INFORMATION-THEORETIC ANALYSIS OF DESYNCHRONIZATION INVARIANT OBJECT IDENTIFICATION

Oleksiy Koval, Sviatoslav Voloshynovskiy, Farzad Farhadzadeh, Taras Holotyak and Fokko Beekhof

University of Geneva
Department of Computer Science
7, route de Drize, CH 1227, Geneva, Switzerland

ABSTRACT

This paper is dedicated to the analysis of object identification under desynchronization distortions. While this class of degradations is almost unavoidable in the functionality of such systems especially at the verification stage via acquisition imperfections, currently available their performance limits do not take its impact into consideration. In this paper we will try to close this gap providing the estimation of the achievable identification rates due to desynchronization distortions. Finally, the impact of the codeword length on the identification performance is justified.

Index Terms— Identification of multimedia objects, desynchronization, capacity, probabilities of error.

1. INTRODUCTION

In recent years, a significant progress achieved in multimedia and modern networking technologies has revolutionized multimedia data generation, transmission, access and distribution. This factor demands the research community for extending and generalizing the main design principles of management and security systems being able efficiently operate with an exponentially increasing amount of multimedia data in large scale applications.

These issues directly concern multimedia object identification that is widely applied in copy detection, content management, biometrics, etc. Typically, establishing an identity is based on robust fingerprinting that represents a unique solution to the identification problem when no content modification is admissible as in identifying art works or medical data.

Digital fingerprints or robust perceptual hashes can be considered as compact and robust representations of multimedia objects/contents designed for its distinctive, computationally efficient and privacy preserving management.

Since mid of 90th, the domain of robust fingerprinting has performed an impressive evolution in two main directions: design of practical algorithms and theoretical performance limit analysis. The progress achieved along the former direction was mostly oriented on design of various robust feature extraction techniques and efficient matching strategies [1],[2]. The efforts along the latter one were mostly spent to analyze the achievable identification rates [3] in the formulation of infinite length codeword transmission over the discrete memoryless channel (DMC). Furthermore, several groups of authors analyzed the identification problem within the information-detection framework for various channel and codeword length assumptions [4, 5, 6].

One of main underlying assumptions for the performance analysis accomplished in most of the referred papers is the perfect synchronization between the query and the content of the multimedia object database. The only exception is a conjecture formulated in [4, 5] characterizing the probability of error of generalized maximum likelihood hypothesis testing strategy.

Such an assumption, that is appropriate in communications over the DMC, could not be valid in multimedia object identification and might finally lead to an inaccurate performance analysis due to the potential mismatch between the identified query and the information enrolled in the database. The importance of the desynchronization consideration is additionally justified by omnipresence of such kind of degradations at both multimedia database enrollment and query identification due to acquisition imperfections.

Therefore, the main goal of this paper is to extend the existing analysis of the identification problem to the channels with desynchronization distortions. It should be pointed out that the results developed in this paper and obtained in [4, 5] are different in follows. In our previous papers, we were targeting justification of universal hypothesis testing in channels with desynchronization distortions while in this paper the achievable rates over such channels are analyzed.

The remaining part of the paper is organized as follows. The problem of geometrically resilient multimedia object identification is formulated in Section 2 and analysed in Section 3. The impact of a database entry length on the system performance is considered in Section 4. Finally, conclusions

The contact author is O. Koval (email: Oleksiy.Koval@unige.ch). http://sip.unige.ch. This work is supported by SNF project 200021-119770 and CRADA

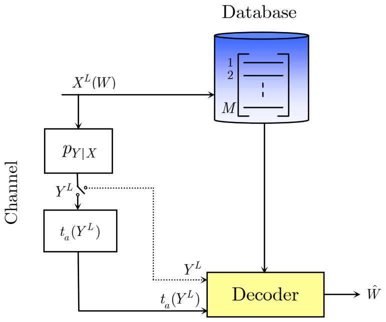

Fig. 1. Multimedia object identification as a communications problem (dashed line) and over channels with desynchronization (solid line).

are formulated in Section 5.

Notations We use capital letters to denote scalar random variables X, X^N to denote vector random variables, x and x^N to denote the realizations of X and X^N, respectively. The superscripts are used to designate length of vectors, i.e., $x^N = [x[1], x[2], ..., x[N]]$ with the i^{th} element $x[i]$. We use $X \sim p(x)$ to indicate that a random variable X is distributed according to $p(x)$. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ with cardinalities $|\mathcal{X}|$.

2. PROBLEM FORMULATION

The setup for multimedia object identification is presented in Fig. 1 (dashed line). We assume that a set of M multimedia objects represented by corresponding indexes $w \in \{1, 2, ..., M\}$ is given. Every object is associated to a corresponding fingerprint sequence $x^L = [x_1, x_2, ..., x_L]$ generated i.i.d. according to a certain distribution $p(x)$:

$$p(x^L) = \prod_{i=1}^{L} p(x[i]), \quad x^L \in \mathcal{X}^L. \tag{1}$$

The vectors $x^L(w), w \in \{1, 2, ..., M\}$, are obtained at the **enrollment stage** according to a predefined feature extraction procedure [1, 2]. Such an assumption corresponds to a symmetric modeling of biometric identification problem [7], where it was supposed that original biometrics are available at the enrollment.

In the **identification stage**, the unknown multimedia object y^L is observed at the output of a memoryless identification channel $\{\mathcal{Y}, p(y|x), \mathcal{X}\}$, where $\mathcal{Y}$ is the channel output alphabet:

$$p(y^L|x^L) = \prod_{i=1}^{L} p(y[i]|x[i]), \tag{2}$$

for all $y^L \in \mathcal{Y}^L$. An identity decoding is accomplished using the following deterministic mapping:

$$g : \mathcal{Y}^L \mapsto \{\oslash, 1, 2, ..., M\} \tag{3}$$

that is based on joint typicality [8]. A symbol $\oslash$ is reserved to indicate that the system processes an enrllment-irrelevant input. It is known [8] that (3) operates with the following average probability of error:

$$P_e = \frac{1}{M} \sum_{w=1}^{M} \Pr[g(Y^L) \neq i | X^L = x^L(w)], \tag{4}$$

defined in our case for a fixed rate of identification $R_{id} = L^{-1} \log_2 M$. Finally, capacity of the identification system C_{id} is defined as a supremum of the identification rates R_{id} such that $P_e \to 0$ for a sufficiently large L [8].

Theorem.[9] The capacity C_{id} of a symmetric identification system operating over a DMC is given by $I(X;Y)$, where $p(x, y) = p(x)p(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $I(X;Y)$ is a mutual information between random variables X and Y.

The proof of this theorem follows a random coding-based strategy similar to one used in [9] and is omitted.

Modeling desynchronization distortions

We assume that distortions between database entries and queries are modeled by a channel that besides the DMC part includes desynchronization that acts as a parametric mapper:

$$t_a : \mathcal{A}^J \times \mathcal{Y}^L \mapsto \mathcal{Y}^L \tag{5}$$

that is parametrized by a vector a^J, $a[i] \in |\mathcal{A}|, i \in \{1, 2, ..., J\}$. It is supposed that t_a impacts only the coordinates of the elements of its input and does not modify this input cardinality. Additionally, we assume that the defined desynchronization preserves input sample independence.

Since, a malicious desynchronization falls outside of the scope of this paper, we assume that not all parameters from $\mathcal{A}^J$ can be used in our analysis due to technological or acquisition constraints. Therefore, we will constraint this set to be a set of ϵ-typical desynchronization transformations defined over a corresponding set $\mathcal{A}_\epsilon^{(J)}$ with $|\mathcal{A}_\epsilon^{(J)}| \leq |\mathcal{A}^J|$.

If the parameters of $a^J = [a[1], a[2], ..., a[J]]$ are i.i.d. distributed according to $p(a)$, then one has [8]:

$$|\mathcal{A}_\epsilon^{(J)}| \leq 2^{J(H(A)+\epsilon)}, \tag{6}$$

where $H(A)$ is the entropy of A.

3. PERFORMANCE ANALYSIS

The identification system under the analysis is presented in Fig 1 (solid line). The main goal of this Section consists in evaluating the impact the desynchronization operator $t_a(.)$ might have on the performance limits of the identification system discussed in Section 2.

In the achievability part of our analysis we will try to limit the average probability of the decoder failure (4) that is modified according to $t_a(.)$ in the following way:

$$P_e^G = \frac{\sum_{w=1}^{M} \sum_{a \in \mathcal{A}_\epsilon^{(J)}} \Pr[g(t_a(Y^L)) \neq w | X^L = x^L(w)]}{M |\mathcal{A}_\epsilon^{(J)}|} \tag{7}$$

taking into account the cardinality of $\mathcal{A}_\epsilon^{(J)}$. Therefore, the following bound on P_e^G exists [8]:

$$P_e^G \quad \leq \quad 2^{L(\frac{1}{L}\log_2|\mathcal{A}_\epsilon^{(J)}|+R_{id}-(I(X;Y)-\epsilon))}, \tag{8}$$

where it is assumed that decoding was performed at all elements of $\mathcal{A}_\epsilon^{(J)}$. Therefore, if R_{id} satisfies

$$R_{id} \leq I(X;Y) - \epsilon - L^{-1}\log_2|\mathcal{A}_\epsilon^{(J)}|, \tag{9}$$

$P_e^G \to 0$ as $L \to \infty$ and $\epsilon \to 0$. Moreover, since according to the made assumption $\frac{1}{L}\log_2|\mathcal{A}_\epsilon^{(J)}| \leq \frac{J(H(A)+\epsilon)}{L}$ that vanishes for $L \to \infty$, one can conclude that

$$R_{id} \leq I(X;Y) - \epsilon' \tag{10}$$

coincides with the direct part of the Theorem in Section 2.

The proof of the converse part of the theorem can be summarized as follows:

$$LR_{id} = H(W) = H(W|t_a(Y^L)) + I(X^L(W);t_a(Y^L))$$
$$\leq H(W|t_a(Y^L)) + I(X^L(W);Y^L) \leq 1 + P_e^G LR_{id} + LC_{id}, \tag{11}$$

where we used Markovianity $X^L \to Y^L \to t_a(Y^L)$ in the first inequality and Fano inequality in the second one [8]. Therefore, normalizing by L and assuming $L \to \infty$, one has:

$$R_{id} \leq L^{-1} + P_e^G R_{id} + C_{id} \tag{12}$$

that approaches C_{id} due to the vanishing character of the first term for $L \to \infty$ and $P_e^G R \to 0$. Thus, the result of the converse part coincides with the one obtained in Section 2.

4. FINITE CODEWORD LENGTH IDENTIFICATION

The capacity achieving argument used in the previous Sections is based on the assumption of the infinite length of database entries. In order to investigate the impact of this design parameter on the system performance, we formulate the identification problem as a multiple hypothesis testing problem. In such a formulation, we are interested in a probability of error analysis of multimedia object identification over channels with desynchronization. For the sake of the quantitativeness of our analysis, we deviate from the identification setup analyzed in earlier Sections to a more particular construction (Fig. 2) [10]. In this system design it is assumed that fingerprint generation is organized in a two-stage procedure. First, a multimedia object relevant data of reduced dimensionality $\tilde{x}^L(w)$ are obtained by projecting the original multimedia vector $x^N(w)$:

$$\tilde{x}^L(w) = \mathbf{\Psi} x^N(w). \tag{13}$$

Here, $\mathbf{\Psi} \in \mathbb{R}^{L \times N}$, where $L \leq N$ and $\mathbf{\Psi} = (\psi_1, \psi_2, \cdots, \psi_L)^T$, is a dimensionality reducing operator. The elements of $\mathbf{\Psi}$, $\psi_{i,j}$, are i.i.d. generated from a Gaussian distribution, i.e., $\Psi_{i,j} \sim \mathcal{N}(0, \frac{1}{N})$. The parameters of the generating distribution are adjusted to guarantee that $\mathbf{W}$ is an approximate orthoprojector, i.e., $\mathbf{\Psi}\mathbf{\Psi}^T \approx \mathbf{I}_L$. The selection of the Gaussian distribution is justified by a desired property of such a projection output to have the Gaussian statistics without the need to satisfy conditions of the Central Limit Theorem.

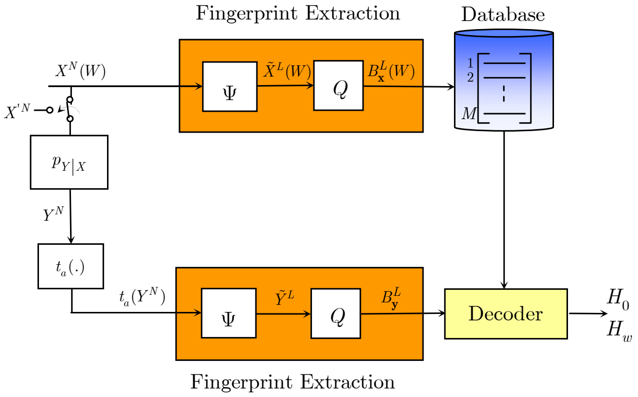

Fig. 2. Multimedia object identification as a multiple hypothesis testing over channels with desynchronization.

To relax storage requirements, complexity issues as well as enforce the system privacy, a random projection output is converted to a length L binary fingerprint according to:

$$b_{\mathbf{x}}[i] = sign(\psi_i^T x^N(w)), i \in \{1, 2, ..., L\}, \tag{14}$$

where ψ_i stays for the i^{th} row of $\mathbf{\Psi}$. Generated in such a way binary fingerprints are stored in the database.

In order to make the query, that is a DMC $p(y|x)$ distorted and desynchronized ($t_a(Y^N)$) version either of one of the enrolled objects or of an input that has no relevance to the database, compatible to the database content, one applies:

$$\tilde{Y}^L = \mathbf{\Psi} t_a(Y^N), \tag{15}$$

$$b_{\mathbf{y}}(a)[i] = sign(\psi_i^T t_a(Y^N)), i \in \{1, 2, ..., L\}. \tag{16}$$

Assuming i.i.d. zero-mean distribution of the elements of X^N, it is possible to show that the output length L fingerprints follow a Binomial distribution with 0.5 probabilities of both binary events due to particularities of their generation, $\mathcal{B}(L, 0.5)$. A similar argument is applicable to $b_{\mathbf{y}}^L(a)$.

The cases of irrelevant input are included in the analysis of this Section since such situations are not rare in identification setups. These inputs were not analyzed in details in the previous Sections due to the capability of the jointly typical decoder to reliably eliminate them from consideration with high probability for the infinite codeword length case.

Finally, the multimedia object identification is formulated as a multiple hypothesis testing with corresponding priors:

$$\begin{cases} H_0: & B_{\mathbf{y}}^L \sim \frac{1}{2^L}, \\ H_w: & B_{\mathbf{y}}^L \sim P_{b_e}^{d_H(b_{\mathbf{y}}^L(a), b_{\mathbf{x}}^L(w))}(1-P_{b_e})^{L-d_H(b_{\mathbf{y}}^L(a), b_{\mathbf{x}}^L(w))}, \end{cases} \tag{17}$$

where $d_H(.,.)$ designates the Hamming distance.

Having access to the database content and $b_{\mathbf{y}}^L(a)$, the decoder should decide which one out of $M+1$ alternatives is present at the input of the identification system. We assume that it operates as the Bounded Distance Decoder (BDD):

$$d_H(b_{\mathbf{y}}^L(a), b_{\mathbf{x}}^L(w)) \leq L\gamma, \tag{18}$$

defined for a certain threshold γ. The optimal selection of γ for the BSC distortion model case was considered in [10].

Thus, our goal is to demonstrate how t_a impacts these asymptotics. As in [10], we analyze probabilities of error of two kinds assuming that decoding acts at every point of $\mathcal{A}_\epsilon^{(J)}$.

For the probability of false acceptance P_f one has:

$$\begin{aligned} P_f &= \Pr[\bigcup_{a \in \mathcal{A}_\epsilon^{(J)}} \bigcup_{w=1}^{M} d_H(b_{\mathbf{x}}^L(w), b_{\mathbf{y}}^L(a)) \leq \gamma L | H_0] \\ &\leq 2^{-L(1-H_2(\gamma)-R_{id}-\frac{J(H(A)+\epsilon)}{L})}, \end{aligned} \tag{19}$$

where the inequality is due to the union bound and the Chernoff bound application on the tail of $\mathcal{B}(L, 0.5)$. The probability of incorrect identification P_{ic} is bounded as:

$$\begin{aligned} P_{ic} &= \Pr[\bigcup_{a \in \mathcal{A}_\epsilon^{(J)}} d_H(b_{\mathbf{x}}^L(w), b_{\mathbf{y}}^L(a)) > \gamma L \\ &\quad \cup \bigcup_{n \neq m}^{M} d_H(b_{\mathbf{x}}^L(s), b_{\mathbf{y}}^L(a)) \leq \gamma L | H_w] \\ &\leq 2^{-L(D(\gamma||P_{b_e})-\frac{J(H(A)\epsilon)}{L}))} \\ &+ 2^{-L(1-H_2(\gamma)-R_{id}-\frac{J(H(A)+\epsilon)}{L}))}. \end{aligned} \tag{20}$$

where $D(\gamma||P_{b_e}) = \gamma \log_2 \frac{\gamma}{P_{b_e}} + (1-\gamma)\log_2 \frac{1-\gamma}{1-P_{b_e}}$ is the divergence and the inequality is justified similarly to (19).

Therefore, the average probability of error for equally likely hypothesis case $P_e = 0.5P_f + 0.5P_{ic}$ is a function of γ. It is easy to show that it is minimized at $\gamma = \gamma_{opt}$, where

$$\gamma_{opt} = [1 - R + \log_2(1-P_{b_e}) - 1/L][\log_2[\frac{1-P_{b_e}}{P_{b_e}}]]^{-1} \tag{21}$$

that coincides with [10]. Therefore, although the probabilities of error are impacted by desynchronization in a fixed length multimedia object identification, the system design (threshold selection) remains unchanged with respect to the case when identification is performed over DMC channels.

5. CONCLUSIONS

In this paper we considered the problem of fingerprint based multimedia object identification over channels with symmetric desynchronizations. We analyzed the achievable rate in such a protocols and concluded that it asymptotically coincides with identification capacity. In order to justify the impact of t_a on the system performance, we investigated its particular implementation based on random projections and binarization. We confirmed that the presence of the considered class of desynchronizations as a part of identification channel model leads to probabilities of error increase as a function of randomness of the desynchronization parameters, when L is finite. However, this impact is not reflected in the system design, i.e., the decision threshold of the BDD coincides with the case when identification is performed over the DMC.

6. REFERENCES

[1] J. Haitsma, T. Kalker, and J. Oostveen, "Robust audio hashing for content identification," in *International Workshop on Content-Based Multimedia Indexing*, Brescia, Italy, September 2001, pp. 117–125.

[2] F. Lefebvre and B. Macq, "Rash : RAdon Soft Hash algorithm," in *Proceedings of EUSIPCO - European Signal Processing Conference*, Toulouse, France, 2002.

[3] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *Proc. 2003 IEEE Int. Symp. Inform. Theory*, Yokohama, Japan, June 29 - July 4 2003, p. 82.

[4] O. Koval, S. Voloshynovskiy, F. Beekhof, and T. Pun, "Decision-theoretic consideration of robust perceptual hashing: link to practical algorithms," in *WaCha2007, /Third WAVILA Challenge/*, Saint Malo, France, 2007.

[5] S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Pun, "Robust perceptual hashing as classification problem: decision-theoretic and practical considerations," in *Proc. of the IEEE 2007 Int. Workshop on Multimedia Sig. Proc.*, Chania, Crete, Greece, October 1–3 2007.

[6] A. L. Varna, A. Swaminathan, and Min Wu, "A decision theoretic framework for analyzing hash-based content identification systems," in *ACM Digital Rights Management Workshop*, Oct. 2008, pp. 67–76.

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

[8] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley and Sons, New York, 1991.

[9] F. Willems, T. Kalker, J. Gosseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *Proc. of the IEEE Int. Symp. on Inf. Theory*, 2003.

[10] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *Proc. IEEE Information Theory Workshop (ITW 2010)*, Dublin, Ireland, August,30 - September, 3 2010.