PHYSICAL OBJECT AUTHENTICATION: DETECTION-THEORETIC COMPARISON OF NATURAL AND ARTIFICIAL RANDOMNESS

Slava Voloshynovskiy, Taras Holotyak

SIP, University of Geneva
Geneva, Switzerland

Patrick Bas

CNRS, CRIStAL Lab, University of Lille
Ecole Centrale de Lille, France

ABSTRACT

In this paper, we compare two methods that can be used by the anti-counterfeiting industry to protect physical objects, which are either based on an object's *natural randomness* or on *artificial randomness* embedded on the object. We show that the considered verification architectures rely either on a comparison between an enrolled fingerprint and an extracted one or between a tag and a fingerprint. We compare these setups from detection-theoretic perspectives for both types of architectures. Authentication performance using false and miss error probabilities of the two systems are analysed and then compared using two practical setups. We highlight the advantages and limitations of each architecture. These theoretical results derived for binary fingerprints are useful to construct and optimise practical methods and to help select the appropriate architecture.

1. INTRODUCTION

Anti-counterfeiting of physical objects based on digital solutions has attracted a lot of attention in the last years. This interest is caused by the urgency for technologies to deal with modern counterfeiting, which implies having access to easy, fast, reliable and user-friendly verification of objects authenticity. Moreover, objects authentication on consumer mobile phones de-facto becomes a standard solution for brand protection. In this respect, there is significant interest in cheap and simple secure methods, which are suitable for protection of various objects such as pharmaceutical products (including both drugs and its packaging), cosmetics, food, luxury goods, spare parts, as well as medical equipment, components and implants.

Recent technologies that address the above demands are based on physical uncloneable properties, *a.k.a. physical uncloneable features (PUF)*, that are easy to verify but difficult to clone, generally referred to as a *randomness*. As a matter of fact, optically visible microscopic features which exhibit random features are of special interest for mobile verification since the advent of high definition of modern imaging sensors.

Depending on the origin of randomness the anticounterfeting technologies can be divided into *natural randomness*, when the system exploits the randomness inherently created by nature, and *artificial randomness*, when the randomness is created on purpose. Natural randomness systems typically use object surface images such as shown in Figure 1a whereas artificial randomness systems are based on various uncontrolled effects occurring during marking (printing, embossing, laser engraving, etc.) of random-like structures such as *graphical codes* (GC) as shown in Figures 1b-1c.

Several papers address the practical and theoretical performance and security of natural randomness [1, 2] and artificial randomness [3–7] systems. However, to our best knowledge, there is little if no work that compares the theoretical performance of these systems using the same assumptions behind the statistical models of randomness and acquisition/clonning processes. Therefore, the goal of this paper is to compare the two systems and to analyse the potential advantages of each system.

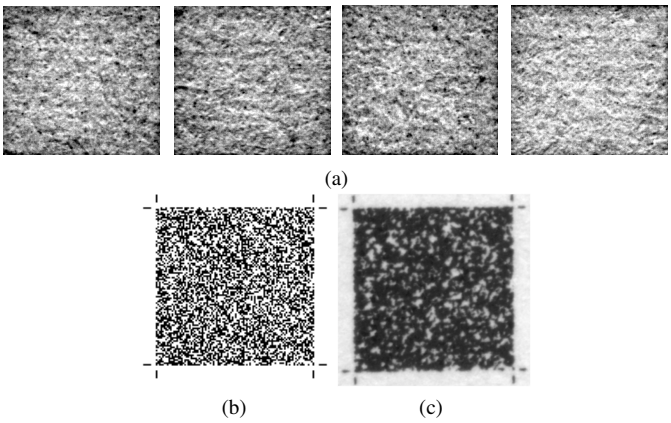


Fig. 1. Example of natural randomness (a) scanned paper fibbers at 800 dpi; and artificial randomness: (b) original digital tag representing a GC, (c) GC printed and scanned at 600dpi.

The paper is organised as follows. The problem formulation as well as the statistical models behind the natural randomness and artificial randomness systems are introduced in Section 2, and section 3 summarises the main detection-theoretic results. The comparison between the two practical setups is given in Section 4.

Notation. We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$, corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively, i.e., $\mathbf{x} = (x_1, x_2, ..., x_N)^T$. We use $X \sim p(x)$ to indicate that a random variable X follows $p_X(x)$. The sign $*$ denotes the convolution of probabilities $p * q = p(1-q) + q(1-p)$.

2. PROBLEM FORMULATION: STATISTICAL MODELS

We present here the models associated with authentication using either natural or artificial randomness.

The block diagram for systems based on natural randomness is shown in Figure 2. At the enrollment stage, the microstructure of object $\mathbf{o}$ corresponding to object index $w \in \{1, 2, \cdots, M\}$ is acquired by a device (e.g. a camera) resulting in the image $\mathbf{x}(w) \in \mathcal{R}^N$ and the extracted fingerprint $\mathbf{f}_{\mathrm{x}}(w) \in \{0,1\}^n$ which is stored in the database for future authentication. The helper data are given in the form of a tag w pointing to the fingerprint $\mathbf{f}_{\mathrm{x}}(w)$ stored in the database. In this paper, we consider only helper data in a form of a

pointer to the concerned fingerprint. However, other forms of helper data encoding developed in biometric applications are possible and the reader is referred to [8, 9] for more details.

The authentication consists in: (i) acquisition of an object microstructure by the acquisition device resulting in image $\mathbf{y}$, (ii) computation of binary $\mathbf{f}_{\mathrm{y}} \in \{0,1\}^n$ or soft $\mathbf{f}_{\mathrm{y}} \in \mathcal{R}^n$ fingerprint and (iii) making a binary decision by comparing $\mathbf{f}_{\mathrm{y}}$ to $\mathbf{f}_{\mathrm{x}}(w)$. Because of this last step, we will refer to these systems as **fingerprint-to-fingerprint** (F2F) architectures.

We will assume that the F2F setup can be modeled as memoryless source and the acquisitions are modeled by memoryless channels. The authentication model can be represented by a chain in the form of a probabilistic graphical model $F_{\mathrm{x}} \leftarrow X \leftarrow O \rightarrow Y \rightarrow F_{\mathrm{y}}$ or by the joint distribution $p(f_{\mathrm{x}}, x, o, y, f_{\mathrm{y}}) = p(o)p(x, y|o)p(f_{\mathrm{x}}|x)p(f_{\mathrm{y}}|y)$ with $p(x, y|o) = p(x|o)p(y|o)$.

The block diagram of systems based on artificial randomness is shown in Figure 3. At the enrollment stage, the object tag $\mathbf{m}(w) \in \{0,1\}^n$ in a form of any random-like modality is generated based on the index w and reproduced on the object surface resulting in object $\mathbf{o}(w)$. In the general case, the tag w is known at the verification stage too, it can be added to the object at the enrollment either as a standalone index encoded in any machine readable form or integrated directly into the object tag $\mathbf{m}(w)$. Additionally, the object tag $\mathbf{m}(w)$ can be generated pseudo-randomly from w or coded from w using error correction codes [1].

The authentication process consists in (i) the acquisition of the object's microstructure by the capturing device resulting in image $\mathbf{y}$, (ii) extraction of the tag estimate $\mathbf{f}_{\mathrm{y}} \in \{0,1\}^n$ or soft $\mathbf{f}_{\mathrm{y}} \in \mathcal{R}^n$ and (iii) making a binary decision by comparing $\mathbf{f}_{\mathrm{y}}$ to $\mathbf{m}(w)$, or decoding $\hat{w}$ based on $\mathbf{f}_{\mathrm{y}}$ and comparing the obtained estimate $\hat{w}$ to the claimed object tag w. We will refer to such artificial randomness systems as **tag-to-fingerprint** (T2F) architectures.

We assume that tag statistics are governed by a Bernoulli distribution with parameter $\Pr\{M_i = 1\} = \theta_m$, $1 \le i \le n$. The above T2F setup can be represented by a Markov chain in the form of a probabilistic graphical model $M \rightarrow O \rightarrow Y \rightarrow F_{\mathrm{y}}$ or by the joint distribution $p(m, o, y, f_{\mathrm{y}}) = p(m)p(o|m)p(y|o)p(f_{\mathrm{y}}|y)$.

Note that the physical protection based on the T2F architecture is based on a fact that any attempt of counterfeiter to clone the original object tag will lead to the additional distortions between the digital tag $\mathbf{m}(w)$ and its cloned version [2], that can be detected using an appropriate test (see section 3.2).

3. DETECTION-THEORETIC ANALYSIS

3.1. Statistical models under consideration

This analysis relies on a set of assumptions that can be summarised as follows: (a) the pmf of sources and channels are known and the sources are assumed to be memoryless and binary and the channels to be binary symmetric channel (BSC) models; (b) the synchronization between all enrolled sequences and probe is perfect.

F2F setup (Figure 4a): We consider a hypothetical model of natural randomness generated according to a Bernoulli distribution

[1] The object tag can additionally be encrypted to prevent the reverse engineering attack.

[2] Several studies [5,10,11] investigate a possibility to produce high quality clones from multiple scans of the same GC and combination of various image processing techniques targeting an accurate estimation of object tag.

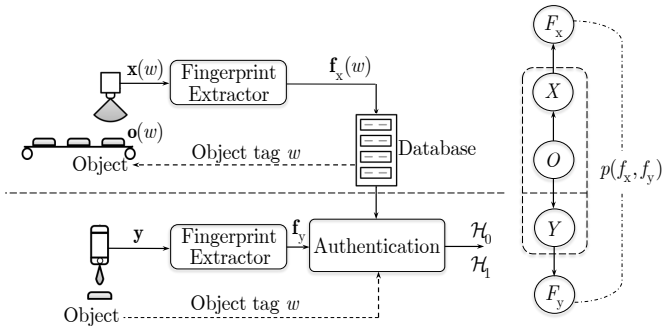


Fig. 2. Block diagram of natural randomness system: Fingerprint-to-Fingerprint (**F2F**) architecture and corresponding graphical probabilistic model.

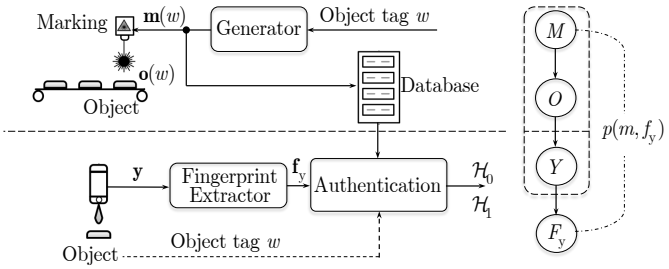


Fig. 3. Block diagram of artificial randomness system: Tag-to-Fingerprint (**T2F**) architecture and corresponding graphical probabilistic model.

with parameter $\theta_o = \Pr[F_{o_i} = 1]$, $1 \le i \le n$. Therefore, the statistical model of the source governing all samples is $F_o \sim Bern(\theta_o)$[3]. Accordingly, we assume that all corresponding components of the model representing enrolled data F_{x}, probe F_{y} for the authentic object and F'_{y} for the fake object, are binary, i.e., with the alphabet $\{0,1\}$.

The *enrollment channel* is assumed to be represented by an additive modulo-2 channel, i.e., F_o is XORed with independent enrollment noise F_{Z_e} resulting into enrolled data F_{x}:

$$F_{\mathrm{x}} = F_o \oplus F_{Z_e},$$

where the noise $F_{Z_e} \sim Bern(P_{b_e})$ and the enrolled fingerprint $F_{\mathrm{x}} \sim Bern(\theta_o * P_{b_e})$. This model also represents the BSC with the cross-over probability P_{b_e}.

The *authentic verification channel* is also assumed to be an additive modulo-2 channel with:

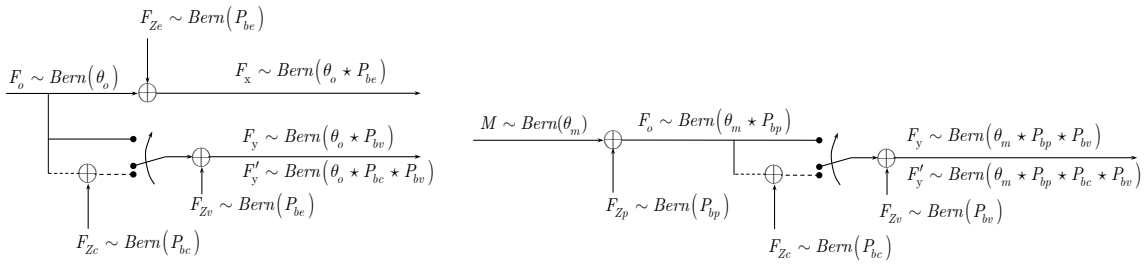
$$F_{\mathrm{y}} = F_o \oplus F_{Z_v},$$

where $F_{Z_v} \sim Bern(P_{b_v})$ and $F_{\mathrm{y}} \sim Bern(\theta_o * P_{b_v})$.

The *opponent verification channel* corresponds to the case when the opponent tries to clone the original F_o and since the process is noisy:

$$F'_{\mathrm{y}} = F_o \oplus F_{Z_c} \oplus F_{Z_v},$$

[3] The statistics of extracted binary fingerprint are determined by the parameter θ_o. In th general case, some data independent or data-dependent (learned) transform can be applied to the image $\mathbf{x}(w)$ with a following binarization to ensure the desirable value θ_o [12, 13], for example $\theta_o = 0.5$.



(a) F2F setup. (b) T2F setup.

Fig. 4. Binary statistical models used for the statistical analysis in the legitimate and opponent channels: (a) F2F setup with the model $F_x \leftarrow F_o \rightarrow F_y$ and (b) T2F setup with the model $M \rightarrow F_o \rightarrow F_y$.

where $F_{Z_c} \sim Bern(P_{b_c})$ models to cloning channel and $F'_y \sim Bern(\theta_o * P_{b_c} * P_{b_v})$.

T2F setup (Figure 4b): The statistical model of the source (the mark here) is also assumed to be generated as $\mathbf{M}(w) \in \{0,1\}^n$ with $\Pr[M_i = 1] = \theta_m$, i.e., it is generated from the Bernoulli pmf as $M_i \sim Bern(\theta_m)$. The marking process is modeled as a BSC with transition probability $\Pr[O_i \neq M_i] = P_{b_p}$. Thus the marked object is modeled as: $F_{o_i} \sim Bern(\theta_m * P_{b_p})$. The reproduction process can be equivalently represented by the additive modulo-2 channel:

$$F_o = M \oplus F_{Z_p},$$

where $F_{Z_p} \sim Bern(P_{b_p})$ stands for the marking noise.

The *authentic verification channel* is also memoryless where for each element:

$$F_y = M \oplus F_{Z_p} \oplus F_{Z_v},$$

where $F_{Z_v} \sim Bern(P_{b_v})$ and $F_y \sim Bern(\theta_m * P_{b_p} * P_{b_v})$.

The *verification channel for the opponent* is represented by:

$$F'_y = M \oplus F_{Z_p} \oplus F_{Z_c} \oplus F_{Z_v},$$

where $F_{Z_c} \sim Bern(P_{b_c})$ model the copying process and $F'_y \sim Bern(\theta_m * P_{b_p} * P_{b_c} * P_{b_v})$.

3.2. Hypothesis testing

We measure the system performance using probability of miss P_M and probability of false acceptance P_{FA}. Probability of miss P_M corresponds to the error event when authentic object with tag w is rejected by the authentication system. Probability of false acceptance P_{FA} reflects a probability of falsely accepting any fake item as an authentic one with index w.

It should be pointed out that a fake object presented for authentication might be chosen *blindly* without any reference to the original object with tag w. However, a fake might also be designed more meticulously using any available information about object $\mathbf{o}(w)$ or its features $\mathbf{f}_x(w)$. We refer to this case as an *informed attack*. The corresponding probability of false acceptance under the informed attack is denoted probability of *successful attack* P_{SA}. An informed attack is more dangerous than a blind attack since it might generate a fake object whose features are considerably closer to the features of an authentic object. For this reason, we focus on informed attacks here.

We consider this authentication problem as a binary hypothesis testing with hypothesis $\mathcal{H}_w^F$, representing the hypothesis that the presented object is not authentic (fake), and $\mathcal{H}_w$ its authentic counterpart. Moreover, we assume the worst case attack, i.e., an informed attack where the object presented for authentication under hypothesis $\mathcal{H}_w^F$ is reproduced from an authentic object. The distributions under the corresponding hypothesis are [2]:

$$\begin{cases} \mathcal{H}_w^F : p(\mathbf{f}_y|\mathbf{f}_x(w), \mathcal{H}_w^F) = P_b'^{d_H(w)}(1-P_b')^{n-d_H(w)}, \\ \mathcal{H}_w : p(\mathbf{f}_y|\mathbf{f}_x(w), \mathcal{H}_w) = P_b^{d_H(w)}(1-P_b)^{n-d_H(w)}, \end{cases} \quad (1)$$

where $d_H(w) = d_H(\mathbf{f}_y, \mathbf{f}_x(w))$. Probability of bit error for the hypothesis $\mathcal{H}_w^F$ characterises the opponent channel $P'_b = P_{b_e} * P_{b_c} * P_{b_v}$ for F2F, and $P'_b = P_{b_p} * P_{b_c} * P_{b_v}$ for T2F systems. Hypothesis $\mathcal{H}_w$ corresponds to the case of legitimate channel and is equal to $P_b = P_{b_e} * P_{b_v}$ for F2F and $P_b = P_{b_p} * P_{b_v}$ for T2F systems. Note that for blind attacks, $P'_b = 0.5$ in both cases. We also assume that the fingerprinting scheme in the case of F2F systems is designed to maximise the entropy of the source, i.e., such that $\theta_0 = 0.5$ for F2F[4] and $\theta_m = 0.5$ for the T2F.

The authentication test is performed based on rule $d_H(\mathbf{f}_y, \mathbf{f}_x(w)) \leq \gamma n$, where γ is a threshold relying on P_{SA} and P_M.

In this paper, we follow the approach proposed in [2] that considered the performance of authentication systems under the informed attacks. The probability of successful attack is defined as [2]:

$$\begin{aligned} P_{SA}(\gamma) &= \Pr\{D_H(w) \leq \gamma n | \mathcal{H}_w^F\} \\ &= \sum_{d_H=0}^{\lfloor \gamma n \rfloor} \binom{n}{d_H} P_b'^{d_H}(1-P_b')^{n-d_H}, \end{aligned} \quad (2)$$

and the probability of miss is [2]:

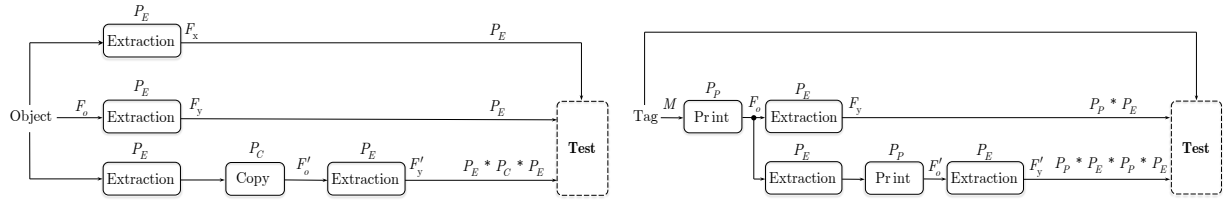
$$\begin{aligned} P_M(\gamma) &= \Pr\{D_H(w) > \gamma n | \mathcal{H}_w\} \\ &= \sum_{d_H=\lceil \gamma n \rceil + 1}^{n} \binom{n}{d_H} P_b{}^{d_H}(1-P_b)^{n-d_H}. \end{aligned} \quad (3)$$

4. COMPARISON BETWEEN T2F AND F2F SCHEMES

4.1. Setups and assumptions

The considered models are generic and allow considering different relationships between the parameters of enrollment and verification for both the legitimate user and the opponent. We consider here a conservative scenario assuming that the counterfeiter uses the same marking and acquisition equipment as a legitimate user, as well as

[4]This can be achieved even for correlated input mages by randomly projecting the input image and binary quantizing the resulting projections as shown in [14].



(a) F2F setup. (b) T2F setup.

Fig. 5. Simplified setups under analysis: (a) F2F and (b) T2F.

the same fingerprinting extraction algorithms. Other possible scenarios could include: (a) High-Quality clones, when the opponent has the high quality scanning-reproduction equipment w.r.t. the genuine enrollment, and (b) Low-Quality clones, when the situation is an inverse one. More particularly, we assume that: (a) $P_{b_e} = P_{b_v} = P_E$ to be a generic extraction probability of error, (b) $P_{b_c} = P_E * P_C$ with P_C to be a probability of reproduction in F2F systems and (c) $P_{b_c} = P_E * P_P$ with P_P to be a probability of reproduction/printing in T2F systems. We also set $\theta_0 = \theta_m = 0.5$ for the above discussed reasons. The setups for the F2F and T2F authentication methods are depicted in Figures 5a and 5b, respectively.

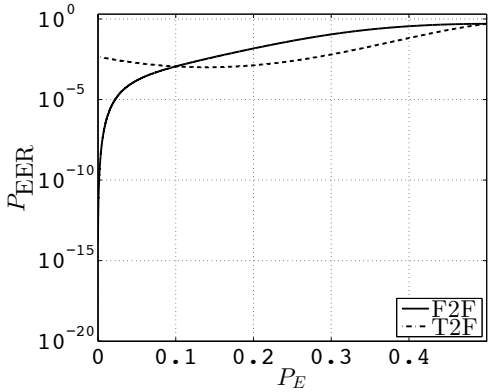
4.2. Performance of authentication systems

We compute P_{SA} and P_{M} according to (2) and (3) with: (a) F2F parameters $P_b = P_E * P_E$ and $P'_b = P_E * P_E * P_C * P_E$ and (b) T2F parameters $P_b = P_P * P_E$ and $P'_b = P_P * P_E * P_P * P_E$.

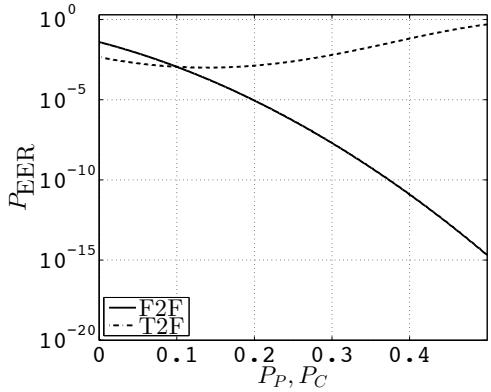
The probabilities $P_{\text{M}}(\gamma)$ and $P_{\text{SA}}(\gamma)$ depend on the selection of threshold γ. We simulated the equal-error-rate strategy popular in biometrics, i.e., $P_{\text{M}} = P_{\text{SA}} \triangleq P_{\text{EER}}$ shown in Figure 6a assuming $P_P = P_C = 0.1$ and $n = 500$. We have observed the same behavior of plots withe the threshold γ selection under the Neyman-Pearson (NP) strategy for the bounded $P_{\text{SA}}(\gamma)$[5] . For both strategies of threshold γ selection, P_{EER} for the F2F systems is lower of those of T2F systems whenever $P_E \leq P_P$. For both schemes there may exist an optimal extraction value P_E that minimizes P_{EER}. Note that the possible existence of an optimal value is due to the fact that: for low P_E, it is easier for the opponent to reproduce an accurate copy; for large P_E, the original and fake fingerprint tend to be equally noisy and are not distinct.

Finally, Figure 6b summarises the behavior of P_{EER} as a function of $P_P = P_C$ for $P_E = 0.1$. In this scenario we assume that counterfeiting devices are comparable. Accordingly, we set the extraction error to a given value $P_E = 0.1$. It is interesting to note that the behaviour of these two schemes w.r.t. the duplication error is completely different. The authentication performance decreases w.r.t. P_P (after reaching a maximum for P_E) for T2F setup but increases w.r.t P_C for F2F setup. The first phenomenon can be explained by the fact that if a printing device is highly noisy, it is difficult to distinguish between two equally noisy fingerprints. The second phenomenon can be explained by the fact that only the opponent has to use a duplication device for the F2F scheme which makes the discrimination between the fake object and the original one growing w.r.t the duplication noise. It is interesting to point out that in this case the F2F has an advantage over the T2F systems due to the independence of the legitimate channel from these parameters which only determine the opponent channel.

[5]These results are not shown due to the paper length restrictions.



(a) $P_P = P_C = 0.1$



(b) $P_E = 0.1$

Fig. 6. Comparison of performance of F2F and T2F systems.

5. CONCLUSIONS AND PERSPECTIVES

In this paper, we have presented and compared the F2F and T2F systems from the detection-theoretic perspectives. These systems can model a large variety of practical scenarios used for object identification and authentication. For conservative scenarios, we demonstrate that F2F authentication systems have lower equal error rate than T2F systems for large reproduction errors. At the same time, we have found that T2F systems may reach a minimum for a given parameters of extraction and printing. In our future research, we intend to consider the impact of security and information theoretic oriented constraints such as the unclonability of the fingerprint, the information leakage of the secret parameter in the T2F setup or the identification capacity associated to the authentification system.

6. REFERENCES

[1] S. Voloshynovskiy, M. Diephuis, F. Beekhof, O. Koval, and B. Keel, "Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (famos)," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5 2012.

[2] F. Beekhof, S. Voloshynovskiy, and F. Farhadzadeh, "Content authentication and identification under informed attacks," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5 2012.

[3] S. Shariati, F. Standaert, L. Jacques, B. Macq, M. Salhi, and P. Antoine, "Random profiles of laser marks," in *Proceedings of the 31st WIC Symposium on Information Theory in the Benelux*, 2010.

[4] A. T. Phan Ho, B. A. Hoang Mai, W. Sawaya, and P. Bas, "Document Authentication Using Graphical Codes: Reliable Performance Analysis and Channel Optimization," *EURASIP Journal on Information Security*, pp. 10.1186/1687–417X–2014–9, Jun. 2014.

[5] C. Baras and F. Cayre, "Towards a realistic channel model for security analysis of authentication using graphical codes," in *Information Forensics and Security (WIFS), 2013 IEEE International Workshop on*. IEEE, 2013, pp. 115–119.

[6] T. Haist and H. J. Tiziani, "Optical detection of random features for high security applications," *Optics communications*, vol. 147, no. 1, pp. 173–179, 1998.

[7] B. Zhu, J. Wu, and M. S. Kankanhalli, "Print signatures for document authentication," in *Proceedings of the 10th ACM conference on Computer and communications security*. ACM, 2003, pp. 145–154.

[8] P. Tuyls, B. Skoric, and T. K. (Eds.), *Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer, 2007.

[9] T. Ignatenko and F. Willems, "Privacy leakage in biometric secrecy systems," in *46th Annual Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, IL, USA, 23-26 Sept. 2008, pp. 850–857.

[10] L. Diong, P. Bas, C. Pelle, and W. Sawaya, "Document authentication using 2D codes: Maximizing the decoding performance using statistical inference," in *Communications and Multimedia Security*, United Kingdom, Sep. 2012, p. tba. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00728161

[11] C. Baras and F. Cayre, "2D bar-codes for authentication: A security approach," *Proceedings of EUSIPCO 2012*, 2012.

[12] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *IEEE Information Theory Workshop, ITW2010*, Dublin, Ireland, Aug.30-Sep.3 2010.

[13] S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Pun, "Conception and limits of robust perceptual hashing: toward side information assisted hash functions," in *Proceedings of SPIE Photonics West, Electronic Imaging / Media Forensics and Security XI*, San Jose, USA, 2009.

[14] F. Farhadzadeh, S. Voloshynovskiy, and O. J. Koval, "Performance analysis of content-based identification using constrained list-based decoding," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1652–1667, 2012.