

New Results on Robustness of Secure Steganography

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

School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4,
Ireland.

ABSTRACT

Steganographic embedding is generally guided by two performance constraints at the encoder. Firstly, as is typical in the field of watermarking, all the transmission codewords must conform to an average power constraint. Secondly, for the embedding to be statistically undetectable (secure), it is required that the density of the watermarked signal must be equal to the density of the host signal. Assuming that this is not the case, statistical steganalysis will have a probability of detection error less than $1/2$ and the communication may be terminated. Recent work has shown that some common watermarking algorithms can be modified such that both constraints are met. In particular, spread spectrum (SS) communication can be secured by a specific scaling of the host before embedding. Also, a side informed scheme called stochastic quantization index modulation (SQIM), maintains security with the use of an additive stochastic element during the embedding. In this work the performance of both techniques is analysed under the AWGN channel assumption. It will be seen that the robustness of both schemes is lessened by the steganographic constraints, when compared to the standard algorithms on which they are based. Specifically, the probability of decoding error in the SS technique increases when security is required, and the achievable rate of SQIM is shown to be lower than that of dither modulation (on which the scheme is based) for a finite alphabet size.

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. This detection procedure is known as *steganalysis*.

Now, consider the nature of Wendy's actions. Typically she can be either *passive* or *active*. If passive then a detection test is all that is performed on the received document. If she is active, then the document is attacked regardless of the outcome of any detection test. In this work we consider that Wendy is basically passive, but we also assume that the transmitted document undergoes a channel distortion before it is decoded. This operation may be interpreted as an active warden or attacker pre-emptively jamming the transmitted signal irrespective of the detection test outcome; for the sake of comparison with prior research, we assume that this distortion is additive white Gaussian noise (AWGN) independent of the host.

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,² where the security of steganography has been defined in terms of the Kullback Leibler distance (D_{KL}) between the densities of the host and stegotext signals. The D_{KL} is equal to zero iff the two distributions are equal. 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. If the embedding is such that $D_{KL} = 0$ between the host and stegotext densities then we have perfect steganography.

Further author information: (Send correspondence to M.H.)

- M.H.: E-mail: markhogan@ihl.ucd.ie, Telephone: +353 (0)1 716 2454.

In this case optimal statistical steganalysis will always have a probability of detection error, P_e^* , no better than $1/2$. In non-perfect steganography a small value for D_{KL} is allowed, such that the results of practical statistical tests are unreliable (c.f. ϵ -secure steganography²). In this work only perfect steganography is considered.

In terms of robustness a work by Moulin and Wang³ showed that the capacity calculation of a steganographic channel is guided by two constraints; namely an embedding power constraint (due to the codebook) *and* a stegotext probability density function (pdf) constraint (also due to the codebook). Given that the equivalent watermarking calculation only adheres to the power constraint it is to be expected that, in general, the additional constraint will lead to a performance degradation; for steganography the set of allowable codebooks is necessarily reduced over that for watermarking, meaning the optimal codebook in a watermarking sense may not be available for use in the steganographic problem. Indeed, in watermarking, for the AWGN channel, distortion compensated dither modulation (DC-DM)⁴ shows a good performance, with an achievable rate close to the channel capacity. However the stegotext pdf for DC-DM—with a non-zero rate—can never equal that of the host,⁵ regardless of whether or not the attacker has (partial) knowledge of any key used in the embedding. Thus, assuming that perfect steganography is required, DC-DM is not a suitable technique for communication.

In other previous work Wang and Moulin⁶ adapted some common embedding algorithms in such a way that the conditions for perfect steganography are met. To the authors' knowledge these are the only schemes proposed in the literature in which statistical transparency is guaranteed when viewed by a third party. Firstly we have a blind* algorithm, based on the common spread spectrum method.⁷ The standard embedding method is modified using a host signal attenuation in such a way as to maintain $D_{\text{KL}} = 0$, for the case in which the host is Gaussian. We will refer to this technique as secure SS. Secondly, a side informed embedding method based on quantization index modulation⁴ (QIM) called stochastic QIM (SQIM) is proposed. Subject to the commonly adopted “flat host assumption,”⁸ this embedding method maintains security through the addition of an extra stochastic element at the encoding stage.

In this work a robustness analysis of both schemes is performed with a view to quantifying the loss in performance—over the standard embedding schemes on which they are based—enforced by the additional constraint at the encoder.³ In the case of secure SS no previous robustness analysis is available. Here, the probability of bitwise decoding error is analysed and compared to a similar analysis for standard SS embedding. It will be seen that to achieve the same robustness to AWGN as standard SS, the watermark power in secure SS must be approximately double that of standard SS, for a given host power.

For SQIM an achievable rate analysis is presented which builds on earlier work^{6,9} on the topic. Previously it was shown that the rate of SQIM is upper bounded by that of dither modulation⁴ (DM) for a finite alphabet size and scalar embedding.⁹ To extend this work, a lattice based analysis of SQIM is presented and asymptotic closed form expressions for the achievable rate are obtained. It will also be seen that for a large alphabet the rate of SQIM approaches that of DM.

The paper is organised as follows. Section 2 contains the preliminaries, including a description of the secure SS and SQIM embedding techniques. Sections 3 and 4 contain the robustness analyses of the two embedding schemes and Section 5 contains the final remarks.

2. PROBLEM SET-UP

2.1. Preliminaries and Notation

All vectors in this work are length N column vectors and denoted by bold letters, e.g. $\mathbf{x} = [x_1, \dots, x_N]^T$, whereas scalar variables are denoted by normal types, e.g. x . If a capital letter refers to a random variable (vector), e.g. X (\mathbf{X}), then the same letter in lower case, e.g. x (\mathbf{x}), is a realisation of the corresponding random variable (vector). The 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 its differential entropy is denoted as $h(X) = -\int f_X(x) \log f_X(x) dx$. The mutual information between X and Y is denoted $I(X; Y) = h(X) - h(X|Y)$.

*Here, “blind” is taken to have the watermarking interpretation (the host signal information is ignored by the embedding algorithm) and not the steganalysis meaning (where a blind test means that Wendy performs a detection test without knowledge of the embedding algorithm which generated the stegotext).

We assume that the host, $\mathbf{x} = [x_1, \dots, x_N]^T$, 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 a message b . In the case of scalar embedding we will assume that one message symbol is embedded per host sample, giving an information vector $\mathbf{b} = [b_1, \dots, b_N]^T$ where b_j , $j = 1, \dots, N$ is a realization of an iid random variable B , with elements uniformly distributed over the alphabet given by the set of real numbers \mathcal{B} , with cardinality $|\mathcal{B}|$. When examining higher dimensional embedding it is assumed that one symbol $b \in \mathcal{B}$ is embedded per host signal vector \mathbf{x} . The embedding produces a stegotext (watermarked) vector $\mathbf{s} = G(\mathbf{x}, \mathbf{b})$, and the watermark \mathbf{w} is then given as $\mathbf{w} \triangleq \mathbf{s} - \mathbf{x}$. 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

$$\text{HWR} \triangleq \frac{\text{E}\{\|\mathbf{X}\|^2\}}{\text{E}\{\|\mathbf{W}\|^2\}} = \frac{\sigma_X^2}{\sigma_W^2} \triangleq \gamma,$$

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

$$\text{WNR} \triangleq \frac{\text{E}\{\|\mathbf{W}\|^2\}}{\text{E}\{\|\mathbf{V}\|^2\}} = \frac{\sigma_W^2}{\sigma_V^2} \triangleq \xi.$$

The achievable rate for a given coding scheme and channel is defined consequently as $R = I(Y; B)$.

Some valuable insights into generic quantization schemes are obtained by means of lattice-based quantization schemes, which make possible a systematic analysis and provide performances reasonably close to those obtained with generic quantizers. Our lattice-based analysis of SQIM in this paper adheres to previous work carried out for DC-DM.¹⁰ For this reason some lattice definitions, that we present next, will be needed. Briefly, we have that any N -dimensional lattice,¹¹ Λ , may be partitioned by another sublattice $\Lambda' \subset \Lambda$ as follows. A coset of Λ' is a translated lattice $\Lambda'_c \triangleq \{\boldsymbol{\lambda} + \mathbf{c} : \boldsymbol{\lambda} \in \Lambda'\}$, where the coset representative $\mathbf{c} \in \Lambda$ and then $\Lambda'_c \subset \Lambda$. The ensemble of all distinct Λ'_c , which are disjoint and whose union gives Λ , is called a partition of Λ induced by Λ' ; the number of such cosets is the order of the partition $|\Lambda/\Lambda'|$. A lattice Λ defines an associated minimum Euclidean distance quantizer $Q_\Lambda(\cdot)$. Then, in data hiding methods based on quantization, a partitioned lattice may be used to embed symbol b by means of the sublattice associated to the coset representative \mathbf{c}_b , which we may denote in short by $\Lambda'_b \triangleq \Lambda'_{\mathbf{c}_b}$; in this case $|\mathcal{B}| = |\Lambda/\Lambda'|$.¹⁰ The fundamental volume of Λ , $V(\Lambda)$, is the volume of any quantisation region of its associated quantizer. Also, its Voronoi region is $\Phi(\Lambda) \triangleq \{\mathbf{x} : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}$. Finally, we define the decision region corresponding to symbol b as

$$\mathcal{V}_b \triangleq \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x} - Q_{\Lambda'_b}(\mathbf{x})\| \leq \|\mathbf{x} - Q_{\Lambda'_q}(\mathbf{x})\|, \quad \forall q \in \mathcal{B}\}. \quad (1)$$

2.2. Optimal Detection

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 $f_{\mathbf{X}}(\cdot)$ or from $f_{\mathbf{S}}(\cdot)$. Assuming that $f_{\mathbf{X}}(\cdot)$ is known, then, given $G(\cdot)$, $f_{\mathbf{S}}(\cdot)$ is also known. This becomes an hypothesis testing problem with two choices, the null hypothesis H_0 (\mathbf{z} is a host), and alternative hypothesis, H_1 (\mathbf{z} is a stegotext). The optimal test is the Bayes likelihood ratio,¹²

$$\text{L}(\mathbf{z}) \triangleq \frac{f_{\mathbf{X}}(\mathbf{z})}{f_{\mathbf{S}}(\mathbf{z})} \underset{H_1}{\overset{H_0}{\geq}} \frac{P_0}{P_1} \cdot \frac{C_{10} - C_{00}}{C_{01} - C_{11}} \triangleq \mu, \quad (2)$$

where 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 . We assume is all further work that the *a priori* probabilities are uniformly distributed. The (optimal) probability of error in (2) is given as

$$P_e^* \triangleq P(L(\mathbf{Z}) < \mu | \mathbf{Z} \sim f_{\mathbf{X}}) \cdot P_0 + P(L(\mathbf{Z}) > \mu | \mathbf{Z} \sim f_{\mathbf{S}}) \cdot P_1 = \frac{P_{fa} + P_m}{2}, \quad (3)$$

where $\mathbf{Z} \sim f_{\mathbf{X}}$ is taken to mean that the random vector \mathbf{Z} follows $f_{\mathbf{X}}(\cdot)$, P_{fa} represents the probability of false alarm and P_m , the probability of a miss.

From Alice's point of view it is desirable to set the P_e^* of the test to be $1/2$. In this way the test result is always unreliable and the communication will pass the warden unhindered. The relationship between (2) and (3) is given through Stein's lemma,¹³ which, for iid elements in \mathbf{x} and \mathbf{s} , and $P_m = 0$, can be written as a bound, giving,

$$P_{fa} > 2^{-ND_{\text{KL}}(f_{\mathbf{X}} \| f_{\mathbf{S}})}, \quad (4)$$

where D_{KL} is the Kullback Leibler distance given as

$$D_{\text{KL}}(f_{\mathbf{X}} \| f_{\mathbf{S}}) \triangleq \int f_{\mathbf{X}}(\mathbf{x}) \log \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{S}}(\mathbf{x})} d\mathbf{x}. \quad (5)$$

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 $f_{\mathbf{X}}(\cdot) = f_{\mathbf{S}}(\cdot)$.

From (4) it can be seen that the only case in which $P_e^* = 1/2$ is if $f_{\mathbf{X}}(\cdot) = f_{\mathbf{S}}(\cdot)$, giving $D_{\text{KL}} = 0$.² Thus for a perfectly secure embedding algorithm the stegotext density must equal the host density. If equality is not met then the test (2) will have a $P_e^* < 1/2$.

In this work we consider only that the attacker does not have access to any secret key used in the embedding process. This poses a problem for the attacker as $f_S(\cdot)$ is not available. One possible approach under this scenario is to take the average expression of $f_S(s|k)$ over K for use in (2). This gives

$$\bar{f}_S(s) = E_K\{f_S(s|k)\} = \int f_S(s|k) \cdot f_K(k) dk. \quad (6)$$

In a recent work Wang and Moulin⁶ have adapted some common watermarking algorithms such that the condition for $P_e^* = 1/2$ is met. These embedding techniques will be reviewed next.

2.3. SS Embedding Techniques

In standard SS embedding the message is modulated onto a pseudo-random sequence which is then added to the host to form the stegotext. We assume for this embedding that the message is binary and given as $b \in \{-1, +1\}$. Also, one message bit is embedded per host vector meaning that repetition encoding is implicitly used in this technique. The embedding power is guided by a parameter vector α , which is typically host signal dependent. For ease of exposition we assume here that the parameter is constant for all host samples, i.e. $\alpha_j = \alpha$, $j = 1, \dots, N$.

The particular type of SS embedding we analyse is given as¹⁴

$$\mathbf{s} = \mathbf{x} + \alpha \cdot b \cdot \mathbf{g} \quad (7)$$

where \mathbf{g} is the key dependent zero mean pseudo random sequence. Letting $g_j \in \{-1, +1\}$ we have that $E\{\mathbf{G}\mathbf{G}^T\} = \mathbf{I}_N$ when viewed by an external observer, where \mathbf{I}_N denotes the $N \times N$ identity matrix. This gives the embedding power, $D_w = \sigma_W^2 = \alpha^2$. For the purposes of analysis this also means all elements of (7) are equally distributed with $f_{\mathbf{S}}(\mathbf{s}) = \prod_{j=1}^N f_S(s_j)$, and then we may focus our study in one such distribution.

For $f_X(\cdot)$ Gaussian, under the procedure outlined in (6), $\bar{f}_S(\cdot) = \frac{1}{2} (\mathcal{N}(\alpha, \sigma_X^2) + \mathcal{N}(-\alpha, \sigma_X^2))$, a Gaussian mixture. After some algebra this gives,

$$\bar{f}_S(s) = \frac{1}{\sqrt{2\pi\sigma_X}} \exp\left(\frac{-(s^2 + \alpha^2)}{2\sigma_X^2}\right) \cosh\left(\frac{s\alpha}{\sigma_X^2}\right). \quad (8)$$

This is of course not equal to $f_X(\cdot)$, unless $\alpha = 0$, in which case no embedding takes place. Therefore, if this technique is adopted for message transmission, we have that $P_e^* < 1/2$, and communication may be terminated by the warden.

In response to this problem a scaling of the host has been proposed⁶ which has the effect of making $f_S(\cdot) = f_X(\cdot)$. Firstly assume that we have two mutually independent random vectors given as $\mathbf{M}_{+1} \sim \mathcal{N}(\mathbf{0}, \sigma_K^2 \mathbf{I}_N)$ and $\mathbf{M}_{-1} \sim \mathcal{N}(\mathbf{0}, \sigma_K^2 \mathbf{I}_N)$ corresponding to $b = +1$ and $b = -1$ respectively. The value of σ_K^2 will be discussed in a moment.

Now the modified formula for embedding b is

$$\mathbf{s} = \nu \mathbf{x} + \mathbf{m}_b, \quad (9)$$

with \mathbf{m}_b drawn from \mathbf{M}_b , as above. Because the two signals on the right hand side of (9) are independent the pdf of \mathbf{s} is the convolution of the two Gaussians, resulting in another Gaussian. As all signals are zero mean, all that remains to achieve security is to maintain $\sigma_S^2 = \sigma_X^2$. This condition gives the value of $\nu = \sqrt{1 - \frac{\sigma_K^2}{\sigma_X^2}}$. Then it can be seen that

$$D_w = \mathbb{E}\{(s - x)^2\} = \mathbb{E}\{((\nu - 1)x + m_1)^2\} = (\nu - 1)^2 \sigma_X^2 + \sigma_K^2. \quad (10)$$

This of course must equal σ_W^2 to meet the HWR constraint with the result that the power of the message can be written as

$$\sigma_K^2 = \sigma_W^2 \left(1 - \frac{\sigma_W^2}{4\sigma_X^2} \right). \quad (11)$$

Under this procedure, all detection tests will have $P_e^* = 1/2$.

It is evident from this embedding that the key to security lies in the scaling of the host before embedding. A similar method, used to improve robustness is that of improved spread spectrum (ISS).¹⁵ The embedding formula for the linear version of this technique can be written in the following form,

$$\mathbf{s} = (\mathbf{I}_N - \kappa \mathbf{g} \mathbf{g}^T) \mathbf{x} + b \cdot \alpha \cdot \mathbf{g}, \quad (12)$$

where κ is an optimisable constant. This method is shown to have better robustness to noise characteristics than standard SS (7), when κ is optimised for a given channel noise power. This is similar to the embedding method distortion compensated dither modulation (DC-DM),⁴ where the so called Costa parameter is optimised for a given channel to maximise the achievable rate of the scheme. However, it was also shown that manipulating the Costa parameter can be used to increase the security of DC-DM.^{6, 9, 16}

The question then arises as to whether κ in (12) can be optimised for security, as opposed to robustness, similarly to DC-DM. Firstly, it should be noted that the matrix $\mathbf{I}_N - \kappa \mathbf{g} \mathbf{g}^T$ is not diagonal, implying that the elements of \mathbf{s} in (12) are not strictly independent. However, assuming that κ is small, the matrix is predominantly diagonal and an approximation of independence between elements can be made.

If we consider that $g \in \{+1, -1\}$, the (approximate) scalar embedding formula can be written as $s = (1 - \kappa)x + b \cdot \alpha \cdot g$. It can be seen that, in this case $\bar{f}_S(s)$ is given as

$$\bar{f}_S(s) = \frac{1}{|1 - \kappa| \sqrt{2\pi} \sigma_X} \exp\left(-\frac{s^2 + \alpha^2}{2\sigma_X^2(1 - \kappa)^2}\right) \cosh\left(\frac{s\alpha}{(1 - \kappa)^2 \sigma_X^2}\right). \quad (13)$$

This shows that the only case in which security is guaranteed is when $\kappa = 0$ and $\alpha = 0$, i.e., no embedding takes place. This particular type of ISS is therefore clearly not suitable for steganography. If we take $G \sim \mathcal{N}(0, 1)$, (12) becomes $s = (1 - \kappa g^2)x + b \cdot \alpha \cdot g$. This transformation is much more involved but a brief glance indicates that again it is required that $\alpha = \kappa = 0$, for $\bar{f}_S(\cdot) = f_X(\cdot)$.

This reasoning indicates that ISS embedding will never give perfect security but manipulation of κ may allow the method to be used in such a way that the performance exceeds that of SS while also maintaining an acceptable P_e^* similar to the argument for DC-DM.⁹ This particular topic is outside the scope of this work and is not pursued here.

2.4. SQIM Embedding Technique

SQIM is a side informed embedding technique⁶ based on quantization of the host signal samples. Lattices provide a natural method of analysis for such schemes, such as the work of Pérez-González on the DC-DM algorithm.¹⁰ Here, we follow this analysis, with specific application to SQIM.

Firstly, the host space is tiled with a lattice Λ partitioned by Λ' , with $|\mathcal{B}| = |\Lambda/\Lambda'|$. We now have that any $b \in \mathcal{B}$ can be encoded using the quantizer $Q_{\Lambda'_b}(\cdot)$. For a given message b there are two possible embedding scenarios depending on the location of \mathbf{x} . Firstly assume that $\mathbf{x} \in \mathcal{V}_b$, with the decision region \mathcal{V}_b given by (1). In this case \mathbf{x} is already a valid codeword, and $\mathbf{s} = \mathbf{x}$ is transmitted. Obviously $Q_{\Lambda'_b}(\mathbf{s}) \in \mathcal{V}_b$ and minimum Euclidean distance decoding will be correct (in a noiseless scenario). Next consider that $\mathbf{x} \notin \mathcal{V}_b$. In this case a watermark must be added to \mathbf{x} to communicate b . This watermark in turn consists of two summands, namely a quantization error given as $Q_{\Lambda'_b}(\mathbf{x}) - \mathbf{x}$, and an additional element \mathbf{d} drawn from a certain random variable \mathbf{D} , which is used to maintain the pdf of \mathbf{S} equal to that of \mathbf{X} for an external observer. The pdf of \mathbf{D} is just given by $f_{\mathbf{X}}(\cdot)$ truncated to the support given by the region $Q_{\Lambda'_b}(\mathbf{x}) + \Phi(\Lambda)$, and normalized to remain a pdf.

From the above discussion the SQIM embedding procedure can be written as

$$\mathbf{s} = \begin{cases} Q_{\Lambda'_b}(\mathbf{x}) + \mathbf{d}, & \mathbf{x} \notin \mathcal{V}_b, \\ \mathbf{x}, & \mathbf{x} \in \mathcal{V}_b, \end{cases} \quad (14)$$

where the pdf of \mathbf{D} is given as

$$f_{\mathbf{D}}(\mathbf{z}) = \begin{cases} \frac{1}{\kappa} \cdot f_{\mathbf{X}}(Q_{\Lambda'_b}(\mathbf{x}) + \mathbf{z}), & \mathbf{z} \in \Phi(\Lambda), \\ 0, & \text{otherwise,} \end{cases} \quad (15)$$

with the normalization factor $\kappa \triangleq \int_{\Phi(\Lambda)} f_{\mathbf{X}}(Q_{\Lambda'_b}(\mathbf{x}) + \mathbf{z}) d\mathbf{z}$. Note that the shape of the pdf of \mathbf{D} depends on \mathbf{x} , but this dependence vanishes with the flat host assumption.

3. SPREAD SPECTRUM COMMUNICATIONS

In this section the probability of decoding error in the SS techniques where it is seen that the insecure technique performs better than the secure embedding, when the noise is AWGN. Let the probability of bitwise decoding error for the standard SS scheme be denoted as P_b^{SS} , and that of secure SS as P_b^{SSS} .

3.1. Probability of Error in Standard SS

We now wish to analyse the probability of error for the system given that $\mathbf{y} = \mathbf{s} + \mathbf{v}$. The pdf of the host is symmetric so the decision threshold is zero. The ML decoding procedure is to set $\hat{b} = +1$ if the correlation value $r = \mathbf{y}^T \cdot \mathbf{g} > 0$. Without loss of generality assume that $g_j = 1$, $j = 1, \dots, N$. By the central limit theorem (CLT) this statistic is Gaussian so it suffices for performance purposes to find the mean and variance. We have that $E\{r\} = E\{\sum_j (x_j + \alpha + v_j)\} = N\alpha$. The variance is given as $\text{Var}\{r\} = E\{r^2\} - E^2\{r\} = N \cdot (\sigma_X^2 + \sigma_V^2)$. We then can write, by symmetry (with $P(b = +1) = P(b = -1) = 0.5$), that $P_b^{\text{SS}} = P(r < 0 | b = +1)$. This gives

$$P_b^{\text{SS}} = Q\left(\frac{\sqrt{N}\alpha}{\sqrt{\sigma_X^2 + \sigma_V^2}}\right) = Q\left(\sqrt{\frac{N}{\gamma + \xi^{-1}}}\right), \quad (16)$$

where $Q(\cdot)$ represents the integral of the tail of a Gaussian pdf[†].

[†] $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{z^2}{2}} dz$

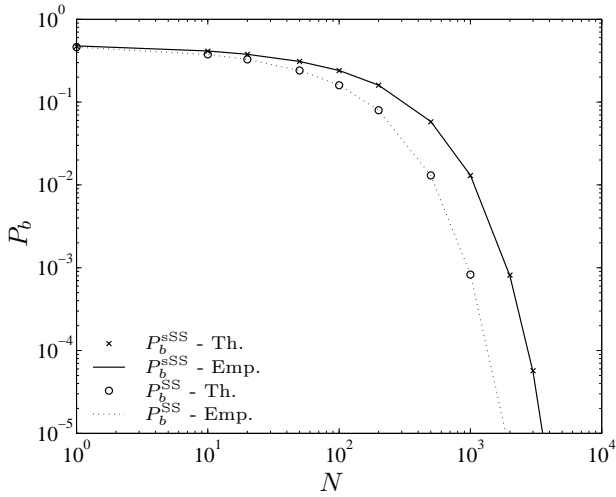


Figure 1. Probability of bit error against host vector length N for standard SS (P_b^{SS}) and secure SS (P_b^{sSS}). The theoretical (Th.) values are shown as with markers and the lines represent the empirical (Emp.) evaluation. HWR = 20 dB, WNR = 0 dB.

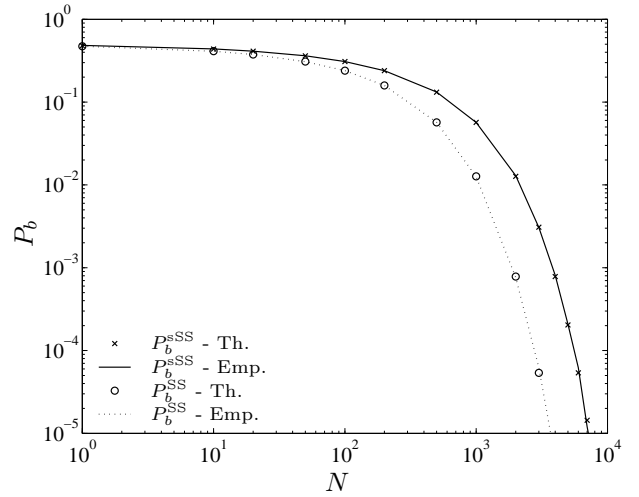


Figure 2. Probability of bit error against host vector length N for standard SS (P_b^{SS}) and secure SS (P_b^{sSS}). The theoretical (Th.) values are shown as with markers and the lines represent the empirical (Emp.) evaluation. HWR = 20 dB, WNR = -20 dB.

3.2. Probability of Error in Secure SS

Now in the case of secure SS we have a different scenario. From (9), assuming that $b = +1$ without loss of generality, we have $\mathbf{y} = \nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v}$. Again using the ML correlation receiver we have that an error occurs when $(\nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v})^T \cdot \mathbf{m}_{+1} < (\nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v})^T \cdot \mathbf{m}_{-1}$. With the same symmetry conditions as standard SS this means that the P_b^{sSS} can be expressed as

$$\begin{aligned} P_b^{\text{sSS}} &= P((\nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v})^T \cdot \mathbf{m}_{+1} < (\nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v})^T \cdot \mathbf{m}_{-1}) \\ &= P((\nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v})^T \cdot (\mathbf{m}_{+1} - \mathbf{m}_{-1}) < 0). \end{aligned} \quad (17)$$

The decision statistic $r = (\nu\mathbf{x} + \mathbf{m}_{+1} + \mathbf{v})^T \cdot (\mathbf{m}_{+1} - \mathbf{m}_{-1})$, is again Gaussian under the CLT, and thus completely specified by the mean and variance. We have $E\{r\} = N\sigma_K^2$. The expression for $\text{Var}\{r\}$ is more longwinded but the result can be seen to be

$$\text{Var}\{r\} = N\sigma_K^2(2\nu^2\sigma_X^2 + \sigma_K^2 + 2\sigma_V^2). \quad (18)$$

This gives

$$P_b^{\text{sSS}} = Q\left(\frac{\sqrt{N}\sigma_K}{\sqrt{2\nu^2\sigma_X^2 + 2\sigma_V^2 + \sigma_K^2}}\right) = Q\left(\sqrt{\frac{N\xi(4\gamma - 1)}{\xi(8\gamma^2 - 4\gamma + 1) + 8\gamma}}\right). \quad (19)$$

3.3. Quantifying the Loss in Performance

A plot of the performance of secure SS is shown in Figures 1 and 2 as a function of the length of the code N . It is clear that there is a loss in performance when the secure SS technique is used, at both WNRs tested. However, gauging this loss in performance directly from (16) and (19) is not immediately obvious. To gain more insight we consider the case in which only the only noise in the communication is the host signal interference, i.e. $\sigma_V^2 = 0$. In this case, we have that the error probability in standard SS is

$$P_b^{\text{SS}} = Q\left(\sqrt{\frac{N}{\gamma}}\right). \quad (20)$$

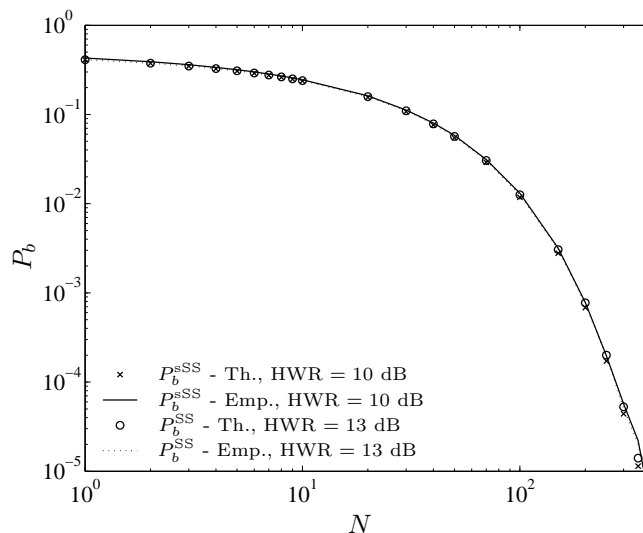


Figure 3. Probability of bit error vs. host vector length N for secure (P_b^{sSS}) and standard SS (P_b^{SS}), illustrating the approximate equality in performance if the power in secure SS is double that of standard SS. $\xi \rightarrow \infty$.

For secure SS with no channel distortion we have that $P_b^{sSS} = Q\left(\sqrt{\frac{N(4\gamma-1)}{8\gamma^2-4\gamma+1}}\right)$. Noting that $\gamma \gg 1$, this can be accurately approximated as

$$P_b^{sSS} \approx Q\left(\sqrt{\frac{N}{2\gamma}}\right). \quad (21)$$

Now, comparison of (20) and (21) illustrates more clearly the impact of the security condition. It can be seen that, to achieve roughly equivalent performances, the HWR must be halved (reduced by 3 dB) in the secure SS scheme. This equates to requiring that the watermark power be doubled for the case in which the channel noise is set to zero. This result is illustrated in Figure 3, where the exact expressions are used for results rather than the approximation from (21).

4. STOCHASTIC QIM

In this section an analysis of lattice based SQIM is presented. It is important to note that this analysis makes use of the flat host assumption simplifying the analysis considerably.⁹ The section begins with a brief summary of previous work on SQIM.

4.1. Binary Scalar Embedding

It has been shown in previous work⁹ that the pdf of the watermark in scalar SQIM, under the flat host assumption and with $b \in \{0, 1\}$, is given as

$$f_W(w) = \frac{1}{2}(\delta(w) + f_E(w) * f_D(w)), \quad (22)$$

where $f_E(\cdot)$ is the pdf of the quantization error,

$$f_E(e) = \begin{cases} \frac{2}{\Delta}, & \frac{\Delta}{2} \geq |e| > \frac{\Delta}{4}, \\ 0, & \text{otherwise.} \end{cases}$$

and the non-null part of the pdf of D is given as $f_D(d) = 2/\Delta$, $d \in (\Delta/4, \Delta/4]$.

We then have that $E\{w^2\} = \sigma_W^2 = \Delta^2/12$ which is the same as the DM watermark power⁴ and leads to a fair comparison between DM and SQIM. In terms of achievable rate it results in the fact that SQIM can never outperform DM. This can be seen by considering the following.

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 $f_D(\cdot)$ and $f_E(\cdot)$. 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. This shows that in SQIM 1/8 of the embedding power is used in the pdf compensation while the remaining 7/8 is actually used in the message transmission.

In order to obtain the achievable rates, the pdfs of the stegotexts are required for each scheme. Let $q_{i,b} = i\Delta + b\Delta/|\mathcal{B}|$ for suitable $i \in \mathbb{Z}$ be a generic quantization point associated with message b . Firstly for SQIM and $s \in (q_{i,b} - \Delta/(2|\mathcal{B}|), q_{i,b} + \Delta/(2|\mathcal{B}|])$ we have that

$$f_{S_i}(s|b) \approx \begin{cases} |\mathcal{B}| \cdot f_X(s), & |s - q_{i,b}| < \frac{\Delta}{2|\mathcal{B}|}, \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Then, the conditional pdfs $f_S(s|b)$ are computed from (23) by summing over $i \in \mathbb{Z}$. It is easy to verify that $f_S(s) = E_B\{f_S(s|b)\} = f_X(s)$. Next, the pdf for DM can be seen to be,⁵

$$f_S(s|k) = \sum_{b \in \mathcal{B}} \sum_{i=-\infty}^{\infty} w_{i,b} \cdot \delta(s - q_{i,b}), \quad (24)$$

where the weights, $w_{i,b} = \frac{1}{|\mathcal{B}|} \int_{q_{i,b} - \frac{\Delta}{2}}^{q_{i,b} + \frac{\Delta}{2}} f_X(z) dz$. The conditioned pdfs follow simply. To conclude, convolving $f_S(\cdot)$ with $f_V(\cdot)$ will give $f_Y(\cdot)$, and the rates are then obtained by numerical evaluation.

The achievable rates of both DM and SQIM are shown in Figure 4. As can be seen from the plots that the rate of DM upper bounds that of SQIM, as expected from the previous considerations. Also note that performance in high noise conditions is very poor for both schemes, unlike the DC-DM performance in that case.⁹ As the pdfs of the stegotext depends on the pdf of X both for DM and SQIM, the performance of both schemes depends on the HWR. This dependence is displayed in Figure 4. The results indicate that lowering the HWR will improve the rate of communication, a result that has been shown to hold for DC-DM also.⁸

4.2. Achievable Rate Analysis of SQIM using Larger Alphabets

Considering multi-dimensional SQIM once again we have that the received signal, for message b , may be seen at the decoder as $\mathbf{y} = Q_{\Lambda'_b}(\mathbf{x}) + \mathbf{d} + \mathbf{v}$, with \mathbf{d} drawn from \mathbf{D} and $\mathbf{x} \notin \mathcal{V}_b$. We will initially ignore the case of $\mathbf{y} = \mathbf{x} + \mathbf{v}$, for reasons that will be discussed later. Without loss of generality, let us assume that $\mathbf{x} \in \Phi(\Lambda')$. Now, as the only term required to convey the message is $Q_{\Lambda'_b}(\mathbf{x})$, we may view the term \mathbf{d} as an additional noise parameter. This means that the total noise added during communication is given as $\mathbf{t} \triangleq \mathbf{v} + \mathbf{d}$ with pdf given as

$$f_{\mathbf{T}}(\mathbf{t}) = f_{\mathbf{V}}(\mathbf{t}) * f_{\mathbf{D}}(\mathbf{t}). \quad (25)$$

Given that the pdf of the message is a δ -function on \mathbf{c}_b we have that the received signal pdf consists of $f_{\mathbf{T}}(\mathbf{t} - \mathbf{c}_b)$, i.e., $f_{\mathbf{Y}}(\mathbf{z}|B=b) = f_{\mathbf{T}}(\mathbf{z} - \mathbf{c}_b)$. To remove the message dependence we can modularise the pdf as follows,

$$\tilde{f}_{\mathbf{Y}}(\mathbf{z}) = \begin{cases} \sum_{\mathbf{w} \in \Lambda'} f_{\mathbf{Y}}(\mathbf{z} - \mathbf{w}), & \mathbf{z} \in \Phi(\Lambda'), \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

Now, it has been shown¹⁰ that the achievable rate can be written as

$$R = D_{\text{KL}} \left(\tilde{f}_{\mathbf{Y}}(\mathbf{z}|b=0) \parallel \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \tilde{f}_{\mathbf{Y}}(\mathbf{z}|b) \right). \quad (27)$$

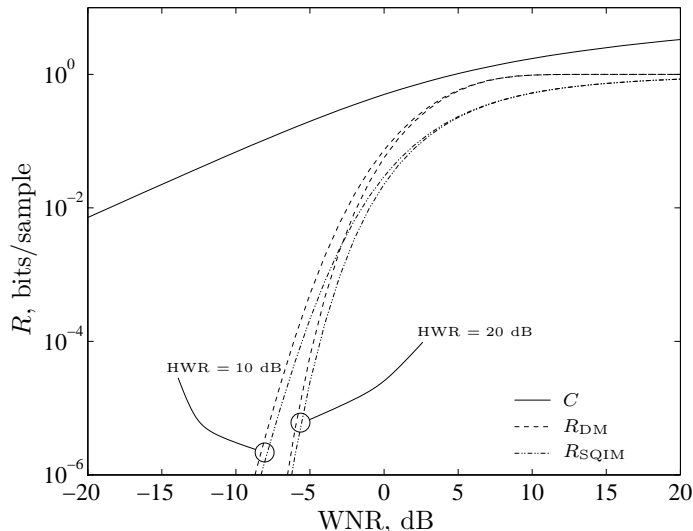


Figure 4. The achievable rate R for SQIM and DM. Two separate HWRs are plotted to emphasize the fact that the rate is a function of the host signal power. The channel capacity $C = \frac{1}{2} \log(1 + \xi)$ is plotted for reference. $\sigma_X^2 = 1.0$.

Before proceeding we let $|\mathcal{B}| \rightarrow \infty$. This provides two simplifications. Firstly, the Voronoi region of Λ approaches a δ -function, implying that $f_{\mathbf{D}}(\mathbf{t}) \rightarrow \delta(\mathbf{t})$, and further that $f_{\mathbf{Y}}(\mathbf{t}) \rightarrow f_{\mathbf{V}}(\mathbf{t})$. The convergence of these distributions is understood loosely as an approximation. This convergence also has the effect that the probability $P(\mathbf{x} \in \mathcal{V}_b) \rightarrow 0$ meaning that the embedding case of $\mathbf{s} = \mathbf{x}$ can be ignored for the purposes of the analysis, as mentioned above. Secondly, the continuous approximation¹⁰ can be applied to solve one of the integrals in (27), giving

$$R_{\text{SQIM}} = \int_{\mathbf{t} \in \Phi(\Lambda')} \tilde{f}_{\mathbf{V}}(\mathbf{t}) \log_2 \tilde{f}_{\mathbf{V}}(\mathbf{t}) d\mathbf{t} + \log_2 V(\Lambda'). \quad (28)$$

It is important to note that this expression is independent of $f_{\mathbf{X}}(\cdot)$. This is as a result of the flat host assumption and the limit in the alphabet size. This implies that comparison of (28) with the exact rate for relatively low HWRs (i.e. below approximately 15 dB) will show inaccuracies as the flat host assumption begins to fail and (28) becomes inaccurate. Also noteworthy is the fact that this expression is equal to that for DC-DM¹⁰ in the case for which $\alpha = 1$, i.e. the rate of DM. The implication is that, as the alphabet size increases, the rate of SQIM converges to that of DM and the effect of \mathbf{D} on the performance disappears.

This result is general for all WNRs but under specific noise conditions, further simplifications are possible. Firstly, in low noise conditions the pdf $\tilde{f}_{\mathbf{V}}(\cdot) \approx f_{\mathbf{V}}(\cdot)$ as most of the area of the pdf lies within $\Phi(\Lambda')$. This means that (28) can be approximated as $R_{\text{SQIM}}^{\text{High}} \approx -h(\mathbf{V}) + \log V(\Lambda')$, giving,

$$R_{\text{SQIM}}^{\text{High}} \approx \log V(\Lambda') - \frac{N}{2} \log(2\pi e \sigma_V^2). \quad (29)$$

To obtain an approximation to the rate in high noise conditions we revert to the scalar lattice. It has been shown that pdfs of the form of $\tilde{f}_{\mathbf{V}}(\cdot)$ in (28) can be represented as a Fourier series.¹⁷ Note that this extended version of $\tilde{f}_{\mathbf{V}}(\cdot)$ can be viewed as the convolution of $f_{\mathbf{V}}(\cdot)$ with a train of δ -functions on the lattice Λ' . In the frequency domain this convolution is of course a multiplication. The Fourier transform (FT) of $f_{\mathbf{V}}(t)$ is $\mathcal{F}_V(f) = \exp(-2\pi^2 \sigma_V^2 f^2)$ for V Gaussian. On the other hand, the FT of the δ train on Λ' is another δ train, this time on the dual lattice Λ'^{\perp} .¹¹

Substituting for $\tilde{f}_{\mathbf{V}}(\cdot)$ this gives

$$\tilde{f}_{\mathbf{V}}(t) = \sum_{\omega \in \Lambda'^{\perp}} \mathcal{F}_V(\omega) \cdot \Pi(\omega) e^{-j2\pi\omega t}, \quad (30)$$

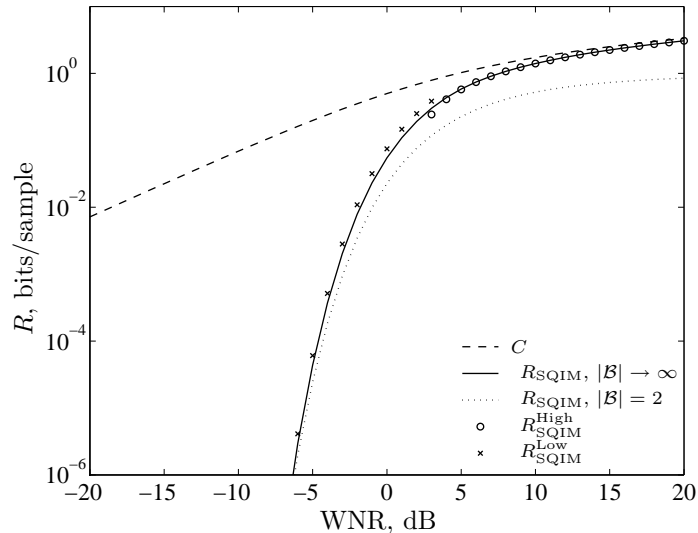


Figure 5. R for scalar SQIM for alphabets with cardinality $|\mathcal{B}| = 2$ and ∞ . The approximations from (32) ($R_{\text{SQIM}}^{\text{Low}}$) and (29) ($R_{\text{SQIM}}^{\text{High}}$) are also plotted. HWR = 20 dB. The channel capacity C is plotted for reference.

where $\Pi(f) = \frac{1}{V(\Lambda')} \sum_{v \in \Lambda'} \delta(f - v)$ is the FT of a δ train on Λ' . Since $\tilde{f}_V(\cdot)$ is slowly changing over Λ' , a convenient approximation is to consider only the low frequency terms,¹⁷ corresponding to $\omega = \{0, \pm \frac{1}{\Delta}\}$. After some algebra we have that

$$\tilde{f}_V(t) \approx \frac{1}{\Delta} \left(1 + 2 \exp\left(\frac{-2\pi^2 \sigma_V^2}{\Delta^2}\right) \cos\left(\frac{2\pi t}{\Delta}\right) \right). \quad (31)$$

Now in the calculation of (28) it is seen that the entropy of (31) is required. This integral as it stands is difficult to solve, but, using the approximation $\log(1+x) \approx x$ for $|x|$ small, which makes sense for high noise conditions, the integral becomes considerably simpler. Substituting into (28) we finally get

$$R_{\text{SQIM}}^{\text{Low}} \approx 2 \exp\left(-\frac{4\pi^2 \sigma_V^2}{\Delta^2}\right). \quad (32)$$

The evaluation of (28) for the scalar lattice with $|\mathcal{B}| \rightarrow \infty$ is plotted in Figure 5 alongside the rate of the binary scheme, for a fixed HWR = 20 dB. As happens with DM, it is interesting to see that at high WNRs the rate is significantly improved by using the $|\mathcal{B}|$ -ary alphabet but this is not the case at lower WNRs where the two rate plots converge.

5. CONCLUSION

The performance of perfectly secure steganographic techniques is examined under the AWGN channel. Two techniques, namely secure spread spectrum and stochastic quantization index modulation are analysed and compared to the standard watermarking algorithms on which they are based.

It has been shown that the increased security comes at the cost of robustness for both schemes. In the case of secure SS it was shown that to achieve approximately the same probability of error performance as standard SS the watermark power must be doubled.

For SQIM, the security condition is met through the addition of a stochastic element during encoding. This requires additional embedding power and reduces the achievable rate of the scheme over that of dither modulation. Explicit formulae for the rate of SQIM are derived using a lattice based analysis of the scheme. Under the assumption that the alphabet size tends to infinity it was seen that the effect of the additive element is removed from the analysis, and rate of SQIM becomes equal to that of DM.

Finally, it is worth noting that the two methods differ in how they achieve undetectability. The SQIM method generates true random variables (the codewords are random for a third party observer and the legitimate decoder) whereas the SS embedding only generates pseudo-random codewords (i.e. random to the attacker but not the decoder). The implication of this is that —assuming the attacker has access to a group of stegotexts— SQIM is completely resilient to an estimation attack whereas the secure SS technique is not.

ACKNOWLEDGMENTS

This work is kindly 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. G. Simmons, “The prisoner’s problem and the subliminal channel,” in *Advances in Cryptology, Crypto ’83*, **20**, pp. 51–67, Plenum Press, 1984.
2. C. Cachin, “An information-theoretic model for steganography,” *Lecture Notes in Computer Science* **1525**, pp. 306–318, 1998.
3. P. Moulin and Y. Wang, “New results on steganographic capacity,” in *Proc. CISS Conference*, (Princeton, USA), March 2004.
4. B. Chen and G. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. on Information Theory* **47**, pp. 1423–1443, May 2001.
5. M. T. Hogan, N. J. Hurley, G. C. M. Silvestre, F. Balado, and K. M. Whelan, “ML detection of steganography,” *Security, Steganography, and Watermarking of Multimedia Contents VII* **5681**(1), pp. 16–27, SPIE, 2005.
6. Y. Wang and P. Moulin, “Steganalysis of block-structured stegotext,” *Security, Steganography, and Watermarking of Multimedia Contents VI* **5306**(1), pp. 477–488, SPIE, 2004.
7. I. Cox, J. Kilian, T. Leighton, and T. Shamoan, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. on Image Processing* **6**, pp. 1673–1687, December 1997.
8. L. Pérez-Freire, F. Pérez-González, and S. Voloshinovskiy, “Revisiting scalar quantization-based data hiding: Exact analysis and results,” *IEEE Trans. on Information Forensics and Security*, 2006. To appear.
9. M. T. Hogan, F. Balado, N. J. Hurley, and G. C. Silvestre, “On the achievable rate of side informed embedding techniques with steganographic constraints,” in *Digital watermarking: 4th international workshop, IWDW 2005, Sienna, Italy, September 15–17, 2005: proceedings*, M. Barni, I. Cox, T. Kalker, and H. J. Kim, eds., *Lecture Notes in Computer Science* **3710**, pp. 387–402, Springer-Verlag Inc., (New York, NY, USA), 2005.
10. F. Pérez-González, “The importance of aliasing in structured quantization index modulation data hiding,” in *Digital Watermarking: Second International Workshop, IWDW, Lecture Notes in Computer Science* **2939 / 2004**, pp. 1–17, Springer-Verlag, (Berlin), October 2003.
11. J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, 2 ed., 1991.
12. H. L. Van Trees, *Detection, Estimation and Modulation Theory*, J. Wiley & Sons, 1968.
13. T. Cover and J. Thomas, *Elements of Information Theory*, J. Wiley & Sons, 1991.
14. F. Balado, *Digital Image Data Hiding Using Side Information*. PhD thesis, Universidade de Vigo, Vigo, Spain, Dec 2003.
15. H. S. Malvar and D. A. Florêncio, “Improved spread spectrum: A new modulation technique for robust watermarking,” *IEEE Trans. on Signal Processing* **51**, pp. 898–905, April 2003.
16. P. Guillon, T. Furon, and P. Duhamel, “Applied public-key steganography,” *Security and Watermarking of Multimedia Contents IV* **4675**(1), pp. 38–49, SPIE, 2002.
17. G. D. Forney, M. D. Trott, and S.-Y. Chung, “Sphere-bound-achieving coset codes and multilevel coset codes,” *IEEE Trans. on Information Theory* **46**, pp. 820–850, March 2000.