A New Biometric Identification Model and the Multiple Hypothesis Testing for Arbitrarily Varying Objects

Ashot Harutyunyan[*], Naira Grigoryan[†], Svyatoslav Voloshynovskiy[‡], and Oleksiy Koval[§]

ashot@iem.uni-due.de

Abstract: We introduce a new interpretation for the biometric enrollment and identification paradigms and show how the problem of multiple hypothesis testing (HT) for arbitrarily varying sources (AVS) in a special case relates to it. The traditional studies on biometric systems from communication perspectives assume the noisy channel model. If suppose that the process of the biometric data enrollment for a person can be performed several times and at each time both the person and the detector have some arbitrary "state", then those observations characterized according to their empirical distributions can be treated as family distributions of an AVS. It means that M persons enrollment indicate M different AVS's. Then the problem of biometric identification based on a new observation turns to be a detection of true AVS with an additional option of rejecting the existing M hypotheses. In this context, the biometric identification over noisy channels converts to one in an arbitrarily varying stochastic environment. We consider the problem within a fundamental framework of HT and information theory. The asymptotic tradeoffs among error probability exponents associated with *false acceptance of rejection decision* and *false rejection of true distribution family* are investigated and the optimal decision strategies are outlined. It is proved that for an optimal discrimination of M hypothetical distribution families/persons the ideal detector permits always lower error than in deciding in favor of the rejection.

1 Introduction

The scientific and technological interest in fundamental frameworks of biometric identification/authentication systems design rapidly grows with security needs of modern society. One of those fundamental frameworks from information-theoretic perspectives was disclosed by Willems *el al.* [5]. Its innovation is that the authors transfer a biometric identification problem to a communication problem over discrete memoryless channels (DMC) and thus reveal the concept of identification capacity (in other words, theoretically achievable maximum number of persons that can be reliably identified within a given system defined by DMC's). The latter is a fundamental limit and a performance target for any such biometric identification system. At the same time, characterization of performance bounds of identification systems in the setting of optimal hypothesis testing (HT) (see [7], [8]) are also highly important both from practical and theoretical considerations. In this

[*] Institute for Informatics and Automation Problems (IIAP), Armenian National Academy of Sciences (NAS), ashot@iem.uni-due.de. The first two authors' research was supported by national grant 11-1b255.

[†] IIAP, Armenian NAS, nar.gri@gmail.com.

[‡] University of Geneva, svolos@unige.ch.

[§] University of Geneva, oleksiy.koval@unige.ch.

context, keeping the general information-theoretic framework, we propose an alternative model of biometric identification (that naturally implies for the authentication) within the multiple HT for information sources. Here, when we make the model transform against the classical views, the discrete arbitrarily varying sources (AVS) play a central role.

The current analysis primarily relies on the paper [10] (see also the other references therein) and on the classical works in HT by Blahut [1], Haroutunian [2], Fu and Shen [4], as well as on the recent developments [6] and [9]. We briefly recall that, in particular, [1] characterizes the optimum relation between two kinds of error exponents in binary HT for discrete memoryless sources (DMS). The papers [2] and [6] study the multiple ($M > 2$) HT for DMS's in terms of logarithmically asymptotically optimality (LAO) and error exponents achievability, respectively. The subjects of [4] and [9] (also [10]) are the binary and M-ary HT for AVS's (a rather more practical model of source than the abstract DMS), respectively.

According to this model it is assumed that the source enrollment or registration is conditioned by a certain parameter further referred to as state. Under the state one can, for instance, imagine geometrical orientation of the object during its registration. It is further agreed that the state remains unchanged during entire registration. In such a scenario we say that that we deal with an arbitrarily varying object, a special case of an AVS. Performance analysis of optimal identification of such objects in terms of the best achievable error exponents represents the main research challenge of this paper.

In Section 2 the proposed model and the relevant mathematical concepts are introduced. Section 3 demonstrates the main result. Its further elaboration in view of optimal identification or HT strategies is the topic of Section 4.

2 Models of biometric identification, information source, and HT

Following modern trends of multi-biometric identification, it is assumed that we are allowed to acquire several samples from the same person (to enhance the accuracy of the identification system benefiting from multiple observations) which can result in different records. It means that at enrollment and identification the person stays at different "states" s (from finite set $\mathcal{S}$) during the registration and those states are arbitrary (coming, for instance, from physical and other conditions of the enrollment device, human interaction with it, etc.). For each person m among M possible, an N-length vector of observations $\mathbf{x} \stackrel{\triangle}{=} (x_1, ..., x_N) \in \mathcal{X}^N$ ($\mathcal{X}$ being the enrollment or information source alphabet, also finite) or $\mathbf{x}(s)$ depending on a state has its own statistics of signals, or, in other words, its own type/emprical distribution (ED) [3], denote it by $G_{m,s} \stackrel{\triangle}{=} \{G_{m,s}(x),\, x \in \mathcal{X}\}$. The latter is computed by an extractor of empirical distributions (Fig. 1). Moreover, the enrollment state does not change during a particular sampling or feature extraction. Those distributions collected for all possible states of the enroller create a family $\mathcal{G}_m \stackrel{\triangle}{=} \{G_{m,s},\, s \in \mathcal{S}\}$ of probability distributions (PD) which is saved in a database. Therefore, each of possible M persons can be characterized by his/her specific family of PD's (over

arbitrarily varying states) which constitute an AVS. Within this model the biometric identification becomes a problem of multiple HT (making a decision on the true distribution family or a person $\hat{m}$ among M or on the rejection of all M's) based on an observation made at an unknown state for an identifiable person (Fig. 2). Note that the genuine statistical characteristics (type family) of an enrolled person remains unknown, denote it by $\mathcal{G}_m^*$ for person m. So in the phase of identification the identifier has to match the outcome of the extractor of distributions for an observation (at unknown state) with the M hypotheses available in the biometric database and make a decision in favor of one of them or the rejection alternative:

$$H_m : \mathcal{G}^* = \mathcal{G}_m, \quad H_R : \text{ none of } H_m\text{'s is true}, \quad m = \overline{1, M}.$$

As a typical HT problem this decision making can be performed applying a test φ_N as a partition of $\mathcal{X}^N$ into $M+1$ disjoint subsets $\mathcal{A}_N^m,\ m = \overline{1, M}$, and $\mathcal{A}_N^R$. If $\mathbf{x} \in \mathcal{A}_N^m$ then the test adopts the hypothesis H_m. If $\mathbf{x} \in \mathcal{A}_N^R$, the test rejects all those M hypotheses. Below we categorize the errors occurring in the decision making. $(M+1)M$ different kinds of errors are possible. We treat the problem in the memoryless formulation. Therefore, the probability of $\mathbf{x}$ according to PD G^* is $G^{*N}(\mathbf{x}) \stackrel{\triangle}{=} \prod_{n=1}^{N} G^*(x_n)$. Furthermore, the probability of a subset $\mathcal{A}_N \subset \mathcal{X}^N$ is measured by the sum $G^{*N}(\mathcal{A}_N) \stackrel{\triangle}{=} \sum\limits_{\mathbf{x} \in \mathcal{A}_N} G^*(\mathbf{x})$.

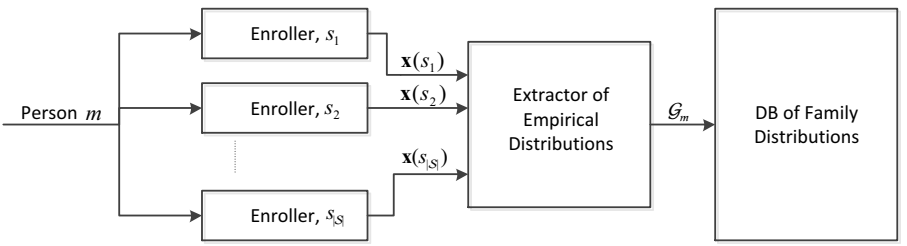

Fig. 1. Enrollment of person m.

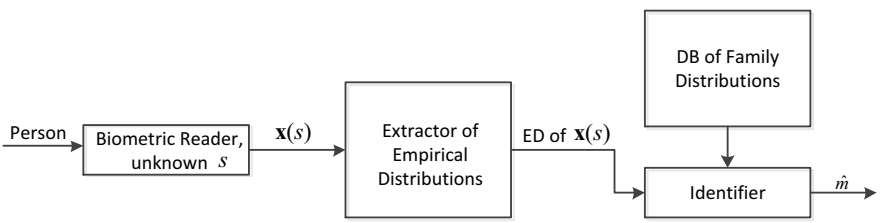

Fig. 2. Identification of a person.

Now, the probability of an erroneous acceptance of hypothesis H_l when H_m was true is

$$\alpha_{l,m}(\varphi_N) \stackrel{\triangle}{=} \max_{s \in \mathcal{S}} G_{m,s}^N(\mathcal{A}_N^l), \quad 1 \le l \ne m \le M. \tag{1}$$

And the error probability of false rejection (a false accept of rejection decision) when H_m was true is defined by

$$\alpha_{R,m}(\varphi_N) \stackrel{\triangle}{=} \max_{s\in\mathcal{S}} G^N_{m,s}(\mathcal{A}^R_N), \quad m = \overline{1,M}. \tag{2}$$

In case of true H_m the probability of wrong decision (false reject of true hypothesis) will be

$$\alpha_m(\varphi_N) \stackrel{\triangle}{=} \max_{s\in\mathcal{S}} G^N_{m,s}(\overline{\mathcal{A}}^m_N) = \sum_{l\neq m}^{M} \alpha_{l,m}(\varphi_N) + \alpha_{R,m}(\varphi_N),\ m = \overline{1,M}. \tag{3}$$

Therefore, the following sort of error probability exponents/reliabilities (log-s and exp-s being to the base 2) of (1) and (2) are of interest:

$$E_{l|m}(\varphi) \stackrel{\triangle}{=} \limsup_{N\to\infty} -\frac{1}{N}\log \alpha^N_{l|m}(\varphi_N), \quad l \neq m = \overline{1,M}, \tag{4}$$

$$E_{R,m}(\varphi) \stackrel{\triangle}{=} \limsup_{N\to\infty} -\frac{1}{N}\log \alpha^N_{R,m}(\varphi_N), \quad m = \overline{1,M}, \tag{5}$$

where $\varphi \stackrel{\triangle}{=} \{\varphi_N\}_{N=1}^{\infty}$. From (3) and (4) it follows that

$$E_m(\varphi) = \min_{l\neq m} \left[E_{l|m}(\varphi), E_{R,m}(\varphi)\right]. \tag{6}$$

Now the question is: which collection (trade-offs) of error exponents can be theoretically achieved for the given identification or HT problem? Consider the $M(M+1)$-dimensional point $\mathbf{E} \stackrel{\triangle}{=} \{E_{R,m}, E_m\}_{m=\overline{1,M}}$ with respect to the error exponent pairs $(-\frac{1}{N}\log \alpha_{R,m}(\varphi_N), -\frac{1}{N}\log \alpha_m(\varphi_N))$, where the decision regions $\mathcal{A}^m_N$ ($m = \overline{1,M}$) and $\mathcal{A}^R_N$ satisfy $\mathcal{A}^m_N \cap \mathcal{A}^l_N = \emptyset$ for $m \neq l$, $\mathcal{A}^m_N \cap \mathcal{A}^R_N = \emptyset$ and $\bigcup\limits_m \mathcal{A}^m_N = \mathcal{X}^N / \mathcal{A}^R_N$.

Definition 1 The collection of error exponents (reliabilities) $\mathbf{E}$ is called achievable if for all $\varepsilon > 0$ there exists a decision scheme $\{\mathcal{A}^m_N\}_{m=1}^M$ plus $\mathcal{A}^R_N$ such that

$$-\frac{1}{N}\log \alpha_{R,m}(\varphi_N) > E_{R,m} - \varepsilon,\ -\frac{1}{N}\log \alpha_m(\varphi_N) > E_m - \varepsilon$$

for N large enough. Let $\mathcal{R}_{\rm b}(M)$ denotes the set of all achievable reliabilities.

3 Error exponents: achievable tradeoffs

The method of typical sequences [3] is underlying for proofs of achievable error bounds. Let $\mathcal{G}(\mathcal{X}) \stackrel{\triangle}{=} \{G(x),\ x \in \mathcal{X}\}$ be the collection of all PD's on $\mathcal{X}$. Each observation $\mathbf{x} \in \mathcal{X}^N$

has a type defined by its composition or empirical PD $G_{\mathbf{x}}(x) \stackrel{\triangle}{=} G \stackrel{\triangle}{=} \frac{1}{N}N(x|\mathbf{x})$, where $N(x|\mathbf{x})$ is the number of occurrences of x in $\mathbf{x}$. Denote the set of all possible types of such N-length vectors by $\mathcal{G}^N(\mathcal{X})$. Additionally, denote by $\mathcal{T}_G^N(X)$ the type class of G, the set of G-type vectors $\mathbf{x}$. Let $H(G)$ stands for the Shannon entropy of G and $D(G \parallel G_m)$ for the KL divergence between distributions G and G_m. In the sequel we use several properties of types. First of all,

$$|\mathcal{G}^N(\mathcal{X})| < (N+1)^{|\mathcal{X}|}, \tag{7}$$

$$|\mathcal{T}_G^N(X)| \leq \exp\{NH(G)\}. \tag{8}$$

For a PD $G_{m,s} \in \mathcal{G}(\mathcal{X})$, the sequence $\mathbf{x} \in \mathcal{T}_G^N(X)$ has the probability

$$G_{m,s}^N(\mathbf{x}) = \exp\{-N[H(G) + D(G \parallel G_{m,s})]\}. \tag{9}$$

The equations (8) and (9) imply estimates for the probability of a type class:

$$G_{m,s}^N(\mathcal{T}_G^N(X)) \geq (N+1)^{-|\mathcal{X}|}\exp\{-ND(G \parallel G_{m,s})\}, \tag{10}$$

$$G_{m,s}^N(\mathcal{T}_G^N(X)) \leq \exp\{-ND(G \parallel G_{m,s})\}. \tag{11}$$

In the theorem and its proof below we show that the following collection of exponents characterizes the unknown region $\mathcal{R}_{\mathrm{b}}(M)$:

$$\begin{aligned}\mathcal{E}_{\mathrm{b}}(M) \quad \stackrel{\triangle}{=} \quad & \{\mathbf{E} : \forall\, G\ \exists\, m \text{ s. t. } \min_{s\in\mathcal{S}} D(G \parallel G_{m,s}) > E_m \text{ and} \\ & \exists\, G \text{ s. t. } \min_{s\in\mathcal{S}} D(G \parallel G_{m,s}) > E_{R,m} \text{ for all } m\}.\end{aligned} \tag{12}$$

Theorem 1 *The set $\mathcal{E}_b(M)$ is an achievable region of reliabilities: $\mathcal{E}_b(M) \subset \mathcal{R}_b(M)$. Additionally, if $\mathbf{E} \in \mathcal{R}_b(M)$, then for any $\delta > 0$ it follows that $\mathbf{E}_\delta \in \mathcal{E}_b(M)$, where $\mathbf{E}_\delta \stackrel{\triangle}{=} \{E_{R,m} - \delta, E_m - \delta\}_{m=\overline{1,M}}$.*

The proof of the theorem consists of direct and converse parts. For the direct part, we observe that if $\mathbf{E} \in \mathcal{E}_{\mathrm{b}}(M)$, then from (8), (9), and (11) for any $s \in \mathcal{S}$ we have

$$\begin{aligned} G_{m,s}^N(\overline{\mathcal{A}}_N^m|s) \quad &= \quad \sum_{\mathbf{x}\in\overline{\mathcal{A}}_N^m} G_{m,s}^N(\mathbf{x}|s) \\ &\leq \sum_{\mathcal{T}_G^N(X)\subset\overline{\mathcal{A}}_N^m} \exp\{-ND(G \parallel G_{m,s})\} \\ &\leq |\mathcal{G}^N(\mathcal{X})|\exp\{-ND(G \parallel G_{m,s})\}. \end{aligned} \tag{13}$$

Applying (13) and (7) we derive

$$\alpha_m(\varphi_N) \leq |\mathcal{G}^N(\mathcal{X})|\exp\{-N\min_{s\in\mathcal{S}} D(G \parallel G_{m,s})\} \leq \exp\{-N(E_m - \delta)\}.$$

Similar steps can lead us to other desirable inequalities:

$$\alpha_{R,m}(\varphi_N) \le \exp\{-N(E_{R,m} - \delta)\}. \tag{14}$$

In the converse part we assume that $\mathbf{E} \in \mathcal{R}_{\rm b}(M)$. It means that for every $\varepsilon > 0$ there exists a decision scheme $\{\mathcal{A}_N^m, \mathcal{A}_N^R\}_{m=1}^M$ that provides the following inequalities for all m's with large enough $N > N_0(\varepsilon)$:

$$-\frac{1}{N}\log \alpha_{R,m}(\varphi_N) > E_{R,m} - \varepsilon,\ -\frac{1}{N}\log \alpha_m(\varphi_N) > E_m - \varepsilon, \tag{15}$$

We pick a $\delta > 0$ and show that

$$\forall\, G\ \exists\, m\ \text{ s. t.}\min_{s\in\mathcal{S}} D(G \parallel G_{m,s}) > E_m - \delta, \tag{16}$$

$$\exists\, G\ \text{s. t.}\min_{s\in\mathcal{S}} D(G \parallel G_{m,s}) > E_{R,m} - \delta\ \text{for all}\ m. \tag{17}$$

For the equation (16), by the continuity of $D(\cdot \parallel G_{m,s})$ there exists a type $Q \in \mathcal{G}^N(\mathcal{X})$ that for $N > N_1(\varepsilon)$ and a fixed m satisfies

$$D(Q \parallel\ G_{m,s}) \le\ D(G \parallel\ G_{m,s}) +\ \delta/2. \tag{18}$$

Let $\overline{G}_m \stackrel{\triangle}{=} \arg\min_{s\in\mathcal{S}} D(Q \parallel G_{m,s}) > E_m - \delta/2$, then in light of (8) we have

$$\begin{aligned}
\alpha_m(\varphi_N) &\ge \overline{G}_m^N(\overline{\mathcal{A}}_N^m) \\
&\ge \overline{G}_m^N(\overline{\mathcal{A}}_N^m \cap \mathcal{T}_Q^N(X)) \\
&= \sum_{\overline{\mathcal{A}}_N^m \cap \mathcal{T}_Q^N(X)} \exp\{-N[H(Q) + D(Q \parallel \overline{G}_m)]\} \\
&\ge |\overline{\mathcal{A}}_N^m \cap \mathcal{T}_Q^N(X)| \exp\{-NH(Q)\} \exp\{-ND(Q \parallel \overline{G}_m)\}.
\end{aligned}$$

Note that for $N > N_2(\delta)$,

$$|\overline{\mathcal{A}}_N^m \cap \mathcal{T}_Q^N(X)| \exp\{-NH(Q)\} \ge \exp\{-N\delta/4\}, \tag{19}$$

Whence, for $N > \max\{N_1(\delta), N_2(\delta)\}$ we conclude that

$$\alpha_m(\varphi_N) \ge \exp\{-N[D(Q \parallel \overline{G}_m) - \delta/4]\} \ge \exp\{-N[D(G \parallel \overline{G}_m) + \delta/4]\}$$

which with (15) and $\varepsilon = 3\delta/4$ gives $E_m - \delta < -\frac{1}{N}\log \alpha_m(\varphi_N) < D(G \parallel \overline{G}_m)$ for $N > \max\{N_0(\varepsilon), N_1(\delta), N_2(\delta)\}$ and for every $m = \overline{1, M}$.

Now we proceed to the equation (17). Pick a $\delta > 0$. If $\mathbf{E}_\delta \notin \mathcal{E}_{\mathrm{b}}(M)$ then for arbitrary G there exists m satisfying $D(G \parallel \overline{G}_m) \le E_{R,m} - \delta$. In view of (8), (18), and (19) we get

$$\begin{aligned}
\alpha_{R,m}(\varphi_N) &\ge \overline{G}_m^N(\mathcal{A}_N^R) \\
&\ge \overline{G}_m^N(\mathcal{A}_N^R \cap \mathcal{T}_Q^N(X)) \\
&= \sum_{\mathcal{A}_N^R \cap \mathcal{T}_Q^N(X)} \exp\{-N[H(Q) + D(Q \parallel \overline{G}_m)]\} \\
&\ge |\mathcal{A}_N^R \cap \mathcal{T}_Q^N(X)| \exp\{-NH(Q)\} \exp\{-ND(Q \parallel \overline{G}_m)\} \\
&\ge \exp\{-N[D(G \parallel \overline{G}_m) - \delta/4]\} \\
&\ge \exp\{-N[E_{R,m} - \delta/4]\}.
\end{aligned}$$

However, the last inequality contradicts to (15) for $\varepsilon < \delta/4$ and N large enough.

4 Optimal decision schemes

Theorem 1 specifies all possible reliability trade-offs for the identification system of Figs. 1-2. It contains also optimal relations between those error exponents in sense of LAO testing of hypotheses. In other words, let $E_m,\ m = \overline{1, M}$, be fixed: what are the "maximum" values $\{E_{l,m}^*, E_{R,m}^*\}_{l \ne m = \overline{1,M}}$ for the rest of reliabilities such that there is no other collection $\{E'_{l,m}, E'_{R,m}\}_{l \ne m = \overline{1,M}}$ satisfying $E'_{l,m} > E_{l,m}^*$ and $E'_{R,m} > E_{R,m}^*$ for all $l \ne m = \overline{1, M}$?

Let φ^* be a test sequence defined by the following decision regions:

$$\mathcal{B}_R \stackrel{\triangle}{=} \{G : \min_{s \in \mathcal{S}} D(G \parallel G_{m,s}) > E_m \text{ for all } m\}, \tag{20}$$

$$\mathcal{B}_m \stackrel{\triangle}{=} \{G : \min_{s \in \mathcal{S}} D(G \parallel G_{m,s}) < E_m\}, \quad m = \overline{1, M}. \tag{21}$$

For $l \ne m = \overline{1, M}$ we define:

$$E_{R,m}(\varphi^*) \stackrel{\triangle}{=} E_{R,m}^* \stackrel{\triangle}{=} \min_{G \in \mathcal{B}_R} \min_{s \in \mathcal{S}} D(G \parallel G_{m,s}), \tag{22}$$

$$E_{l,m}(\varphi^*) \stackrel{\triangle}{=} E_{l,m}^* \stackrel{\triangle}{=} \min_{G \in \mathcal{B}_l} \min_{s \in \mathcal{S}} D(G \parallel G_{m,s}). \tag{23}$$

A detailed analysis of this decision scheme results in the next assertion.

Theorem 2 *Let the inequalities*

$$E_1 < \min_m \{ \min_{s,s' \in \mathcal{S}} D(G_{m,s} \parallel G_{1,s'}) \},$$

$$E_m < \min_{l \neq m}\{ \min_{l=\overline{1,m-1}} E_{l,m}, \min_{l=\overline{m+1,M}} \min_{s,s' \in \mathcal{S}} D(G_{l,s} \parallel G_{m,s})\}, \quad m = \overline{1,M},$$

hold, then the optimum collection of error exponents are defined by (20)–(23).

Theorem 2 implies an interesting observation.

Remark 1 *Further analysis shows that* $\min\limits_{l=\overline{1,M},\, l \neq m} \left[E^*_{l,m}, E^*_{R,m}\right] = E^*_{R,m}$, *for all* $m = \overline{1,M}$. *This statement means that discriminating among* M *families is always easier than voting for the rejection. Its biometric reflection within the above introduced identification model is that the persons can be recognized easier than claimed unfamiliar.*

Conclusion. We introduced a novel mathematical interpretation and model for the biometric identification and showed its relation to the multiple HT for arbitrarily varying objects within an information-theoretic framework. The achievable performance bounds for this identification system are specified including special optimality tradeoffs.

References

[1] R.E. Blahut, "Hypothesis testing and information theory", *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.

[2] E.A. Haroutunian, "Logarithmically asymptotically optimal testing of multiple statistical hypotheses", *Problems of Control and Inform. Theory,* vol. 19, no. 5-6, pp. 413–421, 1990.

[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[4] F.-W. Fu and S.-Y. Shen, "Hypothesis testing for arbitrarily varying source with exponential-type constraint", *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 892–895, 1998.

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

[6] E. Tuncel, "On error exponents in hypothesis testing", *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2945–2950, 2005.

[7] A.L. Varna, A. Swaminathan, and M. Wu, "A decision theoretic framework for analyzing hash-based content identification systems", *Proc. ACM Digital Rights Management Workshop*, pp. 67–76, Oct. 27, 2008.

[8] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification", *Proc. IEEE Workshop Inform. Theory*, Dublin, Ireland, August 30 – September 3, 5 p., 2010.

[9] N.M. Grigoryan and A.N. Harutyunyan, "Error exponents in multiple hypothesis testing for arbitrarily varying sources," *Proc. IEEE Workshop Inform. Theory*, Dublin, Ireland, August 30 – September 3, 5 p., 2010.

[10] N. Grigoryan, A. Harutyunyan, S. Voloshynovsky, and O. Koval, "On multiple hypothesis testing with rejection option," accepted, *IEEE Workshop Inform. Theory*, Paraty, Brazil, October 16–20, 5 p., 2011.