Multimodal object authentication with random projections : a worst-case approach

Koval, Oleksiy; Voloshynovskyy, Svyatoslav

Multimodal object authentication with random projections: a worst-case approach

Oleksiy Koval and Sviatoslav Voloshynovskiy*

ABSTRACT

In this paper, we consider a forensic multimodal authentication framework based on binary hypothesis testing in random projections domain. We formulate a generic authentication problem taking into account several possible counterfeiting strategies. The authentication performance analysis is accomplished in the scope of Neyman-Pearson framework as well as for an average probability of error for both direct and random projections domains. Worst-case attack/acquisition channel leading to the worst performance loss in terms of Bhattacharyya distance reduction is presented. The obtained theoretical findings are also confirmed by results of computer simulation.

Keywords: multimodal authentication, fusion, performance analysis, worst case distortions.

1. INTRODUCTION

In the last years, digital reproduction tools have performed an impressive evolution, providing professional solutions to various groups of users. Besides the obvious advantages, these tools offer at the same time unprecedent possibilities for the counterfeiters that can virtually reproduce authentic items, i.e., objects, documents, IDs, packaging or even biometrics. Thus, the item authentication becomes a critical issue demanding an urgent solution for various applications. This urgency is also caused by the fundamental inability to satisfy the security requirements by the currently used proprietary (mostly material-science based) technologies and classical crypto-based techniques. Moreover, the particularities of modern markets, characterized by distributed manufacturing and distribution, require new authentication technologies oriented on the end-users. A practically attractive protocol of item authentication is based on the mobile phones of end-users. This protocol is schematically shown in Figure 1. As secure forensic features that can not be copied or cloned we will consider here a random surface microstructure image known to be a powerful discriminative and difficult to duplicate structure.[1] The user acquires the random surface microstructure image and sends it to the server with the accompanying authentication data. The server, possibly connected to the database of enrolled images, makes the binary decision about requested item authenticity and communicates the decision back to the end-user portable device.

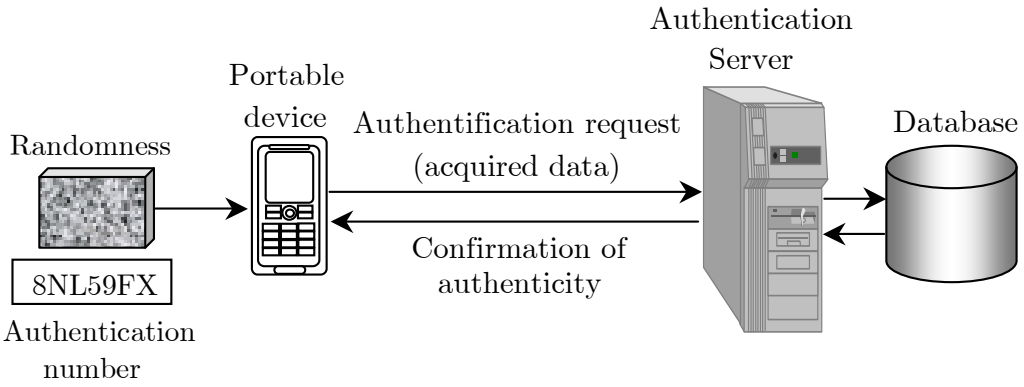
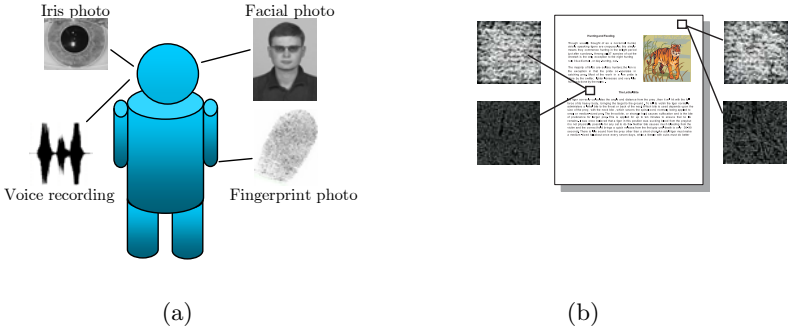


Figure 1. Authentication architecture.

O. Koval and S. Voloshynovskiy are with CUI-University of Geneva, Stochastic Information Processing Group, Battelle Batiment A, 7 route de Drize, 1227 Carouge, Switzerland. The contact author is O. Koval (email: Oleksiy.Koval@unige.ch). http://sip.unige.ch

Unique forensic features are suitable for authentication of both mentioned real world items where they are represented by the mentioned random surface microstructure images and humans where biometrics are exploited (Figure 2). This makes possible to consider both problems from common positions of object authentication.

As it is widely discussed in the domain of biometric person identification,[2] selection of a particular forensic feature to be used for authentication is not an easy task. The reason is that the successful candidate should satisfy certain requirements that include performance (authentication accuracy) of the corresponding binary or multiple hypothesis classification system, universality (presence of the feature in all objects), distinctiveness (all objects should be differentiated based on them), *etc* .[3]



(a) (b)

Figure 2. Forensic features: belonging to the same person (a) and describing the paper acquired at 1200 dpi by an optical scanner (top) and by CO.sub.2 laser (bottom).

A natural way of satisfying these requirements is to fuse various modalities that can be implemented at different levels.[4,5] While, fusion is mostly performed at match score and decision levels, such an approach is suboptimal in terms of the attained performance accuracy due to the data processing inequality[6] that suggests to combine the available data at sensor level. However, due to various practical constraints, not so many cases are known that follow this strategy.[7] To fulfill the gaps, fusion at the sensor level with data dimensionality reduction based on random projections will be used in this paper in order to approach authentication performance limits.

Most of forensic features like a person iris and fingerprint or paper scans obtained from different locations (Figure 1) can be considered as statistically independent. However, in certain[8] some intermodal statistical dependence might be assumed. Nevertheless, in the scope of this paper we assume the former case of independent modality fusion. This choice is further justified by the ambiguity concerning the impact of the dependence between multimodal signals on the accuracy of multimodal authentication/identification.[9,10]

While adding modalities hopefully leads to the improvement of performance accuracy, requirements of protocol security/privacy and computational complexity should be additionally satisfied.[11] Since cryptography-based solutions are not fully applicable here due to inherent noisy nature of the observations, an alternative solution should be proposed capable of resolving this performance/security/privacy/complexity trade-off. Classically, several solutions in unimodal authentication of objects are proposed via various feature extraction techniques.[12–15] As it was already mentioned in,[16] all this operations are non-invertable and due to the data processing inequality lead to a certain information loss. The first attempt to quantify this loss was performed in[16] for a unimodal case. In the scope of this paper we are trying to extend the earlier obtained results to a multimodal case.

The analysis in[16] was accomplished under a very conservative assumption about the statistics of the authentication channel. Specifically, it was supposed that distortions introduced by this channel follow the Gaussian distribution. However, this assumption appears to be too conservative for the authentication application. Indeed, assuming that authentication is performed within the Kerckhoffs security framework,[17] selection of the channel might be performed based on all available prior information about the protocol design in order to maximally deteriorate its performance. Possible channel configurations are broad and crucially depend on the amount of a priori information. Moreover, during the life cycle of an authentified object, the acquisition conditions might

significantly vary due to, for instance, wide diversity of acquisition devices and introduced physical degradations that makes applicability of the mentioned analysis doubtful in general. That is why the second research goal of this paper is to perform analysis of authentication accuracy under the certain freedom of the statistics of the acquisition channel.

Notations We use capital letters to denote scalar random variables X and $\mathbf{X}$ to denote vector random variables, x and $\mathbf{x}$ denote the realizations of scalar and vector random variables, respectively. All vectors without sign tilde are assumed to be of the length N and with the sign tilde of length L with the corresponding subscripts. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$. We use $\mathbf{X} \sim p(\mathbf{x})$ to indicate that $\mathbf{X}$ is distributed according to $p_{\mathbf{X}}(\mathbf{x})$. Statistical hypotheses are denoted as H_i, $i = \{0, 1\}$, and the distributions under these hypothesis as $p(\mathbf{v}|H_i)$. $\mathcal{N}(\mu, \sigma_X^2)$ stands for the Gaussian distribution with mean μ and variance σ_X^2.

2. OBJECT AUTHENTICATION: PROBLEM FORMULATION AND ARCHITECTURE

A generic object authentication problem can be considered as a hypothesis testing[18] that requires the selection of authentication criteria and assumptions behind the statistics of alternatives. In the case one has to authenticate one out of $\{1, 2, ..., m, ..., |\mathcal{M}|\}$ objects associated with an index m, it can be formulated as a decision making that the observed length-N feature vector $\mathbf{v}$ representing the object under authentication is in some proximity to the genuine vector $\mathbf{x}(m)$, $1 \leq m \leq |\mathcal{M}|$ for a given m. The decision might be taken under conditions of conservative or relaxed acquisition that refer to the assumptions of degradations that are introduced into $\mathbf{v}$ versus the genuine data $\mathbf{x}(m)$. In the former case, acquisition imperfectness is characterized by a certain statistical model of distortions introduced by the authentication channel (for instance, additive white Gaussian noise[19]). Definitely, this approach is too restrictive and is not fully capable of adequately representing all potential degradations that might happen. Oppositely, in the latter case such an assumption is dropped and the acquisition has a compound nature and can rather be represented as a family of possible distortion models.

Evidently, the second scenario is much more involved for the performance of object authentication systems and will be considered in this paper. The main expected result that we are targeting here is to provide some tractable performance limits of such systems under the certain freedom of acquisition additive authentication channels. In particular, we target finding the worst statistical channel model that will provide the lower bound on the object authentication system accuracy. This strategy often appears in security applications, where the goal of the opponent is to design an attack leading to the worst possible performance attained by the authentication system taking into account all available prior knowledge about the system design as well as particularities of object manufacturing or forensic features.

We start our analysis by the formulation of the main goal of object authentication. It consists in reliable discrimination between two following alternatives:

$$\begin{cases} H_0: & \mathbf{V} \sim p(\mathbf{v}|H_0) = p(\mathbf{v}|\mathbf{x}(/m))); \\ H_1: & \mathbf{V} \sim p(\mathbf{v}|H_1) = p(\mathbf{v}|\mathbf{x}(m)), \end{cases} \tag{1}$$

where $p(\mathbf{v}|\mathbf{x}/m)$ stays to indicate statistics of feature vectors of all objects excluding the genuine one $\mathbf{x}(m)$. In this case it is assumed that zero hypothesis is composite and for $|\mathcal{J}|$ such objects $\mathbf{x}'(i), i \in \{1, 2, ..., |\mathcal{J}|\}$, is represented as:

$$p(\mathbf{v}|H_0) = \begin{cases} p(\mathbf{v}|\mathbf{x}'(1)) & \text{with prob. } p_1, \\ ... & ... \\ p(\mathbf{v}|\mathbf{x}'(\mathcal{J})) & \text{with prob. } p_{|\mathcal{J}|}, \end{cases} \tag{2}$$

where $\sum_{i=1}^{|\mathcal{J}|} p_i = 1$ and p_n can be assumed to be uniform.

In the analysis of performance of this test we will use a technique that is widely exploited in digital communications to limit performance of practical codes that is justified by the closest codebook entries.[20] Thus, to design the decision rule for the binary hypothesis test, we will use the worst case condition for the selection of the alternative hypothesis assuming that one needs to ensure the desired performance characteristics for the

closest possible to $\mathbf{x}(m)$ feature vector $\mathbf{x}(n)$ in a collection $\mathbf{x}'(i), i \in \{1, 2, ..., |\mathcal{J}|\}$. It should be noticed that it can be done taking into account all feature vectors according to the model (2). However, to avoid cumbersome integrations that reduces tractability, we will follow the worst case approach. Alternatively, one can also consider generalized maximum likelihood approach that is tractable but not always optimal.

Then, (1) can be reformulated as follows:

$$\begin{cases} H_0: & \mathbf{V} \sim p(\mathbf{v}|H_0) = p(\mathbf{v}|\mathbf{x}(n)), \\ H_1: & \mathbf{V} \sim p(\mathbf{v}|H_1) = p(\mathbf{v}|\mathbf{x}(m)). \end{cases} \tag{3}$$

We use the Neyman-Pearson decision rule maximizing probability of correct acceptance P_D subject to the constraint probability of false alarm P_F that can be formulated as the likelihood ratio test:

$$\Lambda(\mathbf{v}) = \frac{p(\mathbf{v}|H_1)}{p(\mathbf{v}|H_0)} \lessgtr \eta, \tag{4}$$

with the threshold η chosen to satisfy the constraint on $P_F = \int_{\Lambda(\mathbf{v})>\eta} p(\mathbf{v}|H_0)d\mathbf{v} = \alpha$.

As it was already mentioned in the introductory part of this paper, in order to implement such an object authentication system in practice additional complexity, memory storage, privacy and security requirements should be taken into account. In most cases, the solution is based on channel coding fundamentals and principles of robust perceptual hashing that are known as a robust version of classical cryptographic hashes. Unfortunately, these solutions have a fundamental shortcoming related to collusions. To overcome this problem, authentication based on source coding principles was proposed. Such an approach is adopted in this paper (Figure 2) where authentication accuracy is defined by the properties of the acquisition and by the rate-distortion properties of the corresponding source code.

It should be pointed out that practical implementation of the above schemes was envisioned by the conversion of the continuous vectors to the discrete representations. These conversions are not invertible and obviously lead to the information loss due to data processing inequality.[6] On the other side, non-invertability of this operations that can be referred as feature extraction that plays a particular role of securing original feature vectors. This generic feature extraction operation is modeled in our case by a random projection operator $\Phi_{\mathbf{x}}$ ($\Phi_{\mathbf{y}}$) that transforms input data to a lower-dimensional representation (Figure 2). Moreover, as it will be shown in this paper, this transform has nice robustness properties for a certain class of acquisition distortions.

3. MULTIMODAL AUTHENTICATION: BINARY HYPOTHESIS TESTING FORMULATION

As it was mentioned earlier, multimodal fusion might be considered as a practical way of performance optimization of object authentication system (Figure 2).

The problem under analysis might be considered as a binary hypotehsis testing between the following two alternatives:

$$\begin{cases} \tilde{H}_0: & \tilde{\mathbf{v}} = \mathbf{\Phi}_{\mathbf{x}}(\mathbf{x}(n) + \mathbf{z}_X) = \tilde{\mathbf{x}}(n) + \tilde{\mathbf{z}}_X, \\ & \tilde{\mathbf{w}} = \mathbf{\Phi}_{\mathbf{y}}(\mathbf{y}(n) + \mathbf{z}_Y) = \tilde{\mathbf{y}}(n) + \tilde{\mathbf{z}}_Y, \\ \tilde{H}_1: & \tilde{\mathbf{v}} = \mathbf{\Phi}_{\mathbf{x}}(\mathbf{x}(m) + \mathbf{z}_X) = \tilde{\mathbf{x}}(m) + \tilde{\mathbf{z}}_X, \\ & \tilde{\mathbf{w}} = \mathbf{\Phi}_{\mathbf{y}}(\mathbf{y}(m) + \mathbf{z}_Y) = \tilde{\mathbf{y}}(m) + \tilde{\mathbf{z}}_Y, \end{cases} \tag{5}$$

where $\mathbf{x} \in \mathbb{R}^{N_X}$, $\mathbf{y} \in \mathbb{R}^{N_Y}$, $\tilde{\mathbf{x}} \in \mathbb{R}^{L_X}$, $\tilde{\mathbf{y}} \in \mathbb{R}^{L_Y}$, $\mathbf{\Phi}_{\mathbf{x}} \in \mathbb{R}^{L_X \times N_X}$, $\mathbf{\Phi}_{\mathbf{y}} \in \mathbb{R}^{L_Y \times N_Y}$, $L_X \leq N_X$ and $L_Y \leq N_Y$ and $\mathbf{z}_X$ and $\mathbf{z}_Y$ are noise components in each modality corresponding to the discrete memoryless channels (DMC) $p_{V|X}$ and $p_{W|Y}$. It is assumed that elements of $\mathbf{\Phi}_{\mathbf{x}}$ and $\mathbf{\Phi}_{\mathbf{y}}$ are generated i.i.d. from a zero-mean Gaussian distribution with variances $\frac{1}{N_X}$ and $\frac{1}{N_Y}$, respectively in order to approximate orthogonality of the projection, i.e., $\mathbf{\Phi}_{\mathbf{x}}\mathbf{\Phi}_{\mathbf{x}}^T = \mathbf{I}_{L_X}$ and $\mathbf{\Phi}_{\mathbf{y}}\mathbf{\Phi}_{\mathbf{y}}^T = \mathbf{I}_{L_Y}$, where $\mathbf{I}_L$ is an $L \times L$ identity matrix.

Furthermore one has:

$$\begin{cases} H_0: & p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_0) = p(\tilde{\mathbf{v}}|H_0)p(\tilde{\mathbf{w}}|H_0), \\ H_1: & p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_1) = p(\tilde{\mathbf{v}}|H_1)p(\tilde{\mathbf{w}}|H_1), \end{cases} \tag{6}$$

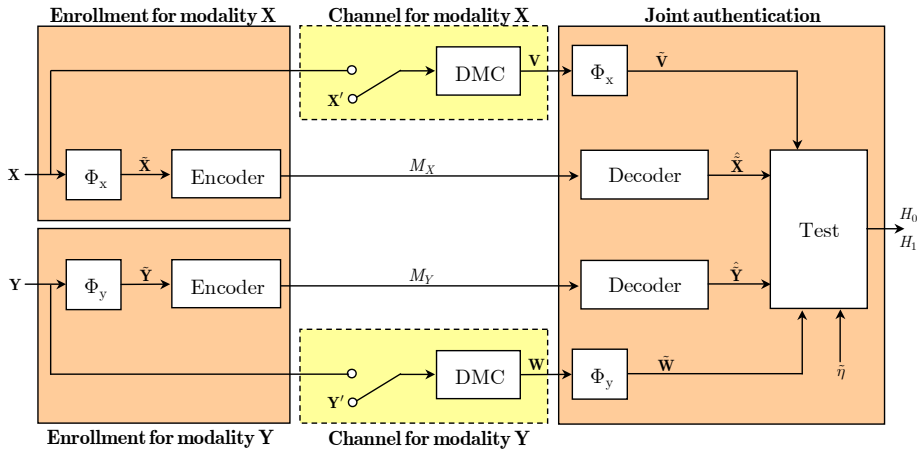


Figure 3. Multimodal authentication architecture based on random projections.

with the assumption of independence between modalities justified in the introductory part of the paper.

We will use the Neyman-Pearson decision rule that maximizes P_D subject to the constraint $P_F \leq \alpha$ according to the unimodal analog (4):

$$\Lambda(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}) = \frac{p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_1)}{p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_0)} = \frac{p(\tilde{\mathbf{v}}|H_1)p(\tilde{\mathbf{w}}|H_1)}{p(\tilde{\mathbf{v}}|H_0)p(\tilde{\mathbf{w}}|H_0)} \lessgtr \tilde{\eta}, \tag{7}$$

with the threshold η chosen to satisfy the constraint $P_F = \int_{\Lambda(\tilde{\mathbf{v}},\tilde{\mathbf{w}})>\tilde{\eta}} p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_0) d\tilde{\mathbf{v}} d\tilde{\mathbf{w}} = \alpha$.

In our analysis, the signals pairs $(\tilde{\mathbf{x}}(m), \tilde{\mathbf{y}}(m))$ and $(\tilde{\mathbf{x}}(n), \tilde{\mathbf{y}}(n))$ are assumed to be known according to the assumptions of the worst case authentication. We begin our considerations assuming that the DMC parts of acquisition channels are Gaussian due to the highest possible differential entropy, i.e., largest sphere of ambiguity, in the class of all distributions with the bounded variance.

Assuming that noise in the coordinate domain follows i.i.d. Gaussian, i.e., $\mathbf{Z}_X \sim \mathcal{N}(\mathbf{0}, \sigma^2_{Z_X}\mathbf{I}_{N_X}), \mathbf{Z}_Y \sim \mathcal{N}(\mathbf{0}, \sigma^2_{Z_Y}\mathbf{I}_{N_Y}))$, in the projected domain one has $\tilde{\mathbf{Z}}_X \sim \mathcal{N}(\mathbf{0}, \sigma^2_{Z_X}\mathbf{C}_\mathbf{x})$ and $\tilde{\mathbf{Z}}_Y \sim \mathcal{N}(\mathbf{0}, \sigma^2_{Z_Y}\mathbf{C}_\mathbf{y})$, where $\mathbf{C}_\mathbf{x} = \mathbf{\Phi}_\mathbf{x}\mathbf{\Phi}_\mathbf{x}^T$ and $\mathbf{C}_\mathbf{y} = \mathbf{\Phi}_\mathbf{y}\mathbf{\Phi}_\mathbf{y}^T$. Then, for $\mathbf{Z}_X$ and $\mathbf{Z}_Y$ and known signal pairs $(\tilde{\mathbf{x}}(m), \tilde{\mathbf{y}}(m))$ and $(\tilde{\mathbf{x}}(n), \tilde{\mathbf{y}}(n))$, we obtain:

$$\begin{cases} H_0 : p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|\tilde{H}_0) = \mathcal{N}(\tilde{\mathbf{x}}(n), \sigma^2_{Z_X}\mathbf{C}_\mathbf{x})\mathcal{N}(\tilde{\mathbf{y}}(n), \sigma^2_{Z_Y}\mathbf{C}_\mathbf{y}), \\ H_1 : p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|\tilde{H}_1) = \mathcal{N}(\tilde{\mathbf{x}}(m), \sigma^2_{Z_X}\mathbf{C}_\mathbf{x})\mathcal{N}(\tilde{\mathbf{y}}(m), \sigma^2_{Z_Y}\mathbf{C}_\mathbf{y}). \end{cases} \tag{8}$$

It is important to note that the mapping based on the orthoprojectors does not change the variances of noise, i.e., $\sigma^2_{Z_X}\mathbf{C}_\mathbf{x} = \sigma^2_{Z_X}\mathbf{I}_{L_X}$ and $\sigma^2_{Z_Y}\mathbf{C}_\mathbf{y} = \sigma^2_{Z_Y}\mathbf{I}_{L_Y}$. Under such kind of mapping the impact is only reflected on the transformed mean vectors: $\tilde{\mathbf{x}}(m) = \Phi_\mathbf{x}\mathbf{x}(m), \tilde{\mathbf{x}}(n) = \Phi_\mathbf{x}\mathbf{x}(n), \tilde{\mathbf{y}}(m) = \Phi_\mathbf{y}\mathbf{y}(m), \tilde{\mathbf{y}}(n) = \Phi_\mathbf{y}\mathbf{y}(n)$.

This setup was intensively studied in literature (see, for instance[19, 21]) and the performance measures in terms of probabilities of error and their error exponents are available. For the purpose of this paper we will use the results of[21] limiting error probabilities in terms of Chernoff distance.

The Chernoff distance between two densities is defined as:

$$D_s(p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_1), p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_0)) = -\mu(s) = -\log \int_{\tilde{\mathcal{V}}^{L_X}} \int_{\tilde{\mathcal{W}}^{L_Y}} p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_1) \left(\frac{p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_1)}{p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_0)}\right)^s d\tilde{\mathbf{v}} d\tilde{\mathbf{w}}, \tag{9}$$

for $0 \leq s \leq 1$.

The Chernoff distance provides an upper bound on both P_F and probability of miss $P_M = \int_{\Lambda(\mathbf{v})<\eta} p(\mathbf{v}|H_1) d\mathbf{v}$[22]:

$$P_F \leq e^{\mu(s) - s\dot{\mu}(s)}, P_M \leq e^{\mu(s) + (1-s)\dot{\mu}(s)}, \tag{10}$$

where $\eta = \dot{\mu}(s)$ and $\dot{\mu}(s)$ denotes the first derivative of $\mu(s)$ with respect to s. The fastest convergence rate of the exponential terms is achieved for the optimal selection of s.

For the average probability of error[22]:

$$P_e \leq 0.5 \exp(\mu(s_m)), \tag{11}$$

where s_m is the value for which $\dot{\mu}(s) = 0$.

For the independent modalities, the Chernoff distance satisfies the additivity property and reduces to:

$$D_s(p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_1), p(\tilde{\mathbf{v}}, \tilde{\mathbf{w}}|H_0)) = D_s(p(\tilde{\mathbf{v}}|H_1), p(\tilde{\mathbf{v}}|H_0)) + D_s(p(\tilde{\mathbf{w}}|H_1), p(\tilde{\mathbf{w}}|H_0)). \tag{12}$$

Thus, presence of the second modality increases the convergence rate of the exponential terms to zero for any distribution. One can also extend these bounds to the considered Gaussian case (8) for which $s = s_m = 0.5$ and:

$$\begin{cases} P_F \leq \exp\left(-\frac{1}{8}\left(\frac{\tilde{d}_X^2}{\sigma_{Z_X}^2} + \frac{\tilde{d}_Y^2}{\sigma_{Z_Y}^2}\right)\right); \\ P_M \leq \exp\left(-\frac{1}{8}\left(\frac{\tilde{d}_X^2}{\sigma_{Z_X}^2} + \frac{\tilde{d}_Y^2}{\sigma_{Z_Y}^2}\right)\right), \end{cases} \tag{13}$$

and the average probability of error is:

$$P_e \leq \frac{1}{2} \exp\left(-\frac{1}{8}\left(\frac{\tilde{d}_X^2}{\sigma_{Z_X}^2} + \frac{\tilde{d}_Y^2}{\sigma_{Z_Y}^2}\right)\right), \tag{14}$$

where $\tilde{d}_X^2 = (\mathbf{x}(m) - \mathbf{x}(n))^T \mathbf{\Phi}_\mathbf{x}^T \mathbf{C}_\mathbf{x}^{-1} \mathbf{\Phi}_\mathbf{x} (\mathbf{x}(m) - \mathbf{x}(n))$, $\tilde{d}_Y^2 = (\mathbf{y}(m) - \mathbf{y}(n))^T \mathbf{\Phi}_\mathbf{y}^T \mathbf{C}_\mathbf{y}^{-1} \mathbf{\Phi}_\mathbf{y} (\mathbf{y}(m) - \mathbf{y}(n))$.

In order to verify the impact of random transforms on the multimodal binary authentication performance, one can assume that $\Phi_\mathbf{x} = \mathbf{I}_{N_X}, \Phi_\mathbf{y} = \mathbf{I}_{N_Y}$ that allows to rewrite (13)-(14) as follows:

$$\begin{cases} P_F \leq \exp\left(-\frac{1}{8}\left(\frac{d_X^2}{\sigma_{Z_X}^2} + \frac{d_Y^2}{\sigma_{Z_Y}^2}\right)\right); \\ P_M \leq \exp\left(-\frac{1}{8}\left(\frac{d_X^2}{\sigma_{Z_X}^2} + \frac{d_Y^2}{\sigma_{Z_Y}^2}\right)\right), \end{cases} \tag{15}$$

$$P_e \leq \frac{1}{2} \exp\left(-\frac{1}{8}\left(\frac{d_X^2}{\sigma_{Z_X}^2} + \frac{d_Y^2}{\sigma_{Z_Y}^2}\right)\right), \tag{16}$$

where $d_X^2 = (\mathbf{x}(m) - \mathbf{x}(n))^T (\mathbf{x}(m) - \mathbf{x}(n))$ and $d_Y^2 = (\mathbf{y}(m) - \mathbf{y}(n))^T (\mathbf{y}(m) - \mathbf{y}(n))$.

Thus, the impact of $\Phi_\mathbf{x}$ and $\Phi_\mathbf{y}$ is introduced through modification of the distance between vectors in $\mathbb{R}^{N_X}$, $\mathbb{R}^{N_Y}$ and $\mathbb{R}^{L_X}$, $\mathbb{R}^{L_Y}$, respectively. In fact, this impact is negative in sense that $\tilde{d}_X^2 < d_X^2$ and $\tilde{d}_Y^2 < d_Y^2$ for $L_X < N_X$ and $L_Y < N_Y$ leading to a performance loss. The loss is not unbounded and the corresponding limit is introduced by the Johnson-Lindenstrauss lemma.[23]

Therefore, to introduce the bounds on the distance $\tilde{d}^2$ we will use the results of Johnson-Lindenstrauss lemma (JLL).[23] It states that, with high probability, the geometry of a point cloud is not disturbed by certain Lipschitz mappings onto a space of dimension logarithmic in the number of points. In particular, the mapping $\mathbf{\Phi}$ can be taken as a linear mapping represented by an $L \times N$ matrix whose entries are randomly drawn from certain probability distributions. More particularly, $|\mathcal{M}|$ vectors in the Euclidean space can be projected down to $L \geq \frac{4 \log_2 |\mathcal{M}|}{\zeta^2/2 - \zeta^3/3}$ dimensions while incurring a distortion of at most $1 + \zeta$, $0 < \zeta < 1$, in their pairwise distances. In principle, this can be achieved by a dense $L \times N$ matrix and such a mapping takes $O(N \log_2 |\mathcal{M}|)$ (for fixed ζ). We refer interested readers to[24] for more details.

According to the JLL[23] one has:

$$(1-\zeta)\sqrt{\frac{L_X}{N_X}}||\mathbf{x}|| \leq ||\mathbf{\Phi_x x}|| \leq (1+\zeta)\sqrt{\frac{L_X}{N_X}}||\mathbf{x}||,$$
$$(1-\zeta)\sqrt{\frac{L_Y}{N_Y}}||\mathbf{y}|| \leq ||\mathbf{\Phi_y y}|| \leq (1+\zeta)\sqrt{\frac{L_Y}{N_Y}}||\mathbf{y}||, \quad (17)$$

where $0 < \zeta < 1$. Therefore, the random projections reduce the information distances in the corresponding error exponents due to the data processing inequality. Using the result of JLL (17) this loss can be approximately quantified as:

$$\mu(s) \approx \frac{s(s-1)}{2}\left(\frac{L_X}{N_X}\frac{d_X^2}{\sigma_{Z_X}^2} + \frac{L_Y}{N_Y}\frac{d_Y^2}{\sigma_{Z_Y}^2}\right). \quad (18)$$

The observed situation will be exemplified in the experimental part of the paper.

4. WORST CASE ADDITIVE ACQUISITION DISTORTIONS

The analysis performed in the previous Section was based on a conservative assumption that the acquisition distortions introduced by the authentication channel are modeled as i.i.d. Gaussian distribution. However, for the authentication performed within the Kerckhoffs security framework such an assumption is too restrictive. Moreover, it is highly questionable how accurate such an approximation of realistic distortions could be. From another point of view, realistic distortion modeling in the presented compound sense might be too complex and not tractable. Therefore, instead of following these approaches, we will consider an authentication protocol where similarly to data-hiding applications,[25] some freedom exists in selecting authentication channel that is the most harmful in terms of authentication accuracy loss. The main goal is to investigate a class of possible additive authentication channels that can be modeled in the class of i.i.d. distributions with a zero-mean and a limited variance. The results of such an analysis are the limits that object authentication systems could achieve in such settings in the terms of bounds on the probability of authentication error.

In our analysis we assume that the performance of the authentication protocol is defined only by the closest feature vectors. Then, the goal of our study will be to deduce an acquisition distortion statistical model that leads to the worst performance loss in the terms of the Bhattacharyya distance D_{bh} that is a special case of the Chernoff distance for $s = 0.5$. It is known to provide bounds on the average probability of error in discrimination between two equally likely classes.[19] The problem under consideration can be formulated similarly to.[26] The difference with the referred approach consists in the relaxation of the Gaussian masking constraint (in order to find an optimal distortion model) as well as in the change of the cost function to the Bhattacharyya distance. We will formulate and consider the problem only for the case of $\mathbf{x}$ modality since for the second one it is fully symmetric for the independent case studied in this paper.

Thus, we will be seeking for the solution to the following constrained optimization problems in the direct domain:

$$p(\mathbf{z_X}_{opt}) = \arg\max_{p(\mathbf{z}_X)} D_{bh}(p(\mathbf{z}_X); \mathbf{x}(m), \mathbf{x}(n));$$
$$s.t. \int ... \int_{-\infty}^{+\infty} p(\mathbf{z}_X) d\mathbf{z}_X = 1; Cov[\mathbf{Z}_X] = \sigma_{Z_X}^2 \mathbf{I}_{N_X}, \quad (19)$$

and in the projected domain:

$$p(\tilde{\mathbf{z}}_{X\,opt}) = \arg\max_{p(\tilde{\mathbf{z}}_X)} D_{bh}(p(\tilde{\mathbf{z}}_X); \mathbf{\Phi_x x}(m), \mathbf{\Phi_x x}(n));$$
$$s.t. \int ... \int_{-\infty}^{+\infty} p(\tilde{\mathbf{z}}_X) d\tilde{\mathbf{z}} = 1; Cov[\tilde{\mathbf{Z}}_X] = \sigma_{\tilde{Z}_X}^2 \mathbf{I}_{L_X}, \quad (20)$$

where $Cov[\mathbf{Z}]$ denotes a covariance matrix of $\mathbf{z}$. According to the results presented in,[26] one can expect that the seeking distribution $p(\mathbf{z_X}_{opt})$ will be defined on a discrete set. The solution to this problem is obtained by numerical optimization and can be found in Section 5.

In the case of (20) one can immediately note that:

$$\tilde{\mathbf{z}}_X = \mathbf{\Phi_x z}_X, \quad (21)$$

will be a Gaussian vector of length L_X with zero-mean vector and covariance matrix $\sigma_Z^2 \mathbf{\Phi_x \Phi_x^T}$ according to the Central Limit Theorem.[6] Therefore, the solution to (20) will be a multivariate Gaussian density with zero-mean and covariance $\sigma_Z^2 \mathbf{\Phi_x \Phi_x^T}$ disregarding distribution of $\mathbf{z}_X$. Thus, such a random projection performs a mapping of arbitrary statistics to Gaussian ones. Remarkably, that in the considered case the assumption about Gaussianity is valid and one can use the previous results.

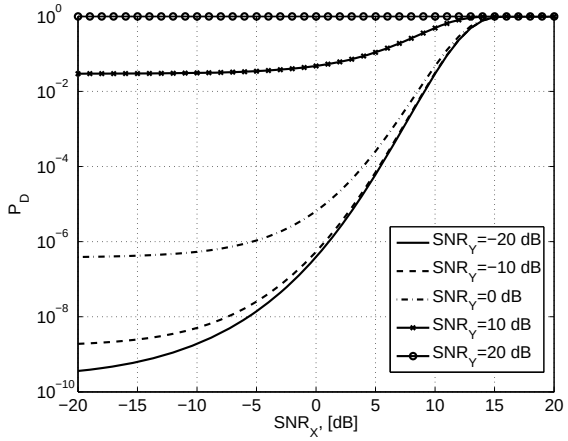


Figure 4. Probability of correct detection P_D for $P_F = 10^{-10}$ for $\frac{L_X}{N_X} = \frac{L_Y}{N_Y} = 1$.

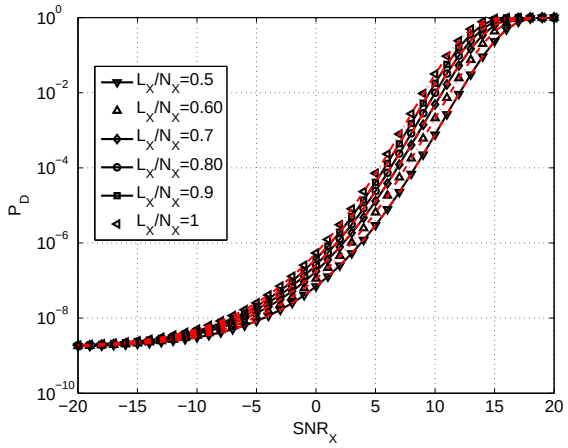


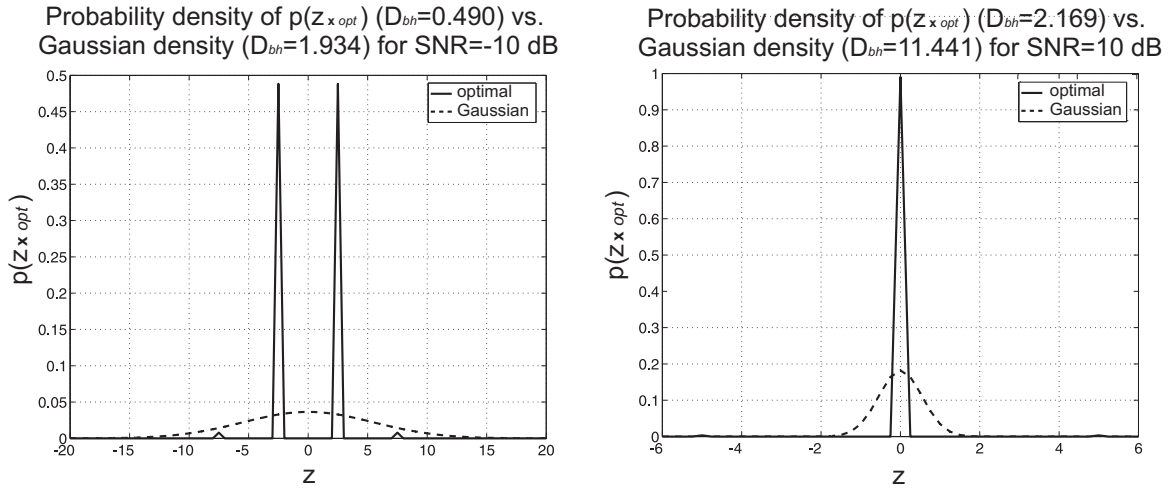
Figure 5. Probability of correct detection P_D for $P_F = 10^{-10}$ and various $\frac{L_X}{N_X}$ and $\frac{L_Y}{N_Y} = 1$ for $SNR_Y = -10$ dB. The dash lines show the corresponding approximations.

5. EXPERIMENTAL RESULTS AND CONCLUSIONS

In this section we will demonstrate the main obtained results. The impact of the second modality $\mathbf{Y}$ on authentication performance based on modality $\mathbf{X}$ is demonstrated in Figure 4 for $P_F = 10^{-10}$ without dimensionality reduction. The presence of noisy modality $\mathbf{Y}$ increases P_D. The impact of dimensionality reduction based on the random projections and approximation accuracy were investigated using both analytical formulas and Monte Carlo simulation for Gaussian data of lengths $N_X = N_Y = 2^{15}$. This selection is justified by the requirements of Johnson-Lindenstrauss lemma to produce a stable projections for the dimensionality reduction ratio $\frac{L}{N} = \{0.5; 0.6; , 0.7; 0.8; 0.9\}$ for $\zeta = 0.1$. The orthoprojectors $\mathbf{\Phi_x}$ and $\mathbf{\Phi_y}$ have been generated according to the Gaussian distribution with the parameters described in Section 3. The results of simulation for the probability P_D are shown in Figure 5. The dimensionality reduction of modality $\mathbf{X}$ for the fixed length of modality $\mathbf{Y}$ in the indicated ranges revealed the loss in performance bounded by ~ 3 dB. This loss is relative small price for the reduced complexity of processing and especially for the memory storage reduction and security/privacy enhancement.

The final set of experiments represents the solutions of the numerical problem defined in (19).

Since, according to the observation presented in Section 4, the acquisition distortions are well defined in the case of projected domain authentication that is Gaussian with a limited variance, the situation appears to be



(a) (b)

Figure 6. Worst case additive attack pdf for worst case direct domain authentication (modality $\mathbf{x}$) for various constraints on attack variance: 4-δ attack (a) and 3-δ attack (b).

different in the case of direct domain authentication. The distributions $p(\mathbf{z}_{\mathbf{x}opt})$ are discrete in this case and are significantly more dangerous in terms of D_{bh} loss versus Gaussian distribution of the same variance (Figure 6).

The analysis of the obtained results allows to reveal the following. While mapping from a high- to low-dimensional space leading to the original distance reduction from d^2 to $\tilde{d}^2$, the impact of this mapping on the binary classification error clearly depends on the statistics of the acquisition distortions. For instance, in the case of Figure 6,(a) the D_{bh} in the case of direct domain authentication are 1.934 and 0.490 for the case of Gaussian distortions and $p(\mathbf{z}_{\mathbf{x}opt})$, respectively, while in the projected domain it is equal to 1.914 for $\frac{L_X}{N_X} = 0.9$. Thus, one can state that reducing dimensions besides breaking curse of dimensionality provides a significant resistance of the authentication protocol to various acquisition distortion models.

6. CONCLUSIONS

In this paper we consider the problem of multimodal authentication with random projections in the worst-case formulation. Similarly to digital communication, we propose to limit the performance of teh considered authentication system based on the probability of error caused by two classes that are located at the smallest pairwise distance in the given class set. We demonstrated that the involvement of additional source of information is beneficial for the overall system performance accuracy. Contrarily to the existing conservative assumptions on acquisition channel distortions, we consider the opponent that has a relative freedom of selecting a distortion model in the class of additive channels with a bounded variance. We revealed that in such settings an optimal attacking distribution leading to the highest probability of misclassification is not continuous and differ significantly from the commonly assumed Gaussian statistics. Finally, we demonstrated that for the given choice of a random projector, the output will follow Gaussian distribution that provides a certain universality of the considered protocol to various additive distortion models.

In our future research, we are planning to extend the performed analysis to the issues of security and privacy.

7. ACKNOWLEDGMENTS

This work is supported by SNF projects 111643 and 1119770 and SNF IM2 project.

REFERENCES

1. F. Beekhof, S. Voloshynovskiy, O. Koval, R. Villan, and T.Pun, "Secure surface identification codes," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, San Jose, CA, USA, Jan 27–31 2008.
2. A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, June 2006.
3. S. Kung, M. Mak, and S. Lin, *Biometric authentication: a machine learning approach*, Prentice Hall Press, Upper Saddle River, NJ, USA, 2004.
4. B. Ulery, W. Fellner, A. Hicklin P. Hallinan and, and C. Watson, "Evaluation of selected biometric fusion techniques," Tech. Rep., http://www.itl.nist.gov/iad/894.03/pact/ir_7346_C.pdf.
5. R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 10, pp. 955–966, 1995.
6. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley and Sons, New York, 1991.
7. A. Ross, R. GovindarajanL. Hong, A. K. Jain, and S. Pankanti, "Feature level fusion using hand and face biometrics," in *SPIE Conf. Biometric Technology for Human Identification II*, 2005, pp. 196–204.
8. K. Kryszczuk and A. Drygajlo, "Credence estimation and error prediction in biometric identity verification," *Signal Processing*, vol. 88, pp. 916–925, April 2008.
9. K. Kryszczuk and A. Drygajlo, "Impact of feature correlations on separation between bivariate normal distributions," in *International Conference on Pattern Recognition (ICPR) 2008*, Tampa FL, USA, December 2008.
10. O. Koval, S. Voloshynovskiy, and T. Pun, "Error exponent analysis of person identification based on fusion of dependent/independent modalities," in *SPIE: Security and Watermarking of Multimedia Contents*, San Jose, California, January 2007.
11. A. Vetro, S. Draper, S. Rane, and J. Yedidia, *DISTRIBUTED SOURCE CODING: Theory, Algorithms and Applications (P.L. Dragotti and M. Gastpar ed.), chapter Securing Biometric Data*, Elsevier, 2009.
12. E. Martinian, S. Yekhanin, and J.S. Yedidia, "Secure biometrics via syndromes," in *43rd Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, USA, October 2005.
13. Y.-C. Lin, D. Varodayan, and B. Girod, "Image authentication based on distributed source coding," in *IEEE International Conference on Image Processing (ICIP2007)*, San Antonio, USA, April 2007, p. September.
14. T. Ignatenko and F.M.J. Willems, "On the security of xor-method in biometric authentication systems," in *The twenty-seventh symposium on Information Theory in the Benelux*, Noordwijk, The Netherlands, June 8-9 2006, pp. 197–204.
15. P. Tuyls, B. Skoric, and T. Kevenaar (Eds.), *Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting*, Springer, 2007.
16. S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Pun, "Random projections based item authentication," in *Proceedings of SPIE Photonics West, Electronic Imaging / Media Forensics and Security XI*, San Jose, USA, 2009.
17. A. Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires*, vol. IX, 1883.
18. U. Maurer, "A unified and generalized treatment of authentication theory," in *Proc. 13th Symp. on Theoretical Aspects of Computer Science (STACS'96)*. Feb 1996, vol. 1046 of *Lecture Notes in Computer Science*, pp. 387–398, Springer-Verlag.
19. S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall Signal Processing Series, 1993.
20. H.Herzberg and G.Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," *IEEE Trans. Information Theory*, vol. 44, no. 4, pp. 427–433, April 1996.
21. Sviatoslav Voloshynovskiy, Oleksiy Koval, and Thierry Pun, "Multimodal authentication based on random projections and distributed coding," in *Proceedings of the 10th ACM Workshop on Multimedia & Security*, Oxford, UK, September 22–23 2008.
22. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, New York: John Wiley and Sons, 1968.
23. W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into Hilbert space," *Contemporary Mathematics*, , no. 26, pp. 189–206, 1984.

24. D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *JOURNAL OF COMPUTER AND SYSTEM SCIENCES*, vol. 66, no. 4, pp. 671–687, 2003.
25. J. E. Vila-Forcen, S. Voloshynovskiy, O. Koval, F. Prez-Gonzlez, and T. Pun, "Quantization-based methods: Additive attacks performance analysis," *LNCS Transactions on Data Hiding and Multimedia Security*, May 2008.
26. A. McKellips and S. Verdu, "Worst case additive noise for binary-input channels and zero-threshold detection under constraints of power and divergence," *IEEE Transactions on Information Theory*, vol. 43, no. 4, pp. 1256–1264, July 1997.