COSTA PROBLEM UNDER CHANNEL AMBIGUITY

J.E. Vila-Forcén, S. Voloshynovskiy, O. Koval and T. Pun

University of Geneva, Department of Computer Science.
24 rue Général-Dufour, CH 1211, Geneva, Switzerland

ABSTRACT

In this paper, we address the analysis of the Costa setup under channel uncertainty. Since the Costa setup was entirely considered under the Gaussian assumptions about host and channel statistics, we assume that the channel is an additive white Gaussian noise (AWGN) with unknown variance defined on some interval. Firstly, we solve the problem of average achievable rate optimization assuming that the distribution of noise variances is known. Secondly, we withdraw the previous assumption and consider that the variance distribution is unknown. The corresponding criteria are formulated and the performance of the Costa under these criteria is demonstrated.

1. INTRODUCTION

Robust data-hiding has appeared as an emerging tool for copyright protection. In this application, the main requirement is to provide reliable communications of the information hidden in the body of a media file via some aggressive environment created by the attacker. The issue of maximization of the achievable rate is closely related to the problem of host interference cancellation. This problem was extensively investigated in digital communications by Gel'fand and Pinsker [1]. They were the first who demonstrated how, using a random binning codebook construction, to relax the host interference for the case of a known fixed channel with random parameter.

Costa [2] considered the Gel'fand-Pinsker problem in the Gaussian formulation and showed that the maximal achievable rate for such a channel coincides with the capacity of the ideal AWGN channel without any interference. For this purpose, Costa applied a specific assumption about the codeword construction that crucially relies on the knowledge about the channel statistics (the AWGN variance) available at the encoder prior to the transmission.

Evidently, this assumption can be rarely met in practical situations. Furthermore, in robust data-hiding, where usually the opposite is supposed, i.e., that the attacker can know the data-hider strategy and that the data-hider cannot predict the attacker strategy. The problem of the achievable rate maximization in the scenario with an aggressive adversary can be found in [3]. The results obtained there also might be used for the benchmarking of various practical robust data-hiding methods.

Another widely used benchmarking strategy, introduced in [4], consists in evaluation of the achievable rate in the AWGN channel assuming that its variance varies within a predefined interval, in particular in terms of watermark-to-noise ratio (WNR) it is defined to be $[-5, 10]$dB. The rate evaluation here is performed assuming a fixed encoder structure that is optimal only for the given communications conditions, i.e., for the fixed channel variance. However, some rate loss is observed in the performance [2] while deviating from this optimal noise variance. Moreover, one can conclude that for some practical encoding strategies this loss can be significant [5]. This loss can be justified since most of the existing practical embedding techniques can be considered as practical approximations to the Costa coding for particular communications conditions.

Therefore, contrarily to the existing encoding strategies for practical robust data-hiding attempting encoder optimization for the particular state of the attacking AWGN channel (i.e., fixed variance), we would like to address the problem of optimal encoder design for the Costa communications setup assuming that the channel variance is varying within some interval. Optimality here is measured in a twofold manner, in terms of average loss of the achievable rate from the AWGN channel capacity and when the minimum rate of reliable communication on the interval is specified.

The paper has the following structure. A brief overview of the Costa communications protocol is given in Section 2. The framework for average-achievable-rate-optimal communications is analyzed in Section 3. The analysis of the achievable rates of the Costa setup for unknown variance probability density function (pdf) is presented in Section 4. Finally, Section 5 concludes the paper.

This paper was partially supported by SNF Professorship grant No PP002-68653/1, Interactive Multimodal Information Management (IM2) project and by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT and FP6 - 507609 - SIMILAR. The authors are thankful to the members of the Stochastic Image Processing group at University of Geneva for many helpful and interesting discussions. The information in this document reflects only the author's 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.

1-4244-0469-X/06/$20.00 ©2006 IEEE

Notations We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$ and corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively. The variance of a random variable X is designated as σ_X^2. Finally we define the WNR as $\mathrm{WNR} = 10\log_{10}\frac{\sigma_W^2}{\sigma_Z^2}$ and the watermark-to-image ratio (WIR) as $\mathrm{WIR} = 10\log_{10}\frac{\sigma_W^2}{\sigma_X^2}$, where X, W and Z denotes the host, watermark and noise random variables, respectively.

2. COSTA DATA-HIDING

As it was mentioned in the Introduction, Costa [2] considered the Gel'fand-Pinsker [1] problem in the Gaussian formulation, i.e., assuming that $X \sim \mathcal{N}(0,\sigma_X^2)$, $Z \sim \mathcal{N}(0,\sigma_Z^2)$ and the encoder power constraint defined according to the mean-square error distance, meaning $E[W^2] \leq \sigma_W^2$ (Figure 1).

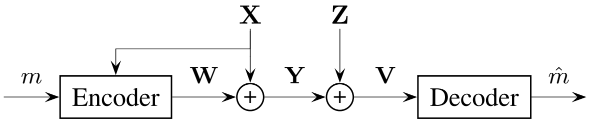


Fig. 1. Costa communications setup.

The auxiliary random variable was chosen in the form $U = W + \alpha X$ with optimization parameter α that leads to the following rate of reliable communications:

$$R(\alpha,\sigma_X^2,\sigma_W^2,\sigma_Z^2) = \frac{\sigma_W^2(\sigma_W^2+\sigma_X^2+\sigma_Z^2)}{\sigma_W^2\sigma_X^2(1-\alpha)^2+\sigma_Z^2(\sigma_W^2+\alpha^2\sigma_X^2)}, \quad (1)$$

where α should be fixed in advance at the encoder. It was shown that the optimization parameter can be selected as $\alpha_{opt} = \frac{\sigma_W^2}{\sigma_W^2+\sigma_Z^2}$ that requires the knowledge of the noise variance σ_Z^2 at the encoder. In this case the achievable rate does not depend on the host variance and:

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

where C^{AWGN} is the capacity of the AWGN channel without host interference.

3. AVERAGE ACHIEVABLE RATE OPTIMAL COSTA CODING: ENCODING STRATEGIES

In order to achieve optimal performance, the codebook in the Costa communication protocol should be designed assuming that $\alpha = \alpha_{opt}$. Since this selection can be made only for the case when the AWGN channel variance is known at the encoder prior to transmission, it can be hardly satisfied in practice. According to the motivation presented in the Introduction, the main goal of this Section is to propose a Costa-based communications protocol that performs optimally in terms of average achievable rate on the interval of attacking channel variances that correspond to the WNR $\in$ [WNR$_{\text{min}}$, WNR$_{\text{max}}$]. Optimization is performed with respect to the parameter α under the assumption that the attacking variances are distributed according to some $f_{\Sigma_Z^2}(\sigma_Z^2)$ of interest.

This problem can be formulated as the minimization of the rate loss $R_L(\alpha)$ on the given WNR interval:

$$\hat{R}_L = \min_{\alpha} R_L(\alpha) = \min_{\alpha}\int_{\text{WNR}_{\text{min}}}^{\text{WNR}_{\text{max}}}\left(C^{\text{AWGN}} - R(\alpha,\sigma_Z^2)\right) f_{\Sigma_Z^2}(\sigma_Z^2)d\sigma_Z^2, \quad (3)$$

where $f_{\Sigma_Z^2}(\sigma_Z^2)$ is uniform on the interval WNR $\in$ [WNR$_{\text{min}}$, WNR$_{\text{max}}$].

Unfortunately, no close analytical solution to this problem was found and the optimal parameter α in terms of average achievable rate was determined using numerical optimization. The optimization was performed for the interval defined by WNR$_{\text{min}} = -5$dB and WNR$_{\text{max}} = 10$dB for two WIR, WIR$_1 = -6$dB and WIR$_2 = -16$dB (here it was assumed that the variance of the watermark σ_W^2=10 in order to satisfy the embedding distortion constraint according to the Stirmark benchmark [6]).

The solution to this problem is shown in Figure 2 and in Table 1. The results of optimization are given versus those ones obtained under the assumption that the data-hider is targeting to optimize the performance for the favorable conditions for the attacker, meaning, for the maximal variance of the noise. The obtained results allow to conclude that the average achievable rate for the proposed optimization case is higher than in the latter case at least by factor of 1.8 for both WIRs.

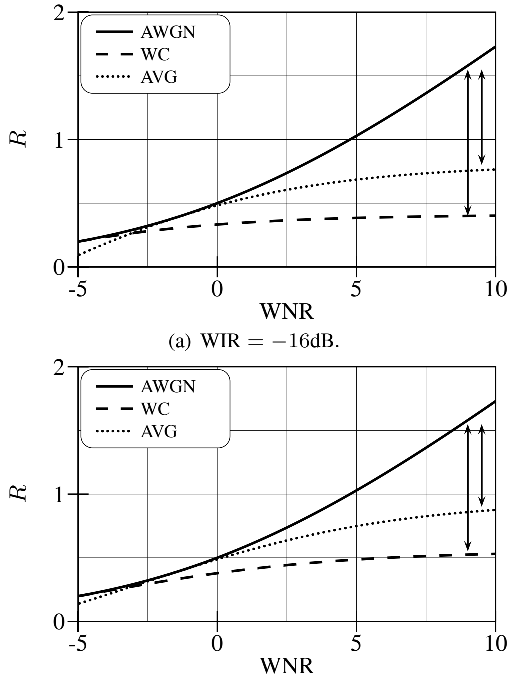
WIR	α	$\hat{R}_L$	$R_L(\alpha_{\text{WC}})$	$R_L(\alpha_{\text{WC}})/\hat{R}_L$
-16dB	0.42	0.0938	0.169	1.8
-6dB	0.446	0.0698	0.136	1.94

Table 1. Average rate loss analysis of on-average-optimal Costa coding versus worst-case-adopted encoding strategy.

4. COSTA CODING PERFORMANCE UNDER UNCERTAINTY ABOUT THE CHANNEL VARIANCE

It is evident that the optimization performed in the previous Section is not possible if the distribution of the attacking variances is unknown. In this case one can consider the rate optimization problem from the min-max performance perspective:

$$\min_{\alpha}\max_{\text{WNR}\in[\text{WNR}_{min},\text{WNR}_{max}]}\left(C^{\text{AWGN}} - R(\alpha,\sigma_Z^2)\right). \quad (4)$$



(a) WIR $= -16$dB.

(b) WIR $= -6$dB.

Fig. 2. Achievable rate performance comparison of on-average-optimal Costa coding (AVG) versus worst-case-adopted encoding strategy (WC).

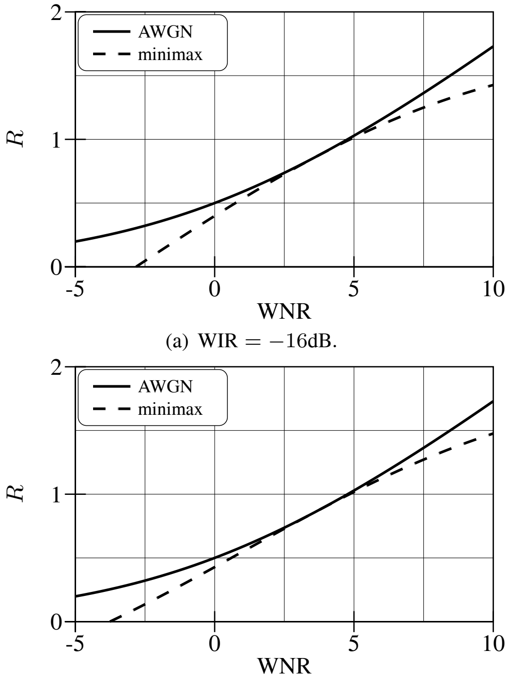
It can be shown that for this case the optimal value of α parameter is given by:

$$\alpha = \begin{cases} \frac{\sigma_W^2+\sigma_{Z\min}^2-\sqrt{\sigma_W^2\sigma_{Z\min}^2+(\sigma_{Z\min}^2)^2}}{\sigma_W^2+\sigma_{Z\min}^2}, \\ \qquad \sigma_{Z\max}^2 \geq \frac{\sigma_W^2\sigma_a^2(\sigma_W^2+\sigma_{Z\min}^2)}{\sigma_X^2(\sigma_W^2+\sigma_{Z\min}^2-\sqrt{\sigma_{Z\min}^2(\sigma_W^2+\sigma_{Z\min}^2)})}, \\ \frac{\sigma_W^2(\sigma_W^2\sigma_a^2+\sigma_{Z\max}^2(\sigma_W^2+\sigma_{Z\min}^2))-\sqrt{(\sigma_W^2)^2\sigma_{Z\max}^2\sigma_{Z\min}^2\sigma_a^2\sigma_b^2}}{(\sigma_W^2)^3+\sigma_{Z\max}^2\sigma_{Z\min}^2(\sigma_W^2-\sigma_X^2)+\sigma_W^2(\sigma_X^2+\sigma_{Z\max}^2+\sigma_{Z\min}^2)}, \\ \qquad \text{otherwise,} \end{cases} \tag{5}$$

where $\sigma_a^2 = \sigma_W^2 + \sigma_X^2 + \sigma_{Z\min}^2$ and $\sigma_b^2 = \sigma_W^2 + \sigma_X^2 + \sigma_{Z\max}^2$. $\sigma_{Z\min}^2$ and $\sigma_{Z\max}^2$ correspond to the WNR$_{\max}$ and WNR$_{\min}$, respectively.

The results presented in Figure 3 results allow to conclude that for some particular low-WNRs only zero-rate communications are possible. Obviously, such a solution is not acceptable for the robust data-hiding. In order to overcome the revealed problem, one can apply the modified worst case attacking scenario where it supposed to guarantee not maximum but a predefined achievable rate of reliable communications. The solution to this problem with respect to the α parameter is easily obtained solving the equation (1):

$$\alpha = \frac{\sigma_W^2\sigma_X^2 + \sqrt{2^{-2r}\sigma_W^2\sigma_X^2\sigma_b^2(\sigma_W^2 - \sigma_{Z\max}^2(2^{2r}-1))}}{\sigma_X^2(\sigma_W^2+\sigma_{Z\max}^2)}, \tag{6}$$



(a) WIR $= -16$dB.

(b) WIR $= -6$dB.

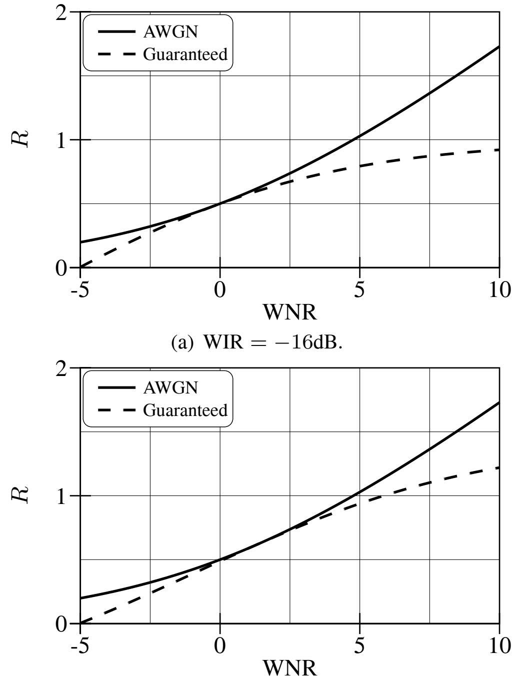
Fig. 3. Achievable rates in the Costa setup when the distributions of the attacking channel variances is unknown.

where r denotes the minimum achievable rate. One can compute r supposing that 64 bits should be reliably communicated through an 512×512 pixel image.

The achievable rates for such a coding for the WNR $\in [-5; 10]$dB and WIR$_1 = -6$dB and WIR$_2 = -16$dB are presented in Figure 4.

5. CONCLUSIONS

In this paper we have analyzed the Costa communications setup assuming different level of prior information about the statistics of the attacking channel. We proposed rather than optimizing the protocol performance for a certain WNR to maximize the average achievable rate on the given interval of WNRs assuming that the attacking channel variances are distributed uniformly. The result of the formulated problem gives the value of the optimization parameter α to be used in the codebook design. Furthermore, we considered the case when the distribution of these variances is unknown and performed the analysis within the minimax framework. The obtained result revealed the problem of zero-rate communications for the low-WNR regime. The disclosed problem solution was obtained assuming that fixed rate of reliable communication should be guaranteed for the worst communication conditions.



(a) WIR $= -16$dB.

(b) WIR $= -6$dB.

Fig. 4. Achievable rates in the Costa communications setup with the fixed lower bound.

6. REFERENCES

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

[2] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[3] P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 563–593, Oct. 2003.

[4] J. Eggers, J. K. Su, and B. Girod, "Performance of a practical blind watermarking scheme," in *Proceedings of IS&T/SPIE 13th Annual Symposium: Security and Watermarking of Multimedia Contents*, San Jose, CA, Jan. 2001.

[5] F. Pérez-González, F. Balado, and J. R. Hernández, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Transactions on Signal Processing, Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery*, vol. 51, no. 4, pp. 960–980, Apr. 2003.

[6] F. A. P. Petitcolas, "Stirmark benchmark 4.0," `http://www.cl.cam.ac.uk/\%7efapp2/watermarking/stirmark`, 2002.