ON SECURITY OF GEOMETRICALLY-ROBUST DATA-HIDING

Emre Topak, Sviatoslav Voloshynovskiy, Oleksiy Koval, José Emilio Vila Forcén and Thierry Pun

CUI, University of Geneva
24, rue du General-Dufour, CH-1211 Geneve 4, Switzerland

ABSTRACT

In this paper we analyze security of robust data-hiding in channels with geometrical transformations. We categorize possible decoding strategies for channels with geometrical transformations within the information-theoretic framework for geometrically-robust data-hiding. Furthermore, considering template-based and redundant-based design of geometrically robust data-hiding systems, we provide the analysis of general attacking strategies and particular attacking scenarios for each group of structured codebooks. Finally, reversibility of data-hiding and effect of security leakages on system performance are investigated.

1. INTRODUCTION

Digital data-hiding aims at communicating application-specific data reliably through a specified channel by embedding it into some digital multimedia documents. This data should be reliably extractable even some intentional and unintentional attacks were applied to the marked document.

In general case, digital data-hiding can be considered as a game between data-hider and attacker. O'Sullivan, Moulin and Ettinger were among the first who recognized this game [1]. In the extended version of the previous paper [2], Moulin and O'Sullivan have considered two possible set-ups. In the first one they assumed the availability of host at both encoder and decoder, i.e., the so-called *private game* and in the second one they considered a case, where the host is available only at the encoder, i.e., a *public game*. Moulin and O'Sullivan considered the games with the capacity as a cost function. Moreover, they assumed that the decoder is informed of the attack channel, and thus, applied *maximum likelihood (ML) decoding*.

The knowledge of attack channel at the decoder is not a very common case for most practical applications. More realistic set-up was considered by Somekh-Baruch and Merhav [3],[4] in assumption of *arbitrary varying channel* (AVC), whose state is not available to the decoder. In the first paper [3] Somekh-Baruch and Merhav have considered private game, where both capacity and error exponent were analyzed as the cost functions. The channel capacity is a good measure of performance, if one is interested to know the maximum rate of reliable communications. The error exponent provides the lowest achievable probability of error at a given information rate. From the practical point of view, the error exponents seem to be more attractive since they bring out clear and simple relationship between error probability, data rate, constraint length, and channel behaviour [5]. A remarkable result has been achieved since the attack channel was not known at the decoder [3] using *maximum mutual information* (MMI) decoding. This decoding strategy can be considered as a *universal decoding* for the class of AVC. Such a decoder can be regarded as a two part system that consists of channel state estimation (CSE) and decoder for the particular CSE output. These two procedures are iterated to guarantee the reliable communications at rates below the channel capacity defined by the max-min game.

In [4], Somekh-Baruch and Merhav have considered capacity of a public game using the same MMI decoding set-up. Being theoretically justified, this approach meets some difficulties in practical applications dealing with geometrical channels. In such kind of channels, the attacker applies some desynchronization transform to the watermarked data from a set of parametric transforms with large cardinality. On the data-hider side, the applied transform can be regarded as a random one with the uniform probability of appearance over the set of chosen cardinality.

To simplify the task of the decoder, most of data-hiding systems use certain simplifications that lead to the suboptimal performance of universal decoder. First, the CSE-decoding is implemented in the sequential two-step manner rather than in iterative way. Once one obtains the CSE, the channel state compensation (CSC) is applied and the message decoding is based directly on the recovered data. Second, to simplify the task of CSE, most of data-hiding techniques are exploiting specially structured codebooks instead of random coding. This is closely related to the use of special *pilot* or *template* signals that facilitate estimation problem often used in digital communications. We will refer to these codebooks as *geometrically structured codebooks*. Depending on the particular codebook design, they might be classified into two main groups:

- *template-based structured codebooks* in which a specially designed template or a pilot data is used to perform CSE and CSC [6];
- *redundant-based structured codebooks* in which codewords have special construction or statistics to aid CSE and CSC [7].

A thorough theoretical analysis of this geometrical synchronization framework is given in [8]. This analysis can be also quite indicative while considering security leakages of robust data-hiding schemes based on the structured codebooks.

The rest of the paper is organized as follows. In Section 2, problem formulation is presented. In Section 3, possible decoding strategies are considered. Afterwards, in Section 4, the information-theoretic framework to data-hiding synchronization is provided. Section 5 contains the analysis of attacking strategies and particular attacking scenarios for each group of structured codebooks. In Section 6, reversibility of data-hiding and the effect of security leakages on the cardinality of the decoding space are investigated. Finally, Section 7 concludes the paper.

Notations: 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 designate the realization of scalar and vector random variables, respectively. The superscript N is used to denote length-N vectors $\mathbf{x} = x^N = \{x[1], x[2], \ldots, x[N]\}$ with i^{th} element $x[i]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. Calligraphic fonts $\mathscr{X}$ designate sets $X \in \mathscr{X}$ and $|\mathscr{X}|$ denotes the cardinality of the set $\mathscr{X}$. $\mathbb{Z}$ and $\mathbb{R}$ stand for the set of integers and the set of real numbers, respectively. $H(X)$ denotes the entropy of a random variable X and $I(X;Y)$ designates the mutual information between random variables X and Y.

2. PROBLEM FORMULATION

Block diagram of a generic data-hiding system is presented in Fig. 1.

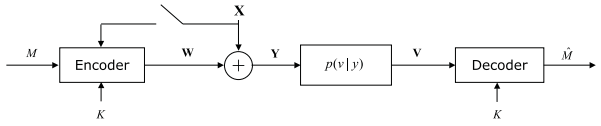

Figure 1: Communication set-up for data-hiding.

A stego data $\mathbf{y} \in \mathscr{Y}^N$ of length N is obtained by adding a watermark sequence $\mathbf{w} \in \mathscr{W}^N$ to a cover data $\mathbf{x} \in \mathscr{X}^N$ according to:

$$\mathbf{Y} = \mathbf{W} + \mathbf{X}. \tag{1}$$

$\mathbf{W}$ is generated by the encoder based on the message index $M \in \mathscr{M}$, the key $K \in \mathscr{K}$, and, possibly, the cover data $\mathbf{X}$.

The realization of key determines a particular codebook to be used at both encoder and decoder during communications. The codebooks are generated randomly and revealed to the encoder and the decoder with the knowledge of corresponding keys.

The watermark sequence combined with the host data is sent to the discrete memoryless channel (DMC) that converts the input $\mathbf{Y}$ to the output $\mathbf{V}$ in a probabilistic manner according to the channel transition probability $p(\mathbf{v}|\mathbf{y}) = \prod_{i=1}^{N} p(v_i|y_i)$. Afterwards, $\widehat{M}$ is decoded from $\mathbf{V}$ at the decoder with the knowledge of K.

When a geometrical transformation $\mathbf{A} = \mathbf{a}$ is applied to $\mathbf{Y}$, $\mathbf{Y}$ is transformed to the attacked data $\mathbf{V}$ as:

$$\mathbf{V} = T_A(\mathbf{Y}), \tag{2}$$

where the properties of T_A are determined by the applied geometrical transformation $\mathbf{A} = \mathbf{a} = \{a_1, a_2, \ldots, a_{J_a}\}$ from the space $\mathscr{A}^{J_a}$ of all possible geometrical transformations.

3. DECODING STRATEGIES

In general case, the technical implementation of channel state compensation at the decoder can be organized in three different ways depending on the combination of CSE, CSC and message decoding:

1. **Joint CSE-CSC-decoding:** In this case, decoding of M, estimations of $\mathbf{A}$ and $\mathbf{Y}$ are performed simultaneously as follows:

$$(\hat{m}, \hat{\mathbf{a}}, \hat{\mathbf{y}}) = \arg\max_{m \in \mathscr{M}, \mathbf{a} \in \mathscr{A}, \mathbf{y} \in \mathscr{Y}^N} f(m, \mathbf{a}, \mathbf{y}|\mathbf{v}). \tag{3}$$

   The solution of this joint optimization problem is quite involved and is outside of the scope of this problem.

2. **Iterative CSE-CSC-decoding:** In this case, decoding of M, estimations of $\mathbf{A}$ and $\mathbf{Y}$ are performed iteratively as follows (Fig. 2):

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a} \in \mathscr{A}} f(\mathbf{a}|\mathbf{v}, \hat{\mathbf{y}}, \hat{m}), \tag{4}$$

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y} \in \mathscr{Y}^N} f(\mathbf{y}|\mathbf{v}, \hat{\mathbf{a}}, \hat{m}), \tag{5}$$

$$\hat{m} = \arg\max_{m \in \mathscr{M}} f(m|\mathbf{v}, \hat{\mathbf{y}}, \hat{\mathbf{a}}). \tag{6}$$

   The universal decoding for the class of AVC is based on similar decoding framework. Although it is favorable from the performance point of view, the complexity of its design is higher than for hierarchical decoding.

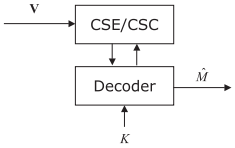

Figure 2: Iterative CSE-CSC-decoding.

3. **Hierarchical CSE-CSC-decoding:** In this case, decoding of M, estimations of $\mathbf{A}$ and $\mathbf{Y}$ are performed hierarchically as follows:

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a} \in \mathscr{A}} f(\mathbf{a}|\mathbf{v}), \tag{7}$$

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y} \in \mathscr{Y}^N} f(\mathbf{y}|\mathbf{v}, \hat{\mathbf{a}}), \tag{8}$$

$$\hat{m} = \arg\max_{m \in \mathscr{M}} f(m|\mathbf{v}, \hat{\mathbf{a}}, \hat{\mathbf{y}}), \tag{9}$$

   Hierarchical decoding is used mostly in practical data-hiding applications due to its low design complexity (Fig. 3). In the following section, a geometrically-robust data-hiding set-up based on this type of decoding will be given.

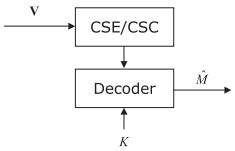

Figure 3: Hierarchical CSE-CSC-decoding.

4. PRACTICAL FRAMEWORK FOR INFORMATION-THEORETIC CONSIDERATION OF GEOMETRICALLY-ROBUST DATA-HIDING CODES

As a practical framework for geometrically-robust data-hiding, we propose the information-theoretic set-up presented in Fig. 4 that is based on a memoryless Multiple Access Channel (MAC) with side information (SI) about the host state $\mathbf{X}$ non-causally available at one of the encoders.

Inputs to the channel, $\mathbf{W_1}$ and $\mathbf{W_2}$, are parts of the watermark $\mathbf{W}$, where $\mathbf{W_1}$ is dedicated to pure message communication and $\mathbf{W_2}$ is additionally used for geometrical synchronization purposes. Message M to be communicated is split into two parts, M_1 and M_2, depending on the rate pair (R_1, R_2) and they are encoded into $\mathbf{W_1}$ and $\mathbf{W_2}$ using corresponding encoders.

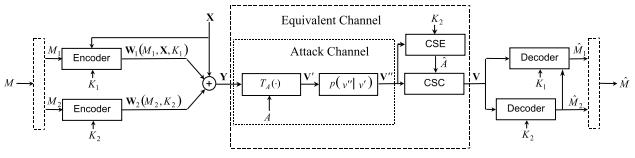

Figure 4: MAC framework to geometrically robust data-hiding.

In practice, only a small fraction of energy/space will be spent for $\mathbf{W}_2$ communications. This means that R_2 will be very small and asymptotically it will tend to zero as $N \to \infty$ and all rate will be assigned to R_1.

The achievable rates for practical scenarios, i.e., N is finite, are given by Harotunian [9] using error exponents in assumption of AVC.

5. ANALYSIS OF ATTACKING STRATEGIES

The objective of attacker that operates between the encoder and the decoder would be to exploit all available prior information about the data-hiding scheme and all security leakages from the observed stego data $\mathbf{Y}$ to destroy reliable communications. In order to comply with *Kerckhoff principle* [10] in the design of a *secure data-hiding protocol*, it is assumed that the attacker has access to encoding and decoding algorithms and has the knowledge of codebooks used at both encoders and decoders as the prior information. Furthermore, it is supposed that the attacker does not know:

- secret keys K_1 and K_2 or particular codebooks that are exploited by encoders and decoders for ongoing communications,
- indexes M_1 and M_2 that are sent by corresponding encoders,
- the original host image $\mathbf{X}$ that carries communicated watermark codewords $\mathbf{W}_1$ and $\mathbf{W}_2$.

Under given conditions, the attacker may apply one of the following *attacking strategies*:

- Statistical signal processing attacks: the attacker exploiting the knowledge of statistics of the watermark and of the host data may estimate the watermark, subtract the estimate from the stego data and add noise, thus avoiding inverse mapping, to decrease the rate of reliable communications;
- Geometrical attacks: the attacker may apply a geometrical transformation to the stego data to desynchronize the communication between encoder and decoder of the data-hiding system;
- Key space search attacks: the attacker with an access to the decoder and with the knowledge of codebooks may prefer to perform "cryptographic like" attack by decoding through all possible codebooks, i.e., *exhaustive search*, and to subtract the decoded codeword from the stego data to destroy the communications. Due to the equivocation, every codebook has some security leaks that could simplify the search of attacker [11]. Moreover, for robustness to geometrical attacks, we further introduce redundancy into the codebook structure. Thus, the attacker may try to benefit from the particular codebook design in reducing the search space.

In the following sections, the cardinalities of the search spaces for attacking scenarios that are inspired by the given strategies for each group of structured codebooks based on the proposed MAC framework will be investigated for theoretical set-ups, i.e., for $N \rightarrow \infty$.

5.1 Attacks against Template-Based Structured Codebooks

Attacks against template-based structured codebooks benefit from the fact that template $\mathbf{W}_2$ is only key-dependent and unique for a particular key $K_2 = k_2$. Thus, the attacker with the access to codebooks would look for a jointly-typical pair $(\widehat{\mathbf{W}}_2, \mathbf{Y})$. The cardinality of the $\mathbf{W}_2$ decoding space for the attacker is $|\mathscr{K}_2|$, where $|\mathscr{K}_2|$ represents the total number of codebooks for $\mathbf{W}_2$.

One may encounter with following scenarios depending on the key management protocol for K_1 and K_2 [8]:

- The data-hider uses the same key at both encoders, i.e., $K_1 = K_2 = K$, and there is a one-to-one correspondence between the codebooks of $\mathbf{W}_1$ and $\mathbf{W}_2$ for a given key K (scenario 5.1.1),
- The data-hider has different keys for each encoder, i.e., $K_1 \neq K_2$, and there is no relationship between the codebooks of $\mathbf{W}_1$ and $\mathbf{W}_2$ (scenario 5.1.2),
- The data-hider has different keys for each encoder, i.e., $K_1 \neq K_2$, but K_2 is fixed and is the same for all users (scenario 5.1.3).

In Table 1, the upper bound on the cardinality of the $\mathbf{U}$ decoding space for the attacker is given for each scenario:

Scenario	The upper bound on the cardinality of the $\mathbf{U}$ decoding space for the attacker
5.1.1	$\lvert\mathscr{K}_2\rvert + 2^{N(R_1+I(U;X\lvert K_1))}$
5.1.2	$\lvert\mathscr{K}_2\rvert + \lvert\mathscr{K}_1\rvert 2^{N(R_1+I(U;X\lvert K_1))}$
5.1.3	$1 + \lvert\mathscr{K}_1\rvert 2^{N(R_1+I(U;X\lvert K_1))}$

Table 1: The upper bound on the cardinality of the $\mathbf{U}$ decoding space for the attacker for various design scenarios of template-based structured codebooks

Therefore, it is beneficial for the data-hider to keep different keys for each encoder.

5.2 Attacks against Redundant-Based Structured Codebooks

In the case of redundant-based structured codebooks, codewords are generated having special features or statistics to facilitate the geometrical synchronization at the decoder. Therefore, one would expect the attacker to benefit from these statistics in the search of $\mathbf{W}_2$ part. By observing the stego data $\mathbf{Y}$, the attacker could learn the statistics of $\mathbf{W}_2$ even when the key K_2 is not available. Furthermore, the knowledge of statistics for $\mathbf{W}_2$ reduces the ambiguity in finding $\mathbf{W}_2$. For the attacker with an access to the codebooks, the upper bound for the cardinality of the $\mathbf{W}_2$ decoding space is $|\mathscr{K}_2|2^{NR_2}$.

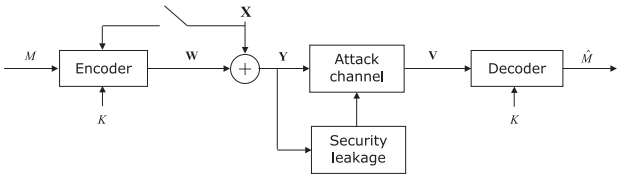

Figure 5: Attack design based on security leakages.

One may encounter with following scenarios depending on the statistical codebook design strategy for $\mathbf{W}_2$ [8]:

- The statistics of $\mathbf{W}_2$ are the same for all codebooks (scenario 5.2.1),
- The statistics of $\mathbf{W}_2$ are different for all user codebooks and there is a one-to-one relationship between the codebooks of $\mathbf{W}_1$ and $\mathbf{W}_2$ (scenario 5.2.2),
- The statistics of $\mathbf{W}_2$ are different for all user codebooks and there is no relationship between the codebooks of $\mathbf{W}_1$ and $\mathbf{W}_2$ (scenario 5.2.3).

In Table 2, the upper bound on the cardinality of the $\mathbf{U}$ decoding space for the attacker is given for each scenario:

Scenario	The upper bound on the cardinality of the $\mathbf{U}$ decoding space for the attacker
5.2.1	$\lvert\mathscr{K}_2\rvert 2^{NR_2} + \lvert\mathscr{K}_1\rvert 2^{N(R_1+I(U;X\lvert K_1))}$
5.2.2	$\lvert\mathscr{K}_2\rvert 2^{NR_2} + 2^{N(R_1+I(U;X\lvert K_1))}$
5.2.3	$\lvert\mathscr{K}_2\rvert 2^{NR_2} + \lvert\mathscr{K}_1\rvert 2^{N(R_1+I(U;X\lvert K_1))}$

Table 2: The upper bound on the cardinality of the $\mathbf{U}$ decoding space for the attacker for various design scenarios of redundant-based structured codebooks

6. REVERSIBILITY OF DATA-HIDING AND THE EFFECT OF SECURITY LEAKAGES ON THE CARDINALITY OF THE DECODING SPACE

After decoding $\mathbf{U}$ in the scenarios given in the previous Section, it is possible for the attacker also to obtain $\mathbf{X}$. For example, in the Costa set-up [12], which is proposed for the Gaussian formulation of the Gel'fand-Pinsker problem [13], $\mathbf{U} = \mathbf{W}_1 + \alpha\mathbf{X}$. Since $\mathbf{Y} - \widehat{\mathbf{W}}_2 = \mathbf{X} + \mathbf{W}_1$, $\mathbf{X}$ can be calculated if the jointly-typical $\left(\mathbf{U}, \left(\mathbf{Y} - \widehat{\mathbf{W}}_2\right)\right)$ pair is found [14]. The possibility for the attacker to obtain $\mathbf{X}$ means the total failure of the communications.

When codebooks for $\mathbf{W}_1$ are generated by distributing all possible $\mathbf{U}$ sequences to the codebooks uniquely, the cardinality of decoding space depends on the ambiguity $2^{H(\mathbf{U})}$ about $\mathbf{U}$. Therefore, one would expect the cardinalities $|\mathscr{K}_1|2^{N(R_1+I(U;X|K_1))}$ and $2^{H(\mathbf{U})}$ to be equal in the limit case.

However, attacker's knowledge about the stego data $\mathbf{Y}$ reduces this ambiguity to $H(\mathbf{U}|\mathbf{Y})$ as:

$$\begin{aligned} H(\mathbf{U}|\mathbf{Y}) &= H(\mathbf{U}|\mathbf{X}) - [I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{X})], \quad (10) \\ &\leq H(\mathbf{U}) - [I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{X})], \quad (11) \end{aligned}$$

where the inequality follows since conditioning reduces the entropy [15]. Thus, if $I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{X}) \neq 0$, then the cardinality of search trials for the attacker can be decreased based on the observed $\mathbf{Y}$. We refer interested readers to [14] for more details on this subject.

7. CONCLUSION

In this paper, security of robust data-hiding in channels with geometrical transformations is analyzed. Decoding strategies for channels with geometrical transformations are given with particular examples. Among those strategies, the MAC framework based on hierarchical decoding is considered for the design of practical rate maximizing data-hiding codes that are robust to geometrical transformations. The corresponding methods based on this MAC framework are classified into groups of template-based and redundant-based codebooks depending on the particular codebook design. The analysis of security leaks of each codebook structure is performed in terms of the upper bound on the cardinality of the decoding space for the attacker to design the worst case attack. Finally, reversibility of data-hiding is introduced and the effect of security leakages on the system performance is demonstrated.

8. ACKNOWLEDGMENT

This paper was partially supported by SNF Professeur Boursier grant PP002–68653, by the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT and Swiss IM2 projects. The authors are also thankful to the members of SIP group for many helpful discussions during group seminars.

The information in this document reflects only the authors views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

REFERENCES

[1] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger. Information-theoretic analysis of steganography. In *Proc. IEEE Symp. on Information Theory*, Boston, MA, August 1998.

[2] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory*, 49(3):563–593, 2003.

[3] A. Somekh-Baruch and N. Merhav. On the error exponent and capacity games of private watermarking systems. *IEEE Transactions on Information Theory*, 49(3):537–562, 2003.

[4] A. Somekh-Baruch and N. Merhav. On the capacity game of public watermarking systems. *IEEE Transactions on Information Theory*, 20(3):511–524, 2004.

[5] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, 11:3–17, 1965.

[6] S. Pereira and T. Pun. Fast robust template matching for affine resistant image watermarking. In *International Workshop on Information Hiding*, volume LNCS 1768 of *Lecture Notes in Computer Science*, pages 200–210, Dresden, Germany, 29 September –1 October 1999. Springer Verlag.

[7] F. Deguillaume, S. Voloshynovskiy, and T. Pun. Method for the estimation and recovering of general affine transforms in digital watermarking applications. In *IS&T/SPIE's 14th Annual Symposium, Electronic Imaging 2002: Security and Watermarking of Multimedia Content IV*, volume 4675, pages 313–322, San-Jose, CA, USA, January 20-25 2002.

[8] E. Topak, S. Voloshynovskiy, O. Koval, M.K. Mihcak, and Thierry Pun. Security analysis of robust data hiding with geometrically structured codebooks. In *Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents*, San Jose, CA, USA, January 2005.

[9] M. E. Haroutunian. New bounds for e-capacities of arbitrarily varying channel and channel with random parameter. *Mathematical Problems of Computer Science*, 22:44–59, 2001.

[10] A. Kerckhoff. La cryptographie militaire. *Journal des sciences militaires*, 9:5–38, 1883.

[11] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, October 1949.

[12] M. Costa. Writing on dirty paper. *IEEE Trans. on Information Theory*, 29(3):439–441, May 1983.

[13] S.I. Gel'fand and M.S. Pinsker. Coding for channel with random parameters. *Probl. Control and Inf. Theory*, 9(1):19–31, 1980.

[14] S. Voloshynovskiy, O. Koval, E. Topak, J. Vila, and T. Pun. On reversibility of random binning techniques: Security and multimedia perspectives. In *International Workshop on Information Hiding (submitted)*, Barcelona, Catalonia (Spain), June 6–8 2004.

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