On Reversibility of Random Binning Techniques: Multimedia Perspectives

Sviatoslav Voloshynovskiy†, Oleksiy Koval†, Emre Topak†,
José Emilio Vila-Forcén†, Pedro Comesaña Alfaro‡, and Thierry Pun†

† - CUI-University of Geneva, Stochastic Image Processing Group,
24, rue du Général-Dufour, 1211 Genève 4, Switzerland
‡ - Signal Processing in Communications Group, Signal Theory & Communications
Department, University of Vigo, 36200 Vigo, Spain
† - {svolos, Oleksiy.Koval, Emre.Topak, Jose.Vila,
Thierry.Pun}@cui.unige.ch,
‡ - pcomesan@gts.tsc.uvigo.es

Abstract. In this paper, we analyze a possibility of reversibility of data-hiding techniques based on random binning from multimedia perspectives. We demonstrate the capabilities of unauthorized users to perform hidden data removal using solely a signal processing approach based on optimal estimation as well as consider reversibility on the side of authorized users who have the knowledge of key used for the message hiding.

1 Introduction

Digital data-hiding appeared as an emerging tool for multimedia security, processing and management. A tremendous amount of possible applications have been recently reported that include copyright protection, tamper proofing, content integrity verification, steganography and watermark-assisted media processing such as multimedia indexing, retrieval and quality enhancement [1].

Most of these applications are facing an important problem of host interference. The related issue in communications under the assumption of a fixed channel was considered by Gel'fand and Pinsker [2]. Costa considered the Gel'fand-Pinsker problem in a Gaussian formulation and mean squared distortion criteria and demonstrated that the capacity of the Gaussian channel with the Gaussian interfering host can be equal to the capacity of interference-free communications using *random binning*-based codebook design [3]. Recent advantages in the design of practical capacity achieving codes makes this technique even more attractive for various purposes [4].

The wide practical use of the Gel'fand-Pinsker set-up has raised a number of problems related to its performance and reversibility in various multimedia applications. Although these aspects seem to be unrelated from the first point of view, there exist a lot in common among these issues that can throw more light on the optimal design of binning-based techniques.

Therefore, the goal of this paper is to reveal these relationships on the side of data-hider in multimedia applications. Similar framework for the case of descrete alphabets was considered by Eggers *et al.* [5].

The paper has the following structure. The basic information-theoretic set-up of side information-assisted data-hiding is considered in Section 2. Section 3 presents the analysis of reversibility problem from multimedia perspectives for both unauthorized and authorized users. The experimental results demonstrating the validity of presented theoretical analysis are given in Section 4. Finally, Section 5 concludes the paper and presents some future research perspectives.

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 denote the realizations of scalar and vector random variables, respectively. The superscript N is used to designate length-N vectors $\mathbf{x} = x^N = [x[1], x[2], ..., x[N]]^T$ with k^{th} element $x[k]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. The mathematical expectation of a random variable $X \sim p_X(x)$ is denoted by $E_{p_X}[X]$ or simply by $E[X]$ and $Var[X]$ denotes the variance of X. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. We also define the watermark-to-image ratio (WIR) as WIR$=10\log_{10}\frac{\sigma_W^2}{\sigma_X^2}$ and the watermark-to-noise ratio (WNR) as WNR $= 10\log_{10}\frac{\sigma_W^2}{\sigma_Z^2}$, where σ_X^2, σ_W^2, σ_Z^2 represent the variances of host data, watermark and noise, respectively.

2 Gel'fand-Pinsker Set-up: Random Binning in Data-Hiding

In this section we consider the Gel'fand-Pinsker problem in data-hiding formulation. The generalized block-diagram of this set-up is shown in Figure 1.

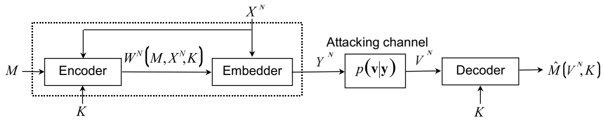


Fig. 1. Generalized Gel'fand-Pinsker channel coding with side information at the encoder: data-hiding formulation.

In this scenario, the data-hider has access to the uniquely assigned secret key $K = k$, uniformly distributed over the set $\mathcal{K} = \{1, 2, ..., |\mathcal{K}|\}$ of cardinality $|\mathcal{K}|$, and to the non-causal interference $\mathbf{x} \in \mathcal{X}^N$. A message $m \in \mathcal{M}$ is uniformly distributed over $\mathcal{M} = \{1, 2, ..., |\mathcal{M}|\}$, with $|\mathcal{M}| = 2^{NR}$, where R is the data-hiding rate. It is assumed that the stego and attacked data are defined on $\mathbf{y} \in \mathcal{Y}^N$ and $\mathbf{v} \in \mathcal{V}^N$, respectively. The length N vector distortion function is defined as:

$$d^N(\mathbf{x}, \mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N} d(x_i, y_i), \qquad (1)$$

where $d(x_i, y_i)$ denotes element-wise distortion between x_i and y_i.

Definition 1: A *discrete memoryless data-hiding channel* consists of five alphabets $\mathcal{X}$, $\mathcal{K}$, $\mathcal{W}$, $\mathcal{Y}$, $\mathcal{V}$ and a transition probability matrix $p_{V|W,X}(v|w,x) = p_{Y|W,X}(y|w,x)p_{V|Y}(v|y)$. The attack channel is subject to the distortion constraint D^A:

$$\sum_{\mathbf{y}\in\mathcal{Y}^N}\sum_{\mathbf{v}\in\mathcal{V}^N} d^N(\mathbf{y},\mathbf{v})p_{\mathbf{V}|\mathbf{Y}}(\mathbf{v}|\mathbf{y})p_{\mathbf{Y}}(\mathbf{y}) \leq D^A, \tag{2}$$

where $p_{\mathbf{V}|\mathbf{Y}}(\mathbf{v}|\mathbf{y}) = \prod_{i=1}^{N} p_{V|Y}p(v_i|y_i)$.

Definition 2: A $(2^{NR}, N)$ code for data-hiding channel consists of a *message set* $\mathcal{M} = \{1, 2, ..., 2^{NR}\}$, an *encoding function*: $f^N : \mathcal{M} \times \mathcal{X}^N \times \mathcal{K} \to \mathcal{W}^N$, an *embedding function*: $\varphi^N : \mathcal{W}^N \times \mathcal{X}^N \to \mathcal{Y}^N$, subject to the embedding distortion constraint D^E: $\frac{1}{|\mathcal{K}||\mathcal{M}|}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\sum_{\mathbf{x}\in\mathcal{X}^N} d^N(\mathbf{x}, \varphi^N(f^N(m,\mathbf{x},k),\mathbf{x}))p_{\mathbf{X}}(\mathbf{x}) \leq D^E$ and a *decoding function*: $g^N : \mathcal{V}^N \times \mathcal{K} \to \mathcal{M}$.

We define the *average probability of error* for a $(2^{NR}, N)$ code as:

$$P_e^{(N)} = \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} Pr[g^N(\mathbf{V}, K) \neq m|M = m]. \tag{3}$$

Definition 3: A rate $R = \frac{1}{N}\log_2|\mathcal{M}|$ is achievable for the distortions (D^E, D^A), if there exists a sequence $(2^{NR}, N)$ codes with $P_e^{(N)} \to 0$ as $N \to \infty$.

Definition 4: The capacity of the data-hiding channel is the supremum of all achievable rates.

Theorem 1 (Data-hiding capacity for the fixed channel): A rate R is achievable for the distortion D^E and the attack channel $p(v|y)$, with the bounded distortion D^A, iff $R < C$, where:

$$C = \max_{p(u,w|x,k)}\left[I(U;V|K) - I(U;X|K)\right], \tag{4}$$

and U to be a random variable $u \in \mathcal{U}$, with $|\mathcal{U}| \leq \min\{|\mathcal{W}|, |\mathcal{Y}|\} + |\mathcal{X}| - 1$. We also assume that $p(u,w|x,k) = p(u|x,k)p(w|u,x,k)$.

The details of this theorem proof in more general form of active attacker are provided in [6]. The main difference with our set-up is the codebook construction and the corresponding interpretation of the user key. In the scope of this paper, the key K is considered uniquely as the index that defines the codebook of a particular user. Contrarily, in [6] the key represents a side information shared between the encoder and the decoder and can be in some relationship with X. Therefore, we assume that K is solely an independent of X cryptographic key.

2.1 Costa set-up: Gaussian assumption

Costa considered the Gel'fand-Pinsker problem for the Gaussian context and mean-square error distance [3]. The corresponding fixed channel $p_{V|W,X}(v|w,x)$ is the Gaussian one with $X \sim \mathcal{N}(0, \sigma_X^2)$ and additive $Z \sim \mathcal{N}(0, \sigma_Z^2)$ (Figure 2). The auxiliary random variable was chosen in the form $U = W + \alpha X$ with optimization parameter α to maximize the rate:

$$R(\alpha) = \frac{1}{2}\log_2\frac{\sigma_W^2(\sigma_W^2+\sigma_X^2+\sigma_Z^2)}{\sigma_W^2\sigma_X^2(1-\alpha)^2+\sigma_Z^2(\sigma_W^2+\alpha^2\sigma_X^2)}. \tag{5}$$

Costa has shown that if $\alpha = \alpha_{opt} = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_Z^2}$ that requires the knowledge of σ_Z^2 at the encoder, $R(\alpha_{opt})$ does not depend on the host variance and:

$$R(\alpha_{opt}) = C^{\text{AWGN}} = \frac{1}{2}\log_2\left(1 + \frac{\sigma_W^2}{\sigma_Z^2}\right) \tag{6}$$

that corresponds to the capacity of AWGN channel without host interference.

It is important to note that the number of codewords in each message bin of the Gel'fand-Pinsker set-up is approximately equal to $2^{NI(U;X|K)}$. In the Costa set-up, $I(U;X|K) = \frac{1}{2}\log_2\left(1 + \alpha^2\frac{\sigma_X^2}{\sigma_W^2}\right)$. Thus, the larger variance of the host σ_X^2, the larger number of codewords is needed at the encoder at each bin.

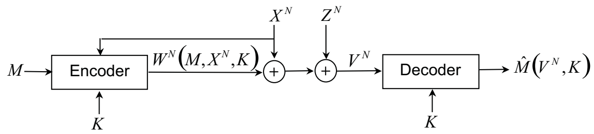


Fig. 2. Costa channel coding with the host state information at the encoder.

3 Reversibility of Random Binning

3.1 Unauthorized user reversibility

In multimedia applications, the unauthorized users are considered not to have access to the secret key used for the data-hiding. Nevertheless, these users might be motivated in certain circumstances [7] to estimate the original image $\mathbf{X}$ given noisy version of stego data $\mathbf{V}$ (Figure 3).

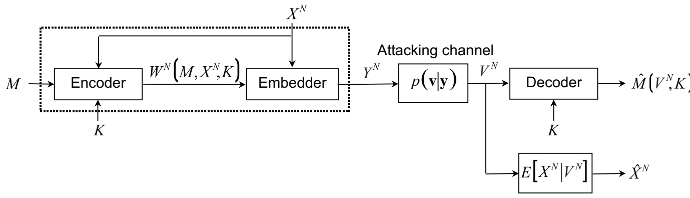


Fig. 3. Reversibility set-up for the unauthorized user.

Assume that in this set-up $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2\mathbf{I}_N)$, $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma_W^2\mathbf{I}_N)$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2\mathbf{I}_N)$. In this case, the embedding distortion is $D^E = \sigma_W^2$ and the attacker distortion corresponds to the variance of AWGN, i.e., $D^A = \sigma_Z^2$.

To estimate $\mathbf{X}$, one can use either minimum mean squared error (MMSE) or maximum a posteriori probability (MAP) estimators that coincide in the case of Gaussian set-up. The MMSE estimate of the unauthorized user is obtained as:

$$\hat{\mathbf{X}} = E[\mathbf{X}|\mathbf{V}]. \tag{7}$$

Assuming $\mathbf{v} = \mathbf{x} + \mathbf{w} + \mathbf{z}$, one obtains the following MMSE estimate [8]:

$$\hat{\mathbf{X}} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2 + \sigma_Z^2}\mathbf{V}. \tag{8}$$

The variance of this estimate D^r_{MMSE} is given by:

$$D^r_{\mathrm{MMSE}} = E[d^N(\hat{\mathbf{X}}, \mathbf{X})] = \frac{\sigma_X^2(\sigma_W^2+\sigma_Z^2)}{\sigma_X^2+\sigma_W^2+\sigma_Z^2}. \quad (9)$$

It is important to note that $\widehat{\mathbf{X}}$ depends on the variances of original image, watermark and noise. In the asymptotic case of infinitely large image variance ($\sigma_X^2 \to \infty$) that corresponds to the highly textured regions in images or edges, no reliable estimate is possible. Moreover, perfect host restoration is not possible in this set-up even in the noiseless case ($\sigma_Z^2 = 0$) due to the watermark presence that reflects the price of lack of information for the unauthorized users.

3.2 Authorized user reversibility

In the case of authorized user, the secret key used for data-hiding at the encoder is available at the decoder side. The knowledge of the key considerably extends the possibilities of image restoration on the decoder side in comparison with the unauthorized user. The block-diagram of this set-up is shown in Figure 4.

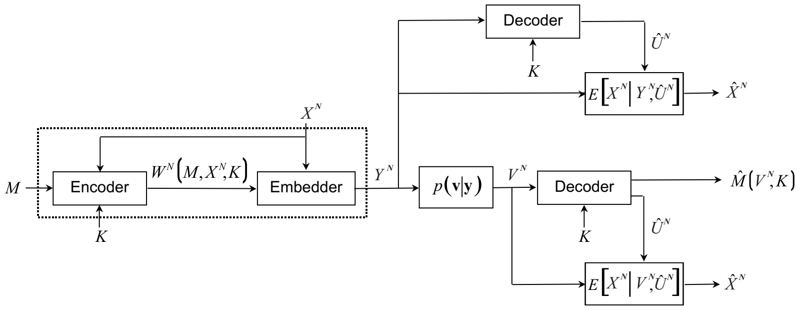


Fig. 4. Reversibility set-up for the authorized user.

Several scenarios are possible. *Scenario (A) (noisy case)*. This scenario refers to the situation when the data-hider designs the scheme for a particular fixed channel $p(v|y)$, certain achievable rate and corresponding codebook construction. The decoder should properly estimate the sent message based on $\mathbf{V}$ and K. At the same time, the authorized user is interested to estimate the host based on the available possibly distorted data $\mathbf{V}$ using the mapping $\psi^N : \mathcal{V}^N \times \mathcal{K} \to \hat{\mathcal{X}}^N$. The criterion that judges the performance of above mapping is defined based on the mean-squared estimation error similarly to the previous case, i.e.:

$$D^r = E[d^N(\hat{\mathbf{X}}, \mathbf{X})]. \quad (10)$$

Therefore, the problem is to design the estimator ψ that produces the minimum mean-squared estimation error.

Scenario (B) (noiseless case). The second scenario of interest (upper part of Figure 4) refers to the situation when the data-hider designs the codebook for a particular fixed channel $p(v|y)$, performs data-hiding procedure and stores the data in the form of $\mathbf{Y}$ for himself and at the same time makes it available via the channel $p(v|y)$. After certain time, the data-hider finds it necessary to recover the original host data due to some reasons caused by the loss of original data, its unavailability due to the time or access restrictions. In these circumstances, the authorized user knows the key and has the undistorted watermarked data $\mathbf{Y}$. The problem now is formulated as the design of a proper mapper ψ that can produce MMSE estimation of $\mathbf{X}$ based on $\mathbf{Y}$. Moreover, it is of particular interest to establish a possibility to perfectly restore the original data $\mathbf{X}$, i.e., to achieve restoration distortion equal to zero.

We split our analysis in two parts. Firstly, we consider the reversibility of Gel'fand-Pinsker problem for the authorized user. Secondly, we analyze the Costa set-up to have a fair comparison with the previously considered scenario of unauthorized user reversibility.

The problem formulation that will be a common basis for the set-ups below can be given as follows. In the case of authorized user it is supposed that the distorted version of the watermarked data $\mathbf{V}$ and the key K are available. The problem is to design the estimate $\hat{\mathbf{X}}$ of the original data $\mathbf{X}$ based on $\mathbf{V}$ using all information about the data-hiding scheme design and corresponding codebook of the user defined by the key K. The quality of the obtained estimate should be validated by the restoration distortion D^r.

Reversibility of the Gel'fand-Pinsker set-up. In the analysis of Gel'fand-Pinsker set-up, we assume that the conditions of reliable message communications provided by the Theorem 1 are satisfied and $\hat{m} = m$ with $P_e^{(N)} \to 0$ as $N \to \infty$. This implies that given the distorted data v^N and the key k, the decoder can uniquely find a jointly typical pair $(u^N(m, j, k), v^N) \in A_\delta^{*(N)}(U, V)$[1], where $j \in \{1, 2, ..., J\}$, $J = 2^{NR'}$, $R' = I(U; X)$ is the number of bits that are used to represent the host, and it can declare that $\hat{m} = m$ and $\hat{u}^N = u^N$, where $\hat{u}^N$ is the estimate of u^N.

In the noisy case (scenario A) (Figure 4) one can design a proper estimator of $\hat{x}^N$ based on v^N for the fixed channel $p(v|y)$ and errorless knowledge of u^N. The decoder forms the MMSE estimate $\hat{x}^N$ given v^N and u^N:

$$\hat{\mathbf{X}} = E[\mathbf{X}|\mathbf{V}, \mathbf{U}(\mathbf{W}(M, \mathbf{X}, K), \mathbf{X})], \tag{11}$$

where we emphasize that u^N is a function of the known message m, key k and the host realization x^N itself.

In the noiseless case (scenario B), $v^N = y^N$ and $y^N = \varphi^N(x^N, w^N)$. Since $w^N = f^N(m, x^N, k)$ and assuming that $\hat{m} = m$ is correctly decoded that is obviously a case for the noiseless transmission and k is known, one can substitute w^N into y^N obtaining $y^N = \varphi^N(x^N, f^N(m, x^N, k))$ and find $\hat{x}^N$ assuming invertibility of functions $\varphi^N(.)$ and $f^N(.)$. In this case, $\hat{x}^N = x^N$ and the authorized user can obtain the perfect estimate of the original data at the decoder.

[1] Here and in the following we assume that the set $A_\delta^{*(N)}(U, X)$ is defined for a particular realization of the key $K = k$. Typical and jointly typical sets are defined in the strong sense, see [9], pp. 288 and 434.

Reversibility of the Costa set-up. To practically validate the above framework, we consider reversibility of the Costa set-up assuming $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathbf{I}_N)$, $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma_W^2 \mathbf{I}_N)$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2 \mathbf{I}_N)$. The distorted version of the watermarked data $\mathbf{v} = \mathbf{x} + \mathbf{w} + \mathbf{z}$ is available at the decoder as well as the authorized user key k. This makes possible to find $\hat{\mathbf{u}}$ based on the jointly typical decoding in the k-specified codebook. Moreover, we assume that $\hat{\mathbf{u}} = \mathbf{u}$ meaning that the sent codeword can be recovered at the decoder. From the Costa assumption about the auxiliary random variable, one can express the watermark as:

$$\mathbf{W} = \mathbf{U} - \alpha \mathbf{X}. \tag{12}$$

Substituting $\mathbf{W}$ into $\mathbf{V}$, one obtains:

$$\mathbf{V} = (1-\alpha)\mathbf{X} + \mathbf{U} + \mathbf{Z}, \tag{13}$$

because of $\hat{\mathbf{u}} = \mathbf{u}$ according to our assumption. The MMSE estimate of $\mathbf{X}$, $\hat{\mathbf{X}} = E[\mathbf{X}|\mathbf{V}, \mathbf{U}]$, assuming Gaussian data statistics is given by:

$$\hat{\mathbf{X}} = a\mathbf{V} + b\mathbf{U}, \tag{14}$$

where $a = \sigma_X^2 \sigma_W^2 (1-\alpha)(-2\alpha\sigma_W^2\sigma_X^2 + \sigma_X^2\sigma_W^2 + \alpha^2\sigma_W^2\sigma_X^2 + \alpha^2\sigma_Z^2\sigma_X^2 + \sigma_Z^2\sigma_W^2)^{-1}$, $b = \sigma_X^2(\sigma_W^2\alpha + \alpha\sigma_Z^2 - \sigma_W^2)(-2\alpha\sigma_W^2\sigma_X^2 + \sigma_X^2\sigma_W^2 + \alpha^2\sigma_W^2\sigma_X^2 + \alpha^2\sigma_Z^2\sigma_X^2 + \sigma_Z^2\sigma_W^2)^{-1}$.

The variance of this estimator is:

$$D^r(\alpha) = E[d^N(\hat{\mathbf{X}}, \mathbf{X})] = \frac{\sigma_X^2\sigma_W^2\sigma_Z^2}{\alpha^2\sigma_X^2\sigma_Z^2 + \sigma_W^2(\sigma_X^2(1-\alpha)^2 + \sigma_Z^2)}. \tag{15}$$

In the noiseless case (scenario B), $\sigma_Z^2 = 0$, using (12) and assuming that $\alpha \neq 1$ and $\mathbf{V} = \mathbf{Y} = \mathbf{X} + \mathbf{W}$, the estimate (14) is reduced to:

$$\hat{\mathbf{X}} = \frac{1}{1-\alpha}(\mathbf{V} - \mathbf{U}) = \frac{1}{1-\alpha}(\mathbf{Y} - \mathbf{U}) = \frac{1}{1-\alpha}(\mathbf{X} + \mathbf{W} - \alpha\mathbf{X} - \mathbf{W}) = \mathbf{X} \tag{16}$$

that leads to $D^r = 0$ and provides the perfect reversibility.

In the above analysis we have referred to the generic selection of the parameter α. However, it depends on the variance of the watermark and the noise, i.e., maximum allowed embedding and attacking distortions. Normally in the practice of the digital data-hiding, the actual value of the applied attack variance is rarely known in advance at the encoder. Thus, α is selected keeping in mind some critical, the least favorable, or average conditions of system applications. This definitely provides the mismatch between the optimal parameter and the actual one that leads to some decrease in the system performance in terms of maximum achievable rate that will be shown by the results of our simulation.

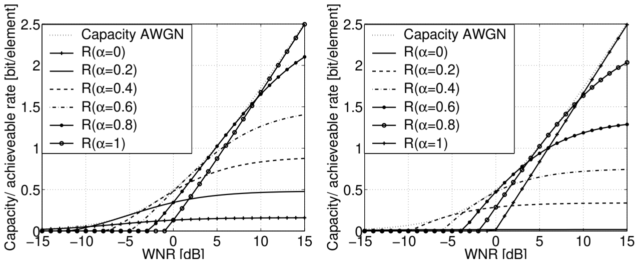
Nevertheless, it is interesting to investigate the hypothetical system performance in terms of reversibility, if one assumes the perfect knowledge of the operational scenario at the encoder that makes possible to choose the optimal parameter $\alpha = \alpha_{opt}$ according to the Costa result. In this case, substituting $\alpha_{opt} = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_Z^2}$ into (15), one obtains:

$$D^r(\alpha_{opt}) = \frac{\sigma_X^2(\sigma_W^2 + \sigma_Z^2)}{\sigma_W^2 + \sigma_X^2 + \sigma_Z^2} \tag{17}$$

that coincides with the estimation variance of the unauthorized user (9). The Gel'fand-Pinsker/Costa set-ups are designed to maximize the rate of reliable communications but not to minimize possible distortion of the host communicated via the noisy channel. This justifies that side information-assisted host estimation accuracy in this set-up cannot exceed one provided by estimation without side information. Thus, this scheme is not the optimal one when two constraints are imposed simultaneously. The option of reversibility was considered as a granted one along the main line of reliable message communications.

4 Results of Computer Simulation

To confirm the theoretical findings, we have performed the experimental validation of different reversibility scenarios for the Gaussian set-up. Figure 5 summarizes the known results for the achievable rates of the Costa set-up (6) with different values of optimization parameter α for the WIR equal to -6 dB and -16 dB to underline the critical dependence of the achievable rates on the selection of α. While capacity of the AWGN channel is achieved for α_{opt} (6), the fixed α is not optimal for all WNRs in terms of achievable rate and one observes the rate loss. It is a natural price for the lack of prior information about attack variance at the encoder.



(a) (b)

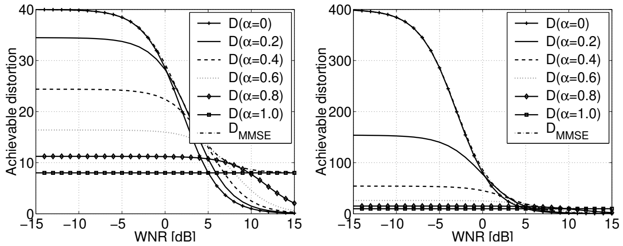
Fig. 5. Costa rate: WIR=-6 dB and WIR=-16 dB.

To investigate the impact of α on the restoration distortion, we have performed a number of simulations for different types of users. Firstly, assuming unauthorized user, who is aware only of the host, watermark and noise statistics, we have applied the MMSE estimation (8). The variance of this estimate D^r_{MMSE} (9) equals to the variance of the authorized user $D^r(\alpha_{opt})$ (17) and is plotted in Figure 6 for both WIRs. Secondly, assuming the authorized user with the knowledge of the key (consequently, we suppose the knowledge of $\mathbf{U}$), we have computed the variance of the restored host $D^r(\alpha)$ according to (15) for various values of α (Figure 6).

The obtained results confirm the non-optimality of the optimal Costa α selection for host communications. They demonstrate the estimation accuracy improvement at low WNRs in comparison with unauthorized user/authorized user with $\alpha = \alpha_{opt}$ when α parameter increases. However, at high WNRs the situation is the opposite one. This behavior is justified by the fact that for $\alpha = 0$ (spread sprectrum communications) $\mathbf{U} = \mathbf{W}$ and it represents additional interference source for host communications. In this case the input for the optimal MMSE estimator of $\mathbf{X}$ will be $(\mathbf{V} - \mathbf{U})$. Therefore, $D(\alpha = 0) = \frac{\sigma_X^2 \sigma_Z^2}{\sigma_X^2 + \sigma_Z^2}$ and asymptotically perfect host recovery at high WNRs ($\sigma_Z^2 \to 0$) is possible.

When $\alpha = 1$, $\mathbf{V} = \mathbf{X} + \mathbf{W} + \mathbf{Z}$, $\mathbf{U} = \mathbf{X} + \mathbf{W}$ and the optimal MMSE estimate is obtained based on $\mathbf{U}$ only. Thus, $D(\alpha = 1) = \frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}$ and it is independent of σ_Z^2. Therefore, at low WNRs this selection of α provides the smallest possible variance of the host estimation while at the high WNRs presence of $\mathbf{W}$ leads to the performance loss in comparison with the previous case ($\alpha = 0$).

The performed analysis allows to conclude that knowledge of the auxiliary random variable plays a crucial role for accurate host estimation in the Costa communications set-up. However, in order to provide satisfactory solution for both high and low WNRs one needs to design a communications protocol for the properly selected value of α (for instance, $\alpha \in [0.4, 0.6]$, Figure 6).



(a) (b)

Fig. 6. Distortion: WIR=-6 dB and WIR=-16 dB.

Finally, it can be observed (Figure 6) that as WNR $\to \infty$ or $\sigma_Z^2 \to 0$ that corresponds to the noiseless case within the considered scenario B, D^r for the authorized user tends to 0 for all values of α. This corresponds to the case of perfect reversibility and confirms our theoretical analysis. At the same time, the unauthorized user distortion asymptotically tends to $\frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}$, i.e., it is non-decreasing with σ_Z^2 that prevents the perfect reversibility for the unauthorized user in this signal processing set-up.

5 Conclusions and Future Perspectives

In this paper, the problem of reversibility of random binning-based data-hiding was analyzed from multimedia perspectives. Estimation-based reversibility was generally formulated within the Gel'fand-Pinsker framework and qualitatively analyzed in the Costa set-up. We demonstrated that in the noisy case the unauthorized user is capable to remove the hidden data using optimal MMSE with the same host reconstruction distortion than the authorized one with the perfect knowledge of the attacking noise variance. Contrarily, non-optimal in the communications sense selection of α together with the access to the proper codeword $\mathbf{U}$ provide significant estimation performance improvement. In the noiseless case ($\sigma_Z^2 \longrightarrow 0$), the knowledge of $\mathbf{U}$ allows the authorized user to completely recover the host data ($\sigma_Z^2 = 0$) that is never possible for the unauthorized user.

As a possible extension of the presented results we are going to consider the problem of maximization of the rate of reliable communications for a given target distortion D^{*r} and WNR regime that can be formulated as a joint optimization of achievable rate and restoration distortion. Finally, the same set-up will be analyzed from the security perspective.

6 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 European Commission through sixth framework program under the number FP6-507609 (SIMILAR) and Swiss IM2 projects.

References

1. Cox, I.J., Miller, M.L., Bloom, J.A.: Digital Watermarking. Morgan Kaufmann Publishers, Inc., San Francisco (2001)
2. Gel'fand, S., Pinsker, M.: Coding for channel with random parameters. Probl. Control and Inf. Theory **9** (1980) 19–31
3. Costa, M.: Writing on dirty paper. IEEE Trans. on Inf. Th. **29** (1983) 439–441
4. Perez-Freire, L., Perez-Gonzalez, F., Voloshynovskiy, S.: Revealing the true achievable rates of Scalar Costa Scheme. In: IEEE International Workshop on Multimedia Signal Processing (MMSP), Siena, Italy (2004) 235–238
5. Eggers, J., Buml, R., Tzschoppe, R., Girod, B.: Inverse mapping of scs-watermarked data. In: Eleventh European Signal Processing Conference (EUSIPCO'2002), Toulouse, France (2002)
6. Moulin, P., O'Sullivan, J.: Information-theoretic analysis of information hiding. IEEE Trans. on Information Theory **49** (2003) 563–593
7. Voloshynovskiy, S., Koval, O., Deguillaume, F., Pun, T.: Visual communications with side information via distributed printing channels: extended multimedia and security perspectives. In: Proceedings of the SPIE Int. Conf. on Security and Watermarking of Multimedia Contents III, San Jose, CA, USA (2004)
8. Kay, S.M.: Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall Signal Processing Series (1993)
9. Cover, T., Thomas, J.: Elements of Information Theory. Wiley and Sons, New York (1991)