RANDOMLY DRIVEN FUZZY KEY EXTRACTION OF UNCLONABLE IMAGES

Saloomeh Shariati[1], *Laurent Jacques*[1], *François-Xavier Standaert*[1], *Benoit Macq*[1], *Mohamed Amin Salhi*[2], *Philippe Antoine*[2].

[1]Information and Communication Technologies, Electronics and Applied Mathematics, [2] Institute of Condensed Matter and Nanosciences. Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.

ABSTRACT

In this paper, we develop an adjustable Fuzzy Extractor using the Physical Unclonable Functions (PUF) obtained by a common laser engraving method to sign physical objects. In particular, a string (or helper data) is generated by XORing a binary reduction of the PUF observation with the encoding of a randomly generated key, or *identifier*. Since the binary reduction (or hash) relies on keeping the sign of few random projections of the observation, a measure concentration property bounds, with a controlled accuracy, the distance between two different hashes in function of this of the original images. The error correcting code used to encode the identifier stabilizes therefore both the observation noise and the hashing distortion. In a verification stage, reobserving the PUF with the helper data in hand allows one to authenticate the object if the identifier can be exactly recovered. We conclude this work by calibrating and challenging the system on a database of laser-written PUFs, balancing helper data size, that is, hashing dimensions, and system security.

Index Terms— Fuzzy Extractor, Unclonability, Random Projection, Measure Concentration, Hamming Distance.

1. INTRODUCTION

Physical Unclonable Functions, or PUFs, have been proposed by Pappu et al. in 2002 [11] as a cost-effective way to produce unclonable tokens for identification. Several physical systems are known on which PUFs can be based. The main types are optical PUFs [11], coating PUFs [7], and silicon PUFs [6]. The common point is that the function (such as a signal or an image) can only be evaluated with the physical system, and is unique for each physical instance. The identification information is contained in an inexpensive, randomly produced, and highly complicated piece of material.

In this work, we design a Fuzzy key Extractor [5] on the observations of a new *laser-written PUF* (or LPUF) technique, or more precisely, from the 2-D topography (or *mark*) of the physical engraving obtained by a laser beam. In short, during the registration of an engraved object, a randomly generated identifier is masked by laser mark observation to produce the *helper data*. Any further authentication (or verification) of the object can be realized by checking if a new observation of the PUF together with the helper data may recover the identifier.

We propose here to build the helper data from a binary dimensionality reduction (or hashing) of the laser marks. It proceeds by keeping only the sign of few random projection of the observed mark. Thanks to a particular measure concentration phenomenon [9], the (Hamming) distance between hashes of two different images is close with high probability to the angle made by these images, with a proximity and a probability accurately controlled by the number of projections. The impact of this hashing method is therefore to provide an adjustable system where a balance between the helper data size and the extractor security may be realized.

The paper is structured as follows. In Section 2, a simple Unclonability model is introduced to characterize PUF observations. Section 3 explains the laser engraving method and assesses its unclonability. Section 4 describes the whole Fuzzy Selector and the particular binary dimensionality reduction that drives it. Finally, Section 5 calibrates and tests the whole system on a database of LPUFs.

2. OBSERVATION AND UNCLONABILITY MODELS

Let us define a simplified *discrete observation model* of *resolution* N where the reading of a real PUF provides a vector $x \in \mathbb{R}^N$ of N samples (or pixels for 2-D PUF[1]). This vector is assumed to be the recording of a *pure* PUF observation $\tau \in \mathbb{R}^N$ corrupted by an additional unknown noise $n \in \mathbb{R}^N$, that is, $x(\tau) = \tau + n$.

In our observation space of resolution N, we select a metric $\mu : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^*_+$ that serves us to compare different vectors $x, y \in \mathbb{R}^N$. We assume it unnormalized with the dimension N. This metric could be the Euclidean distance $\mu(x,y) = \|x-y\|$, with $\|x\|^2 = \sum_i |x_i|^2$ the ℓ^2-norm, or, for vectors restricted to the unit sphere $S^{N-1} = \{x \in \mathbb{R}^N : \|x\| = 1\}$, the *normalized angle* $\mu(x,y) = \widehat{xy} \triangleq$

SS and MAS are funded by the Walloon Region. LJ is a Postdoctoral Researcher of the Belgian National Science Foundation (F.R.S.-FNRS). LJ thanks Dr Abdellatif Zaidi (TELE/UCL) for his help in ECC characterization.

[1]Images are represented as vectors, for instance, by concatenating their rows.

$\frac{1}{\pi}\arccos\langle x,y\rangle \in [0,1]$, with $\langle x,y\rangle = \sum_i x_i y_i$ the common scalar product in $\mathbb{R}^N$.

In this paper, we characterize unclonability with respect to the observation space and the metric μ by the following model:

(i) the pure observation τ of the PUF at a resolution N is the realization of a random vector $\mathcal{T} \in \mathbb{R}^N$ of probability density function $f_{\mathcal{T}} : \mathbb{R}^N \to \mathbb{R}$,

(ii) the *probability of coincidence* at radius $\epsilon > 0$ is bounded by $\mathcal{P}_{\rm co}(\epsilon) = \mathbb{P}[\mu(\mathcal{T}_a, \mathcal{T}_b) \leq \epsilon] \leq \eta$, where $\mathcal{T}_a, \mathcal{T}_b \sim \mathcal{T}$ and $\eta = \eta(\epsilon; N, \mu)$ is a non-decreasing function of ϵ with $\eta(0) = 0$.

The point *(i)* $\mathcal{T}$ represents actually the variability of the (continuous) PUF creation from its subsequent pure observation, while *(ii)* really models the unclonability by considering the possibility of coincidence. Since this latter should be avoided, we expect that a efficient PUF must minimize $\eta(\epsilon; N, \mu)$ over a large range $\epsilon \in [0, \epsilon^*]$.

3. UNCLONABILITY BY LASER MARKING

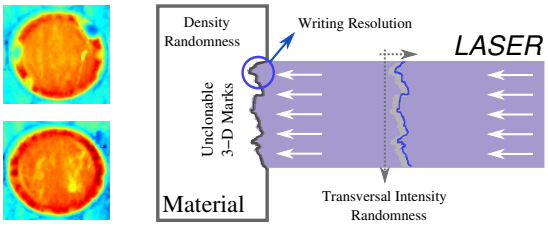

Fig. 1*:* (a) Two laser marks. (b) Laser engraving principle

Physical systems that are produced by an uncontrolled production process, *i.e.*, one that contains some intrinsic randomness, turn out to be good candidates for PUFs [13] since they tend to decrease $\mathcal{P}_{\rm co}$. In this paper, as a technological achievement of the TOMO3D [14], we propose a PUF based on the 3-D profile (or *topography*) of laser marks, with a diameter of 60μm, engraved on the surface of a physical object.

The uncontrollability of the laser marking process is mainly caused by laser instability and characteristic of the object material. The mark profile shows therefore a spatial variability that cannot be reproduced, at least with reasonably inexpensive technology. Fig. 1(b) shows the laser engraving principle illustrating the two main sources of randomness. To exploit randomness, it requires to measure the profile with a (reading) resolution finer than the laser beam diameter. In our scheme, this reading (not illustrated) is performed by White Light Interferometry (WLI) that achieves a sub-micrometer transverse resolution and a nanometer longitudinal resolution [10]. Typical WLI observations of two different laser marks are given in Fig. 1(a).

For this paper, we have recorded a database $\mathcal{S} = \{x_{pq} = \tau_p + n_{pq},\ 1 \leq p \leq P,\ 1 \leq q \leq Q\} \subset \mathbb{R}^N$ of $P = 20$ different marks τ_p of $N = 115\,600$ pixels (340×340) observed $Q = 10$ times each in x_{pq} with an unknown noise $n_{pq} \in \mathbb{R}^N$. Notice that each image has been preprocessed to reduce misalignment, misorientation and gain variations. We have also renormalized all images so that $\|x_{pq}\| = 1$.

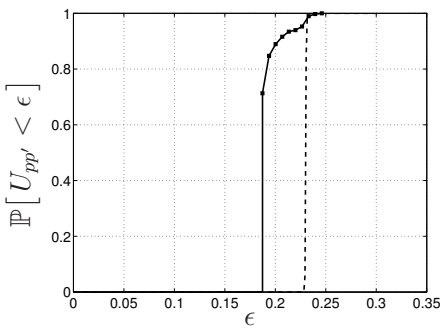

Fig. 2*:* (plain) Cumulative histogram of $U_{pp'}$, *i.e.*, $\mathbb{P}[U_{pp'} < \epsilon] \simeq \mathcal{P}_{\rm co}$. (dashed) Cumulative histogram for $\mathbb{P}[\mathcal{R} < \epsilon]$.

Even if the database is relatively small ($\#\mathcal{S} = 200$), we may estimate $\mathcal{P}_{\rm co}$ for a given metric μ by computing the cumulative histogram of $U_{pp'} = \frac{1}{L}\sum_{q,q'} \mu(x_{pq}, x_{p'q'})$ for the $L = \binom{P}{2} = 190$ choice of distinct pairs of marks $p \neq p'$. The averaging in this estimation reduces the effect of the observation noise, that is, $U_{pp'} \simeq \mu(\tau_p, \tau_{p'})$. In Fig. 2, the plain curve represents $\mathbb{P}[U_{pp'} < \epsilon] \simeq \mathcal{P}_{\rm co}(\epsilon)$ in function of ϵ under the observation conditions defined above and with $\mu(x,y) = \widehat{x\,y}$. For comparison, the dashed curve represent the $\mathbb{P}[\mathcal{R} < \epsilon]$ where $\mathcal{R} = \mu(\mathcal{T}_a, \mathcal{T}_b)$ where $\mathcal{T}_a, \mathcal{T}_b \sim \mathcal{T}$ and $\mathcal{T} \in \mathbb{R}^N$ is a random vector such that all its components are identically and independently (*iid*) drawn from a uniform random variable in $[0,1]$. The sharp (phase) transition of the dashed curve around 0.23 may be explained by concentration measure arguments [9] and actually the slope of the transition increases with N. We see that a slight discrepancy exists between the plain and the dashed curve. The LPUF are indeed not purely random and certain structures arise in their generation. However, a similar gap without any occurrence of $U_{pp'} < \epsilon$ is detected between 0 and $\epsilon' = 0.18$. For larger database, ϵ' would for sure decrease, but we expect that $\mathcal{P}_{\rm co}$ will remain negligible over a stable interval $[0, \epsilon^*]$. A closer inspection of this behavior will be realized in the future with a larger database of laser marks.

4. RANDOMLY DRIVEN FUZZY KEY EXTRACTION

The simple image of a PUF cannot be used as an identifier of the marked object. First, as described previously, different observations of the same PUF are subject to noise (for LPUFs, this is due to small variations in the WLI measurement setup or because of slight mark degradations) and therefore they cannot be perfectly reproduced. Second, any observation needs to be reduced and digitized in a limited binary string, or *fingerprint*, for further comparison, storage or transmission of data for object authentication. Finally, the ex-

tracted fingerprints from a set of similar objects may not produce the uniform distributions required by most of the cryptographic applications.

In this Section we present a global scheme build around a Fuzzy Selector [5] that solves these different issues. This scheme relies on the use of a specific binary dimensionality reduction explained hereafter.

4.1. Binary Dimensionality Reduction

The measure concentration phenomenon [9] provides a certain number of random constructions of linear mappings $f : \mathbb{R}^N \to \mathbb{R}^M$ from a high dimensional space $\mathbb{R}^N$ to a smaller space of dimension M such that f preserves approximately and with a certain probability the Euclidean norm of the projected vectors. Mathematically, there exist a constant $c > 0$ and a non-decreasing function η (with $\eta(0) = 0$) such that, for any $x \in \mathbb{R}^N$,

$$\mathbb{P}_\Phi\big[\,\big|\,\|f(x)\| - \|x\|\,\big| \geq \epsilon\,\|x\|\,\big] \;\leq\; c\,e^{-\eta(\epsilon)M}. \qquad (1)$$

Interestingly, a random matrix $\Phi = (\Phi_{ij}) \in \mathbb{R}^{M\times N}$ such that $\Phi_{ij} \sim_{\text{iid}} N(0, 1/M)$ (Gaussian matrix) or $\Phi_{ij} \sim_{\text{iid}} \pm 1/\sqrt{M}$, both with $\eta(\epsilon) = \epsilon^2/4 - \epsilon^3/6$, and many other sub-Gaussian distributions respect this concentration property.

The concentration (1) provides linear dimensionality reduction between real-valued domains. In our application, since we develop a binary identifier extraction, we search for a mapping between $\mathbb{R}^N$ and the space of M-bits string $B^M \triangleq \{0,1\}^M$. Andoni and Indyk have proposed the Locally Sensitive Hashing methods (LSH) for that purpose [1]. We use however a simpler method, the *binary dimensionality reduction*, defined by

$$h : \mathbb{R}^N \to B^M,\; x \mapsto \overline{x} = h(x) = \text{sign}_\text{b}(\Phi x),$$

where $\Phi \in \mathbb{R}^{M\times N}$ is a Gaussian random matrix, $\text{sign}_\text{b}(\lambda) = 1$ if $\lambda > 0$ and 0 else, and for $u \in \mathbb{R}^M$, $\text{sign}_\text{b}(u) \in B^M$ with $(\text{sign}_\text{b}(u))_i = \text{sign}_\text{b}(u_i)$ for $1 \leq i \leq M$.

Interestingly, the non-linear mapping h satisfies also a certain measure concentration with respect to the angle made by two distinct vectors and the (normalized) Hamming distance of their projections, that is, the distance $d_H(u, v) = \frac{1}{M}\sum_{i=1}^M u_i \oplus v_i \in [0, 1]$ that counts the number of distinct bits between two strings $u, v \in B^M$ with the XOR operation $\oplus$. Indeed, for any $x, y \in \mathbb{R}^N$ and $\epsilon > 0$,

$$\mathbb{P}_\Phi\big[\,\big|\,d_H(\overline{x},\overline{y}) - \widehat{x\,y}\,\big| \geq \epsilon\,\widehat{x\,y}\,\big] \;\leq\; 2\,e^{-2\epsilon^2 M}. \qquad (2)$$

This may be proved by first showing that, thanks to the isotropy of the Gaussian distribution and whatever the component $1 \leq i \leq m$, $\mathbb{P}[h_i(x) \oplus h_i(y) = 1] = \mathbb{P}[h_i(x) \neq h_i(y)] = \pi^{-1}\,\widehat{x\,y}$, as given by the probability that x and y be separated by a plane normal to $\varphi_i = (\Phi_{ij})_{1\leq j\leq N}$ and containing the origin, and second, by observing therefore that $M\,d_H(\overline{x},\overline{y})$ follows a binomial distribution $B(M, p)$ of M trials of success probability $p = \widehat{x\,y}$ that exponentially concentrates around its mean Mp when M increases [8, 3].

Another useful reading of Eq. (2) is as follows. Given the dimensions N and M, if we fix a level of failure $\eta > 0$, for two point $x, y \in \mathbb{R}^N$, we have $\big(1 - \epsilon(M)\big)\,\widehat{x\,y} \leq d_H(\overline{x},\overline{y}) \leq \big(1 + \epsilon(M)\big)\,\widehat{x\,y}$, with a probability higher than $1 - \eta$ if we accept the distortion

$$\epsilon(M) = \sqrt{(\ln 2/\eta)/(2M)}. \qquad (3)$$

4.2. Fuzzy Key Extraction

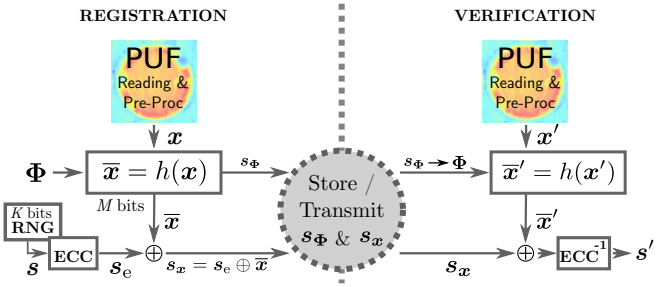

Fig. 3*:* Randomly Driven Fuzzy Extraction

We propose to combine the binary dimensionality reduction of Section 4.1 with a Fuzzy Extraction procedure composed of a *registration* and a *verification* steps. This association creates an adjustable, robust and secure object signature from the unclonability of a LPUF.

The pipeline of this procedure is summarized in Fig. 3. During the *registration* a LPUF is first engraved on the object to be marked. A random matrix Φ is generated in a reproducible way, that is, there exists a string $s_\Phi \in B^{M'}$ such that $\Phi = \Phi(s_\Phi)$ can be reproduced from it (*e.g.*, by using few seeds in a pseudo-random generation procedure). A WLI observation $x \in \mathbb{R}^N$ is obtained from the PUF with some preprocessing aiming at reducing observation noises (such as realignment and reorientation from some landmarks). This image is hashed into a M bits string $\overline{x} = h(x)$ using Φ. In parallel, an identifier s of $K < M$ bits is generated by a Random Number Generator (RNG) and *extended* into a string $s_\text{e} \in B^M$ by an Error Correcting Code encoding (ECC) of error-correction capability $T < K < M$. Finally, $\overline{x}$ is XORed ($\oplus$) with s_e to form $s_x = s_\text{e} \oplus \overline{x} \in B^M$. The helper data of $M + M'$ bits is composed of (s_Φ, s_x). This data constitutes, together with the PUF and the identifier s, the information required to authenticate the object.

During the *verification* stage, the PUF is reobserved into a image $x' \in \mathbb{R}^N$ undergoing the same preprocessing. From the helper data (s_Φ, s_x), Φ is regenerated and used to compute $\overline{x}'$. After XORing this later with s_x and sending the output to the ECC decoder (ECC^{-1}), a final K-bits string s' is produced. Since $(u \oplus v) \oplus v = u$ for any two strings $u, v \in B^M$, if the ECC capability T is set higher than the maximal hashing distortion of different observations of the same PUF, and if we

do observe the same object with the same unclonable PUF, then the system guarantees $s = s'$ [5]. The security of the whole system under a copy-paste attack (namely against the unclonability) may be assessed by evaluating the probability that $s = s'$ conditionally to the observation of different PUFs.

5. EXPERIMENTS

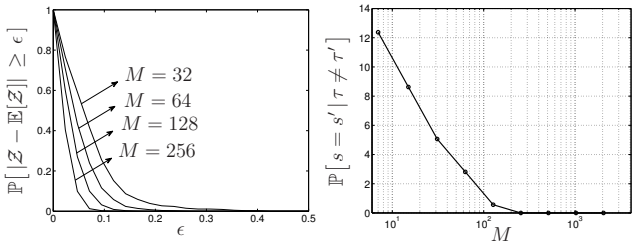

Fig. 4*:* (left) Concentration of $\mathcal{Z} = \widehat{x\,y}/d_H(\overline{x}, \overline{y})$. (right) False Acceptance Rate (in $\%$) of the Fuzzy Selector.

The system of Section 4 has been tested with the database $\mathcal{S}$ of LPUFs described in Section 3. In particular, we have first selected the versatile BCH Error Correcting Code procedure [12]. Second, the hashing method has been realized with another random matrix construction, the Random Noiselet Ensemble (RNE) [4, 2]. The matrix Φ corresponds to picking uniformly at random M "frequencies" in a orthonormal Noiselet transform of $\mathbb{R}^N$ [4]. This induces both a fast evaluation of $\overline{x}$ from $x \in \mathbb{R}^N$ in $O(N \log N)$ computations (compared to $O(MN)$ for Gaussian hashing) and a reproducibility of Φ by recording in s_Φ the M selected frequencies[2]. Moreover, Noiselets are optimal for signals x described with few coefficients in standard Wavelet basis (like Haar or Daubechies 7/9) [2], a property that the laser marks respects due to the structure they contain (see Fig. 1(a)). The left plot on Fig. 4 confirms by a Montecarlo simulation over 1000 randomly generated Φ and $x, y \in \mathbb{R}^N$ that the random variable $\mathcal{Z} = \widehat{x\,y}/d_H(\overline{x}, \overline{y})$ is well centered around its mean. This confirms the existence of a concentration property for the RNE hashing with a distortion respecting (3). Third, the level of observation noise in $\mathcal{S}$ has been estimated by computing $H_0 = \max_{p,q \neq q'} \widehat{x_{pq}\,x_{pq'}} = 0.0775$. Finally, by setting a probability failure $\eta = 0.01$, the BCH-ECC capability has been tuned to $T = \lceil M\,(1 + \epsilon(M))\,H_0 \rceil$, while the length K of the random identifier has been set to the maximal value allowed by BCH given M and T. This last point guarantees that, with a probability higher than $1 - \eta = 0.99$, the same PUF presented at the registration and at the verification leads to recover the identifier, *i.e.*, $s = s'$. This fact has been confirmed experimentally.

An interesting challenge amounts to test when different PUFs may produce equal identifiers. This possibility exists since for small M the hashing distortion can produce close hashes even if they are generated from different PUFs. The identifier may then be erroneously recovered since in addition the capacity of the ECC decoding increases when M decreases. This phenomenon is measured by the False Acceptance Rate $\mathbb{P}[\,s' = s \,|\, \tau' \neq \tau\,]$. It is estimated with the laser mark database $\mathcal{S}$ in the right part of Fig. 4 in function of the reduced dimensionality M. Clearly, for M bigger than $M_0 = 128 \simeq N/900$, this rate is considered as negligible (smaller than $< .01\,\%$). A future study will have to estimate if the trade-off value M_0 between helper data size and system security is stable when the size of the testing database $\mathcal{S}$ increases.

[2]Or alternatively the seed of their pseudorandom selection.

6. CONCLUSION AND FURTHER WORK

In this work we have presented how unclonability of laser marks may be both modeled mathematically and used as a possible foundation for robust material object authentication. This motivated the construction of an adjustable fuzzy key extractor relying on a binary dimensionality reduction of controlled distortion. In the future, we will examine deeper the cryptographic properties of that scheme, compared to other methods like LSH [1]. Different random hashing methods connected to the recent field of Compressed Sensing [2] will also be assessed..

7. REFERENCES

[1] A. Andoni and P. Indyk. "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions." *Comm. ACM*, **51**(1):117–122, 2008.

[2] E. Candès and J. Romberg. "Sparsity and incoherence in compressive sampling." *Inverse Problems*, **23**(3):969–985, 2007.

[3] M. Charikar. "Similarity estimation techniques from rounding algorithms." *Proc. 34*$^{\text{th}}$ *ACM Symp. Th. Comp.*, p. 388, 2002.

[4] R. Coifman, F. Geshwind, and Y. Meyer. "Noiselets." *App. Comp. Harm. Anal.*, **10**(1):27–44, 2001.

[5] Y. Dodis, L. Reyzin, and A. Smith. "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data." *Lect. Not. Comp. Sc.*, pp. 523–540, 2004.

[6] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas. "Silicon physical random functions." *Proc. ACM Conf. Comp. Comm. Sec.,* Nov. 2002.

[7] B. Gassend, D. Clarke, M. van Dijk, S. Devadas, P. Tuyls, and E. Torlak. "Controlled physical random functions." *Proc. Ann. Comp. Comm. Sec. Appl. Conf.*, 2002.

[8] M. Goemans and D. Williamson. "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming." *Journ. ACM* , **42**(6):1145, 1995.

[9] M. Ledoux. "The concentration of measure phenomenon," *American Mathematical Society*, 2001.

[10] D. Malacara. *Optical shop testing*. Wiley, 2$^{\text{nd}}$ ed., 1992.

[11] R. Pappu, B. Recht, and N. G. J. Taylor. "Physical one-way functions." *Science*, **297**:2026–2030, 2002.

[12] J. G. Proakis. *Digital communications*. McGraw-Hill, 2001.

[13] P. Tuyls and B. Skoric. "Secret key generation from classical physics." *Proc. Hard. Tech. Dr. Amb. Intel. Symp.,* 2005.

[14] TOMO3D project, WIST2 Program contract n°616444, Wallon Region, Belgium.