# Applied public-key steganography

Pierre Guillon, Teddy Furon, Pierre Duhamel

# Applied public-key steganography

Pierre Guillon[a], Teddy Furon[b] and Pierre Duhamel[c]

[a]THOMSON multimedia R&D Security lab, Cesson Sevigné, France
[b]TELE lab of Université Catholique de Louvain, Louvain-la-neuve, Belgique
[c]LSS lab of SUPELEC, Gif sur Yvette, France

## ABSTRACT

We consider the problem of hiding information in a steganographic framework, i.e. embedding a binary message within an apparently innocuous content, in order to establish a 'suspicion-free' digital communication channel. The adversary is passive as no intentional attack is foreseen. The only threat is that she discovers the presence of a hidden communication. The main goal of this article is to find if the Scalar Costa Scheme, a recently published embedding method exploiting side information at the encoder, is suitable for that framework. We justify its use assessing its security level with respect to the Cachin's criterion. We derive a public-key stego-system following the ideas of R. Anderson and P. Petitcolas. This technique is eventually applied to PCM audio contents. Experimental performances are detailed in terms of bit-rate and Kullback-Leibler distance.

**Keywords:** passive steganography, scalar costa scheme, public-key system

## 1. INTRODUCTION

Steganography is the art and science of hiding data into innocent-looking cover-data so that no one can detect the very existence of the hidden data. The study of this subject may be schemed by Simmons's prisoner's problem: Alice and Bob are in prison, and want to finalize an escape plan. All their communication pass through the warden, Wendy. Both prisoners have therefore to communicate invisibly, in order not to arouse Wendy's suspicion. The crucial point of stego-systems is hence invisibility, but a great capacity is also desired.

To establish a communication channel, Alice sends to Bob some innocuous contents. Alice is said to be active when she hides a message $m$ modifying these cover-contents $X$ into stego-contents $Y$. Alice is not active when she sends really innocuous contents $X$. The goal of Wendy is to intercept these data and know whether they are innocuous (i.e. $X$) or stego-contents (i.e. $Y$). Wendy is a passive opponent.

To transform the cover-content $X$ into stego-content $Y$, Alice embeds a watermark signal $W$ at a watermark to cover-content power ratio $WCR$. While transmitted some channel noise $N$ is added. This disturbs the communication between Alice and Bob. An important parameter is the watermark to noise power ratio $WNR$.

This paper presents a new public-key steganographic protocol. It uses a recently published embedding scheme developed by J. Eggers and *al.*: The Scalar Costa Scheme (SCS), which embeds information at a very high bit-rate. Section 2 gives a short overview of the principles and practices of this scheme. Section 3 describes how a security level can be defined in the framework of the passive steganography. There are three criteria to pay attention to. The first one is perceptual quality of the stego-contents. Their quality has to be as good as the cover-contents. But, a watermark that is not perceptible does not imply that Wendy cannot create a test that distinguishes cover-contents from stego-contents. The Cachin's criterion gives a bound of the efficiency of Wendy's test whatever its structure. At last, we review the rationale of R.Anderson and *al.* who imagined a way to derive a public-key stego-system. In section 4, we wonder if the SCS is compatible with the later criteria. We stress some conditions on its use and derive a full stego-system divided into two phases. In the initialisation phase, Alice sends Bob a secret parameter via a public-key stego-system at a low bit-rate. In the permanent phase, thanks to this secret, Alice sends Bob information at a very high bit-rate. Section 5 deals with practical experiments when cover-contents are audio clips in CD quality.

(Send correspondence to T.F.)
T.F.: E-mail: teddy.furon@ieee.org, address: IRISA/TEMICS, Campus universitaire de Beaulieu, 35042 Rennes, France
P.D.: E-mail: pierre.duhamel@lss.supelec.fr

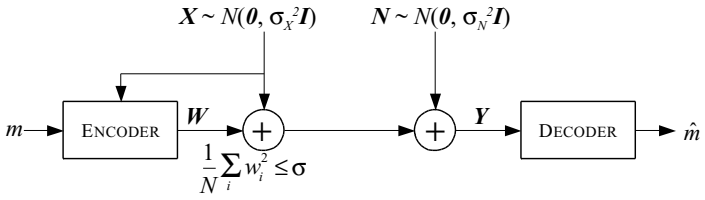


**Figure 1**: Costa communication scheme: the channel state $\mathbf{X}$ is a a side information known at the encoder.

## 2. ACHIEVING A HIGH CAPACITY

This section is a quick presentation of the role played by the side information at the embedding stage from a theoretical point of you to a practical implementation. The stego-channel is not a regular channel as the main source of noise, i.e. the cover-content, is known at the embedding stage. This section is mainly based on the works of J. Su, J. Eggers and B. Girod. We advise to read the following references for further details.[1–3] Others possibilities are articles from P. Moulin,[4] from J. Chou and *al.*,[5] and from B. Chen and *al.*[6]

### 2.1. Costa's theory

M. Costa published in 1983 an article entitled *Writing on dirty papers* proving the positive effect of a side information at the encoding stage of a digital communication over an additive white Gaussian noise channel.[7] The framework is sketched in Fig.1. In order to transmit the symbol $m$ in $n$ channel uses, the informed encoder emits the signal $\mathbf{W}$ whose power is constrained by $\sigma_W^2$. The transmitted signal is polluted by two independent Gaussian white sources: the state $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2\mathbf{I})$ and the noise $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma_V^2\mathbf{I})$. The fact that the encoder knows the values taken by the state $\mathbf{X}$ is called the side information. The decoder received the signal $\mathbf{Y} = \mathbf{W} + \mathbf{X} + \mathbf{N}$ and *without* the knowledge of the state $\mathbf{X}$, takes a decision $\widehat{m}$ about the transmitted symbol. M. Costa proved that the capacity of this informed encoder only channel equals $C^*$ bits, which is the capacity when the encoder and the decoder share the side information.

$$\text{Costa's equations:} \qquad C^* = \frac{1}{2}\log_2(1 + \frac{\sigma_W^2}{\sigma_N^2}) \qquad \text{for} \qquad \alpha^* = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_V^2} \tag{1}$$

The proof is based on a clever strategy to take advantage of the side information, i.e. to create the signal $\mathbf{W}$ knowing $m$ and $\mathbf{X}$. Let us consider that the encoder and the decoder share the knowledge of a huge set $\mathcal{U}$ of reference signals $\mathbf{U}$. $\mathcal{U}$ is then evenly partitioned into (almost) $2^{nC^*}$ code-books $\mathcal{U}_m$. When the encoder has to transmit the symbol $m$, he emits the signal $\mathbf{W}$ which will 'push' the state $\mathbf{X}$ towards the nearest reference signal $\mathbf{U}^*$ of the code-book $\mathcal{U}_m$. The encoding is optimal if $\mathbf{W} = \mathbf{U}^* - \alpha^*\mathbf{X}$. The goal of the detector is then to take a decision to which code-book $\mathcal{U}_{\widehat{m}}$ the received signal $\mathbf{Y}$ is more likely to belong.

Our extremely simplified explanation of the Costa's scheme helps understanding his basic idea: it requires too much power to 'fight against' the effect of the state $\mathbf{X}$ (i.e. to set $\alpha = 1$), a better strategy is to play with it. Costa illustrated this principle by the issue of writing on dirty papers. It takes too much energy to clean-up the paper and then to write on it. A better strategy is to compose with the marks. Costa wisely sums up his philosophy by the expression 'Do the best with what you have'.

### 2.2. Su's interpretation

Costa proved the benefit of the side information in terms of information theory, but he didn't provide a practical implementation. J.Su and *al.* tackle this lack with a geometrical interpretation, that we will briefly describe.

In their article,[1] J.Su and *al.* represent all Costa's random vectors by points in $\mathbb{R}^n$, where orthogonality between Gaussian random vectors stands for their independence. Bin-encoding spheres, centered on each code-vector $\mathbf{U}$ depict Costa's encoding process. The radius of these spheres is proportional to the power constraint on $\mathbf{W}$. The choice of the appropriate code-vector then relies on a very simple decision process: knowing the

message $m_0$, the code vector $\mathbf{U}_0 \in \mathcal{U}_{m_0}$ to be chosen is the one whose encoding sphere contains the scaled state vector $\alpha^*\mathbf{X}_0$. This is judiciously relating Costa's encoding process with the quantization of $\alpha^*\mathbf{X}_0$.

J.Su and *al.* also design a supplementary constant $c^*$, so that the transmitted vector $\mathbf{X}+\mathbf{W}$ equals the vector $c^*\mathbf{U}+\mathbf{T}$, where $\mathbf{T} \perp \mathbf{U}$. The value of $c^*$ is calculated this way: assuming that $\mathbf{X} \perp \mathbf{W}$, we can express $\mathbf{U}$ and $\mathbf{T}$ according to the orthonormal basis $(\mathbf{X}' = \mathbf{X}/\sigma_X, \mathbf{W}' = \mathbf{W}/\sigma_W)$.

$$\text{In this basis:} \qquad \mathbf{U} = \alpha^*\sigma_X.\mathbf{X}' + \sigma_W.\mathbf{W}' \quad \text{and} \quad \mathbf{T} = (1-\alpha^*c^*)\sigma_X.\mathbf{X}' + (1-c^*)\sigma_W.\mathbf{W}'$$

$\mathbf{T} \perp \mathbf{U}$ if and only if their scalar product on this basis is null. This leads to the following definition of $c^*$.

$$c^* = \frac{\alpha^* + \frac{\sigma_W^2}{\sigma_X^2}}{(\alpha^*)^2 + \frac{\sigma_W^2}{\sigma_X^2}} = \frac{1}{\alpha^*}\frac{1+\frac{WCR}{\alpha^*}}{1+\frac{WCR}{\alpha^{*2}}} \tag{2}$$

Whereas the encoding process takes place in the basis $\{\alpha^*\mathbf{X}, \mathbf{W}\}$, the decoding process takes place in the basis $\{c^*\mathbf{U}, \mathbf{T}\}$. In so doing, they explicitly give a way to process the decoding function of Costa's theoretical scheme. Indeed, due to this orthogonality, the decoding process is now similar to the encoding process. The vector $\mathbf{Y}_0 = \mathbf{X}_0 + \mathbf{W}_0 + \mathbf{N}_0$ is received by the decoder. The decision process is to find the appropriate reference vector $\mathbf{U}$, using bin-decoding spheres centred on $c^*\mathbf{U}$, whose radius depend on the powers of each vector $\mathbf{X}$, $\mathbf{W}$ and $\mathbf{N}$. The code-vector $\mathbf{U}$ to be chosen is the one whose corresponding bin-decoding sphere contains the received vector $\mathbf{Y}_0$. The decision $\widehat{m}$ is eventually the index of the code-book this code-vector belongs to.

### 2.3. Eggers' Scalar Costa Scheme

In J.Su's interpretation of Costa's scheme, there are two different kinds of spheres: ones for the encoding stage, centred on $\mathbf{U}$, and the others for the decoding stage, centred on $c^*\mathbf{U}$. In the following scheme called Scalar Costa Scheme (SCS), J.Eggers reduces the whole Costa's scheme to only one type of coding spheres. To use the same coding spheres at the decoding stage is equivalent to assume that the scaling constant $c^*$ can be approximated by $1/\alpha^*$ and that coding $\alpha^*\mathbf{X}$ along $\mathbf{U}$ is equivalent to coding $\mathbf{X}$ along $\mathbf{U}/\alpha^*$. This approximation generally holds, assuming that the $WCR$ is not too high, and $WNR$ is not too low as shown in Eq.(2). These are the needed conditions to claim that $\alpha, \alpha^2 \gg WCR$. Steganography frameworks perfectly meet these conditions.

J.Eggers's issue is also to practically substitute Costa's theoretical huge random code-books. He presented a suboptimal scheme employing a lattice-structured code-book. The first idea is to decompose the code-book $\mathcal{U}$ into a product of one dimension code-books $\mathcal{U}^1$. Assuming a message $m$ of $N$ D-ary letters $d_n \in \{0, 1, \ldots, D-1\}$ to be sent, a $N$-dimensional code book $\mathcal{U}^N$ is set up, structured as the following code-book product:

$$\mathcal{U}^N = \mathcal{U}^1 \times \mathcal{U}^1 \times \ldots \times \mathcal{U}^1 \qquad \text{with} \qquad \mathcal{U}^1 = \bigcup_{d=0}^{d=D-1} \mathcal{U}_d^1$$

Hence, the encoding and the decoding are proceeded component by component. The second idea is to give a strong structure on $\mathcal{U}^1$ to avoid the parsing of the whole code-book to find the nearest reference signal.

$$\mathcal{U}_d^1 = \left\{ u = k\alpha\Delta + \frac{d\alpha\Delta}{2D} | k \in \mathbb{Z} \right\} \qquad \rightsquigarrow \qquad \forall n \quad \frac{u_{n,0}}{\alpha} = \mathcal{Q}\left(x_n, \frac{\mathcal{U}_{d_n}^1}{\alpha}\right)$$

At the encoding process, given a signal vector $\mathbf{X_0}$ and the message $m$ to transmit (a sequence of letters $\{d_n\}$), the problem of finding the appropriate $\mathbf{U_0}$ is reduced to a uniform product quantizer process of the above equation where $\mathcal{Q}(.,\mathcal{U})$ denotes quantization to the code-book $\mathcal{U}$. The scaled code-book corresponds here to a uniform quantization with a step $\Delta$. The embedded signal is given by $\mathbf{W} = \mathbf{U_0} - \alpha\mathbf{X} = \alpha\mathbf{E}$. $\mathbf{E}$ is a quantization error. It is known to be almost orthogonal to the quantizer input $\mathbf{X}$, as required by Costa's scheme, assuming an almost uniform host signal probability density function (pdf) in the range of one quantization bin, which is

the case under the high resolution quantization assumption. The power of the quantization error $\mathbf{E}$ is given by $\Delta^2/12$. In order to control the distortion power $\sigma_W^2$, the two parameters $\alpha$ and $\Delta$ are related by

$$\alpha = \sqrt{\frac{\sigma_W^2}{E\{\mathbf{E}^2\}}} = \sqrt{\frac{12\sigma_W^2}{\Delta^2}} \tag{3}$$

Costa previously found an optimal value $\alpha^*$ of the constant $\alpha$ for his own theoretical system. The suboptimal system that J.Eggers propose leads to another value, which he numerically optimised for an AWGN channel.

$$\alpha^* = \sqrt{\frac{\sigma_W^2}{\sigma_W^2 + 2.71\sigma_V^2}} \tag{4}$$

Experiments we made show that Eggers's formula of $\alpha$ is a statistically robust solution for this scheme, which achieves optimality for most of channel noise pdf. The embedding process is summed up by Eq.(5):

Let us now consider the decoding process. In the SCS, each component of the received vector $\mathbf{Y}_0$ is re-quantized along the code-books $\mathcal{U}^1$, and the indices of the corresponding bins stand for the estimated letter $\hat{d}_n$. Practically, it is defined by the following formula, where $\mathbf{P}$ is a pre-processed data.

$$\text{Embedding: } s_n = x_n + \alpha^* \left( \mathcal{Q}_\Delta \left\{ x_n - \Delta \frac{d_n}{D} \right\} - \left( x_n - \Delta \frac{d_n}{D} \right) \right) \qquad \text{Detection: } p_n = \mathcal{Q}_\Delta\{y_n\} - y_n \tag{5}$$

Depending on $\mathbf{P}$, an estimation of $\mathbf{d}$ can be evaluated. E.g., for a binary SCS, $p_n$ is close to zero if $d_n = 0$, and close to $\pm\Delta/2$ if $d_n = 1$. For a $D$-ary SCS, a similar rounding process can be specified. The probability of detection error against the $WNR$ ratio is given in Fig.2. Note that for a $WNR < -0.47$ dB, Eq. (4) yields $\alpha^* < 0.5$. It means that the decoding spheres are not disjointed as illustrated in Fig. 2 so that errors occurs due to the method and due to the channel noise. It can be wise to fix $\alpha^* = 0.5$ if $WNR < -0.47$dB.

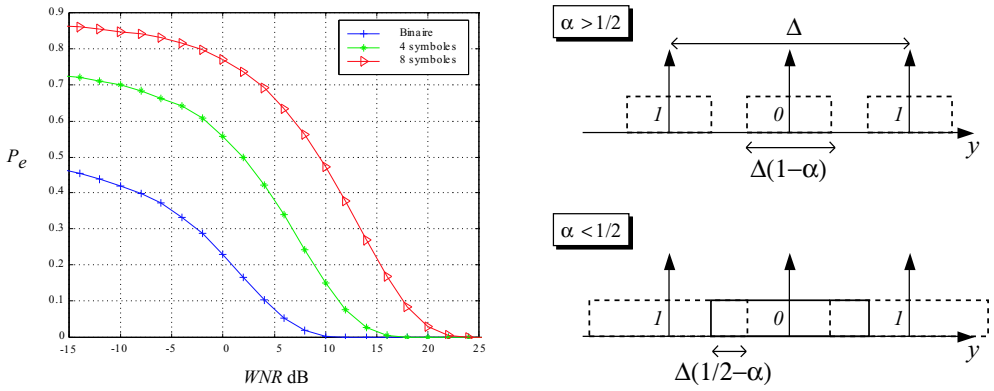


**Figure 2.** a) Probability of decoding errors for a AWGN channel. b) Decoding spheres for D=2. If $y$ belongs to a dashed (resp. solid) line box, then $\hat{d} = 1$ (resp. $\hat{d} = 0$). If $\alpha < 0.5$ the decoding boxes intersect leading to decoding errors.

## 3. ASSESSING SECURITY

In this part, we expose three criteria assessing that a stego-system is secure.

### 3.1. The embedding distortion

The stego-content should be perceptually clean. This is a very subjective criterion especially since Wendy does not know the covert-content. If Alice can not make the difference between the two contents, then Wendy is likely not to suspect stego-contents. This constraint is usually tackled by fixing the watermark to covert-content power ratio $WCR$.

### 3.2. Anderson's criterion

This subsection focuses on the rationale detailed in the article.[8] An encryption method is assumed to produce cipher-texts, which are statistically indistinguishable from random words. Alice encrypts the information and embeds this random-like message in a channel known to Bob and Wendy. Bob, receiving this data, cannot a priori decide whether it contains information or not. He simply extracts and decrypts it using his secret key. If Alice was actually active, this decrypted data is meaningful and is the hidden information.

Since we assume that Wendy knows the detection function and parameters, she can extract the hidden message too. Nevertheless, given that the encryption algorithm produces a random-like cipher-text, she thus cannot decide if the extracted message contains any hidden information because she hasn't got the key to decrypt it. The security therefore rests on the reliability of the crypto-system. The encryption is actually achieved by either a public-key or a private-key crypto-system. However, the use of a private-key scheme needs a secure preliminary phase where Alice and Bob exchange this secret key. A better solution is therefore to use a public-key crypto-system avoiding the secret exchange.

We conclude this subsection on Anderson's criterion by clearly stating its fundamental hypothesis. The stego-content must looks like innocuous. R. Anderson and F. Petitcolas assume this is possible diluting the message to be hidden into a large amount of content. Flipping the least significant bit of one cover-content sample out of a hundred is surely not noticeable for Wendy, but this spoils the capacity of the stego-system. Our goal is to show that the Anderson strategy is feasible with a high bit-rate thanks to the SCS.

### 3.3. Cachin's criterion

The fundamental hypothesis of the Anderson criterion can be easily analysed. The binary stream issued from the detection may not be the most informative data about the presence of a stego-channel. The raw samples of the received content may leak more information for that purpose. Cachin's criterion stems from this rationale.

#### 3.3.1. Definition and theorem

Cachin uses the relative entropy $D(.||.)$ (also called the Kullback-Leibler (KL) distance or the discrimination) between the pdf of the cover-content $X$ and the stego-content $Y$:

$$D(p_X||p_Y) = \sum_{c\in\mathcal{C}} p_X(c) \log \frac{p_X(c)}{p_Y(c)} \tag{6}$$

where $c$ is a content which belongs to the set of possible contents $\mathcal{C}$, distributed as $P_X(c)$ when Alice is passive and as $P_Y(c)$ when she is active.

DEFINITION 1. *A stego-system as defined above $\epsilon$-secure against passive adversaries if*

$$D(p_X||p_Y) < \epsilon \tag{7}$$

The security assessment is based on Wendy's test, a deterministic process whose input is the content $c \in \mathcal{C}$ and whose output $d$ is defined on $\{0, 1\}$. $d$ equals one if the content is considered as a stego-content (i.e. Alice is active), zero if the content is considered as normal. This binary random variable is distributed as $(P_{fa}, 1-P_{fa})$ when Alice is not active, as $(1 - P_{mis}, P_{mis})$ when she is. $P_{fa}$ is the probability of false alarm (Wendy accuses Alice whereas she is not guilty), $P_{mis}$ is the probability of a miss (Wendy doesn't detect a stego-content).

THEOREM 3.1 (DATA PROCESSING). *If a stego-system is $\epsilon$-secure against passive adversaries, it satisfies*

$$D_b(P_{fa}||P_{mis}) = P_{fa} \log \frac{P_{fa}}{1-P_{mis}} + (1-P_{fa}) \log \frac{1-P_{fa}}{P_{mis}} \leq \epsilon \tag{8}$$

*where $D_b$ is the relative entropy of a binary random variable.*

This theorem gives a limit to the efficiency of Wendy's test. E.g., if Wendy does not want to accuse Alice when not guilty (i.e. $P_{fa} = 0$), then $P_{mis} > e^{-\epsilon}$ (logarithm are to the base $e$). More illustrative is the Receiver Operating Curve (ROC) ($P_p = 1 - P_{mis} = P_p(P_{fa})$) of Fig. 5, which shows the possible operating points $(P_{fa}, P_p)$ when the KL distance is bounded.

#### 3.3.2. Application to Costa's theory

We define a content as a sequence of $N$ samples, so that $\mathcal{C} = \mathbb{R}^N$. The relative entropy is, from now on, the entropy of continuous processes and integrals replace the sums of Eq. (6). We calculate this relative entropy in the framework of Costa's theory explained in subsection 2.1. Hence, all the random variables are white and Gaussian. This calculus is then extremely easy.

$$D(p_X||p_Y) = \frac{1}{2}\left(\frac{\sigma_X^2}{\sigma_Y^2} - 1 + \log\frac{\sigma_Y^2}{\sigma_X^2}\right) \tag{9}$$

When Alice is active, $\mathbf{Y}$ is the sum of three independent r.v. $\mathbf{Y} = \mathbf{X} + \mathbf{W} + \mathbf{N}$ and $\sigma_Y^2 = \sigma_X^2 + \sigma_W^2 + \sigma_N^2$. In the case of a passive adversary, $\mathbf{N}$ is only a source coding noise like the distortion due to the quantization of the variable $\mathbf{X} + \mathbf{W}$: adding the signal $\mathbf{W}$, $\mathbf{X} + \mathbf{W}$ is then a real signal which needs to be quantified back. We model this noise by a simple and classical relation with the used bit-rate $R$ under the high resolution assumption: $\sigma_N^2 = k2^{-2R}\sigma_{X+W}^2$. Of course, this rationale is flawed by the fact that the quantization noise is uniformly distributed and not Gaussian as in the Costa's paper. But, our goal is only to roughly feel the impact of these parameters on Cachin's criterion. The following properties hold:

$$\frac{\partial}{\partial WCR}D(p_X||p_Y) > 0 \qquad \text{and} \qquad \frac{\partial}{\partial R}D(p_X||p_Y) < 0$$

This leads to some interpretations. The bigger is the power of the watermark signal, the more efficient is the adversary test. The bigger is the bit-rate, the less efficient is the adversary test. It means that Alice should choose a kind of cover-contents that needs the finest quantizer as possible. In other words, it should be simpler to build a $\epsilon$-secure stego-system based on PCM audio clips at 16bits/sample than on grayscale images at 8bits/sample. However, this statement has to be balanced with the fact that such contents do not support much distortion, hence, much watermark strength. At last, an amazing fact is that the quantization does not improve the security level as $D_{k>0}(p_X||p_Y) > D_{k=0}(p_X||p_Y)$. This seemingly does not respect the data processing theorem. This paradox is simply explained: when Alice is not active, $\mathbf{X}$ is not quantified.

## 4. AN EFFICIENT PUBLIC-KEY STEGANOGRAPHIC SCHEME

As we use the SCS and a product of one dimension code-books (cf. subsection 2.3), we restrict our analysis to one dimension. This is the reason why boldface font disappears in this section.

### 4.1. Anderson's criterion: cryptographic-keyed system

#### 4.1.1. A problem

Plugging the Anderson's criterion into the *binary* SCS framework, anyone may decode the hidden message, simply knowing the quantizer, i.e. the parameter $\Delta$. But the pdf of the stego-data is usually strongly noticeable. Anderson's strategy can not be applied.

The SCS actually causes periodical noticeable gaps of width $\Delta|(\alpha - 1/2)|$ on the stego-data pdf. Fig. 3 shows this phenomenon. These gaps stand either for impossible values of $Y$ ($\alpha > 1/2$), or on the contrary for values more likely to appear ($\alpha < 1/2$). Each value of the cover-data is replaced, by adding it a given fraction $\alpha$ of its quantization error. The possible values for the stego-data therefore spread in periodical boxes of width $(1-\alpha)\Delta$, centred on the quantization peaks. According to the value of $\alpha$, and considering the bit-wise dithering process, these boxes are either disjoint either intersecting, along a width of $\Delta|(\alpha - 1/2)|$ (cf. Fig. 2). When disjoint, these particular gaps stand for impossible values of $Y$, and induce holes in the pdf $p_Y$. When intersecting, the gaps contain values almost twice more likely to appear inducing peaks on the pdf.

Choosing $\alpha = 1/2$ produces a pdf without any of these gaps. However the scaling process, coupled with the message bit-wise dithering, generates a strange shape of the cover-signal pdf (cf. Fig. 3). Nevertheless this contraction is indiscernible if and only if the cover-data pdf is uniform.

Such phenomena imply that Wendy, noticing the particular shape of the pdf, can suspect the existence of the hidden communication, although she cannot decrypt the embedded information. This is a security flaw

and Anderson's criterion can not be applied in general to the SCS. The only solution is the following strategy: choose a cover-content with a uniform pdf, or, if necessary, equalize it with a compressor (a monotonic non-linear smooth function), then embed the information with $\alpha$ fixed to $1/2$, and finally recover the former pdf, by the inverse compressor. But $\alpha = 1/2$ is not the optimal parameter of Eq. (4) except if WNR=-0.47dB. Hence, this strategy leads to bigger bit error rate (BER). We sacrifice a part of the high SCS capacity to security reasons.

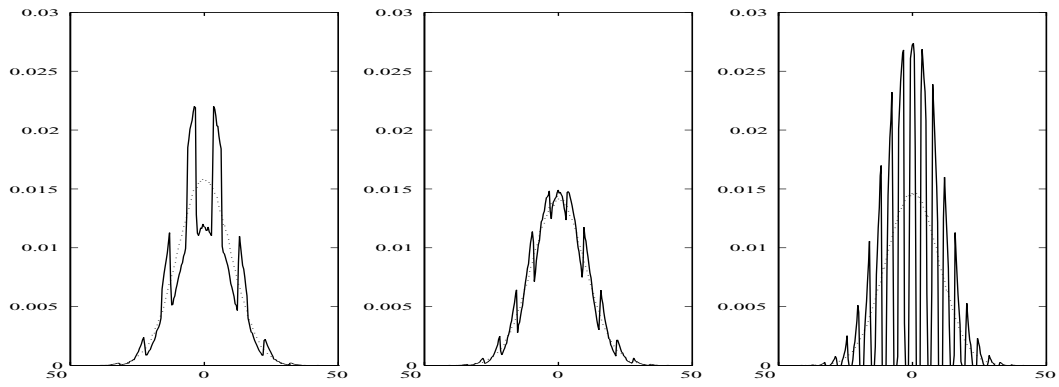


**Figure 3.** SCS (no dithering) for a Gaussian cover data pdf: stego-content data pdf for $\alpha < 1/2, \alpha = 1/2$, and $\alpha > 1/2$.

#### 4.1.2. The compressor

The compressor is a classical histogram equalization assured by a non-linear function $f$, depending on the cover-content pdf $p_X(.)$, and defined on the range $[X_{Min} = \min x, X_{Max} = \max x]$.

$$\xi = f(x) = (N_{bin} - 1)G(x) = (N_{bin} - 1)\left(\int_{X_{Min}}^{x} p_X(u)du - \frac{1}{2}\right) \tag{10}$$

The inverse compressor $f^{-1}$ is applied to the result of the embedding process, in order to recover the cover-content pdf. Side effects are cancelled scaling the output signal, so that its range of possible values corresponds to a multiple of the quantization bins width. This is done in Eq. (10) denoting $N_{bin}$ the number of bins.

We obtain a structure similar to the classical source coding compander.[9] The compressor followed by a uniform quantizer is equivalent to a non-uniform quantizer. Whereas in source coding, the compressor is tuned to minimise the distortion (i.e., by the Panter-Dite formula), it is set here to flatten the pdf of the cover-signal to enable Anderson's strategy (i.e. a public-key stego-system).

#### 4.1.3. The distortion control

The r.v. $\Xi$ defined in Eq. (10) has a constant pdf on the interval $[-(N_{bin} - 1)/2, (N_{bin} - 1)/2]$. The SCS mechanism is applied on this r.v. with $\alpha = 1/2$ and $\Delta = 1$. It outputs a r.v. $\Psi$, whose pdf is constant on the same interval. The SCS mechanism adds a noise H whose power is $\sigma_{\mathrm{H}}^2 = 1/48$ according to Eq.(3) and if the figure of bins is large enough to verify the high resolution assumption. Bennett's formula is then valid.[9]

$$\sigma_W^2 = \frac{\sigma_{\mathrm{H}}^2}{(N_{bin} - 1)^2} \int_{X_{Min}}^{X_{Max}} \frac{p_X(x)}{g(x)^2} dx = \frac{1}{48(N_{bin} - 1)^2} \int_{X_{Min}}^{X_{max}} \frac{1}{p_X(x)} dx \tag{11}$$

with $g(x) = dG(x)/dx$. For a given non singular pdf, the embedding distortion is only tuned by $N_{bin}$.

#### 4.1.4. The KL distance

For memory-less random processes and a monotonic increasing function $G(.)$, we have the following equation.

$$D(p_X||p_Y) = D(p_{G(X)}||p_{G(Y)})$$

If $D(p_\Xi||p_\Psi) < \epsilon$, then the stego-system is $\epsilon$-secure. This is always true because $D(p_\Xi||p_\Psi) = 0$ as the SCS with $\alpha = 0.5$ transforms a uniformly distributed r.v. into a uniformly distributed r.v defined on the same range.

#### 4.1.5. Protocol part: the initialization phase

The capacity is not has high as possible with the SCS, due to the fact that $\alpha = 0.5$ is not the optimum parameter. This choice may lead to a bigger bit error rate. On the other hand, no secret parameter has to be exchanged between Alice and Bob before the communication starts. Bob creates a public-key pair $(PubKey_{Bob}, PriKey_{Bob})$ and publishes his public-key $PubKey_{Bob}$. This parameter is then known from Alice and Wendy. We only use this public-key stego-system as an initialisation stage where Alice transmits to Bob a secret parameter $L$ that she will use after on. Hence, this scheme is only used to transmit a very short message, which is the encryption of $L$ with Bob's public key: $m = \text{Encrypt}_{PubKey_{Bob}}(L)$. The size of the keys and message is at least 1024 bits if a classical public-key crypto-system is used such as RSA.[10] A DH-EC (Diffie Hellman with Elliptic Curves) may also be used with keys' size of 160 bits. Once this short message is transmitted, Bob uses his private key $PriKey_{Bob}$ to decrypt and retrieve $L$. Now, Alice and Bob do share a secret parameter that will be used in a so-called permanent phase, with the very efficient scheme described in the next subsection.

### 4.2. Cachin's criterion: steganographic-keyed system

#### 4.2.1. Dithering

J.Eggers includes in the SCS an interesting process enabling the secrecy of the communication. It is assured by the random dithering function of a vector-key $\mathbf{K}$. Its random components $k_n$ are independently and uniformly chosen in the interval $(-1/2, 1/2]$. This changes the embedding and detection formula in:

$$s_n = x_n + \alpha\left(\mathcal{Q}_\Delta\left\{x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right\} + \Delta\left(\frac{d_n}{D} + k_n\right) - x_n\right) \quad \text{and} \quad p_n = \mathcal{Q}_\Delta\{y_n - k_n\Delta\} + k_n\Delta - y_n$$

The key shifts the position of the quantization bins from a vector component to another. Without its knowledge, there is no way to decode the stego-signal, i.e. to retrieve the hidden message. This dithering efficiently protects the decoding stage, and no encryption is any longer needed.

The statistical phenomena previously exposed for a no-key stego-system no longer exist. Indeed, the former gaps are not at the same position for different values of the key $k_n$, thanks to the dithering. Since it occurs according to a uniform i.i.d process, the global pdf is given by the mean of all possible pdf, with $k_n \in [-1/2, 1/2]$:

$$p_Y(y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} p_Y(y|k)dk$$

It therefore leads to a compensation of the previous holes and peaks, whatever the value of $\alpha$ (cf. Fig. 4). The resulting pdf of the stego-data is then very close to the cover-data's one. No compressor is then needed. $(\alpha^*, \Delta)$ are the optimal parameters given by equations (4) and (3).

#### 4.2.2. Key vector generation

If the vector-key is repeated, a simple statistical study over a stego-signal, component by component, estimates the values of the dithering. The vector-key has therefore to be as long as the signal itself, as required by Shannon's cryptographic theorem (i.e., as a one-time pad).[10] $\mathbf{K}$ is produced by a pseudo-random generator seeded by $L$. This generator produces extremely long sequences like Yarrow-160[11] (160 is the number of bits composing seed $L$). This implies Alice and Bob share secret parameter $L$. This sharing has previously occurred when Wendy was not present, or Alice sent $L$ to Bob via the public-key stego-system of subsection 4.1.

#### 4.2.3. The KL distance

In our framework, $\mathbf{W}$ and $\mathbf{N}$ modify the covert-signal. The first distortion is the noise of a quantization with a subtractive dithering. The quantization step is $\Delta$ from Eq. (3). This subtractive dithering usually improves quantizer performances. As the dithering signal $\mathbf{K}$ respects the Schuchman condition,[9] the quantization noise (resp. the watermark signal $\mathbf{W}$) is independent from the input signal and it is uniformly distributed on $(-\Delta/2, \Delta/2]$ (resp. $(-\alpha\Delta/2, \alpha\Delta/2]$). The additional dithering due to $\mathbf{d}$ does not change this condition.

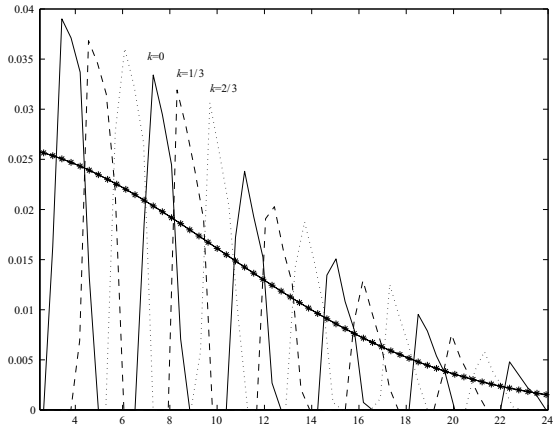


**Figure 4.** Conditional pdf of the stego-data for $\alpha > 1/2$ knowing k=-1/3, 0, 1/3; Resulting pdf from the keyed dithering

The channel noise $\mathbf{N}$ is equivalent of a quantization noise with a non-subtractive dithering. The quantization step is $\Delta_{SC}$, i.e. the step used in the cover-content source coding. The dithering signal is indeed the watermark signal $\mathbf{W}$. But, the non-subtractive quantization does not share the same properties than the subtractive version. For example, the quantization noise is no more independent from the input signal. Yet, in our case, we can argue that the original signal $\mathbf{X}$ is already quantified, so that $\mathbf{N}$ only depends on $\mathbf{W}$, which is independent from $\mathbf{X}$. Moreover, in the next section, a different strategy embeds the watermark signal in a transform domain: $\mathbf{X}$ is not any more quantified, but the quantization noise taking place in the temporal domain is modelized by a Gaussian independent noise $\mathbf{N}$ in the transform domain. Whatever the embedding strategy, we justify that the stego-signal is the sum of two independent r.v.: $Y = X + V$, such that $p_Y = p_X \otimes p_V$:

$$p_Y(y) = \int_{-\infty}^{\infty} p_X(y - v) p_V(v) dv \tag{12}$$

This expression may not lead to a tractable expression of the KL distance $D(p_X||p_Y)$. This is the discrimination Wendy can measure, i.e. ignoring the r.v. $V$. Hence, in its definition, the pdf of $Y$ is defined as in Eq. (12). We define the *condional discrimination* $D(p_X||p_Y|v)$ as the KL distance knowing $v$[12]:

$$D(p_X||p_Y|v) = \sum_{c \in \mathcal{C}} p_X(c) \log \frac{p_X(c)}{p_Y(c|v)}$$

and the *expected discrimination* as $D(p_X||p_Y|V) = E_V\{D(p_X||p_Y|v)\}$.

THEOREM 4.1 (CONDITIONING[12]). *Expected discrimination is nondecreasing under conditioning:*

$$D(p_X||p_Y) \leq D(p_X||p_Y|V)$$

If $D(p_X||p_Y|V) < \epsilon$, then the stego-system is $\epsilon$-secure. For example, suppose $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, then $D(p_X||p_Y|v) = \frac{E\{(v^2 - vx)\}}{2\sigma_X^2}$. Because $V$ and $X$ are independent, $D(p_X||p_Y|v) = (v^2 - v\mu_X)/2\sigma_X^2$ and, as $V$ is a centred process, $D(p_X||p_Y|V) = \sigma_V^2/2\sigma_X^2$. In the same way, if $X$ is Laplacian distributed, $D(p_X||p_Y|V) = \sigma_V^2/\sigma_X^2 + o(\sigma_V^2/\sigma_X^2)$. $\sigma_V^2$ is up-bounded by $(\sigma_W + \sigma_N)^2 = (\alpha\Delta + \Delta_{SC})^2/12$.

## 5. APPLIED STEGANOGRAPHY

### 5.1. Extension to colored cover signal

The stego-systems exposed so far assume that all data vectors were i.i.d. The second order statistics were hence not an issue as both the cover-signal and the embedded signal were white. If the cover-data is colored (e.g.

PCM audio samples), the addition of a white watermark signal becomes a noticeable feature of the stego-signal. This constitutes a security flaw. In this case, a transform $T$ has to whiten it. After the SCS embedding process, the white stego-data must be re-colored as the cover-data were, by the inverse transform.

We now have two domains: the PCM domain and the transform domain. The watermark signal is embedded with a given security level (i.e., a bounded KL distance and a given WCR) in the transform domain where the cover-data are white. The KL distance is invariant if $T$ is a linear invertible transform:

$$D(p_{\mathbf{TX}}||p_{\mathbf{TY}}) = D(p_{\mathbf{X}}||p_{\mathbf{Y}}) \tag{13}$$

An appropriate transform is the Karhunen-Loeve Transform (KLT) if the cover-data are Gaussian distributed. Yet, for non Gaussian random vector, the KLT only decorrelates the samples but do not render them independent. Nevertheless, for practical use, we select the DCT matrix as the $T$ transform pretending this approximation does not spoil our rationale. The structure of the embedding process looks like a transform source encoder.[9] Note that this structure is studied in terms of parallel Gaussian channels by P. Moulin.[4]

### 5.2. Experimental works on PCM audio contents

The cover data are PCM audio samples 44.1kHz - 16 bits. We work with analysis windows of 512 samples lasting 11.6ms. The DCT maps the 512 PCM samples onto 512 coefficients. To impose a given $WCR$, we estimate the cover-data power gathering the coefficients into 32 subbands during 10 consecutive analysis windows. This gives 160 coefficients to estimate the power $\sigma^2_{X,s,t}$ in one subband $s$ for one short period of time $t$. We call this set of coefficient the slot $(s,t)$. The power of the channel noise is $\sigma^2_{N,s,t} = \Delta^2_{SC}/12 \quad \forall(s,t)$, i.e. we assume the quantization in the PCM domain spreads its noise's power uniformly on each coefficients ($T$ is unitary). These 160 coefficients come in one SCS mechanism tuned with $(\alpha^*_{s,t}, \Delta_{s,t})$ calculated from $(\sigma^2_{X,s,t}, \sigma^2_{N,s,t}, WCR)$, except in the initialisation phase (cf. 4.1.5), where $\alpha_{s,t} = 0.5$ and the coefficients within a slot $(s,t)$ are supposed to be Laplacian distributed so that the compressor is defined by:

$$f_{s,t}(x) = \frac{(N_{bin}-1)}{2} \frac{1 - e^{-|x|\frac{\sqrt{2}}{\sigma_{X,s,t}}}}{1 - e^{-X\frac{\sqrt{2}}{\sigma_{X,s,t}}}} \text{sign}(x) \quad \text{with} \quad X = \max(|\max(x)|, |\min(x)|)$$

### 5.3. The spike model strategy

Parameter $\alpha^*$ is optimal in the sense that it leads to the minimum BER achievable for a given $WNR$. When this ratio is too low, the minimum BER is so high that even errors correcting codes (e.c.c.) do not enable a reliable communication. It is not worth spending some energy transmitting in this slot. This is the so-called spike model strategy.[4] For a fixed $WCR$, the $WNR$ is varying from slot to slot following the cover-signal power's evolution. Denote $WNR_T$ a threshold under which no embedding is proceeded. In practice, for the permanent phase, this threshold is 12dB if $D = 2$, 18dB $D = 4$ and 24dB for $D = 8$. The probability of detection error is then less than $5.10^{-3}$. For the initialisation phase, $D = 2$ and $WNR_T = 24$dB.

### 5.4. Detection errors

We assume Bob and Alice can be synchronised, so that Bob knows when Alice starts and stops the initialisation and permanent phases. Bob estimates the power of the cover-content $\sigma^2_{X,s,t}$ for each slot $(s,t)$ via Eq. (14).

$$\widehat{\sigma}^2_{X,s,t} = \frac{\sigma^2_{Y,s,t} - \Delta^2_{SC}/12}{1 + WCR} \tag{14}$$

He creates the compressor function in the initialisation phase or the parameters $(\alpha_{s,t}, \Delta_{s,t})$ in the permanent phase. He also selects the slots containing hidden information as their corresponding $WNR$ is higher than $WNR_T$. But, Bob's estimations are not exactly what Alice has estimated at the embedding stage leading to some detection errors. Bob's estimations are not enough accurate because Eq. (14) is true on expectation, i.e. asymptotically when $N$ goes larger. Experiences show that a set of 160 coefficients is not large enough. But, if we gather more coefficients in a slot decreasing the figure of subbands or increasing the figure of consecutive

analysis windows, then, the spectrum of the watermark signal is not shaped like the cover signal's one. We prefer to keep a small granularity in time and frequency domains even if it leads to a mismatch between Alice's and Bob's estimations.

We can not use any e.c.c. in the initialisation phase, as it will structure the decoded data and arouse Wendy's suspicion. A possibility is to embed cipher-texts corresponding to the seed $L$ attended with a random word $R$ : $m = \text{Encrypt}_{PubKey_{Bob}}([L, R])$. This alea ensures that cipher-texts are independent. This is done several times, so that it leads to a repetition e.c.c. If an error occurs in this block, it is not possible to retrieve the original message. Bob decodes and decrypts to retrieve some messages (throwing away the random words). He chooses the one he retrieves several times (at least twice) as the most likely message. This e.c.c. can not prevent a lot of transmission errors. It works if at least two cipher-texts are transmitted with no error. A smart alternative is to use a crypto-system that bears some communications errors like the Mc Eliece public-key crypto-system.[10] Yet, the size of the public and private keys is so big that this crypto-system is never used in practice.

The last possibility is to create a feedback loop. When Alice embedded a message in a slot $(s, t)$, she check if Bob will correctly decode all the symbols (the channel noise is not taken into account for the moment). Indeed, as Bob does not retrieve exactly the same compressor function (initialisation phase) or the same SCS parameters (permanent phase), errors occur only with large amplitude coefficients. The embedding in these coefficients is processed again based, this time, on Bob's power estimation. As this is done on very few coefficients, the power estimation is almost not changed and this does not spoil the security level. This loop can be done as long as errors are detected. It always converges in practice although we can not prove it theoretically.

The channel noise is now added to the stego-content. In average, one out of 200 coefficients are not detected properly. We create another feedback loop to correct these errors. The only thing we can do is to move these coefficients towards the center of the detection bin. Because this is done on very few coefficients, this does not change the global pdf and it does not spoil our security level. The two loops are finally interlaced. We can not prove why but experiments show that it converges. The main point is that the convergence is experimentally assessed thanks to the spike models strategy, embedding symbols within slots whose cover-content power is high enough to provide a good channel. There is no need of channel coding thanks to a non-causal embedding. Yet, we did not prove this strategy provides a better capacity-security trade-off than embedding in all slots using dedicated e.c.c.

### 5.5. experimental performances assessment

As the initialisation phase lasts a few slots, the experimental performances only address the permanent phase. We plot the bit-rate against KL distance curve. The bit-rate is more meaningful expressed as the average ratio of the figure of hidden bits to the number of PCM samples. The KL distance is not easy to measure for non-stationary colored processes. We assume these processes are stationary during some consecutive analysis windows and the DCT renders r.v. from different subbands independent. We choose a Laplacian distribution with a variance calculated from the coefficients of the cover-content. Hence, the KL distance measures changes of pdf and of variance. Obviously, Wendy can not do this measure because she can not estimate the original variance to compare to the observed one. Hence, our measures of the KL distance up-bound what Wendy is able to distinguish. Fig. 5 shows our results for two different kinds of contents and for different $WCR$. For a given KL distance, more bits are embeeded in a 'rock' style content than in a 'violin' piece of music. The 'violin' content is a narrow-band signal so that $WNR_{(s,t)}$ is greater than $WNR_T$ in fewer slots.

We plot the frontier between non-perceptible and perceptible embedding. This subjective test is made via a comparison between a stego-content and its cover-content. We take precautions imposing to ourself a stronger challenge than Wendy can take up.

Finally, we embed safely and reliably 2 bits by PCM sample, i.e. 64MB in a CD-audio of 640MB.

## 6. CONCLUSION

The Scalar Costa Scheme is an extremely powerful data hiding technique. It achieves very good performances in the framework of passive steganography: a high bit-rate for a fair security level. We derive a practical

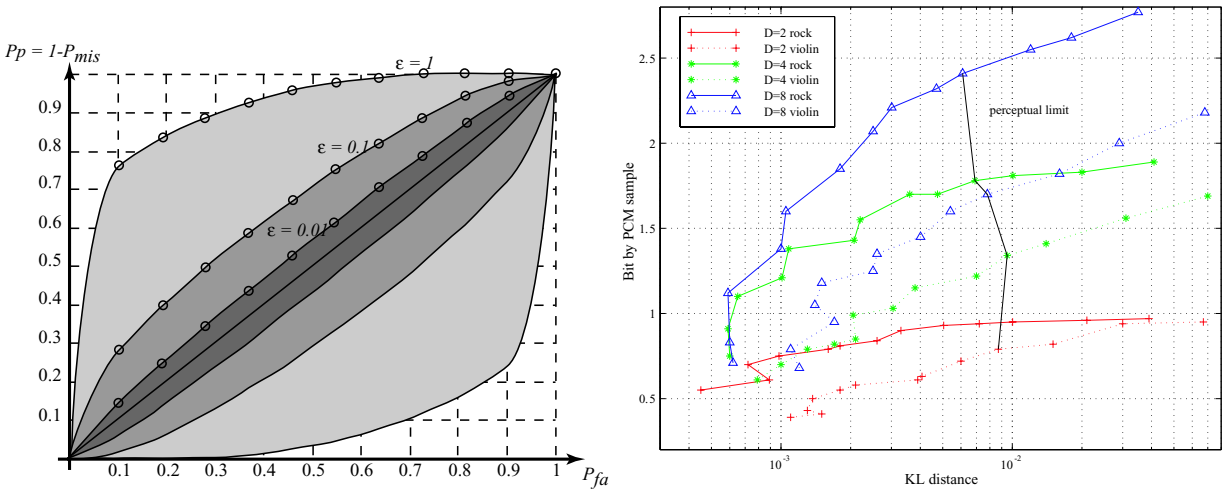


**Figure 5.** a) ROC curve. Wendy's test operating point $(P_{fa}, P_p)$ belongs to an area as narrow as the KL distance is small. b) Bit-rate against security level. An approximation of the average performances for two kinds of contents.

public-key stego-system following Anderson criterion. Our analysis of the security level is driven by Cachin's criterion based on a KL distance. In the initialisation phase and in the permanent phase, this distance can be up-bounded. Experimentally, the distance measured is not the Cachin's criterion but a bigger estimation. In the same way, subjective watermark perception tests have been done comparing the original and the stego-audio clip. This provides stronger constraints yet a more secure system.

## REFERENCES


1. J. Su, J. Eggers, and B. Girod, "Channel coding and rate distortion with side information: Geometric interpretation and illustration of duality," *IEEE trans. on Information Theory* , 2001. accepted.
2. J. Eggers, J. Su, and B. Girod, "Robustness of a blind image watermarking scheme," in *Proc. of Int. Conf. on Image Processing*, IEEE, (Vancouver, Canada), Sept. 2000.
3. J. Eggers, J. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," in *IEE Colloquium: SECURE IMAGES AND IMAGE AUTHENTICATION*, (London, UK), Apr. 2000.
4. P. Moulin, "The role of information theory in watermarking and its application to image watermarking," *Signal Processing* **81**, pp. 1121–1139, 2001.
5. J. Chou, S. Pradhan, and K. Ramchandran, "On the duality between data hiding and distributed source coding," in *Proc. 33rd ann. Asilomar conf. on signals, systems, and computers*, (Asilomar, USA), 1999.
6. B. Chen, *Design and analysis of digital watermarking, information embedding, and data hiding systems*. PhD thesis, Massachusetts Institute of Technology, 2000.
7. M. Costa, "Writing on dirty paper," *IEEE Trans. on Information theory* **29**, May 1983.
8. R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE Journal of Selected Areas in Communications* **16**, pp. 474–481, May 1998. Special issue on copyright & privacy protection.
9. R. Gray and D.L. Neuhoff, "Quantization," *IEEE Trans. on information theory* **44**, Oct. 1998.
10. A. Menezes, P. Van Oorschot, and S. Vanstone, *Handbook of applied cryptography*, Discrete mathematics and its applications, CRC Press, 1996.
11. J. Kelsey, B. Schneier, and N. Ferguson, "Yarrow-160: Notes on the design and analysis of the yarrow crytographic pseudorandom number generator," in *6th Ann. Work. on Sel. Areas in Cryptography*, 1999.
12. R. Blahut, *Principes and practice of information theory*, Addisson-Wesley, 1987.