A NEW PERSPECTIVE FOR EMBEDDING-DETECTION METHODS WITH DISTORTION COMPENSATION AND THRESHOLDING PROCESSING TECHNIQUES

Husrev T. Sencar[1], Mahalingam Ramkumar[2], Ali N. Akansu[1]

[1] New Jersey Institute of Technology
Electrical and Computer Engineering Department
Newark, New Jersey 07102
[2] Polytechnic University
Computer Information Science Department
Brooklyn, New York 11201

ABSTRACT

In this paper, we analyze oblivious (blind) information hiding methods from a new perspective. In [1], Costa introduced a communications framework that also applies to oblivious information hiding. We present an alternate and equivalent framework by carrying out the channel dependent nature of the optimal encoder in a different manner. Within the proposed framework, decoder structure is simplified, although in effect it's a slender advantage compared to overall complexity. This interpretation provides a better connection between the analytical results and practical designs. We evaluate the practical embedding-detection schemes employing scalar quantization procedures along with thresholding and distortion compensation types of processings from this perspective. Furthermore, we justify the assumptions for the optimality of the two types of processings.

1. INTRODUCTION

Information hiding addresses the achievable limits and design of methods for conveying a message signal through a host signal in an imperceptible and reliable way. Ultimately, information hiding formulation reduces to a trade-off among the goals of hiding rate, imperceptibility and robustness. One categorization of information hiding methods is based on the availability of the host signal at the extraction of the host signal. Our focus in this study is the oblivious one where extractor has no access to the host signal.

The analysis of oblivious information hiding was initially inspired by earlier studies that considered communications with side information [2] [3]. Costa, [1], from a communications standpoint introduced a framework that also applies to oblivious information hiding. He showed that the capacity of a power constrained additive white Gaussian noise (AWGN) channel with the side information at the encoder, in the form of channel's state, is the same with that of an AWGN channel under Gaussian distributed side information and input signal. He also described the optimal encoding and decoding that achieves the corresponding capacity. When his results and the framework were translated into oblivious information hiding by interpreting the side information as the host signal, power constraints as the perceptual limitations, channel distortion as the attack and the input as the codeword corresponding to a message to be conveyed, they laid the groundwork for the upcoming formulations of the problem and the practical designs.

Refs. [4] [5] presented the complementary formulation of the problem from a game-theoretic viewpoint that also considers the interactions between encoder-decoder and attacker in terms of the knowledge of one party on the other's strategy. They showed that the setting in [1] refers to upper bound on the coding capacity of all versions of the game since attacker has a fixed strategy known to both encoder and decoder. In [6] [7] [8] [9], schemes that follow the framework of [1] are introduced. Among these methods [7] [8] [9] proposed the use of scalar quantization procedures for embedding and detection that also employ thresholding and distortion compensation processings as a part of the embedding and those methods are of concern to us.

In this paper, we present an alternate framework derived based on Costa's framework with the corresponding embedding-detection scheme. In the proposed framework, the channel dependent nature of the optimal encoder is applied through a newly introduced term that generates the codeword. At the same time, decoding structure is simplified by removing the need for channel noise information, however the added complexity due to acquisition of this information at the decoder is very small compared to the amount of information shared between the encoder and decoder. This interpretation of the oblivious information hiding leads to a better connection between the analytical results and practical designs. Therefore, methods employing processings such as thresholding and distortion compensation are better evaluated. We compare the embedding-detection schemes relying on the two types of processings based on their detection performances.

2. COSTA'S FRAMEWORK AND THE ALTERNATE FRAMEWORK

Costa considered a communications scenario with the random channel state information made available to the encoder, but not to the decoder, in a non-causal manner. The goal of the communication is to transmit the index of a message to the receiver in N uses of the channel. The channel output at the receiver is defined in terms of the input X, the random state S and the AWGN Z as $Y = X + S + Z$. The channel input X is power limited and corresponds to the codeword for the particular message to be transmitted. The variables X, S and Z are assumed to be identical independently distributed (i.i.d.) random vectors of size N with the Gaussian distributed marginals $\mathcal{N}(0, P)$, $\mathcal{N}(0, \sigma_S^2)$ and $\mathcal{N}(0, \sigma_Z^2)$, respectively. Extending the results of [3] for continuous alphabets with the design of $U = X + \alpha S$, $0 < \alpha < 1$, for the auxiliary variable, Costa computed the transmission rate for the assumed channel. Furthermore, he showed that for $\alpha = P/(P+\sigma_Z^2)$

the rate is maximized to $1/2\log_2\left(1+P/\sigma_Z^2\right)$ bits per transmission which is the capacity of the same AWGN channel with the side information available to both encoder and decoder. Therefore, the presence of the channel state information at the encoder but not at the decoder does not reduce the capacity as long as encoder is modified to adapt the codeword X to the host signal S.

The encoding and decoding scheme that achieves the capacity is derived based on random coding techniques. A very large number of U sequences of length N with individual distributions $\mathcal{N}(0, P+\alpha^2\sigma_S^2)$ are generated and divided into bins so that each bin is associated with a separate message. Then, these sequences and the α parameter are made available to both encoder and decoder. The codeword corresponding to a message is obtained by finding the U sequence, that's jointly typical with the given S, in the bin pointed by the message index. In other words, the codeword X is the resulting signal sequence due to $U-\alpha S$ while satisfying the power constraint, $\frac{1}{N}||X||^2 \leq P$, and the orthogonality constraint, $E[XS]=0$. At the decoder, the U sequence that is jointly typical with the received signal Y is searched over all sequences in all bins such that $E[Y(U-\alpha Y)]=0$ is satisfied. The message index associated with bin containing the particular U sequence is defined as the sent message.

Under the same scenario, in the alternate framework, the design of $U = X+\alpha S$ in Costa's framework is reduced to the design of $U = X+S$ for all channel noise levels. The channel dependent nature of the encoder is carried out by the newly introduced term X_t which is a function of X. For the general case, the relation between X and X_t is defined in terms of the normalized correlation $\rho = \frac{E[XX_t]}{\sqrt{E[X^2]E[X_t^2]}}$. Correspondingly, the channel output at the receiver is re-defined as $Y = X - X_t + S + Z$. The codeword corresponding to a message index to be transmitted is $X_n = X - X_t$ and the power constraint on the channel input is imposed on X_n.

The transmission rate of this channel can be calculated in the same manner for the assumed distributions of X, S, Z and the dependency between X and X_t. When X, X_t, S, Z are distributed according to $\mathcal{N}(0,\sigma_X^2)$, $\mathcal{N}(0,\sigma_{X_t}^2)$, $\mathcal{N}(0,\sigma_S^2)$, $\mathcal{N}(0,\sigma_Z^2)$, respectively, and X_t is a linear function of X with $\sigma_X = (P+\sigma_Z^2)/\sqrt{P}$ and $\sigma_{X_t} = \sigma_Z^2/\sqrt{P}$, the transmission rate achieves the capacity derived by Costa. Moreover, all transmission rates, depending on the selection of α, in Costa's framework can be equivalently achieved in the alternate framework by the proper selection of σ_X. Therefore, two frameworks are equivalent and can be translated into each other through $\sigma_X = \frac{\sqrt{P}}{\alpha}$.

Encoding and decoding for the alternate framework follows that of [1] with the difference that at the encoding the power constraint is imposed as $\frac{1}{N}||X_n||^2 \leq P$. When compared with the Costa's framework, the decoding within the alternate framework is simplified. This is because finding the jointly typical (U,Y) pair does not require the channel noise information to be present in advance (*i.e.* α). (It should be noted that, in Costa's framework, the sent message is decoded by a search over all U sequences so that $Y^T(U-\alpha Y)\approx 0$ is satisfied.) The codebook design in the alternate framework follows $U = X+S$ irrespective of the channel noise level, or equivalently $\alpha = 1$ for all the cases. The channel dependent nature of the encoding, however, is reflected on both the inputs X and X_t. Thus, host interference rejection at the detector is achieved solely by the embedder's ability to properly select σ_X and σ_{X_t}, depending on σ_Z. Correspondingly, when the channel noise variance is changed, in Costa's framework, either a new set of U sequences (with variance $P+\alpha^2\sigma_S^2$) is generated and transferred to the decoder along with the α, or the existing U sequences are left untouched but P is adjusted so that $P+\alpha^2\sigma_S^2$ is the same for the new α which is made known to the decoder. However in the alternate framework, if the channel noise level changes, only encoder has to set the σ_X and σ_{X_t} to their optimal values.

Encoding and decoding of a message m within Costa's framework and alternate framework is depicted in Fig. 1-a and Fig. 1-b. With Gaussian assumption on all variables, the optimal encoder-decoder design in the former relies on choosing α properly whereas in the latter it depends on choosing σ_X when X_t is a linear function of X ($\rho = 1$). Distortion compensation and thresholding are the two types of processings that can be evaluated within the alternate framework.

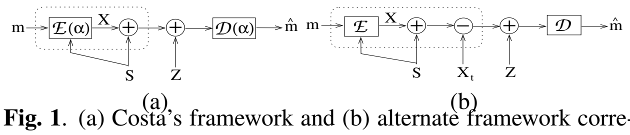


Fig. 1. (a) Costa's framework and (b) alternate framework corresponding to codebook designs of $U = X+\alpha S$ and $U = X+S$, respectively.

3. EMBEDDING AND DETECTION

Evaluating the communication frameworks given in Sec. 2 in terms of information hiding, S, X and Z are respectively interpreted as the host signal, the embedding distortion (or codeword corresponding to the message to be conveyed), and the additive attack. The encoder-decoder pair is the functional equivalent of the embedder-detector pair. The message is a sequence of information symbols which is conveyed by embedding into the host signal. The perceptual constraint used to limit and quantify the distortions due to embedding and attack is the squared error measure (power of the difference signal) because of the assumed power limited communications scenario. Described optimal encoding-decoding schemes also apply to information hiding, however, a practical implementation is not possible. Therefore, schemes described in [6] [7] [8] [9] devised practical methods for embedding and detection. Corresponding hiding rate versus robustness performances of these methods can be judged based on how closely their codebook designs approximate the optimal codebook.

The optimal codebook design in Costa's framework is achieved through the design of $U = X+\alpha S$ for $\alpha = \frac{P}{P+\sigma_Z^2}$ under Gaussian distributed X, S and Z. In the alternate framework, $U = X+S$ with $\sigma_X = \frac{P+\sigma_Z^2}{\sqrt{P}}$ and $\sigma_{X_t} = \frac{\sigma_Z^2}{\sqrt{P}}$ correspond to the optimal codebook design. In Ref. [6], optimal codebook is generated by the use of optimal quantizers. Whereas Refs. [7] [8] [9] employed scalar quantization procedures for codebook design, therefore they are more suitable for practical implementations. These techniques are of special interest due to thresholding and distortion compensation processing that follows embedding quantization. Effects of these processings on the rate versus robustness performance can be better explained in terms of alternate framework by representing them through X_t in the formulation which is defined as the processing distortion.

Chen *et al.* in [10] introduced the Quantization Index modulation (QIM) and an implementation based on dithered scalar quantizers, dither modulation (DM). Later in [7], they identified the capacity achieving variant of QIM as Distortion Compensated QIM (DC-QIM) with perturbating the quantization index modulated signal by a subtractive distortion which is a scaled version of the embedding distortion. Ramkumar *et al.* in [8], based on a similar embedding method considering binary signaling, employed thresholding processing where the embedding distortion introduced to the host signal is thresholded.

3.1. Distortion Compensation and Thresholding

In QIM, the codeword, in terms of the alternate framework, is defined as $X_n = X$ (where X is the distortion introduced to S due to quantization) since no processing is involved, $X_t = 0$. The embedding distortion due to QIM can be expressed as $E\left[X^2\right] = P$ when the the embedding signal size N tends to ∞ and X is zero mean. In DC-QIM, embedding by QIM is followed by subtracting the $1-\alpha$ scaled version of the embedding distortion X from the QIM embedded signal. Thus, processing distortion is defined as $X_t = (1-\alpha)X$, $\rho = 1$, and the codeword is defined as $X_n = \alpha X$. Within DC-QIM, the embedding parameters are the scaling factor α and the distance between the reconstruction points of the embedding quantizers Δ. Since the total distortion introduced to the host signal S due to X and X_t can be expressed as $E[X_n^2] = P$, the variance of the embedding distortion X is computed in the limit as

$$\sigma_X^2 = E\left[X^2\right] = E\left[\left(\frac{X_n}{\alpha}\right)^2\right] = \frac{P}{\alpha^2} = \frac{(P+\sigma_Z^2)^2}{P} \qquad (1)$$

where $\alpha = \frac{P}{P+\sigma_Z^2}$ and X and X_t are considered to be zero mean. It should be noted that, the variance of the X found in Eq. (1) is the same as dictated by the alternate framework and X_t is a linear function of X. Therefore, distortion compensation is the optimal processing when both the host signal S and the embedding distortion X are Gaussian distributed and the attack is AWGN.

When thresholding processing of [8] is considered, the processing distortion takes the form of thresholding noise, $X_t = max(0, |X| - \beta/2) sign(X)$. The embedding parameters in this case are the Δ and the threshold $\pm\beta/2$. The threshold ($0 < \beta \leq \Delta$) used to generate the codeword, $X_n = min(|X|, \beta/2) sign(X)$, through limiting X is set based on the channel noise. However, a capacity achieving embedding-detecting scheme based on thresholding processing is not possible as the relation between X and X_t is not linear. On the other hand, both thresholding and distortion compensation leads to useful practical embedder-detector designs.

3.2. Practical Methods

In the practical extensions, where uniform scalar quantization is employed for embedding along with the squared error distortion measure and AWGN attack, the performance depends on how the embedding distortion X and the corresponding processing distortion X_t are generated. Due to the scalar quantization, X has a non-Gaussian distribution, and therefore distortion compensation is not necessarily the optimal processing.

For the general case where a host signal sequence of length N is considered, the message to be conveyed to the receiver is mapped into a sequence of samples with the same length. Then, the sample sequence is embedded by quantizing each host signal coefficient with the quantizer corresponding to the signal sample which yields X as the embedding distortion. The stego signal, $S + X_n$, is generated by by subtracting the processing distortion X_t from X. Thus, the codeword corresponding to the particular message is $X_n = X - X_t$. Detection of a message is by determining the nearest reconstruction value to the received signal Y. The sample associated with the reconstruction value of the corresponding quantizer is deemed the sent one. Hence, minimum distance decoder is the optimal extraction method and requires N hard decisions to be made in order to detect the hidden sample sequence. When detector is also given the set of message signals, such that the sent sample sequence is constrained to be one of them, a detection method based on soft decisions can be employed. In this case, rather than forcing each of the N signal coefficients to one of the sample values, a real valued sequence of length N is extracted and the sent message is detected directly.

With thresholding as the processing, the embedding distortion X due to quantization and the processing distortion X_t are optimized in terms of the embedding parameters Δ and β. The threshold β, unlike the quantization step size Δ, is not known to the detector. The values of Δ and β change with the embedding distortion power P to the channel noise power, σ_Z^2 (WNR=$\frac{P}{\sigma_Z^2}$). For distortion compensation type of processing, similarly, Δ and α are to be optimized.

Considering binary signaling, each embedded sample can be extracted through soft decisions by mapping the stego signal (or a distorted version) over a periodic triangular function whose peak locations coincide with the quanta values of embedding quantizers, [8]. Therefore, a binary sample sequence is extracted as a real valued signal vector. Then, the sent message is determined by matching the extracted signal to one of the signals in the message signal set available to the detector via computing normalized correlations. Alternately, Euclidean distance decoder can be used to extract each embedded samples by making a hard decision on each sample. Correspondingly, the sent message is determined by combining the extracted samples into a sequence of message signal.

The embedding quantization prior to distortion compensation or thresholding processing is same for both processings except for the value of the parameter Δ. However, the processing distortion X_t and the dependency on X are different for the two types of processing. The performance results for the two processings can be computed in terms of probability of not detecting a hidden sample or in terms of the similarity between a hidden sample sequence and extracted signal sequence. Alternately, one can optimize α through maximizing the mutual information between the embedded sample and the received signal at all WNRs. An approximation to the optimal α value is given in Ref. [9] as $\sqrt{\frac{P}{P+2.71\sigma_Z^2}}$.

From detector's point of view, an erroneous extraction of the embedded message signal is due to two sources of noise. These are the channel noise Z and the processing distortion X_t. It should be noted that an embedded sample will be extracted correctly from the signal $Y = S + X$ since it is a quantized value. Therefore, the effective noise, Z_{eff}, at the detector is $Z_{eff} = Z - X_t$. The statistics of X, X_t and X_n can be obtained based on the type of processing that follows the embedding quantization. For binary signal embedding and small distortions scenario ($\sigma_S^2 \gg \sigma_Z^2$, P) the pdfs of X, X_t and X_n are as shown in Fig. 2-a and Fig. 2-b. Since AWGN attack Z is independent of X_t, Z_{eff} is defined as $f_{Z_{eff}}(z_{eff}) = \int_{-\infty}^{\infty} f_Z(z_{eff} - x) f_{X_t}(x) dx$. The pdf of Z_{eff}, $f_{Z_{eff}}(z_{eff})$, due to embedding with thresholding for $Z \sim \mathcal{N}(0, \sigma_Z^2)$ is derived as

$$f_{Z_{eff}}(z_{eff}) = \frac{\beta}{\Delta\sqrt{2\pi\sigma_Z^2}} \exp^{-\frac{z_{eff}^2}{2\sigma_Z^2}} \qquad (2)$$
$$+ \quad \frac{1}{2\Delta}\left(\text{erf}\left(\frac{z_{eff} + \frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_Z}\right) - \text{erf}\left(\frac{z_{eff} - \frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_Z}\right)\right)$$

where $\text{erf}(z) = \frac{2}{\pi}\int_0^z \exp^{-x^2} dx$ is the Gaussian error function.

Similarly for distortion compensation it is expressed as

$$f_{Z_{eff}}(z_{eff}) = \frac{1}{2(1-\alpha)\Delta}\left(\operatorname{erf}\left(\frac{z_{eff}+\frac{(1-\alpha)\Delta}{2}}{\sqrt{2}\sigma_Z}\right) - \operatorname{erf}\left(\frac{z_{eff}-\frac{(1-\alpha)\Delta}{2}}{\sqrt{2}\sigma_Z}\right)\right). \quad (3)$$

Fig. 3 displays the quanta values (marked by $\times$ and $\circ$ characters) of the embedding quantizers corresponding to binary sample values with decision regions for making hard decisions and the periodic triangular extraction function for making soft decisions. (The stego signal $S + X$ is generated by translating S to one of the quanta values marked in Fig. 3.)

The probability of error, P_e, in detecting a sent binary symbol can be analytically computed, for both processings, for the given pdfs of Z_{eff} and decision regions. Assuming a stego signal coefficient generated by translating the host signal to the quanta value $q_\times$, Z_{eff} determines the deviation of the received signal from $q_\times$ due to processing distortion and channel noise. Therefore, $P_e = \int_{R_\circ} f_{Z_{eff}}(z_{eff} - q_\times)dz_{eff}$. Alternately, the normalized correlation between the embedded binary sample sequence and the extracted signal can be calculated for the given Z_{eff} statistics based on the utilized periodic extraction function. The mean of the normalized correlation m_ρ is derived for large N as

$$\begin{aligned} m_\rho &= \frac{R(1)}{\sqrt{R(2)}}, \\ R(p) &= 2\sum_{i=0}^{i=\infty}\int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}}\left(\left(\frac{(2i+1)\Delta}{4} - z_{eff}\right)(-1)^i\right)^p \\ &\quad f_{Z_{eff}}(z_{eff})dz_{eff}, \end{aligned} \quad (4)$$

where Z_{eff} is as defined in Eqs. (2) and (3)

Fig. 4 displays the detection performance for the two methods obtained for binary signal embedding under squared error distortions and AWGN attack. The α and β values are optimized over all Δ values to yield minimum P_e, Fig. 4-a, and maximum normalized correlation, Fig. 4-b, for WNR range of -15 dB to 0 dB. Both of the results indicate that at WNRs below -9 dB thresholding is marginally better and at higher WNRs distortion compensation is preferable.

4. CONCLUSION

In this paper, we present an alternate framework to that introduced by Costa. The alternate framework provides a better connection with the analytical results and practical hiding methods that utilize some form of post-processing following the embedding quantization. In the corresponding channel model, Fig. 1-b, X and X_t are the distortions introduced to the host signal due to quantization and processing, respectively. We considered distortion compensation and thresholding types of processings where $X_t = (1-\alpha)X$ and $X_t = min(0, |X| - \beta/2)sign(X)$. The probability of detection error and normalized correlation performance results show that both processings lead to useful hiding methods when scalar quantization is used for embedding.

5. REFERENCES

[1] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, 1983.

[2] C. E. Shannon, "Communication theory of secrecy systems," *Bell Systen Technicl Journal*, vol. 28, pp. 656–715, 1949.

[3] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[4] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transactions on Information Theory*, vol. 49, pp. 563–593, Mar. 2003.

[5] A. S. Cohen and A. Lapidoth, "The gaussian watermarking game," *IEEE Transactions on Information Theory*, vol. 48, pp. 1639–1667, June 2002.

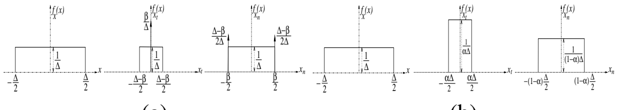
[6] J. Chou, S. S. Pradhan, L. E. Ghaoui, and K. Ramchandran, "A robust optimization solution to the data hiding problem using distributed source coding principles," in *Proc SPIE: Image and Video Communications and Processing*, 2000, vol. 3974.

[7] B. Chen and G. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking," in *Proc SPIE: Security and Watermarking of Multimedia Contents II*, 2000, vol. 3971, pp. 48–59.

[8] M. Ramkumar and A. N. Akansu, "Self-noise suppression schemes for blind image steganography," in *Proc SPIE International Workshop on Voice, Video and Data Communication, Multimedia Applications*, Sept. 1999, vol. 3845.

[9] J. J. Eggers, J. K. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," *IEE Colloq. Secure Images and Image Authentication*, vol. 4, pp. 1–6, Apr. 2000.

[10] B. Chen and G. W. Wornell, "Provably robust digital watermarking," in *Proc SPIE: Multimedia Systems and Applications*, 1998, vol. 3845.



(a) (b)

Fig. 2. Pdfs corresponding to (a) thresholding and (b) distortion compensation processings.

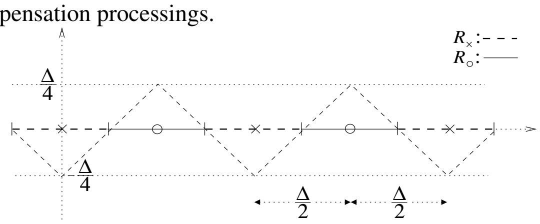
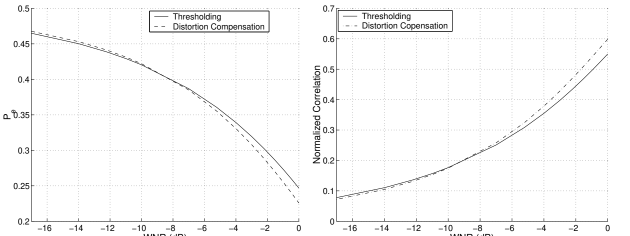


Fig. 3. Quanta values corresponding to embedding quantizers with decision regions and the periodic detection function.



(a) (b)

Fig. 4. (a) P_e and (b) Normalized correlation performances for binary signal embedding with thresholding and distortion compensation processings.