

On the Achievable Rate of Side Informed Embedding Techniques With Steganographic Constraints

Mark T. Hogan, Félix Balado, Neil J. Hurley, and Guéno   C.M. Silvestre.

University College Dublin, Belfield, Dublin 4, Ireland.
{markhogan, fiz, neil.hurley, guenole.silvestre}@ihl.ucd.ie

Abstract. The development of watermarking schemes in the literature is generally guided by a power constraint on the watermark to be embedded into the host. In a steganographic framework there is an additional constraint on the embedding procedure. It states that, for a scheme to be undetectable by statistical means, the pdf of the host signal must be approximately or exactly equal to that of the stegotext. In this work we examine this additional constraint when coupled with DC-DM. An analysis of the embedding scheme Stochastic QIM, which automatically meets the condition under certain assumptions, is presented and finally the capacity of the steganographic channel is examined.

1 Introduction

The term steganography refers to the family of techniques used to hide data within a *host* multimedia signal. Ideally, the corresponding modified signal, referred to as a *stegotext*, is perceptually and statistically indistinguishable from the host. The classical representation of steganographic communication is given by the prisoners' problem [1]. Alice produces a stegotext using the message that she wants to communicate and a given host, and sends it to Bob through an insecure communications channel. Usually, Alice and Bob make use of secret keys for their covert communication. The warden Wendy monitors the channel between Alice and Bob, and performs a detection test to decide if the signal being sent includes hidden information by exploiting potential imperfections of the steganographic method used. In an analogous way to cryptanalysis, this detection procedure is known as *steganalysis*.

Some considerations on the nature of Wendy's tests are necessary. Typically Wendy can be either *passive* or *active*. If the warden is passive then a detection test is all that is performed, but on the other hand, if she is active, then the document is deliberately attacked regardless of the outcome of any detection test. In this work we consider only that Wendy is passive. We also assume that the transmitted document undergoes a channel distortion before it is decoded. The nature of the channel is taken to be additive white Gaussian noise (AWGN), for the sake of comparison with previous results.

The success of detection tests lies in the location of statistical differences between the host signal and the stegotext signal. This idea has been formalised by Cachin in [2], where the security of a steganographic embedding method

has been defined in terms of the Kullback Leibler distance (D_{KL}) between the densities of the host and stegotext signals. D_{KL} is equal to zero iff the two distributions are the same. The implication is that a non-negligible value for D_{KL} for any embedding scheme leads to detectable statistical differences. A major goal of embedding is, therefore, to keep D_{KL} as low as possible, such that the communication passes unhindered.

We now specify two cases of steganographic communication, namely *perfect* and *non-perfect* steganography. The difference lies in the restriction placed on the encoder. If the restriction is such that the $D_{\text{KL}} = 0$ between the host and stegotext densities then the embedding scheme is said to conform to perfect steganography. In this case optimal statistical steganalysis will always be have a probability of error, P_e , no better than 0.5. If a small value for D_{KL} is allowed, such that the results of practical statistical tests are unreliable, then we have non-perfect steganography (c.f. ϵ -secure steganography in [2]).

Statistical differences are, of course, not the only concern when designing embedding schemes. As in the related area of watermarking, it is also desired that the rate of communication be as high as possible. Consider for a moment the case where the host is zero-mean independent identically distributed (iid) Gaussian random vector and the channel noise consists of two sources of additive white Gaussian noise (AWGN), both mutually independent. Assuming power constrained codewords and given knowledge of one of the noise sources at the encoder, the capacity of the channel can be achieved with Costa's codebook [3]. Now, to the authors knowledge, the work of Moulin and Wang in [4], is the only previous work in which the capacity of steganographic communication scenarios is rigorously examined. One of the main starting points in this work is that, for capacity calculations, in addition to the usual power constraint, there is also a pdf constraint which, for $D_{\text{KL}} = 0$ (i.e. perfect steganography), requires that the pdf of the codebook must be equal to the pdf of the host signal.

For the case in which no secret key is used at the encoder, this result implies that Costa's codebook is not suitable for the steganographic channel because the codewords, although Gaussian, are discretely distributed. It can be argued that, if the codewords are unknown to the detector, scaled correctly and only used once, then the system will be perfectly secure. However, if the codewords are used more than once, as is the case in practical methods, then a detector can be designed to exploit this fact and the security is lost.

Considering only the power constraint at the encoder, it has been shown that distortion compensated dither modulation (DC-DM) [5] (or equivalently, the scalar Costa scheme [6]) has an achievable rate close to the capacity of the side informed AWGN channel. It was also shown in [7] that DC-DM can never conform to the restrictions of perfect steganography. However, it is well known that DC-DM requires the optimisation of a parameter, α , for a given noise power over the channel. This parameter can also be tuned for the purposes of reducing the value of D_{KL} , such that non-perfect steganography is still possible. Several authors have adopted this approach in the past [8],[9]. In those works, assuming the key to be leaked to an attacker, the value of α was taken to be 0.5, such

that the D_{KL} is kept as low as possible while allowing errorless communication in the absence of any channel noise. Here, in the case where the key has not been leaked, we will show that this is not necessarily the best value. We indicate, using Stein’s lemma [10], the optimal value for Alice to choose, such that practical statistical tests have a probability of error, P_e , close to 0.5, and the rate of communication is simultaneously maximised over the AWGN channel. We will also use Stein’s lemma to show the penalty in capacity incurred when a non-perfect steganographic constraint is coupled with the power constraint at the encoder.

Given that DC-DM is an approximation to Costa’s discrete codebook, the question arises of whether or not there exists an analogous codebook for the steganographic channel. A promising scheme is that of stochastic quantization index modulation (SQIM), proposed by Wang and Moulin [8]. Unlike DC-DM where the codewords are fixed, SQIM uses non-fixed codewords to improve security. Subject to the flat host assumption, the codebook is then formed from the same density as the host pdf and thus statistical steganalysis on SQIM will fail. Here we present an analysis of SQIM and illustrate its performance compared to that of dither modulation (DM) [5]. We will also show that the achievable rate of scalar SQIM is upper bounded by that of DM.

The essential feature of SQIM, from a perfect steganography point of view, is that every point in the host signal space forms an allowable codeword with probability given according to the host signal pdf. Finally, we extend this philosophy to higher dimensions using a sphere packing argument.

The paper is organised as follows. Section 2 is devoted to setting out the problem and the notation we adopt. An analysis of DC-DM with non-perfect steganographic constraints at the encoder is presented in Sect. 3. Section 4 contains an analysis of SQIM and the capacity of perfect steganographic channels is addressed in Sect. 5. Finally, conclusions are drawn in Sect. 6.

2 Problem Set-up

Notation and Preliminaries. In this work capital letters refer to random variables and vectors respectively, e.g. X , \mathbf{X} , with lower case letters the respective realisations, e.g. x , \mathbf{x} . Individual elements of \mathbf{x} are indexed as x_j . All vectors are of length N . The probability density function (pdf) of a random variable X is denoted as $f_X(\cdot)$ and the corresponding cumulative density function as $F_X(\cdot)$. The statistical expectation of X is denoted $E_X\{X\}$ and the differential entropy of X is denoted as $H(X) = -\int f_X(x) \log f_X(x) dx$. The mutual information between X and Y is denoted $I(X; Y) = H(X) - H(X|Y)$.

We assume that the host, $\mathbf{x} = [x_1, \dots, x_N]$, consists of a realization of a random vector \mathbf{X} formed by independent identically distributed (iid), Gaussian zero-mean random variables for both ease of comparison with previous works, and reasons of analytic tractability. Alice may send either \mathbf{x} to Bob, or modify it before transmission to embed $\mathbf{b} = [b_1, \dots, b_N]$, $b_i \in \mathcal{B}$, where in experiments we take $|\mathcal{B}| = 2$, giving a sequence of binary digits drawn from a uniform distribution. This produces a stegotext (watermarked) vector $\mathbf{s} = G(\mathbf{x}, \mathbf{b})$, and the watermark \mathbf{w} is then given as $\mathbf{w} \triangleq \mathbf{s} - \mathbf{x}$. We assume that only one information

symbol b_j is embedded in one corresponding covertext sample x_j . The embedding process may be secured by using a pseudorandom symmetric key \mathbf{k} , shared by Alice and Bob, and then $\mathbf{s} = G_{\mathbf{k}}(\mathbf{x}, \mathbf{b})$.

Two important parameters for establishing the working point of the steganographic method are the *host to watermark* power ratio (HWR) and the *watermark to noise* power ratio (WNR). The HWR is the average power of the host normalized by the watermark power, which can be written as $\text{HWR} \triangleq E\{\|\mathbf{X}\|^2\}/E\{\|\mathbf{W}\|^2\} = \sigma_X^2/\sigma_W^2$, where σ_X^2 and σ_W^2 refer to the variances of the host signal and watermark, respectively, assuming that W has zero mean. If X and S are independent and zero-mean then $\sigma_W^2 = \sigma_S^2 - \sigma_X^2$. Notice that the perceptual constraints in any data hiding problem impose very high values for the HWR.

The channel noise $\mathbf{v} = [v_1, \dots, v_N]$, is assumed to be AWGN with power σ_V^2 such that the received vector $\mathbf{y} = \mathbf{x} + \mathbf{w} + \mathbf{v}$. Correspondingly, the WNR is defined as the watermark power, normalized by the noise power and is written as, $\text{WNR} \triangleq E\{\|\mathbf{W}\|^2\}/E\{\|\mathbf{V}\|^2\} = \sigma_W^2/\sigma_V^2$.

2.1 Detection Test

Alice transmits either \mathbf{x} or \mathbf{s} . Because Wendy does not know the origin of the document she receives, she can only assume it to be an unclassified document \mathbf{z} . She must decide if \mathbf{z} sent to Bob by Alice has been drawn either from $f_{\mathbf{X}}$ or from $f_{\mathbf{S}}$. Assuming that $f_{\mathbf{X}}$ is known, then, given $G(\cdot)$, $f_{\mathbf{S}}$ is also known. This detection problem is then a hypothesis testing problem with two choices, denoted as the null hypothesis H_0 (\mathbf{z} is a host), and the alternative hypothesis, H_1 (\mathbf{z} is a stegotext). To make a decision on \mathbf{z} , the optimal test based on the Bayes likelihood ratio [11], and is given by

$$\Lambda(\mathbf{z}) \triangleq \frac{f_{\mathbf{X}}(\mathbf{z})}{f_{\mathbf{S}}(\mathbf{z})} \underset{H_1}{\overset{H_0}{\geq}} \mu, \quad (1)$$

where $\mu \triangleq (P_0/P_1) \cdot ((C_{10} - C_{00})/(C_{01} - C_{11}))$. The P_i , $i \in \{0, 1\}$ represent the *a priori* probabilities for the null and alternative hypotheses respectively, and C_{ij} the cost of choosing H_i when the true hypothesis is H_j . Letting $C_{ij} = \delta_{ij}$, with δ_{ij} the Kronecker delta function, and choosing the *a priori* probabilities to be uniformly distributed, gives the maximum likelihood (ML) test.

It is desirable to relate the predicted outcome of (1) to some performance property of the embedding process which is directly measurable. One such interesting property is the probability of error in the detection test, P_e , defined as

$$P_e \triangleq P(\mathbf{Z} \sim f_{\mathbf{X}} | \Lambda(\mathbf{Z}) < \mu) \cdot P_0 + P(\mathbf{Z} \sim f_{\mathbf{S}} | \Lambda(\mathbf{Z}) > \mu) \cdot P_1 = \frac{P_{fa} + P_m}{2}, \quad (2)$$

where $\mathbf{Z} \sim f_{\mathbf{X}}$ is taken to mean that the random vector \mathbf{Z} follows $f_{\mathbf{X}}$, P_{fa} represents the probability of false alarm and P_m , the probability of a miss. We have again assumed that the *a priori* probabilities are uniform as this is worst

case for Wendy. A quantity to relate the P_e and the embedding process, as we will see, is the Kullback-Leibler distance which is defined, in one direction, as

$$D_{\text{KL}}(f_{\mathbf{X}}\|f_{\mathbf{S}}) \triangleq \int f_{\mathbf{X}}(\mathbf{x}) \log \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{S}}(\mathbf{x})} d\mathbf{x}. \quad (3)$$

In general $D_{\text{KL}}(f_{\mathbf{X}}\|f_{\mathbf{S}}) \neq D_{\text{KL}}(f_{\mathbf{S}}\|f_{\mathbf{X}})$ and $D_{\text{KL}}(f_{\mathbf{X}}\|f_{\mathbf{S}}) = 0$ iff $f_{\mathbf{X}} = f_{\mathbf{S}}$.

In [8] the sum of the Kullback Leibler distances in both directions (the J -divergence) was used to lower bound the error probability in Wendy's test. Here, we take a different approach through the use of Stein's lemma [10], which we review next. In [10] the lemma applies to discrete random variables but here we use a direct translation to a continuous domain. Using this lemma the P_e and D_{KL} can be directly related to one another. Firstly let $P_{fa} = \int_{A^c} f_{\mathbf{X}}(\mathbf{z}) d\mathbf{z}$ and $P_m = \int_A f_{\mathbf{S}}(\mathbf{z}) d\mathbf{z}$, where $A \subseteq \mathbb{R}^N$ is an acceptance region for H_0 and A^c its complement. Then for $0 < \nu < 0.5$, let $P_{fa}^\nu = \min_{\substack{A \subseteq \mathbb{R}^N \\ P_m < \nu}} P_{fa}$, which leads to

$$\lim_{\nu \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \log(P_{fa}^\nu) = -D_{\text{KL}}(f_{\mathbf{X}}\|f_{\mathbf{S}}). \quad (4)$$

This gives a direct relationship between the errors in detection of (1) and D_{KL} in the limit as N approaches infinity. Alice can readily monitor D_{KL} and as such can approximately predict the outcome of any optimal detection test on the documents she transmits. Hence suitable parameters can be chosen for the embedding scheme such that the P_e in Wendy's detection test will be close to 0.5. This relationship only holds true in the limit but, here, for practical purposes, we assume that N is large and take (4) to approximately hold in this case. It should also be noted that, in inverting (4), a slowly varying function is required which depends on, among other parameters, the P_m [12]. However, in general, this function is not known and as in [12] we will ignore this function for the purposes of simplicity.

2.2 Capacity

For the purposes of this discussion assume that Alice wishes to embed a message in \mathbf{x} and transmits \mathbf{s} . Ignoring the detection test, this communication scenario is the standard watermarking channel, see e.g. [6]. It has been noted in [3] that the capacity of this channel is given as $C = \max_{f_{U,Y}(u,y|x)} I(U; Y) - I(U; X)$, where U is an auxiliary random variable. Achieving this capacity is then a problem of choosing a suitable codebook U . For the power constrained case where the host signal is iid Gaussian, and the channel distortion is represented as AWGN, Costa [3] showed that, for $U = W + \alpha X$ and $\alpha = \sigma_W^2 / (\sigma_W^2 + \sigma_V^2)$, the capacity of the channel is given as

$$C = \frac{1}{2} \log \left(1 + \frac{\sigma_W^2}{\sigma_V^2} \right). \quad (5)$$

An alternative geometrical approach to calculating the capacity of this channel is contained in [13], which we summarise here because it will be drawn upon

in Sect. 5. Noting that N is large consider the following. Around any given \mathbf{x} , there is an allowable distortion (embedding power) sphere, denoted T_W , with radius given as $\sqrt{N\sigma_W^2}$, which must contain at least one codeword corresponding to each possible message. For a given achievable rate, R , we then wish to position the 2^{NR} codewords within the sphere such that the noise spheres, T_V , of radius $\sqrt{N\sigma_V^2}$ around each codeword have asymptotically vanishing overlap as $N \rightarrow \infty$. To calculate the achievable rate, a ratio of volumes is formed which gives

$$2^{NR} \leq \frac{(N(\sigma_W^2 + \sigma_V^2))^{\frac{N}{2}}}{(N\sigma_V^2)^{\frac{N}{2}}}. \quad (6)$$

A capacity achieving code will achieve this rate for $R = C$ from which it can be seen that (6) reduces to (5).

Finally, for a given coding scheme, host pdf and channel, the achievable rate of communication can be calculated as the following [10],

$$R = I(Y; B) = H(Y) - \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} H(Y|b). \quad (7)$$

3 DC-DM

In this section we analyse DC-DM in respect of the constraints imposed by steganography. Firstly a brief review of DC-DM is presented. Then the achievable rate of DC-DM with steganographic constraints at the encoder is examined.

3.1 Embedding Method

DC-DM with uniform scalar quantizers is a practical implementation of distortion-compensated quantization index modulation (DC-QIM), proposed by Chen and Wornell [5]. It has been shown that, for the AWGN channel with side information at the encoder, DC-DM has an achievable rate acceptably close to the capacity of the channel [6]. The embedding technique is based on the quantization of the host samples with a dithered version of a uniform scalar quantizer $Q_\Delta(\cdot)$. We assume that the quantization step Δ is the same for all covertext samples. For example, in the case of binary messages the embedding takes place with two quantizers shifted by $\Delta/2$. In order to embed a binary symbol b_j at the host sample x_j the corresponding stegotext sample s_j is obtained in DC-DM as

$$s_j = G_{k_j}(x_j, b_j) = x_j + \alpha \left[Q_\Delta \left(x_j - k_j - \Delta \frac{b_j}{2} \right) + k_j + \Delta \frac{b_j}{2} - x_j \right], \quad (8)$$

where the additional dither $k_j \in (-\frac{\Delta}{2}, \frac{\Delta}{2}]$ is the secret key shared by Alice and Bob at the j th sample. The distortion compensation factor $0 < \alpha \leq 1$ allows for tuning the method for optimal robustness to channel noise, assuming that its power is known in advance [6], or alternatively, for tuning its detectability

properties [9],[8], as we will discuss in Sect. 3.2. DM is a particular case of DC-DM for which $\alpha = 1$.

Assuming that the quantization error is approximately uniform and independent from \mathbf{X} , the HWR is for this embedding method is given by

$$\text{HWR} = \frac{12\sigma_X^2}{\alpha^2 \Delta^2}. \quad (9)$$

3.2 Optimal DC-DM Detection

In previous works [7] the optimal detection of DC-DM was presented in the presence and absence of a secret key. Here we limit ourselves to the case of DC-DM in which the secret key has not been leaked to Wendy, as this is the more pertinent case for analysis. In terms of the achievable rate of a particular scheme, the use of a key has no effect, but, of course, the knowledge of the key has implications for the detection of stegotexts. Considering that the key is unavailable to Wendy, then, for her ML detection test she may use the average expression for the pdf, taken over all possible keys. This approach comes down to computing $\tilde{f}_S(s) = E_K\{f(s|K)\} = \int f_S(s|k) \cdot f_K(k) dk$, and to use this average pdf, \tilde{f}_S , in (1). The result for \tilde{f}_S in the case of DC-DM, for a uniformly distributed key variable $K \sim U(-\Delta/2, \Delta/2)$, is as follows (details in [7]),

$$\tilde{f}_S(s) = f_X(s) * U\left(-\frac{\alpha\Delta}{2}, \frac{\alpha\Delta}{2}\right), \quad (10)$$

where $*$ denotes convolution. Noteworthy here, is the fact that (10) illustrates that perfect steganography is never possible using DC-DM. The only case for which $\tilde{f}_S = f_X$ is when α or Δ is set to zero. In either case no embedding takes place and the achievable rate is consequently zero.

3.3 Achievable Rate of DC-DM

It was noted in [7] that for a fixed HWR the choice of α is irrelevant in respect of the secrecy of the communication. It can be seen from (9) that the HWR is directly dependent on the product of α and Δ . Also, \tilde{f}_S from (10), is directly related to the same product. Thus the actual value of α does not matter in respect of the secrecy of the communication. Then, assuming a given performance constraint, an optimal α can be picked for a given channel noise power.

Now, considering the case of fixed Δ it can be seen that the previous rationale is no longer true. Other authors [8],[9], assuming a fixed Δ and that the attacker has access to \mathbf{k} , have proposed the value of $\alpha = 0.5$. This particular value allows for errorless communication in the absence of noise and also has the property that the pdf of the stegotext has full support over \mathbb{R} , assuming that X also does. We consider the case where \mathbf{k} has not been leaked to Wendy and will show that $\alpha = 0.5$ is not necessarily the optimal value in this case.

Now, we will fix the value of Δ and use Stein's lemma to set a limit, α_{\max} , on the maximum value of α such that, if $\alpha \in (0, \alpha_{\max}]$ is used at the encoder,

the P_e in the detection test will be close to 0.5. Firstly, $D_{\text{KL}}(f_X||f_S)$ is calculated using f_X and (10) for all $\alpha \in (0, 1]$ and substituted into (4) to give a value for the P_{fa} as a function of α . These values are then substituted into (2) to give the P_e as a function of α .

It can be seen that in (4), by reversing the probabilities and correspondingly changing the pdfs, the theorem remains essentially unchanged. However the limit now depends on $D_{\text{KL}}(f_S||f_X)$ with the result that the final probability of error may be different. As such, this case is also calculated and the final P_e is taken as the minimum of the two results at each α , as this represents worst case for Alice. Finally α_{max} is chosen according to

$$\alpha_{\text{max}} = \max_{P_e(\alpha) \geq (0.5-\epsilon)} \alpha, \quad (11)$$

where ϵ is an arbitrarily small number. Now, due to the nature of the DC-DM transformation an analytic expression is unavailable for D_{KL} in both directions and the results are only available by numerical evaluation.

The limited range of α values, $\alpha \in (0, \alpha_{\text{max}}]$, is then used to find the constrained achievable rate of DC-DM according to

$$R = \arg \max_{\alpha \in (0, \alpha_{\text{max}}]} I(B; Y|k). \quad (12)$$

This rate is calculated by substituting the pdfs $f_Y(y|k)$, $f_Y(y|k, b = 0)$ and $f_Y(y|k, b = 1)$ into (12). The derivation of these pdfs is performed using (8) and the change of variable theorem, [14] as follows. Assume that message $b \in \mathcal{B}$, corresponds to the reconstruction points denoted as $q_{i,b} = i\Delta + b\Delta/|\mathcal{B}|$ for suitable $i \in \{-\infty, \infty\}$. Then for $x \in (q_{i,b} - \Delta/2, q_{i,b} + \Delta/2]$ we have that $s = x + \alpha(q_{i,b} - x)$. Using the theorem we obtain the following, $f_{S_{i,b}}(s) = (1/(1-\alpha)) \cdot f_X((s_{i,b} - q_{i,b})/(1-\alpha))$, for $s \in (q_{i,b} - (1-\alpha)\Delta/2, q_{i,b} + (1-\alpha)\Delta/2]$. The dependence on the bin can then be removed by summing over i , giving $f_{S_b}(s) = \sum_{i=-\infty}^{\infty} f_{S_{i,b}}(s)$. Finally we can remove the dependence on b by averaging over the message alphabet (uniformity assumption) and then, $f_S(s) = (1/|\mathcal{B}|) \cdot \sum_{b \in \mathcal{B}} f_{S_b}(s)$. Then it can be seen that $f_Y(y|k) = f_S(y) * f_V(y)$. The conditional pdfs follow easily. Now, these pdfs are such that an analytic expression for R is not available so a numerical evaluation of (12) is performed.

For illustration purposes we take $\Delta = 1$ and assume that $N = 10^5$ for the results. A higher value of N with a corresponding lower value of Δ will lead to similar results. Now consider Fig. 1. Here the achievable rate of DC-DM is plotted as a function of $\alpha \in (0, 1]$ for a range of WNRs. The equivalent value of P_e as a function of α is plotted on the x -axis. First examine the cases of high WNRs. Here it can be seen that the highest value of R is achieved when the P_e is approaching 0. This implies that at these values of WNR there is a significant loss in the achievable rate of the scheme due to the steganographic constraint, as the optimum α in terms of rate cannot be used. A sub-optimal value must be used lowering the rate of the scheme. However, when the low WNRs are examined it can be seen that the maximum value of R is obtained within the region where Wendy's detection test will have a probability of error

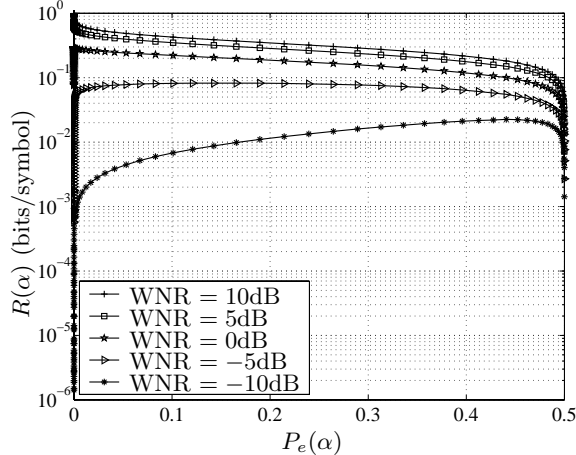


Fig. 1. $R(\alpha)$ for DC-DM plotted against $P_e(\alpha)$ in Wendy’s detection test for a range on WNRs. $\sigma_X^2 = 1.0$, $\Delta = 1.0$, $\alpha \in (0, 1]$, $N = 10^5$, HWR varies with α .

close to 0.5. Now, assuming that a given $P_e > 0.5 - \epsilon$ is acceptable to Alice, it can be observed that there will be no loss in the achievable rate due to the steganographic constraint, and choosing the optimal value of α based on this rate will not allow any significant advantage in the detection test.

Fig. 2 shows plots of (12) alongside the value of R for unconstrained values of α , as discussed above. There is an evident loss in the achievable rate in the high WNRs whereas at low WNRs the rate is equal for both cases. This is due to the fact that at very low WNRs the optimal value of α is close to zero while it approaches 1 as the WNR increases. The increasing α increases the D_{KL} eventually passing the threshold set as α_{max} .

Finally, in Fig. 3 the optimal values of α (i.e. α^*) are plotted for a number of scenarios. Firstly Costa’s α [3] is plotted, alongside the α numerically optimised for DC-DM and finally the α which maximises (12).

4 Stochastic QIM

We have seen in Sect. 3 that the achievable rate of the embedding method is reduced under steganographic conditions. Recently however a data hiding scheme has been proposed which approximately conforms to the perfect steganographic channel. Stochastic QIM is a side informed embedding technique [8] which, under the flat host assumption, maintains $f_S = f_X$. The main idea is that every $x \in \mathbb{R}$ forms a valid codeword s , with probability drawn from f_X . This is of course unlike DC-DM where only certain subsets of \mathbb{R} form the codewords with probability determined by the transformation $G(\cdot)$. In this section an analysis of SQIM is presented. Then the achievable rate of the coding scheme is analysed more closely than previously [8]. This analysis is performed under the assumption that $\mathbf{k} = \mathbf{0}$.

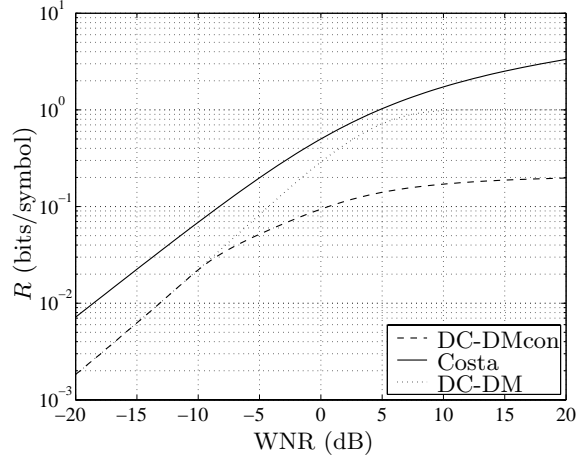


Fig. 2. Achievable rate R for binary, scalar DC-DM under two scenarios. The first plot (DC-DM) gives R , maximised for all $\alpha \in (0, 1]$ while the second (DC-DMcon) gives the constrained value of R maximized over $\alpha \in (0, \alpha_{\max}]$, where the value of ϵ in (11) is 0.05. The channel capacity (Costa) is plotted for reference.

4.1 Analysis of SQIM Watermark

The binary one dimensional SQIM embedding scheme can be summarised as follows. The host space \mathbb{R} is tiled into disjoint regions of length $\Delta/2$. Each region contains codewords corresponding to either $b = 0$ or $b = 1$. Let the union of all the regions of \mathbb{R} corresponding to $b = 0$ be denoted A_0 and similarly for A_1 . Now assume that $x \in A_0$ and that $b = 0$. In this case x already forms a required codeword so $s = x$.

Now again assume that $b = 0$ but that this time $x \in A_1$. Then s is formed as follows. Firstly the nearest correct code region to x (in a Euclidean sense) is chosen. Then s is chosen randomly from this region with probability given by the host pdf truncated to said region and scaled accordingly. Thus $w = s - x$ is the watermark in this case. Now assuming that x has equal probability of lying in either A_0 or A_1 we have, with probability, $P(x \in A_0, b = 0) + P(x \in A_1, b = 1) = P(x \in A_0)P(b = 0) + P(x \in A_1)P(b = 1) = 0.5$, that $s = x$ and with probability 0.5 a watermark w is added to x to form s .

We must now consider the effect of the above embedding on the pdf of the stegotext. Without loss of generality consider a generic quantization point $i\Delta$ with corresponding decision region $(i\Delta - \Delta/4, i\Delta + \Delta/4]$. It is clear that, for equiprobable symbols, the weight of $f_S(s)$ over this region should be

$$a_i \triangleq P\left(x \in \left(i\Delta - \frac{\Delta}{4}, i\Delta + \frac{\Delta}{4}\right]\right) = F_X\left(i\Delta + \frac{\Delta}{4}\right) - F_X\left(i\Delta - \frac{\Delta}{4}\right). \quad (13)$$

This weight is composed of three components, namely a portion equal to $a_i/2$ formed from host points falling in the region already associated with the correct corresponding message bit (i.e. $s = x$) and two other portions formed by trans-

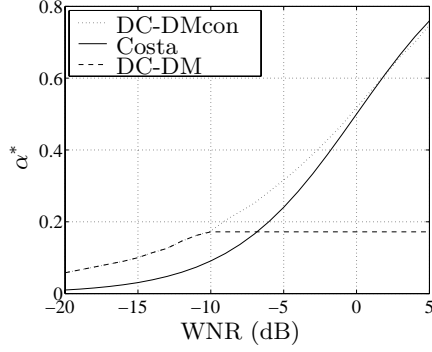


Fig. 3. α^* for DC-DM without steganographic constraints (DC-DM) and the optimal value constrained by Stein's lemma (DC-DMcon). Costa's α [3], is plotted for reference. $\sigma_X^2 = 1.0$, $\Delta = 1.0$, $\epsilon = 0.05$.

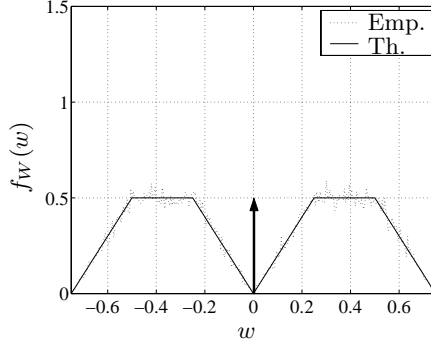


Fig. 4. The empirical histogram of the SQIM watermark (Emp.) plotted alongside the approximation to the derived pdf (Th.). $\Delta = 1$.

formations from the adjacent decision regions, which, it can be seen, are equal in expectation to $a_{i-1}/4$ and $a_{i+1}/4$.

We now have that for the embedding to be perfect, the following must hold, $a_i = a_{i-1}/2 + a_{i+1}/2 \forall i \in \mathbb{Z}$. In general, for a given f_X this is not the case with the result that $D_{\text{KL}} > 0$. If f_X is uniform then the problem is avoided but this, of course, is a generally not the case. Now, under the flat host assumption the difference in weights between adjacent bins is zero and this holds as a good approximation if $\sigma_X^2 \gg \sigma_W^2$. As such we adopt this approximation in all further analysis with the result that $f_S(s) = f_X(s)$.

In practice the problem can be circumvented by calculating a_i for each bin. When the normalised number of samples in the bin reaches this amount any remaining points must be either embedded in a bin further away (increasing the embedding distortion) or embedded with the wrong bit (increasing the errors in the communication). Either option results in a loss in achievable rate for the scheme. Hence the flat host assumption gives an upper bound for calculation of the rate of this steganographic method.

Now, to analyse this watermark we first assume that x is not lying in the correct region for the corresponding b . The embedding can be considered as a standard DM embedding with an additional stochastic element which we denote as D . We can therefore write the following transformation for the j th sample, s_j ,

$$s_j = x_j + \left(Q_\Delta \left(x_j - \Delta \frac{b_j}{|\mathcal{B}|} \right) - x_j + \Delta \frac{b_j}{|\mathcal{B}|} \right) + d_j, \quad (14)$$

where d_j has support range $(-\Delta/4, \Delta/4]$ and where we assume from here on that, without loss of generality, $b_j = 0$. Let $Q_\Delta(x) = i\Delta$ with the appropriate value of i , and the probability of the host lying in the region around this reconstruction

point be given as a_i , from (13). Then the pdf of D can be seen to be

$$f_D(d) = \frac{1}{a_i} \cdot f_X(i\Delta + d), \quad d \in \left[-\frac{\Delta}{4}, \frac{\Delta}{4}\right). \quad (15)$$

Next we have the quantization error, $e = Q_\Delta(x) - x$, usually taken to be uniform over a quantization bin. Here however we have a slightly different scenario. It has been noted that quantization only takes place if $\frac{\Delta}{2} \geq |x - i\Delta| > \frac{\Delta}{4}$. Then the pdf of the quantization error is given as the following,

$$f_E(e) = \begin{cases} \frac{2}{\Delta}, & e \in \begin{cases} [i\Delta - \Delta/2, i\Delta - \Delta/4) \\ (i\Delta + \Delta/4, i\Delta + \Delta/2] \end{cases} \\ 0, & \text{otherwise.} \end{cases}$$

Given that the quantization error is independent of D , this portion of the watermark pdf is given as $f_E(w) * f_D(w)$. This calculation is straightforward but the resulting pdf depends on the absolute value of x . However, the effect is minimal if $\Delta^2 \ll \sigma_X^2$. In this case an approximation of uniformity in the quantization bin is adopted in (15), simplifying the analysis considerably. To finalise, the pdf of the watermark the case when $s = x$ must be considered. It can be easily seen that this contributes a Dirac δ -function to f_W . We therefore obtain f_W as $f_W(w) = \frac{1}{2} (\delta(w) + f_E(w) * f_D(w))$.

4.2 Comparison of SQIM and DM

Here we discuss the tradeoff in achievable rate and statistical transparency between DM and SQIM. In Fig. 4 an example of the pdf $f_W(w)$ for SQIM is presented for the theoretical simplification alongside an empirically obtained histogram. This pdf has the shape of two trapeziums either side of a Dirac δ -function centred at zero. Using the approximate pdf the power of the watermark can be easily obtained as $E\{w^2\}$ which after some calculus gives $\sigma_W^2 = \Delta^2/12$. This power is the same as DM embedding power (i.e. the power of the quantization noise) and leads to a direct, fair comparison between DM and SQIM. In terms of capacity it results in the fact that SQIM can never outperform DM. This is due to the fact that in DM all of the embedding power is used to transmit the message but in SQIM only a proportion of the same total power is used for the message. The remaining power is used to compensate the shape of the pdf. To see what this proportion is we can simply compare f_D and f_E above (the δ -function in f_W contributes no energy). It can be seen that $\sigma_D^2 = \Delta^2/96$ while $\sigma_E^2 = 7\Delta^2/96$. The total power is of course the addition of the two variances as both signals are independent. It can now be seen that in SQIM 1/8 of the embedding power is used in the pdf compensation while the remaining 7/8s is actually used in the message transmission. Thus the capacity of DM acts as an upper bound to that of SQIM.

Another important observation is the fact that DM does not achieve complete host signal rejection in the presence of noise whereas DC-DM almost does. It can be seen that under moderately high noise conditions the performance of

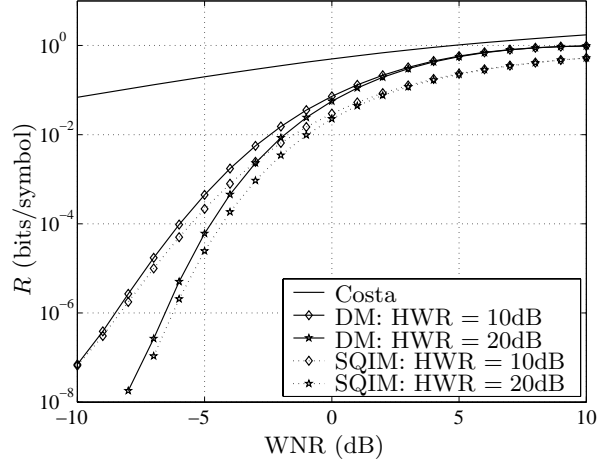


Fig. 5. The achievable rate R , calculated using (7), for SQIM (dashed lines) and DM (continuous lines). Two separate HWRs are plotted to emphasize the fact that the rate is a function of the host signal power. $\sigma_X^2 = 1.0$.

DM falls off considerably in comparison to DC-DM. This indicates that there is a threshold noise level below which the achievable rate of DM tends to zero. It also indicates that reducing the HWR for DM and SQIM will increase the achievable rate of these embedding techniques. This fact can be seen in Fig. 5. The implication of this is that it is not possible to communicate using SQIM in high noise scenarios. In the next section we will see how the same effect can be seen to apply to a generic N dimensional scheme designed along the same lines without the use of a secret key.

5 Capacity of Perfectly Secure Steganography

In previous sections we have seen that the achievable rate of some scalar side-informed embedding schemes is reduced when steganographic security is required. Here we will examine the capacity of the steganographic channel with the use of a sphere packing argument in the case where non-fixed codewords are used at the encoder.

5.1 Stochastic QIM Analogy in N Dimensions

As described in Sect. 4 the basic premise of one dimensional SQIM is that every point in \mathbb{R} is a possible codeword with a probability of each $s \in \mathbb{R}$ given by f_X (i.e. the codewords are random, not fixed). The transformation (14) contained a stochastic element, D , which compensated the pdf of the stegotext such that $f_S = f_X$. We now extend this idea to a higher dimensional space. Consider $\mathbf{X} \in \mathbb{R}^N$. For $f_{\mathbf{X}}$ iid Gaussian, the distribution of \mathbf{X} is uniform on the sphere of radius $\sqrt{N\sigma_X^2}$ with high likelihood, for large N [15].

Let the sphere of radius $\sqrt{N\sigma_W^2}$ around \mathbf{x} be denoted T_W as before. All points on the surface of T_W have the same probability as \mathbf{x} so any point in T_W can be transmitted without altering the host pdf. Now, applying the procedure

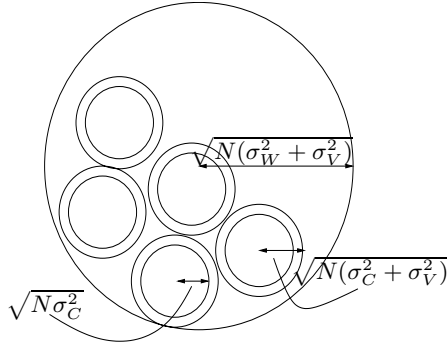


Fig. 6. A schematic representing the sphere packing argument for steganography. All points in T_W are valid codewords but only those contained within the spheres of radius $\sqrt{N\sigma_C^2}$ are reliable.

outlined in Sect. 2.2 we can fill the sphere, T_{W+V} , of radius $\sqrt{N(\sigma_W^2 + \sigma_V^2)}$, with codewords to calculate the capacity. Following the SQIM idea, it can be seen that all points \mathbf{s} lying on T_W must be codewords, with the result that none of said codewords are reliable for a noise variance $\sigma_V^2 > 0$.

To manufacture reliable codewords we do the following for a particular message \mathbf{b} . Consider a noise sphere, T_V , of radius $\sqrt{N\sigma_V^2}$ in T_{W+V} . Now, form a sphere at the centre of T_V such that all codewords in the new sphere, T_C of radius σ_C^2 , are reliable. We now have that T_W is filled with spheres of radius $\sqrt{N(\sigma_V^2 + \sigma_C^2)}$. The idea is illustrated in Fig. 6.

Assume for a moment that σ_C^2 is known such that reliable communication is possible. Then the achievable rate of the channel is given as $2^{NR} = (N(\sigma_W^2 + \sigma_V^2))^{N/2} / (N(\sigma_C^2 + \sigma_V^2))^{N/2}$, which gives $R = \frac{1}{2} \log((\sigma_W^2 + \sigma_V^2) / (\sigma_C^2 + \sigma_V^2))$. This suggests, for $\sigma_C^2 > 0$, that Costa's capacity is not achievable under steganographic conditions when non-fixed codewords are used at the encoder. It also implies that there is a vertical asymptote in the achievable rate curve, similar to those seen in the rate plots for SQIM, Fig. 5.

Now consider the value of the parameter σ_C^2 . It must be chosen such that the probability of error at the decoder vanishes as N tends to infinity. Simple limits on the actual value of σ_C^2 are formed as $0 \leq \sigma_C^2 \leq \sigma_W^2$. It can be seen that this upper limit corresponds to the case of all codewords in the allowable distortion region corresponding to just one message. This will give a zero achievable rate which goes some way to explaining the sharp falloff in the achievable rate of SQIM shown previously.

6 Conclusion

The issue of robust embedding for the steganographic channel has been examined. It was seen that the achievable rate of DC-DM is constrained when steganographic secrecy is required. The optimum value of the Costa parameter α was also seen to be restricted for this channel.

An analysis of an approximately statistically undetectable technique, namely Stochastic QIM was also undertaken. It was shown that the achievable rate of

this technique is bounded by that of DM. Also noteworthy is a vertical asymptote in the achievable rate at a particular WNR. The location of this asymptote was shown to be dependent on the HWR.

Extending the idea behind SQIM to N dimensions indicated that a similar asymptote exists for a more general steganographic embedding scheme utilising non fixed codewords. The location of the asymptote however remains an open topic.

Acknowledgements. The authors wish to thank Pedro Comesaña for interesting remarks and corrections on an early version of this manuscript. This work is supported by Enterprise Ireland under research grant ATRP-2002/230 and the European Commission through the IST Programme under contract IST-2002-507609 SIMILAR.

References

1. Simmons, G.: The prisoner's problem and the subliminal channel. In: *Advances in Cryptology, Crypto '83*. Volume 20., Plenum Press (1984) 51–67
2. Cachin, C.: An information-theoretic model for steganography. In: *Information Hiding: Second International Workshop*. Volume 1525., Springer (1998) 306–318
3. Costa, M.: Writing on dirty paper. *IEEE Trans. on Information Theory* **29** (1983) 439–441
4. Moulin, P., Wang, Y.: New results on steganographic capacity. In: *Proc. CISS Conference, Princeton, USA* (2004)
5. Chen, B., Wornell, G.: Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory* **47** (2001) 1423–1443
6. Eggers, J., Girod, B.: *Informed watermarking*. Kluwer Academic Publishers (2002)
7. Hogan, M.T., Hurley, N.J., Silvestre, G.C., Balado, F., Whelan, K.M.: ML detection of steganography. In: *Security, Steganography and Watermarking of Multimedia Contents*. Volume 5681 of *Proc. Electronic Imaging., SPIE* (2005)
8. Wang, Y., Moulin, P.: Steganalysis of block structured stegotext. In: *Security, Steganography and Watermarking of Multimedia Contents*. Volume 5306 of *Proc. Electronic Imaging., SPIE* (2004)
9. Guillon, P., Furon, T., Duhamel, P.: Applied public-key steganography. In: *Security and Watermarking of Multimedia Contents*. Volume 4675 of *Proc. Electronic Imaging., SPIE* (2002) 38–49
10. Cover, T., Thomas, J.: *Elements of Information Theory*. J. Wiley & Sons (1991)
11. Van Trees, H.L.: *Detection, Estimation and Modulation Theory*. J. Wiley & Sons (1968)
12. Johnson, D.H., Orsak, G.C.: Relation of signal set choice to the performance of optimal non-Gaussian detectors. *IEEE Trans. on Communications* **41** (1993)
13. Barron, R.J., Chen, B., Wornell, G.W.: The duality between information embedding and source coding with side information and some applications. *IEEE Trans. on Information Theory* **49** (2003) 1159–1180
14. Pérez-Freire, L., Pérez-González, F., Voloshinovskiy, S.: Revisiting scalar quantization-based data hiding: Exact analysis and results. *IEEE Trans. on Signal Processing* (2005) To appear.
15. Hamkins, J., Zeger, K.: Gaussian source coding with spherical codes. *IEEE Trans. on Information Theory* **48** (2002) 2980–2988