requirement to protecting a biometric template implies that PIs can be diversified. The ability to diversify is measured in the first place by the theoretical maximum number of PIs that can be generated. Secondly, a theoretical analysis should investigate the degradation in irreversibility, unlinkability, etc. as a function of the number of PTs issued.

4. Protection performance

This section presents a list of criteria related to the protection properties of BTP. We start with a discussion on the concepts of security and privacy and the interpretation of these concepts in [2]. Then we define and elaborate the criteria related to these concepts.

4.1. Concepts of security and privacy

In the context of biometrics, security is often interpreted as the probability of an impostor managing to impersonate

[3] The NIST SHA-3 competition is an example of such equal-condition evaluation activity (`http://csrc.nist.gov/groups/ST/hash/sha-3/Submission_Reqs/ref_and_optim.html`).

a genuine user, which is measured by the false-acceptance-rate (FAR) of a system. Security is, however, more frequently used in a much broader sense because a system can and will be attacked in many more ways. As such, biometric security and privacy refer to a combination of measures at different levels (system, procedures, information, devices).

In [2] security is expressed as requirements at the system level and privacy as requirements on the information level. Security refers to the confidentiality of (biometric) information that is achieved from system-level countermeasure such as access control, the integrity of biometric references, and renewability and revocability as requirements to solve the issue of compromised references. Privacy refers to the irreversibility of PTs, the unlinkability of PTs, and the confidentiality of PTs achieved by applying, *e.g.*, data separation or plain encryption.

The scope of this paper is on the evaluation of algorithms that provide protection at the information-level. The underlying principle of biometric template protection (at template-level) is that protected templates must be self-protecting. Because the terms security and privacy are quite broad and are often used in different ways we will not use them here. Instead we will refer to the specific properties of irreversibility and unlinkability. It should also be noted that there are issues, *e.g.*, spoofing, that cannot be solved by BTP algorithms.

4.2. Irreversibility

Generally speaking, irreversibility refers to the secrecy of the biometric data from which the renewable biometric reference was created. The PI in the renewable reference is often randomly generated, i.e., independent of the enrolment data. The AD is used to diversify PTs but also, in some cases, to assist in compensating for the noise. Therefore, AD are often adjusted to the enrolment data to achieve a higher accuracy. Because of this there is an unavoidable leakage of information about the enrolment data through the AD. It has been proved theoretically by Smith [31] in the fuzzy extractor framework [13] that information leakage, which is modeled as a loss in min-entropy, is unvoidable. This was later exploited in [8] and [30] to link PTs.

Based on these observations we conclude that it is not a criterion if or how much information is leaked by a PT. What is relevant, is the purpose for which this leakage can be exploited. Therefore, we define the following irreversibility criterion, which holds for all BTP methods.

Definition 6 (Full-leakage irreversibility)**.** *The difficulty of determining, exactly or with tolerable margin, from a PT, the biometric sample(s) or features used during enrolment to generate that PT.*

This criterion is particularly relevant in systems where exact secrecy of the enrolment data is required. For example, in [7] a cryptographic key is derived directly from an enrolment sample and the entropy of that key depends on the variability of biometric data.

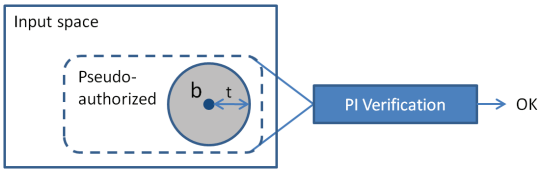

Figure 2. Simplified visualization of a scheme with pseudo-authorized inputs. The enrolment sample is denoted by the dot labeled *b*. The grey-shaded circle represents the authorized inputs and the pseudo-authorized inputs are represented by the region in the dashed rectangle minus the grey shaded circle.

In most systems the main concern is to prevent an attacker to produce a biometric sample that would pass a verification test. Therefore we define the following criterion.

Definition 7 (Authorized-leakage irreversibility)**.** *The difficulty of determining a biometric sample(s) or features from a PT that would "match" the unprotected enrolment data in a disjoint unprotected system.*

Informally, this means that an attacker should not be able to find a sample that is close to the enrolment sample, where close refers to a certain measure of similar, e.g., a distance function, and threshold as defined in an unprotected system. Unfortunatly, due to the inevitable false-match-rate, there will always be a proportion of PTs that are susceptible to offline FMR attacks, *i.e.*, when an attacker runs an entire database against a stolen PT until a matching sample is found. These attacks can to some extent be prevented by data separation.

Some methods accept a larger part of the input space than what an unprotected comparison algorithm would tolerate. This is represented by the dashed rectangle in Figure 2. This "widening" of the authorized input region enhances the full-leakage irreversibility (unconditionally). This phenomenon is typically observed in cancelable biometrics [27] or related methods based on projection [34]. To address this phenomenon we define a third notion of irreversibility.

Definition 8 (Pseudo-authorized-leakage irreversibility)**.** *The difficulty of determining, exactly or to a high degree of similarity, from a PT, the biometric sample(s) or features that match the PT but would not "match" the unprotected enrolment data in a disjoint unprotected system.*

Informally, this means that an attacker is able to distinguish pseudo-authorized inputs as shown in Figure 2 from the authorized inputs. Protected templates that reveal their (widened) matching input range make a biometric system susceptible to spoofing. However, this is a system-level

issue and should not result in a negative evaluation of the method.

Unconditional versus conditional. The term difficulty, as used in the definitions, can be interpreted in two ways. On the one hand irreversibility can be achieved unconditionally. This means that irrespective of the efforts that are put in trying to reverse a PT, there will be always be an amount of uncertainty about the biometric data. In this case, irreversibility can be measured using information-theoretic properties such as conditional min-entropy (cf. [13]), conditional entropy or guessing entropy [11] (See also Ignatenko and Willems [15]). On the other hand, it is sometimes impossible to protect biometric data against an adversary with infinite resources (computing power and time). However, the required resources may be so large, that it is practically infeasible to reverse a PT. In this case irreversibilty should be expressed in terms of computational complexity. A practical and empirical approach to this problem was proposed by Nagar and Jain [22] who defined the "coverage and effort" metric to evaluate non-invertible transformations.

Multi-reference irreversibility. Irreversibility must hold when two or more mated protected templates, i.e., originating from the same characteristic, are available to an attacker. This multi-reference dimension must be taken into account when analyzing irreversibility. This has been demonstrated to be an issue for fuzzy commitment [18] when using different error-correcting codes [30], but also for schemes based on random projections [35] as shown in [38].

4.3. Unlinkability

Unlinkability refers to the classification of renewable references. This is sometimes refered to as cross-matching, but we will use the term cross-comparison. In essence, this means that there should not exist an algorithm that performs well on classifying PTs. If, theoretically, such algorithm exists, it should not be efficiently computable. Hence, unlinkability may not be achieved theoretically, but in practice the classification of PTs is believed to be intractable. The notion of unlinkability is defined as follows.

Definition 9 (Unlinkability)**.** *The difficulty of classifying PTs over time and accross applications.*

Unlinkability is in the first place measured by verifying that two mate PTs, i.e., originating from the characteristic, differ considerably. In the second place, an attacker can try to invert the PT and use the partial information that is revealed about the enrolment data as input to a conventional comparison algorithm. These two approaches were proposed in [23]. Setting the parameters of the comparison algorithm to particular values will result in a certain classification accuracy. Consequently, performance rates called the "false cross-match rate" and "false non cross-match rate" were defined in [23]. We will adopt these names.

More generically, unlinkability should be evaluated using a classification algorithm that works on PTs instead of a conventional algorithm and that uses specific information revealed by the PT as a heuristic. A similar approach of heuristic-based classification of PTs was proposed in [30]. The best heuristic is a function that fully exploits the information that is leaked by the PT. In case this leakage is large, an attacker may be able to construct a very accurate cross-comparison algorithm. The actual metric for evaluating the unlinkablity is a pair of error rates, which reflect the accuracy of the PT classification algorithm.

Let PT_1 and PT_2 denote two protected templates derived from samples b_1 and b_2, respectively. In first instance, $b_1 = b_2$. However, some schemes cannot provide unlinkability if two enrolment samples are equal, hence two different measurements from the same characteristic should be used. Let the binary operator $\sim$ denote that two PTs are a mate pair and $\nsim$ that they are not. Let f be the heuristic function used for evaluation by a cross-comparator CC_f. The CC_f takes as input two PTs and some parameters p, like a decision threshold, and outputs 1 if the input templates are evaluated by CC_f as a mate pair and zero otherwise.

Let DB be a particular database over which a cross-comparator CC_f is evaluated. Then M_{DB} denotes the subset of all mate pairs from DB and NM_{DB} the subset of non-mate pairs:

$$\begin{aligned} M_{DB} &= \{(i,j) \,|\, i,j \in DB \wedge i \sim j\} \\ NM_{DB} &= \{(i,j) \,|\, i,j \in DB \wedge i \nsim j\}\,. \end{aligned}$$

Then we define the false cross-match rate (FCMR) and the false non-cross-match rate (FNCMR) as

$$\begin{aligned} FCMR_f &= \#\{x \in NM_{DB} \colon CC_f(x) = 1\} \,/\, \#NM_{DB} \\ FNCMR_f &= \#\{x \in M_{DB} \,:\, CC_f(x) = 0\} \,/\, \#M_{DB}\,. \end{aligned}$$

We define the equal cross-comparison rate ECCR_f as the point where $\mathrm{FCMR}_f = \mathrm{FNCMR}_f$. Figure 3 demonstrates the expected behaviour of a cross-comparator, i.e., FCMR+FNCMR≈ 1.

This empirical approach to evaluating unlinkability provides an estimate of the unconditional unlinkability under a certain heuristic. However, it might also be impossible to efficiently evaluate a heuristic function. Although theoretically an excellent heuristic may exist, it is practically infeasible to compute it and to conduct the practical experiment described above. In this case, unlinkability is achieved only conditionally.

4.4. Additional properties and remarks

Additional aspects can be considered for the criteria that have been defined above. Instead of defining new criteria for these, we consider them as additional dimensions in which the above defined criteria should be evaluated.

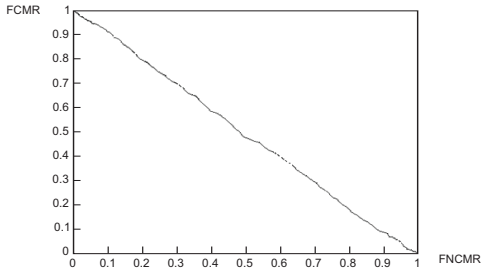

Figure 3. False cross-match rate (FCMR) versus false non-cross-match rate (FNCMR) of a cross-comparator exhibiting expected behaviour.

Confidentiality and integrity. Confidentiality and integrity are defined in [2] as the properties that information is protected against unauthorized access or disclosure and that the accuracy and completeness of assets is safeguarded, respectively. However, as mentioned in Section 4.1 these properties do not directly relate to the BTP algorithm. For example, integrity or authenticity of PTs cannot be achieved at the information level without reference to a larger infrastructure, *e.g*. a PKI, a trusted entity or a key management solution. Confidentiality and integrity are, therefore, not selected as criteria for the evaluation of BTP algorithms.

Revocability. Revocability refers to the ability to prevent verification against a PT in the future [2]. This is in general achieved through system-specific countermeasure, *e.g*., by removing a compromised reference from the system or by blacklisting it. As such, this is not a criteria for the evaluation of BTP methods.

Renewability. Renewability is defined in [2] as an umbrella term for the diversification capacity, and the irreversibility and the unlinkability aspects. Sometimes, renewability is also interpreted as the ability to update PTs, i.e., the ability to generate a new PT from an existing PT without enrolment data. In that case, the two PTs should have the same properties as PTs that are both generated from the same (or close) enrolment data. The ability to update a PT is not considered as an independent criterion, but an additional dimension in which the other criteria, such as diversification or unlinkability, should be evaluated.

Data separation. Evaluation of the criteria defined above implies the assumption that the full PTs are known by the attacker. Besides PI and AD, the PI encoding, recoding, comparison and decision procedures are assumed to be known by the attacker. This is the whitebox attack model. However, a potential separation of the AD and PI cannot be ignored and the BTP criteria should be evaluated in function of the data available to an attacker: the PI alone; the AD alone; and the combination of the PI and the AD. Also, parts of the AD can be separated, e.g., a random seed stored separately in a physical token (e.g. [34]).

Strong and weak variants. Some BTP methods involve some secret during the enrolment and verification procedures and in many schemes the PI is the cryptographic hash of this secret. In correspondence with the properties proposed by Ballard *et al*. [4], we will refer to strong variants of the evaluation criteria if the secret is known to an adversary and weak variants if the secret is not known.

5. Operational performance

Having defined a set of criteria with regards to technical performance and protection performance, we now look at the operational aspects of BTP algorithms.

5.1. Modality independence

Depending on the biometric modality different data representations may be used in a system. This has an impact on the applicability of BTP algorithms. While a fixed-length binary string lends itself easily to be used in combination with traditional cryptographic algorithms and error-correcting codes, many modalities cannot easily be quantized or transformed in a fixed-length vector without a drop in accuracy. We define the following criterion to reflect this.

Definition 10 (Modality Independence)**.** *The flexibility of dealing with different biometric modalities or data representations*

The metric to evaluate this is a simple checklist that presents the supported biometric modalities together with a reference to a proof of implementation.

5.2. Interoperability

Definition 11 (Interoperability)**.** *The degree to which standardized biometric data interchange formats are supported by the BTP algorithm*

At the PT level it is difficult to realize interoperability since BTP algorithms are different in feature extraction and protecting steps. A typical existing standardized biometric data interchange format is the minutiae feature set for fingerprint modality. The metric for interoperability is again a simple checklist.

5.3. Variation of criteria

The performance of certain criteria in a biometric system may vary in function of algorithm parameters and influential (biological, social or environmental) factors in different ways. We define the following criterion to reflect this.

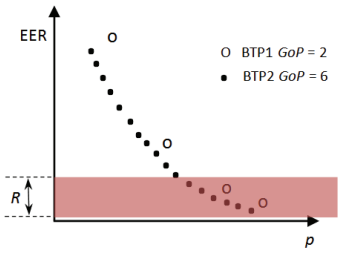

Figure 4. Granularity of performance in the EER range R.

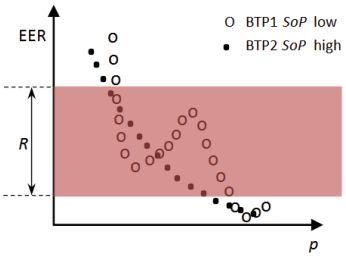

Figure 5. Stability of performance in the EER range R.

Definition 12 (Quality of performance (QoP))**.** *The ability to obtain fine granular and stable performances.*

The variation in performance due to algorithm parameters and influential factors may have two types of quality of performance aspects: granularity of performance (GoP) and stability of performance (SoP).

Definition 13 (Granularity of performance)**.** *Density of the points in a performance curve in a defined dynamic range of an algorithm parameter or influential factor, with the continuous curve as the finest case .*

Granularity of performance can be a criterion for the BTP algorithm's ability to obtain fine performance points (*e.g.*, error rate, template size, security level, etc.) over the target set of biometric subjects. In most applications, a finer curve of performance is desired to make the algorithm more adaptable to variable system requirements, *e.g.*, to achieve a target Equal Error Rate more accurately, or to achieve a precise template size that fully exploits the available storage resources. The main problem with certain BTP methods, such as biometric cryptosystems, is that they operate at a single (or a few disjoint) operating points, *e.g.* FRR/FAR points. Because they do not output a comparison score it may be impossible to generate a continuous ROC curve. For example, schemes based on error-correcting codes are limited by the number of available codes. Figure 4 illustrates GoP.

Definition 14 (Stability of performance)**.** *Degree to which a performance curve varies in a defined dynamic range of an algorithm parameter or influential factor.*

Stability of performance can be a criterion for the BTP algorithm's ability to obtain stably changing performance points (*e.g.*, error rate, template size, security level, etc.) over the target set of biometric subjects. In most applications, a more stable curve of performance is desired to make the algorithm more robust to variable parameter setting or environmental factors, *e.g.*, to achieve a robust accuracy over a wide dynamic range of sample quality, or to achieve a stable thus predictable irreversibility score curve over the interested accuracy range. Figure 5 illustrates SoP.

5.4. Criteria dependencies

The variation of criterion as a function of algorithm parameters can be used to tune a system in function of some application requirements. The best known example is the choice of comparison thresholds to reach a certain FAR. As a consequence, the typical tradeoff is observed between FAR and FRR. In a similar way, tradeoffs could be observed between other criteria and visualized using traditional Detection Error Trade-off (DET) and Receiver Operating Characteristics (ROC) curves. The extension of such approach to the protection performance criteria, like irreversibility and unlinkability, provides insights on the performances that can be achieved.

6. Conclusion

In order to assess BTP algorithms, which claim to be capable of protecting biometric templates, technical efforts for the evaluation of such algorithms are needed. These efforts include the definition of evaluation criteria, metrics and testing methodologies. In this paper we have presented a list of criteria and some metrics that are relevant for BTP and we have defined them in a standardized reference architecture.

It should be noted that metrics for the proposed criteria would only provide a distance function to measure the BTP algorithms' differences in each criterion but do not indicate utility, which should be based on the goals of particular application. Hence, no good or bad evaluation conclusion should be drawn solely by the performance value or score measured using the metrics for each criterion. A score unification strategy may be useful to facilitate direct ranking of different algorithms. The strategy would then provide a final score based on target performances and weights that depend on a particular application context. The further elaboration of metrics and the definition of particular application profiles are the next steps towards benchmarking activities for template protection.

References

[1] FVC Ongoing: Secure Template Fingerprint Verification. https://biolab.csr.unibo.it/fvcongoing/UI/Form/BenchmarkAreas/BenchmarkAreaSTFV.aspx.

[2] ISO/IEC 24745 Information technology - Security techniques - Biometric information protection, 2011.

[3] ISO/IEC DIS 2382-37 Information technology - Vocabulary - Part 37: Harmonized Biometric Vocabulary, 2011.

[4] L. Ballard, S. Kamara, and M. K. Reiter. The practical subtleties of biometric key generation. In *USENIX Security '08*, pages 61–74, 2008.

[5] T. Boult, W. Schdrer, and R. Woodworth. Revocable fingerprint biotokens: Accuracy and security analysis. In *CVPR '07*, pages 1–8. IEEE, june 2007.

[6] J. Bringer, H. Chabanne, D. Pointcheval, and Q. Tang. Extended private information retrieval and its application in biometrics authentications. In *CANS 2007*, volume 4856 of *LNCS*, pages 175–193. Springer, 2007.

[7] J. Bringer, H. Chabanne, D. Pointcheval, and S. Zimmer. An application of the Boneh and Shacham group signature scheme to biometric authentication. In *IWSEC 2008*, volume 5312 of *LNCS*. Springer, 2008.

[8] I. Buhan, J. Breebaart, J. Guajardo, K. de Groot, E. Kelkboom, and T. Akkermans. A quantitative analysis of cross-matching resilience for a continuous-domain biometric encryption technique. In *SPEED 2009*, 2009.

[9] I. Buhan, J. Doumen, P. Hartel, Q. Tang, and R. Veldhuis. Embedding renewable cryptographic keys into noisy data. *Int. J. Information Security*, 9:193–208, 2010.

[10] I. Buhan, J. Doumen, P. H. Hartel, and R. N. J. Veldhuis. Fuzzy extractors for continuous distributions. In *ASIACCS '07*, pages 353–355. ACM, 2007.

[11] C. Cachin. *Entropy Measures and Unconditional Security in Cryptography*. PhD thesis, Swiss Federal Institute of Technology Zürich, 1997.

[12] A. Cavoukian and A. Stoianov. Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy. March 2007. http://www.ipc.on.ca/images/resources/bio-encryp.pdf.

[13] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *EUROCRYPT 2004*, pages 523–540. Springer, 2004.

[14] GenKey. System, portable device and method for digital authenticating, crypting and signing by generating short-lived cryptokeys – US Patent 2006/0198514A1, 2006.

[15] T. Ignatenko and F. M. J. Willems. Biometric systems: Privacy and secrecy aspects. *IEEE Tr. on Inf. For. and Sec.*, 4(4):956–973, 2009.

[16] A. K. Jain, K. Nandakumar, and A. Nagar. Biometric template security. *EURASIP J. on Advances in Signal Processing*, 2008(Article ID 579416):17, 2008.

[17] A. Juels and M. Sudan. A fuzzy vault scheme. In *ISIT 2002*, page 408. IEEE, 2002.

[18] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *CCS '99*, pages 28–36. ACM, 1999.

[19] E. Kelkboom, J. Breebaart, T. Kevenaar, I. Buhan, and R. Veldhuis. Preventing the decodability attack based cross-matching in a fuzzy commitment scheme. *IEEE Tr. on Inf. For. and Sec.*, 6(1):107–121, march 2011.

[20] Q. Li, Y. Sutcu, and N. Memon. Secure sketch for biometric templates. In *ASIACRYPT 2006*, volume 4284 of *LNCS*, pages 99–113. Springer, 2006.

[21] J.-P. M. G. Linnartz and P. Tuyls. New shielding functions to enhance privacy and prevent misuse of biometric templates. In *AVBPA 2003*, volume 2688 of *LNCS*, pages 393–402. Springer, 2003.

[22] A. Nagar and A. Jain. On the security of non-invertible fingerprint template transforms. In *WIFS 2009*, pages 81–85. IEEE, 2009.

[23] A. Nagar, K. Nandakumar, and A. K. Jain. Biometric template transformation: a security analysis. In *SPIE 7541*, 2010.

[24] K. Nandakumar, A. Jain, and S. Pankanti. Fingerprint-based fuzzy vault: Implementation and performance. *IEEE Tr. on Inf. For. and Sec.*, 2(4):744–757, 2007.

[25] S. Prabhakar, S. Pankanti, and A. K. Jain. Biometric recognition: Security and privacy concerns. *IEEE Security and Privacy*, 1(2):33–42, 2003.

[26] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle. Generating cancelable fingerprint templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):561–572, 2007.

[27] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems J.*, 40(3):614–634, 2001.

[28] C. Rathgeb and A. Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP J. on Information Security*, 2011(3), 2011.

[29] W. Scheirer and T. Boult. Cracking fuzzy vaults and biometric encryption. In *Biometrics Symposium*, pages 1–6, 2007.

[30] K. Simoens, P. Tuyls, and B. Preneel. Privacy weaknesses in biometric sketches. In *S&P 2009*, pages 188–203. IEEE, 2009.

[31] A. D. Smith. *Maintaining Secrecy when Information Leakage is Unavoidable*. PhD thesis, Massachusetts Institute of Technology, August 2004.

[32] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. Kumar. Biometric encryption using image processing. In *Optical Security and Counterfeit Deterrence Techniques II*, volume 3314, pages 178–188. SPIE, 1998.

[33] Y. Sutcu, H. T. Sencar, and N. Memon. A secure biometric authentication scheme based on robust hashing. In *MM&Sec '05*, pages 111–116. ACM, 2005.

[34] A. Teoh, A. Goh, and D. Ngo. Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):1892–1901, 2006.

[35] A. Teoh and C. T. Yuang. Cancelable biometrics realization with multispace random projections. *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, 37(5):1096–1106, 2007.

[36] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G. J. Schrijen, A. M. Bazen, and R. N. J. Veldhuis. Practical biometric authentication with template protection. In *AVBPA*, volume 3546, pages 436–446. Springer, 2005.

[37] B. Yang, C. Busch, P. Bours, and D. Gafurov. Robust minutiae hash for fingerprint template protection. In *Media Forensics and Security*, volume 7541. SPIE, 2010.

[38] B. Yang, D. Hartung, K. Simoens, and C. Busch. Dynamic random projection for biometric template protection. In *BTAS 2010*, pages 1–7. IEEE, 2010.