Multimedia Systems (2005) 11(2): 121–132
DOI 10.1007/s00530-005-0196-1

REGULAR PAPER

Emre Topak · Sviatoslav Voloshynovskiy · Oleksiy Koval · M. Kivanc Mihcak · Thierry Pun

Towards geometrically robust data-hiding with structured codebooks

Published online: 15 November 2005

Abstract In this paper we analyze performance of practical robust data-hiding in channels with geometrical transformations. By applying information-theoretic argument we show that performance of a system designed based on both random coding and random binning principles is bounded by the same maximal achievable rate for the cases when communication channel includes geometrical transformations or not. Targeting to provide theoretic performance limits of practical robust data-hiding we model it using a multiple access channel (MAC) with side information (SI) available at one of encoders and present the bounds on achievable rates of reliable communications to such a protocol. Finally, considering template-based and redundant-based design of geometrically robust data-hiding systems, we perform security analysis of their performance and present results in terms of number of trial efforts the attacker needs to completely remove hidden information.

Keywords Security analysis · Robust data-hiding · Structured codebooks · Geometrical synchronization · Multiple access channel

1 Introduction

Digital data-hiding is the art of information communication by embedding it into some digital multimedia documents. Being embedded, this information should be reliably decodable even if some intentional and unintentional attacks were applied to the marked document.

In general case, the digital data-hiding can be considered as a game between the data-hider and the 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 set-up 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., *public game*. Moulin and O'Sullivan considered the games with the coding 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 the 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 that the attacker strategy is not known neither to encoder nor decoder. Moreover, they supposed that any conditional pdf that satisfies certain energy constraint might be a valid attacker choise. In particular, in [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. In fact, one can consider it as the maximum number of distinct messages that can be communicated via a given host data. 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 *universal decoding* for this class of channels. 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 more recent paper [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

E. Topak · S. Voloshynovskiy (✉) · O. Koval · T. Pun
CUI-University of Geneva, Stochastic Image Processing Group, 24 rue General-Dufour, 1211 Geneva, Switzerland
E-mail: svolos@cui.unige.ch, http://sip.unige.ch

M. K. Mihcak
Microsoft Research, Redmond, USA

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. This is closely related to the use of special *pilot* or *template* signals that facilitate estimation problem often used in digital communications. In the following, 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, 7];
- *redundant-based structured codebooks* in which codewords have special construction or statistics to aid CSE and CSC [8–11].

Although the practical usefulness of CSE and CSC was demonstrated in the papers referred above, a thorough theoretical analysis of this geometrical synchronization framework still remains an open and a little studied problem. Therefore, to theoretically justify the design rules of these practical data-hiding techniques, we suggest to consider this problem within information-theoretic framework. Moreover, such a justification might introduce a common basis for the analysis of these techniques using well-established practical communications principles of synchronization via CSE/CSC.

This analysis can be also quite indicative while considering security leakages of robust data-hiding schemes based on the structured codebooks. Security leakages of structured codebooks should be investigated from the position of designing the worst case attacks to destroy the reliable communications. Therefore, the goal of this paper is to put more light on the security and information-theoretical analysis of geometrically robust data-hiding.

The rest of the paper is organized as follows. In Sect. 2, an information-theoretic analysis of data-hiding is presented. In Sect. 3, the impact of geometrical attacks on the communications performance is considered. Afterwards, in Sect. 4, the information-theoretic framework to data-hiding synchronization is provided. Sections 5 and 6 contain the analysis of the template-based structured codebooks and the redundant-based structured codebooks, respectively. In Sect. 7, the security leaks and attacking strategies for each structured codebook group are investigated. Finally, Sect. 8 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 ith 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 $\mathcal{X}$ designate sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of the set $\mathcal{X}$. $\mathbb{Z}$ 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 Information-theoretic analysis of data-hiding

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

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

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

$\mathbf{W}$ is generated by the encoder based on the message index M that is uniformly distributed over the set $\mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\}$, where $|\mathcal{M}| = 2^{NR}$, the key $K \in \mathcal{K}, \mathcal{K} = \{1, 2, \ldots, |\mathcal{K}|\}$, and, possibly, the host $\mathbf{X}$. $R = \frac{1}{N}\log_2|\mathcal{M}|$ is the rate of communications.

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 encoder and decoder with the knowledge of corresponding keys.

Depending on whether or not non-causal host state information $\mathbf{X}$ is taken into account in the watermark sequence generation, *random binning* or *random coding* are used for codebook design, respectively. In the random coding, Fig. 2

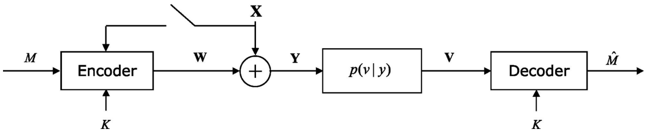


Fig. 1 Communication set-up for data-hiding

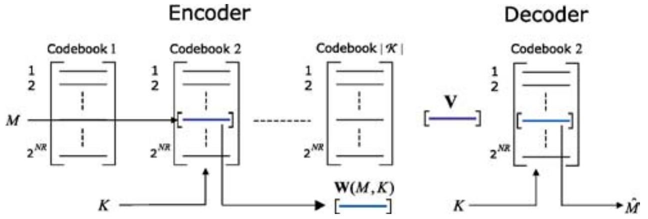


Fig. 2 Communications scenario based on random coding

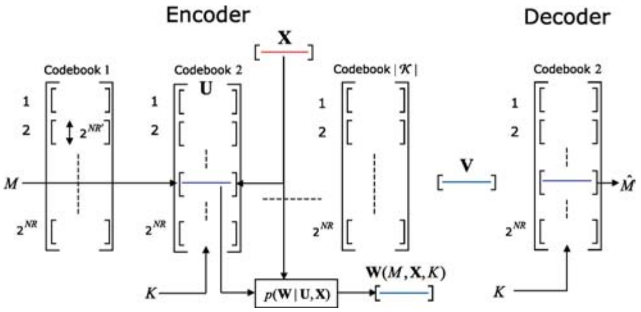


Fig. 3 Communications scenario based on random binning

[12], the encoder sends the codeword $\mathbf{W}(M, K)$, which corresponds to a particular realization of M in the codebook determined by K, as the watermark sequence. In the random binning, Fig. 3 [13], in the codebook defined by K, the encoder seeks a codeword $\mathbf{U}$ in the bin determined by M, which is jointly-typical with $\mathbf{X}$ [13]. After finding the jointly-typical $(\mathbf{U}, \mathbf{X})$ pair, the encoder generates the watermark sequence $\mathbf{W}(M, \mathbf{X}, K)$ according to a probabilistic mapping $p(\mathbf{w} \mid \mathbf{u}, \mathbf{x})$.

The watermark sequence combined with the host data is sent to a 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 $p(\mathbf{v}|\mathbf{y}) = \prod_{i=1}^{N} p(v_i|y_i)$.

At the decoder, $\widehat{M}$ is decoded from $\mathbf{V}$ with the knowledge of K. In the random coding, the decoder looks through the codebook defined by K for the codeword $\mathbf{W}(\widehat{M}, K)$ which is jointly typical with $\mathbf{V}$. When such a unique codeword $\mathbf{W}(\widehat{M}, K)$ is found, the index $\widehat{M}$ is declared as the decoded message. In the random binning, the decoder looks for a codeword $\mathbf{U}$ that is jointly-typical with $\mathbf{V}$ in the K-defined codebook. When such a unique codeword $\mathbf{U}$ is found, the index $\widehat{M}$ of the bin that contains $\mathbf{U}$ is considered as the decoded message.

3 Influence of geometrical attacks on the communications performance

3.1 Modeling of geometrical attacks

When a geometrical transformation $T_A(.)$ is applied to $\mathbf{Y}$, its pixel coordinates are modified accordingly.[1] The result $\mathbf{V}$ of these operations is called the attacked data:

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

where the subscript A represents the type of the geometrical transformation applied to $\mathbf{Y}$. Affine, bilinear and projective transformations are examples that A can take.

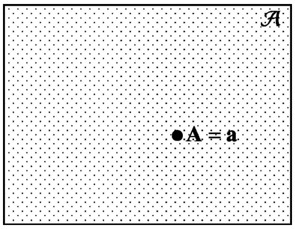


Fig. 4 The space $\mathcal{A}$ of all possible geometrical tranformations and its particular element $\mathbf{A} = \mathbf{a}$

A can be parameterized by a set of J parameters $\mathbf{a} = (a_1, a_2, \ldots, a_J)$ such that $\mathbf{a} \in \mathbb{Z}^J$.[2] For example, when A takes the form of the affine subclass of general geometrical transformations, a pixel at coordinates (n_1, n_2) in $\mathbf{Y}$, i.e., $y[n_1, n_2]$, will be transferred to the new coordinates (n_1', n_2') in $\mathbf{V}$, i.e., $v[n_1', n_2']$, according to:

$$\begin{bmatrix} n_1' \\ n_2' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix}. \tag{3}$$

In this case, $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)$. If we assume that Fig. 4 represents the space $\mathcal{A}$ of all possible geometrical transformations, then a particular transformation $\mathbf{A} = \mathbf{a}$ will be represented by a dot in this space. Total number of elements in this space is defined by the cardinality $|\mathcal{A}|$.

However, in practical data-hiding applications, an intentional geometrical attack space would not include all elements of $\mathcal{A}$ due to the visual acceptability constraint. For example, it is not expected that the attacker would rotate a stego data more than for 10 degrees. Therefore, without loss of generality, we will consider the set of $\epsilon-$typical geometrical transformations $\mathcal{A}_\epsilon^{(J)}(A)$ [12], such that the sample entropy is $\epsilon-$close to the true entropy and $|\mathcal{A}_\epsilon^{(J)}| < |\mathcal{A}|$, as the space of possibly applied geometrical transformations. In the case when $\mathbf{a} \in \mathbb{R}^J$, we refer to the volume of a set instead of cardinality [12].

If the parameters of $\mathbf{a} = (a_1, a_2, \ldots, a_J)$ are distributed independently and identically according to $p(a)$, then, $|\mathcal{A}_\epsilon^{(J)}|$ will be upper bounded as [12]:

$$|\mathcal{A}_\epsilon^{(J)}| \le 2^{J(H(A)+\epsilon)}, \tag{4}$$

where $H(A) = -\sum p(a) \log_2 p(a)$ and the summation is performed over the set of values that a can take.

It should be noticed that most of practically important families of geometrical transformations can be represented by a finite number of parameters, e.g., translation (2), affine transforms (6), bilinear transforms (8), projective transforms (10). At the same time, the local random transformations introduced in Stirmark Benchmark [14] might require larger number of parameters. This is true for the case considered by Voloshynovskiy et al. [10], where it was shown a possibility of random bending attack representation by a set of local affine transforms. In this case, J depens on the image

[1] It should be noticed here that we did not assume memory effects in the channel due to the intersymbol interference caused by the interpolation.

[2] In general case, one can assume $\mathbf{a} \in \mathbb{R}^J$.

size, i.e., it is defined by the number of blocks used for local approximation times number of affine transform parameters. However, besides the drastic increase of J, the upper bound (4) remains valid for our analysis.

3.2 Achievable rates of data-hiding in channels with geometrical transformations

Consider a theoretical communications set-up, where the decoder neither has a geometrical synchronization framework for recovery nor a priori knowledge about the applied geometrical transformation. It is inevitable for this decoder to regard all elements of $\mathcal{A}_\epsilon^{(J)}$ as possibly applied ones and, thus, to perform an exhaustive decoding for each $\mathbf{a} \in \mathcal{A}_\epsilon^{(J)}$.

In the following sections, achievability of reliable communications in channels with geometrical transformations is analyzed for random coding and random binning strategies starting from a communications scenario without any geometrical transformations. The analysis is carried out for the *theoretic set-ups*, where the lengths of communicated sequences asymptotically approach infinity.

Reliability of the communications is measured by a probability of decoding error, P_e, that is the probability that the decoded message $\widehat{M}$ is not equal to the sent message M, i.e., $Pr[\widehat{M} \neq m | M = m]$.

3.2.1 Communication set-ups based on random coding

In the case of random coding, the decoder will make a decoding error in following situations [12]:

- *There is not any codeword* $\mathbf{W}$*, which is jointly-typical with* $\mathbf{V}$ *in the codebook determined by* K: According to the asymptotic equipartition property (AEP) [12], this event is unlikely.
- *Another codeword* $\mathbf{W}'$ *from the codebook such that* $\big(\mathbf{W}' \neq \mathbf{w} | \mathbf{W} = \mathbf{w}\big)$ *is jointly-typical with* $\mathbf{V}$: According to the AEP, any $\mathbf{W}'$ from the $K-$defined codebook and $\mathbf{V}$ constitutes a jointly typical pair with the probability $2^{-N(I(W;V|K)-\epsilon)}$, where ϵ is an arbitrary small positive number, i.e., $\epsilon \to 0$. Since there are $\big(2^{NR_{\mathrm{RC}}} - 1\big)$ different $\mathbf{W}'$ apart from $\mathbf{W}' = \mathbf{w}$ in a particular codebook, the probability of decoding error in the random coding case, $P_\mathrm{e}^{\mathrm{RC}(N)}$, is upper bounded by:

$$P_\mathrm{e}^{\mathrm{RC}(N)} \leq 2^{NR_{\mathrm{RC}}} 2^{-N(I(W;V|K)-\epsilon)}, \tag{5}$$

where R_{RC} is the random coding-based communication rate in channels without geometrical transformations. If R_{RC} satisfies the condition:

$$R_{\mathrm{RC}} \leq I(W;V|K) - \epsilon, \tag{6}$$

then, $P_\mathrm{e}^{\mathrm{RC}(N)} \to 0$, as $N \to \infty$ and $\epsilon \to 0$.

In channels with geometrical transformations, when the decoding is performed at all elements of the space $\mathcal{A}_\epsilon^{(J)}$, the upper bound in (5) becomes:

$$\begin{aligned} P_\mathrm{e}^{\mathrm{RC}(N)} &\leq |\mathcal{A}_\epsilon^{(J)}| 2^{NR_{\mathrm{RC}}^G} 2^{-N(I(W;V|K)-\epsilon)}, \\ &\leq 2^{N\frac{1}{N}\log_2|\mathcal{A}_\epsilon^{(J)}|} 2^{NR_{\mathrm{RC}}^G} 2^{-N(I(W;V|K)-\epsilon)}, \\ &\leq 2^{N(\frac{1}{N}\log_2|\mathcal{A}_\epsilon^{(J)}| + R_{\mathrm{RC}}^G - (I(W;V|K)-\epsilon))}, \end{aligned} \tag{7}$$

where R_{RC}^G is the random coding-based communication rate in channels with geometrical transformations. Therefore, if R_{RC}^G satisfies the condition:

$$R_{\mathrm{RC}}^G \leq I(W;V|K) - \epsilon - \frac{1}{N}\log_2\big|\mathcal{A}_\epsilon^{(J)}\big|, \tag{8}$$

$P_\mathrm{e}^{\mathrm{RC}(N)} \to 0$, as $N \to \infty$ and $\epsilon \to 0$. Moreover, taking into account (4), (8) can be rewritten in the following form:

$$R_{\mathrm{RC}}^G \leq I(W;V|K) - \epsilon - \frac{J(H(A)+\epsilon)}{N}. \tag{9}$$

As $N \to \infty$ and $\epsilon \to 0$ and $\mathcal{A}_\epsilon^{(J)}$ is finite, $\frac{J(H(A)+\epsilon)}{N}$ term in (9) vanishes and the upper bound on R_{RC}^G reduces to:

$$R_{\mathrm{RC}}^G \leq I(W;V|K) - \epsilon, \tag{10}$$

which coincides with (6) that bounds the rate in channels without geometrical transformations.

To prove the converse part of coding theorem for random coding based communications in channels with geometrical transformations one can exploit standard proof technique [12] in order to upper bound the rate $R_{\mathrm{RC}}^G + \frac{1}{N}\log_2|\,\mathcal{A}_\epsilon^{(J)}|$. Due to the imposed constraint on $|\mathcal{A}_\epsilon^{(J)}|$ it is easy to show that the impact of the log term in the sum rate asymptotically vanishes.

Therefore, based on the above consideration one can conclude that in a theoretical set-up based on random coding scenario, the maximum rate of reliable communications is not affected by the applied geometrical transformations.

3.2.2 Communication set-ups based on random binning

In the case of random binning, one encounters with a coding error in following situations:
Encoder:

- *There is not any codeword* $\mathbf{U}$ *in the codebook defined by* K*, which is jointly-typical with* $\mathbf{X}$: According to the AEP, any codeword $\mathbf{U}$ and $\mathbf{X}$ may form a jointly-typical pair with the probability $2^{-N(I(U;X|K)-\epsilon)}$. Since there are $2^{NR'}$ codewords $\mathbf{U}$ for any M in a particular codebook defined by K, the probability of this event will be bounded by

$$\begin{aligned} P_{e_1}^{RB(N)} &\leq \big(1 - 2^{-N(I(U;X|K)-\epsilon)}\big)^{2^{NR'}}, \\ &\leq \exp\big(-2^{N(R'-I(U;X|K)+\epsilon)}\big), \end{aligned} \tag{11}$$

where we used the fact that $(1-x)^n \leq e^{-nx}$. If $R' > I(U;X|K) - \epsilon$, $P_{e_1}^{RB(N)} \to 0$ as $N \to \infty$ and $\epsilon \to 0$.

Decoder:

- *There is not any codeword* $\mathbf{U}$ *in the* $K-$*defined codebook, which is jointly-typical with* $\mathbf{V}$: According to the AEP, this event is unlikely.
- *A codeword* $\mathbf{U}$ *from another bin* $\widehat{M}$ *such that* $\left(\widehat{M} \neq m | M = m\right)$ *is jointly-typical with* $\mathbf{V}$: According to the AEP, arbitrary codeword $\mathbf{U}$ from the codebook defined by K and $\mathbf{V}$ may form a jointly-typical pair with the probability $2^{-N(I(U;V|K)-\epsilon)}$. Since there are $\left(2^{NR_{RB}} - 1\right)$ bins in total with an index $\widehat{M}$ such that $\widehat{M} \neq m$, the probability of decoding error in the random binning case, $P_{e_2}^{RB(N)}$, is upper bounded by:

  $$P_{e_2}^{RB(N)} \leq 2^{N[R_{RB}+R']} 2^{-N(I(U;V|K)-\epsilon)}, \tag{12}$$

  where $R' = I(U;X|K) + \epsilon$ and R_{RB} is the random binning-based communication rate in channels without geometrical transformations.

If the data-hider communicates with the following condition on the rate:

$$R_{RB} \leq I(U;V|K) - I(U;X|K) - 2\epsilon, \tag{13}$$

then, $P_{e_2}^{RB(N)} \to 0$, as $N \to \infty$ and $\epsilon \to 0$.

In channels with geometrical transformations, when the decoder performs the decoding at all elements of the space of typical geometrical transformations, $P_{e_2}^{RB(N)}$ will be upper bounded by:

$$\begin{aligned} P_{e_2}^{RB(N)} &\leq \left|\mathcal{A}_\epsilon^{(J)}\right| 2^{N[R_{RB}^G+R']} 2^{-N(I(U;V|K)-\epsilon)}, \\ &\leq 2^{N\frac{1}{N}\log_2|\mathcal{A}_\epsilon^{(J)}|} 2^{N[R_{RB}^G+I(U;X|K)+\epsilon]} \\ &\quad \times 2^{-N(I(U;V|K)-\epsilon)}, \\ &\leq 2^{N(\frac{\log_2|\mathcal{A}_\epsilon^{(J)}}{N}|+R_{RB}^G+I(U;X|K)-I(U;V|K))}, \end{aligned} \tag{14}$$

where R_{RB}^G is the random binning-based communication rate in channels with geometrical transformations. If R_{RB}^G is such that:

$$\begin{aligned} R_{RB}^G &\leq I(U;V|K) - I(U;X|K) - 2\epsilon \\ &\quad - \frac{1}{N}\log_2\left|\mathcal{A}_\epsilon^{(J)}\right|, \end{aligned} \tag{15}$$

$P_{e_2}^{RB(N)} \to 0$, as $N \to \infty$ and $\epsilon \to 0$. Furthermore, similar to (9), $\frac{1}{N}\log_2|\mathcal{A}_\epsilon^{(J)}|$ term in (15) vanishes as $N \to \infty$ and $\mathcal{A}_\epsilon^{(J)}$ is finite. Then, the upper bound for R_{RB}^G becomes:

$$R_{RB}^G \leq I(U;V|K) - I(U;X|K) - 2\epsilon, \tag{16}$$

which is equivalent to the condition on R_{RB} given in (13) for channels without geometrical transformations.

Similarly to the Sect. 3.2.1, in order to prove the converse of coding theorem for random binning based communications in channels with geometrical transformations one can show using the same argument as in [13] that the log term in $R_{RB}^G + \frac{1}{N}\log_2|\mathcal{A}_\epsilon^{(J)}|$ has not asymptotic impact on the reliable communications.

Thus, in theoretical set-ups with random binning, the reliable communications do not suffer from geometrical transformations.

3.3 Synchronization Impact on the performance of data-diding

Assume that the probability of decoding error for a particular realization of $\mathbf{A} = \mathbf{a}$ is equal to $P_\mathrm{e}^{(N)}(\mathbf{a})$ for watermark codewords of length N. Then, the average probability of decoding error $P_\mathrm{e}^{G(N)}$ over all possible attacks can be computed by averaging $P_\mathrm{e}^{(N)}(\mathbf{a})$ as:

$$P_\mathrm{e}^{G(N)} = \sum_{\mathbf{a}\in\mathcal{A}_\epsilon^{(J)}} P_\mathrm{e}^{(N)}(\mathbf{a}) p_{\mathbf{A}}(\mathbf{a}). \tag{17}$$

As it was shown in the previous section, in the theoretical set-up, when the length of the communicated sequences asymptotically approaches infinity, performance of random coding-based and random binning-based communications in channels without and with geometrical transformations in terms of achievable rates asymptotically coincides.

However, in practical situations with a finite N, the encoding is based on random binning or random coding with expurgating bad codewords depending on whether the host state is taken into account or not in the encoding. The decoding is based on a ML technique [15] and $P_\mathrm{e}^{G(N)}$ is upper bounded by $P_\mathrm{e}^{G(N)} \leq |\mathcal{A}_\epsilon^{(J)}| 2^{-NE_r(R|K)}$, where $E_r(R|K) = \max_{\rho\in[0,1]} \max_{p_{W|K}(w|k)}[E_0(\rho, p_{W|K}(w|k)) - \rho R]$ and $E_0(\rho, p_{W|K}(w|k)) = -\log_2 \sum_y [\sum_x p_{W|K}(w|k) p(y|w)^{\frac{1}{1+\rho}}]^{1+\rho}$. Furthermore, as $|\mathcal{A}_\epsilon^{(J)}|$ gets larger, the upper bound for $P_\mathrm{e}^{G(N)}$ increases. Hence, in the case of practical set-ups, geometrical transformations completely disable the reliable communications.

In many practical applications, it is necessary to decrease the cardinality of search space of the decoder for possible geometrical transformations to decrease the average probability of decoding error in practical cases. A way to accomplish this requirement is to introduce a synchronization framework into the scheme in the expense of dedicating some portion of the rate R, originally used for the message transmission, to the communication of synchronization data that definitely lead to some loss of maximum achievable rate. In the asymptotic case, when $N \to \infty$, $P_\mathrm{e}^{(N)} \to 0$ and there is no need in such a synchronization. However, in practical set-ups, particular N determines $P_\mathrm{e}^{(N)}$ and the relationship between the cardinality of $|\mathcal{A}_\epsilon^{(J)}|$ and $P_\mathrm{e}^{(N)}$ defines the loss in performance. Depending on this relationship it might be beneficial to reduce the rate of communication spending same part for the synchronization as the gain in considerable decreasing $|\mathcal{A}_\epsilon^{(J)}|$, keeping $P_\mathrm{e}^{(N)}$ in the given range.

As an illustrative example, in Fig. 5a, a dot represents a particular geometrical transformation $\mathbf{A} = \mathbf{a}$ in the space $\mathcal{A}_\epsilon^{(J)}$ of typical geometrical transformations. A decoder without a synchronization framework will consider all elements of this space as possibly applied geometrical transformation, i.e., it will perform decoding at each point of this space. However, the use of a geometrical synchronization framework reduces the search space from $\mathcal{A}_\epsilon^{(J)}$ to $\mathcal{A}'$ (Fig. 5b).

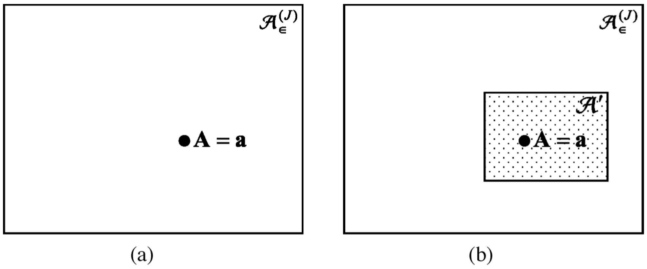


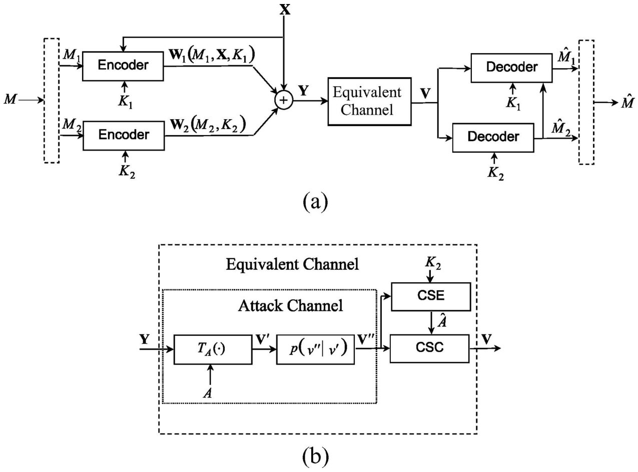
Fig. 5 The original geometrical search space $\mathcal{A}_\epsilon^{(J)}$ (a) and the constrained search space $\mathcal{A}'$ after application of CSE/CSC (b)

The cardinality of $\mathcal{A}'$, i.e., $|\mathcal{A}'|$, is determined by the accuracy of CSE and CSC and depends on a particular design of structured codebook that can be achieved analyzing corresponding Cramer-Rao lower bounds [12]. As the variance of the estimation error goes to zero, constrained search space $\mathcal{A}'$ reduces to $|\mathcal{A}'| = 1$ ($\mathbf{A} = \mathbf{a}$, Fig. 5).

4 Practical framework for information-theoretic consideration of geometrically-robust data-hiding codes

The host interference to the message communication is an essential problem in the design of a practical capacity achieving robust data-hiding. The message encoding based on the random binning dependent on host state provides the solution to this problem. In contrast, robustness to geometrical attacks with an acceptable complexity of protocol design requires the codewords of the synchronization part to have special features that are independent from the statistics of the host data.

In order to resolve these conflicting requirements, we propose the information-theoretic set-up presented in Fig. 6 that is based on a memoryless MAC with side information (SI) about the host state $\mathbf{X}$ non-causally available at one of the encoders. It consists of four alphabets $\mathcal{W}_1, \mathcal{W}_2, \mathcal{X}$ and $\mathcal{V}$, and is denoted by $\{\mathcal{W}_1 \times \mathcal{W}_2, \mathcal{X}, p(v|y), \mathcal{V}\}$. We also assume that the keys K_1 and K_2 are available at corresponding encoders and decoders.



(a)

(b)

Fig. 6 MAC framework to geometrically robust data-hiding: (a) main diagram and (b) equivalent channel

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.

A $(2^{NR_1}, 2^{NR_2}, N)$ code for the MAC with SI consists of two message sets:

$$\mathcal{M}_1 = \{1, 2, \ldots, 2^{NR_1}\}, \tag{18}$$

$$\mathcal{M}_2 = \{1, 2, \ldots, 2^{NR_2}\}, \tag{19}$$

two encoding functions:

$$f_1 : \{1, 2, \ldots, 2^{NR_1}\} \times \{1, 2, \ldots, |\mathcal{K}_1|\} \times \mathcal{X}^N \\ \to \mathcal{W}_1^N, \tag{20}$$

$$f_2 : \{1, 2, \ldots, 2^{NR_2}\} \times \{1, 2, \ldots, |\mathcal{K}_2|\} \to \mathcal{W}_2^N, \tag{21}$$

and a decoding function:

$$g : \mathcal{V}^N \times \{1, 2, \ldots, |\mathcal{K}_1|\} \times \{1, 2, \ldots, |\mathcal{K}_2|\} \\ \to \{1, 2, \ldots, 2^{NR_1}\} \times \{1, 2, \ldots, 2^{NR_2}\}. \tag{22}$$

M_1 and M_2 are chosen randomly from the sets $\{1, 2, \ldots, 2^{NR_1}\}$ and $\{1, 2, \ldots,\ 2^{NR_2}\}$, respectively. The keys K_1 and K_2 determine the particular codebooks that are used by corresponding encoders and decoders. Assuming that joint distribution of messages over the product set $\mathcal{M}_1 \times \mathcal{M}_2$ is uniform, the average probability of error for this code is defined as:

$$P_e^{(N)} = \frac{1}{2^{N(R_1+R_2)}} \\ \times \sum_{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2} \Pr[g(\mathbf{V}, K_1, K_2) \neq (m_1, m_2)| \\ (M_1 = m_1, M_2 = m_2)]. \tag{23}$$

A rate pair (R_1, R_2) is said to be achievable, if there exists a $(2^{NR_1}, 2^{NR_2}, N)$ code with $P_\mathrm{e}^{(N)} \to 0$ as $N \to \infty$. The capacity region of the MAC is the closure of the set of all achievable (R_1, R_2) rate pairs.

Codebook construction: Codebooks for $\mathbf{W}_1$ and $\mathbf{W}_2$ when generated randomly according to random binning [13] and random coding [12] principles, respectively, are revealed to corresponding encoders and decoders.

Encoding: A particular message M is partitioned into (M_1, M_2) depending on the rate pair (R_1, R_2). Then, the encoder for M_1 generates $\mathbf{W}_1(M_1, \mathbf{X}, K_1)$ using random binning by taking into account M_1, non-causal host state information $\mathbf{X}$ and the user-specified key K_1. Similarly, encoder for M_2 produces the codeword $\mathbf{W}_2(M_2, K_2)$ using random coding by considering M_2 and the particular key K_2. Afterwards, these two codewords are combined with the host state $\mathbf{X}$ and sent to the equivalent channel. Information transfer via this channel passes the following stages.

Geometrical Transformation: Attacker applies a geometrical transformation $T_A(.)$ from the set of ϵ-typical geometrical transformations $\mathcal{A}_\epsilon^{(J)}(A)$ to the stego data $\mathbf{Y}$.

Probabilistic Channel: In order to prevent complete inversion of the applied geometrical transformation, the attacker might introduce additional noise to the attacked data $\mathbf{V}'$.[3] Assuming that the noise acts as a discrete memoryless channel (DMC), it converts the input $\mathbf{V}'$ to the output $\mathbf{V}''$ in a probabilistic manner according to the channel transition probability $p(\mathbf{v}''|\mathbf{v}') = \prod_{i=1}^{N} p(v_i''|v_i')$.

Synchronization: The output $\mathbf{V}''$ of the probabilistic channel is provided to CSE and CSC blocks for the synchronization. In fact, this is the part where the cardinality of the search space of the decoder for possibly applied geometrical transformations is reduced from $|\mathcal{A}_\epsilon^{(J)}|$ to $|\mathcal{A}'|$. The output $\mathbf{V}$ of this part is sent to decoders. Geometrical transformation, probabilistic channel and synchronization part form the equivalent channel with the input alphabets $\mathcal{W}_1$, $\mathcal{W}_2$, $\mathcal{X}$ and the output alphabet $\mathcal{V}$. Leaving the problem of intersymbol interference (ISI) outside of the scope of this paper, we assume that the channel output is produced according to the probabilistic mapping $p(\mathbf{v}|\mathbf{y}) = \prod_{i=1}^{N} p(v_i|y_i)$.

Decoding: At the lower decoder (Fig. 6) with the knowledge of the key K_2, $\widehat{M}_2$ is decoded first from $\mathbf{V}$ considering $\mathbf{W}_1$ as interference. Then, the output of this decoder (in assumption of errorless decoding of M_2), $\mathbf{W}_2$, is provided to upper decoder and $\widehat{M}_1$ is decoded from $\mathbf{V}$, with the knowledge of the key K_1, after subtracting $\mathbf{W}_2$ (genie-aided decoding [16]). In this way, the interference of $\mathbf{W}_2$ with respect to $\mathbf{W}_1$ is avoided.

The corresponding achievable rates for the given set-up have been investigated independently for non-watermarking applications in [17]:

$$R_1 \leq \frac{1}{N}\left[I(\mathbf{U};\mathbf{V}|\mathbf{W}_2, K_1) - I(\mathbf{U};\mathbf{X}|K_1)\right], \quad (24)$$

$$R_2 \leq \frac{1}{N}\left[I(\mathbf{W}_2;\mathbf{V}|\mathbf{U}, K_2) - I(\mathbf{U};\mathbf{X}, |K_1)\right], \quad (25)$$

$$R_1 + R_2 \leq \frac{1}{N}[I(\mathbf{U}, \mathbf{W}_2;\mathbf{V}|K_1, K_2) - I(\mathbf{U};\mathbf{X}|K_1)], \quad (26)$$

The achievable rate region for the proposed set-up is presented in Fig. 7. We would like to communicate with the highest possible (R_1, R_2) pair satisfying (25). Actually, all (R_1, R_2) pairs located on the line between the points P_1 and P_2 of the achievable rate region in Fig. 7 agree with (25). However, it is known from [12] that non-corner points can be achieved only by time sharing (or space sharing in our particular application). Therefore, the selection of corners is motivated by the technical design concerns.

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 $N \to \infty$. Therefore, the plot will have notably "asymmetric" character meaning that practically all rate will be assigned to R_1. Therefore, asymptotically the difference between the maximum achievable rate R_1 under $N \to \infty$ and R_1 will be negligibly small for practical usage.

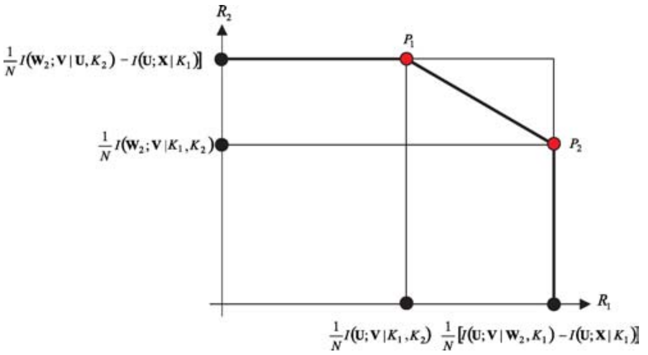


Fig. 7 Achievable rate region of the proposed set-up

According to the particular way of the synchronization part $\mathbf{W}_2$ design, structured codebooks can be divided into two main groups: template-based structured codebooks and redundant-based structured codebooks. In the following sections, the properties of these two groups will be investigated in more details.

5 Template-based structured codebooks

The main idea of template-based synchronization is to use a specially designed pilot to estimate possible geometrical transformations applied to the stego data. Template data itself does not contain any information about the ongoing message transfer, i.e., $R_2 = 0$. It is key-dependent, $\mathbf{W}_2(K_2)$, unique for a given key $K_2 = k_2$ which is shared by encoder and decoder. Once the geometrical transformation is estimated and inverted based on the template $\mathbf{W}_2$, $\mathbf{W}_1$ is decoded from $\mathbf{V}$ after interference of $\mathbf{W}_2$ is canceled by subtraction.

The codebook construction with a template can be considered using code division multiple access (CDMA) and space division multiple access (SDMA) signaling approaches.

In the case of CDMA, $\mathbf{W}_1$ and $\mathbf{W}_2$ are transmitted simultaneously using power sharing since there is a constraint on the power of the total input signal $\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2$ to the channel defined by the distortion $E[d(\mathbf{X},\mathbf{Y})] \leq \sigma_W^2$, where $d(\mathbf{X},\mathbf{Y}) = \frac{1}{N}\sum_{i=1}^{N} d(x_i, y_i)$. It means that, if the total watermark power is σ_W^2, then $\lambda\sigma_W^2$ portion will be assigned to the communication of $\mathbf{W}_1$ and the rest, $(1-\lambda)\sigma_W^2$, will be used to communicate $\mathbf{W}_2$. Obviously, as $N \to \infty$, then $\lambda \to 1$ and negligibly small fraction of total watermark power is spent for the template communication. An example of template-based structured codebook based on CDMA is given in Fig. 8.

In the case of SDMA, transmission of $\mathbf{W}_1$ and $\mathbf{W}_2$ is performed in orthogonal space intervals. Thus, interference

[3] The noise in general can be signal dependent that takes into account interpolation effects.

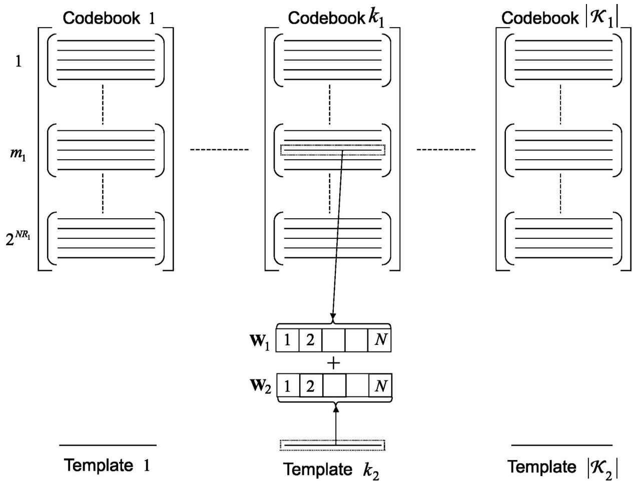


Fig. 8 CDMA template-based structured codebooks

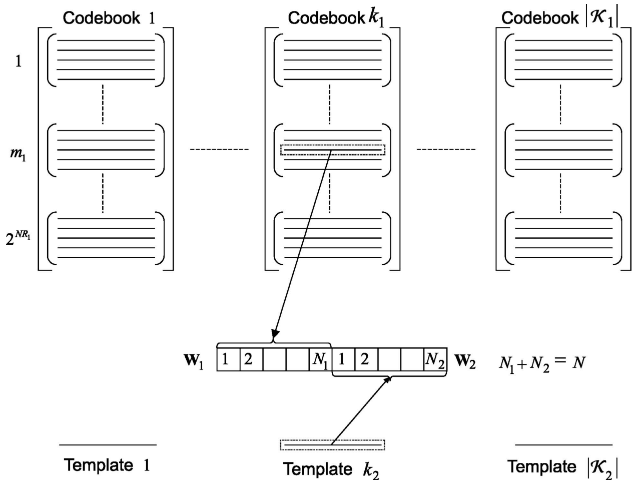


Fig. 9 SDMA template-based structured codebooks

between these two data is avoided. An example of template-based structured codebook based on SDMA is given in Fig. 9.

6 Redundant-based structured codebooks

In redundant-based structured codebooks, $\mathbf{W}_2$ conveys M_2 part of the message M, i.e., $R_2 \neq 0$, using codewords that have a special construction or statistics designed to aid the synchronization. Once the geometrical transformation is inverted using the special structure of $\mathbf{W}_2$ and M_2 is decoded without error from $\mathbf{V}$, then $\mathbf{W}_1$ is decoded after $\mathbf{W}_2(M_2)$ is subtracted from $\mathbf{V}$.

As in the case of template-based structured codebooks, there are CDMA and SDMA approaches for the construction of redundant-based structured codebooks. An example of redundant-based structured codebook using CDMA is given in Fig. 10 and another example based on SDMA is presented in Fig. 11.

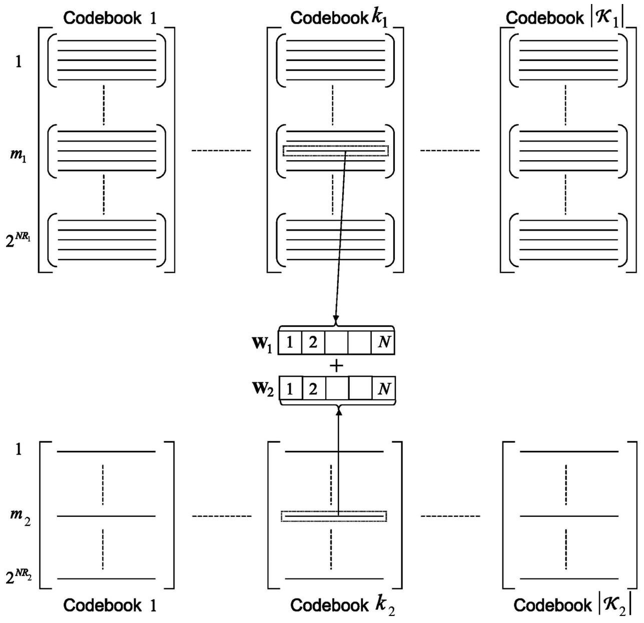


Fig. 10 CDMA redundant-based structured codebooks

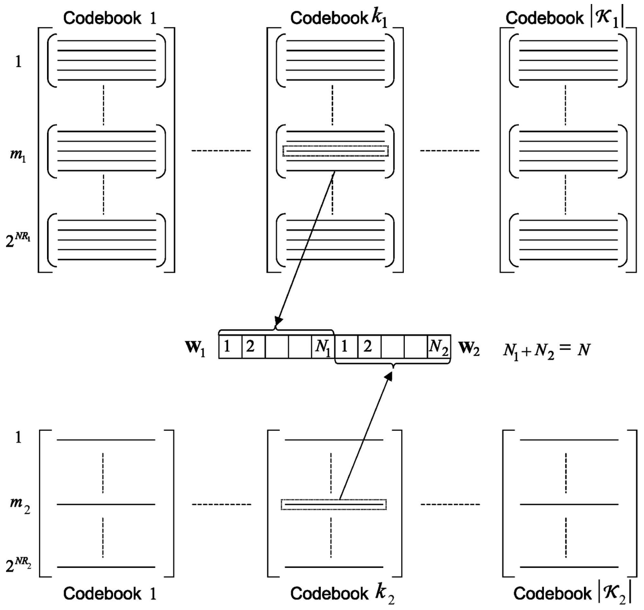


Fig. 11 SDMA redundant-based structured codebooks

7 Analysis of security leaks and attacking strategies

The objective of the 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* [18] in the design of a *secure data-hiding scheme*, 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 find signal processing attacks inefficient since in some cases they are even invertible [19] and may decide to increase the complexity of decoding for the data-hider applying a geometrical attack to the stego data for desynchronization, which is simple in terms of implementation;
- 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 the attacker [20]. 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, attacking scenarios that are inspired by the given strategies for each group of structured codebooks based on the proposed MAC framework will be investigated in details for theoretical set-ups, i.e., for $N \to \infty$.

7.1 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 given in Fig. 8 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 $|\mathcal{K}_2|$, where $|\mathcal{K}_2|$ represents the total number of codebooks for $\mathbf{W}_2$. If $\widehat{\mathbf{W}}_2$ is found, the attacker would subtract it from $\mathbf{Y}$ and apply a geometrical transform $\mathbf{A}$ to $(\mathbf{Y} - \widehat{\mathbf{W}}_2)$. In this case, the data-hider will loose the synchronization framework based on $\mathbf{W}_2$ after this attack.

After $\widehat{\mathbf{W}}_2$ is successfully decoded, instead of applying some geometrical transformation to $(\mathbf{Y} - \widehat{\mathbf{W}}_2)$, the attacker may further develop following attacks based on security leaks, depending on the key management protocol for K_1 and K_2:

- *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*: the knowledge of template $\mathbf{W}_2$ implies the knowledge of corresponding codebook for $\mathbf{W}_1$ in such a design. After revealing $\mathbf{W}_2$ from a decoding space with cardinality $|\mathcal{K}_2|$, the attacker would search in that particular codebook for a $\mathbf{U}$ that is jointly-typical with $(\mathbf{Y} - \widehat{\mathbf{W}}_2)$. The cardinality of this $\mathbf{U}$ decoding space is $2^{N(R_1+R')}$. After finding $\mathbf{U}$, the attacker can also obtain $\mathbf{X}$. For example, in the Costa set-up [21], 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 $(\mathbf{U}, (\mathbf{Y} - \widehat{\mathbf{W}}_2))$ pair is found. The possibility for the attacker to obtain $\mathbf{X}$ means the total failure of the communications. Thus, the cardinality of the decoding space for the attacker is bounded by $|\mathcal{K}_2| + 2^{N(R_1+R')}$.
- *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 $\mathbf{W}_2$* [7]: this time the knowledge of template $\mathbf{W_2}$ does not provide any information about the codebook from which current $\mathbf{W}_1$ in the stego data is coming. The attacker may perform an exhaustive search in all $|\mathcal{K}_1|$ codebooks for the $\mathbf{U}$ that is jointly-typical with $(\mathbf{Y} - \widehat{\mathbf{W}}_2)$. The cardinality of this $\mathbf{U}$ decoding space is $|\mathcal{K}_1|2^{N(R_1+R')}$ that results in total cardinality of $|\mathcal{K}_2| + |\mathcal{K}_1|2^{N(R_1+R')}$.
- *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* [6]: using the same template for all users, i.e., $|\mathcal{K}_2| = 1$, makes the scheme more susceptible to the attacks. Thus, when compared to the previous cases, it is relatively easy for the attacker to find $\mathbf{W}_2$. However, in terms of destroying the reliable communications, the same exhaustive search should be performed in all $|\mathcal{K}_1|$ codebooks for a jointly-typical $(\mathbf{U}, (\mathbf{Y} - \widehat{\mathbf{W}}_2))$ pair with a decoding space of the cardinality $|\mathcal{K}_1|2^{N(R_1+R')}$. Thus, the overall cardinality of the decoding space for the attacker is $1 + |\mathcal{K}_1|2^{N(R_1+R')}$.

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

7.2 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 [8–10]. 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 given in Fig. 10, the upper bound for the cardinality of the $\mathbf{W}_2$ decoding space is $|\mathcal{K}_2|2^{NR_2}$.

An additional analysis should be provided for the probability that the attacker will find another $\widehat{\mathbf{W}}_2$ from the codebooks such that $(\widehat{\mathbf{W}}_2 \neq \mathbf{w}_2|\mathbf{W}_2 = \mathbf{w}_2)$ is jointly-typical with $\mathbf{Y}$. The attacker, who has the access to the stego image $\mathbf{Y}$, but not to the K_2, has to look through all codebooks $|\mathcal{K}_2|$ for a jointly-typical pair $(\widehat{\mathbf{W}}_2, \mathbf{Y})$. The probability of decoding error for the attacker, $P_e^{A(N)}$, will be upper bounded by:

$$\begin{aligned} P_e^{A(N)} &\leq |\mathcal{K}_2|2^{NR_2}2^{-N(I(W_2;Y)-\epsilon)}, \\ &\leq 2^{N\frac{1}{N}\log_2|\mathcal{K}_2|}2^{N(R_2-I(W_2;Y)+\epsilon)}, \\ &\leq 2^{N(\frac{1}{N}\log_2|\mathcal{K}_2|+R_2-I(W_2;Y)+\epsilon)}. \end{aligned} \quad (27)$$

Therefore, if $R_2 \leq I(W_2;Y) - \epsilon - \frac{1}{N}\log_2|\mathcal{K}_2|$, $P_e^{A(N)} \to 0$ as $N \to \infty$.

Moreover, the data-hider having attacked data $\mathbf{V}$ looks for the jointly-typical $(\widehat{\mathbf{W}}_2, \mathbf{V})$ pair in a particular codebook determined by K_2 and the probability of decoding error for the data-hider, $P_e^{DH(N)}$, is bounded by (5). Thus, if $R_2 \leq I(W_2;V|K_2) - \epsilon$, $P_e^{DH(N)} \to 0$ as $N \to \infty$.

In order to prove that the attacker can also decode $\mathbf{W}_2$ reliably from $\mathbf{Y}$ without knowledge of K_2 when the data-hider can do the same from $\mathbf{V}$ with knowledge of K_2, we will try to investigate the relationship between $I(W_2;Y)$ and $I(W_2;V|K_2)$ that bounds R_2 for the attacker and for the data-hider, respectively, in the case of reliable communications. From data processing inequality [12]:

$$I(W_2;V) \leq I(W_2;Y). \quad (28)$$

Furthermore, based on the extension of $I(W_2;V,K_2)$ using chain rule for mutual information:

$$\begin{aligned} I(W_2;V,K_2) &= I(W_2;V) + I(W_2;K_2|V), \\ &= I(W_2;K_2) + I(W_2;V|K_2). \end{aligned} \quad (29)$$

and the fact that $I(W_2;K_2) = I(W_2;K_2|V) = 0$, we can prove that $I(W_2;V) = I(W_2;V|K_2)$. Therefore, $I(W_2;V|K_2) \leq I(W_2;Y)$ according to (27) and (28). If the data-hider selects R_2, which satisfies the condition $R_2 \leq I(W_2;V|K_2) - \epsilon$ to be able to decode $\widehat{\mathbf{W}}_\mathbf{2}$ reliably, then the attacker will also be able to decode $\widehat{\mathbf{W}}_\mathbf{2}$ with $P_{\mathrm{e}}^{A(N)} \to 0$ as $N \to \infty$ with such R_2.

Once the attacker obtains $\mathbf{W}_2$, it is subtracted from $\mathbf{Y}$ and a geometrical transformation is applied to $(\mathbf{Y} - \widehat{\mathbf{W}}_2)$. This causes the data-hider to loose the synchronization framework.

However, instead of applying a geometrical transformation to $(\mathbf{Y} - \widehat{\mathbf{W}}_2)$, the attacker may develop the following attacks in order to destroy the reliable communications, depending on the statistical codebook design strategy for $\mathbf{W}_2$:

- *The statistics of* $\mathbf{W}_2$ *are the same for all codebooks*[8, 9]: in this case the knowledge of $\mathbf{W}_2$ does not have any significance in reaching $\mathbf{W}_1$. Thus, the attacker has to perform an exhaustive search through all $|\mathcal{K}_1|$ codebooks for the jointly-typical $(\mathbf{U}, (\mathbf{Y} - \widehat{\mathbf{W}}_2))$ pair. The cardinality of this $\mathbf{U}$ decoding space is $|\mathcal{K}_1|2^{N(R_1+R')}$. If the jointly-typical pair is found, it is possible to obtain the realization of $\mathbf{X}$. The total cardinality of the decoding space for the attacker is bounded by $|\mathcal{K}_2|2^{NR_2} + |\mathcal{K}_1|2^{N(R_1+R')}$.
- *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* $\mathbf{W}_2$: in such a codebook design scenario, the knowledge of $\mathbf{W}_2$ restricts the search of the attacker for $\mathbf{W}_1$ in a particular codebook. Thus, the cardinality of the decoding space for the jointly-typical $(\mathbf{U}, (\mathbf{Y} - \widehat{\mathbf{W}}_2))$ pair reduces from $|\mathcal{K}_1|2^{N(R_1+R')}$ to $2^{N(R_1+R')}$ when compared to the previous scenario. Thus, the total cardinality of the decoding space for the attacker is reduced to $|\mathcal{K}_2|2^{NR_2} + 2^{N(R_1+R')}$.
- *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* $\mathbf{W}_2$: in this case, the knowledge of $\mathbf{W}_2$ does not facilitate the search of the attacker for $\mathbf{W}_1$. Therefore, the cardinality of the decoding space for the jointly-typical $(\mathbf{U}, (\mathbf{Y} - \widehat{\mathbf{W}}_2))$ pair will be $|\mathcal{K}_1|2^{N(R_1+R')}$. Thus, the total cardinality of the decoding space for the attacker is $|\mathcal{K}_2|2^{NR_2} + |\mathcal{K}_1|2^{N(R_1+R')}$.

7.3 The effect of security leakages on the cardinality of the decoding space

In the random coding scenario, where the decoder looks for a jointly typical $(\mathbf{W}(M,K), \mathbf{Y})$ pair, the attacker, who has the knowledge of the decoding rule (or decoder) and targets destroying reliable communications, has to find $\mathbf{W}(M,K)$. Once $\mathbf{W}(M,K)$ is found, it can be subtracted from $\mathbf{Y}$. Thus, without knowledge of the key K, one will perform an exhaustive search through all codebooks $\{1, 2, \ldots, |\mathcal{K}|\}$ and all messages $M = m$, $m \in \mathcal{M}$ for the jointly typical $(\mathbf{W}(M,K), \mathbf{Y})$ pair. The cardinality of this decoding space will be $|\mathcal{K}|2^{NR}$, where $|\mathcal{K}|$ is the total number of codebooks and 2^{NR} is the number of codewords per codebook.[4]

When the codebooks $\{1, 2, \ldots, |\mathcal{K}|\}$ are generated in the way that each one contains unique codewords and every possible $\mathbf{W}$ is included in only one codebook, the exhaustive search for $\mathbf{W}$ is related to the ambiguity $H(\mathbf{W})$ by $2^{H(\mathbf{W})}$. In this limit case, the cardinalities $|\mathcal{K}|2^{NR}$ and $2^{H(\mathbf{W})}$ will be equal.

However, as it was proposed by Shannon [20], observing $\mathbf{Y}$[5] reduces the ambiguity about $\mathbf{W}$ from $H(\mathbf{W})$ to $H(\mathbf{W}|\mathbf{Y})$

[4] We do not consider here efficient search strategies similar to Viterbi algorithm space [22].

[5] It should be noticed that the attacker operates directly on the stego data $\mathbf{Y}$ contrarily to the data-hider who has access only to the attacked data $\mathbf{V}$.

as:

$$H(\mathbf{W}|\mathbf{Y}) = H(\mathbf{W}) - I(\mathbf{W};\mathbf{Y}), \tag{30}$$

where $I(\mathbf{W};\mathbf{Y})$ is the amount of information that can be learned about $\mathbf{W}$ by observing $\mathbf{Y}$. If $H(\mathbf{W}|\mathbf{Y}) = 0$, the knowledge of the current $\mathbf{Y}$ gives the exact value for $\mathbf{W}$, i.e., the cardinality of the decoding space for the attacker is $2^{H(\mathbf{W}|\mathbf{Y})} = 1$. Therefore, in a communication scenario with $I(\mathbf{W};\mathbf{Y}) \neq 0$, it is possible for the attacker to reduce the cardinality $|\mathcal{K}|2^{NR}$ of the exhaustive search for the jointly typical $(\mathbf{W},\mathbf{Y})$ pair.

In the case of random binning, where the decoder looks for a $\mathbf{U}(M,\mathbf{X},K)$ that is jointly typical with the stego data $\mathbf{Y}$, the attacker will try to find the jointly typical $(\mathbf{U}(M,\mathbf{X},K),\mathbf{Y})$ pair through all $\{1,2,\ldots,|\mathcal{K}|\}$ codebooks and all message bins $M = m, m \in \mathcal{M}$ to be able to destroy reliable communications. The cardinality of this decoding space is given by $|\mathcal{K}|2^{N(R+R')}$ where $2^{NR'}$ is the total number of sequences $\mathbf{U}$ in each message bin $M = m$. The knowledge of $\mathbf{U}$ enables the attacker to get the host state $\mathbf{X}$, the message M and the key K.

Similarly to the random coding scenario, when the codebooks are generated by distributing all possible $\mathbf{U}$ sequences to the codebooks uniquely, the cardinality of the decoding space depends on the ambiguity $2^{H(\mathbf{U})}$ about $\mathbf{U}$. Therefore, one would expect the cardinalities $|\mathcal{K}|2^{N(R+R')}$ 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:

$$\begin{aligned} H(\mathbf{U}|\mathbf{Y}) &= H(\mathbf{U}|\mathbf{X}) - [I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{X})] \\ &= H(\mathbf{U}) - I(\mathbf{U};\mathbf{Y}). \end{aligned} \tag{31}$$

Thus, as in the random coding case, if $I(\mathbf{U};\mathbf{Y}) \neq 0$, then the cardinality of the decoding space for the attacker can be decreased from $|\mathcal{K}|2^{N(R+R')}$ based on the observed $\mathbf{Y}$.

8 Conclusions

In this paper, the conditions of reliable communications based on structured codebooks in channels with geometrical transformations are analyzed from an information-theoretic point of view. Structured codebooks include codewords that have some features or statistics designed for synchronization purposes.

The MAC framework is developed to design the capacity achieving data-hiding codes that are robust to geometrical transformations. The corresponding methods based on the CSE/CSC that are used for reliable communications in channels with geometrical transformations are classified into two main groups depending on the particular codebook design: template-based codebooks and redundant codebooks. The analysis of security leaks of each codebook structure is performed in terms of cardinality of the decoding space for the attacker to design the worst case attack.

As a continuation of our research, we will consider collusion attacks, when there are several stego data copies produced from different hosts, keys or messages, and will emphasize the role of the host data statistics on the security. We will also extend the proposed set-up to real scenarios, when the data lengths N are finite, the decoding is performed using the MMI technique and the probability of error is bounded in terms of error exponents. The particular search algorithms reducing the cardinality of the decoding space for the attacker based on the security leakages $I(\mathbf{W};\mathbf{Y})$ and $I(\mathbf{U};\mathbf{Y})$ are also a subject of our ongoing study.

Acknowledgements 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. They acknowledge interesting discussions during several ECRYPT project meetings with Prof. F. Pérez-González and Prof. M. Barni. They are also thankful to Prof. P. Moulin for the insightful discussion on the subject of coding in arbitrary varying channels.

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. O'Sullivan, J.A., Moulin, P., Ettinger, J.M.: Information-theoretic analysis of steganography. In: Proc. IEEE Symp. on Information Theory. Boston, MA (1998)
2. Moulin, P., O'Sullivan, J.A.: Information-theoretic analysis of information hiding. IEEE Trans. Inf. Theory **49**(3), 563–593 (2003)
3. Somekh-Baruch, A., Merhav. N.: On the error exponent and capacity games of private watermarking systems. EEE Trans. Inf. Theory **49**(3), 537–562 (2003)
4. Somekh-Baruch, A., Merhav. N.: On the capacity game of public watermarking systems. EEE Trans. Inf. Theory **20**(3), 511–524 (2004)
5. Gallager, R.G.: A simple derivation of the coding theorem and some applications. IEEE Trans. Inf. Theory **11**, 3–17 (1965)
6. Rhoads, G.B.: Steganography systems. International Patent WO 96/36163 PCT/US96/06618 (1996)
7. Pereira, S., Pun, T.: Fast robust template matching for affine resistant image watermarking. In: International Workshop on Information Hiding, volume LNCS 1768 of Lecture Notes in Computer Science, pp. 200–210. Dresden, Germany, 29 September–1 October (1999). Springer Verlag
8. Kutter, M., Petitcolas, F.A.P.: A fair benchmark for image watermarking systems. In: IS&T/SPIE's 11th Annual Symposium, Electronic Imaging 1999: Security and Watermarking of Multimedia Content I, vol. 3657, pp. 219–239. San Jose, CA, USA (1999)
9. Voloshynovskiy, S., Deguillaume, F., Pun, T.: Content adaptive watermarking based on a stochastic multiresolution image modeling. In: 10th European Signal Processing Conference EUSIPCO2000. Tampere, Finland (2000)
10. Voloshynovskiy, S., Deguillaume, F., Pun, T.: Multibit digital watermarking robust against local nonlinear geometrical distortions. In: IEEE Int. Conf. On Image Processing ICIP2001, pp. 999–1002. Thessaloniki, Greece (2001)
11. Deguillaume, F., Voloshynovskiy, S., Pun, T.: 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, vol. 4675, pp. 313–322. San-Jose, CA, USA, January 20–25 (2002)

12. Cover, T., Thomas, J.: Elements of Information Theory. Wiley and Sons, New York (1991)
13. Gel'fand, S.I., Pinsker, M.S.: Coding for channel with random parameters. Probl. Control and Inf. Theory **9**(1), 19–31 (1980)
14. Petitcolas, F.A.P.: Stirmark benchmark 4.0. 2002. http://www.cl.cam.ac.uk/ fapp2/watermarking/stirmark/
15. Gallager, R.G.: Information Theory and Reliable Communication. Wiley, New York (1968).
16. Rimoldi, B.: Time-splitting multiple-access. Technical report, Mobile Communications Lab, EPFL, Switzerland (1999)
17. Haroutunian, M.E.: Bounds on e-capacity for multiple access channel with random parameters. Transactions of Institute for Informatics and Automation Problems of NAS RA, Math. Prob. Comput. Sci. **24** (2005)
18. Kerckhoff, A.: La cryptographie militaire. J. Sci. Militaires **9**, 5–38 (1883)
19. Mihcak, M.K., Venkatesan, R., Kesal, M.: Cryptanalysis of discrete-sequence spread spectrum watermarks. In Proceedings of the 5th International Information Hiding Workshop (IH 2002). Noordwijkerhout, The Netherlands (2002)
20. Shannon, C.E.: Communication theory of secrecy systems. Bell Syst. Tech. J. **28**, 656–715 (1949)
21. Costa, M.: Writing on dirty paper. IEEE Trans. Inf. Theory **29**(3), 439–441 (1983)
22. Proakis, J.G.: Digital Communications. McGraw-Hill (1995)