SELF-NOISE SUPPRESSION SCHEMES IN BLIND IMAGE STEGANOGRAPHY

Mahalingam Ramkumar and Ali N Akansu
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
New Jersey Center for Multimedia Research
University Heights, Newark, NJ, 07102.

ABSTRACT

Blind or oblivious data hiding, can be considered as a *signaling* method where the *origin* of the signal constellation is not known. The origin however, can be *estimated*, by means of *self-noise suppression* techniques. In this paper, we propose such a technique, and present both theoretical and numerical evaluations of its performance in an additive noise scenario. The problem of optimal choice of the parameters of the proposed technique is also explored, and solutions are presented. Though the cover object is assumed to be an image for purposes of illustration, the proposed method is equally applicable for other types of multimedia data, like video, speech or music.

1. INTRODUCTION

Data Hiding or *Steganography* is the art of hiding a *message signal* in a *host signal*, without any perceptual distortion in the host signal.[1] The process of data hiding in still images consists of an embedder E, and a detector D. If $\mathbf{I}$ is the original or *cover* image, and $\mathbf{b}$ is a sequence of bits to be embedded in the image, the *stego* image $\hat{\mathbf{I}}$ (the image with the embedded data) is obtained as

$$\hat{\mathbf{I}} = E(\mathbf{I}, \mathbf{b}, K) \tag{1}$$

where K is a *key*. We expect the image $\hat{\mathbf{I}}$ to undergo some modification (like lossy compression) before it reaches the receiver (detector D), where the hidden bit sequence is extracted. Let $\tilde{\mathbf{I}} = \hat{\mathbf{I}} + \mathbf{N}$ be the received image.

Data hiding can be broadly classified into two categories depending on whether the original image is needed for extraction of the hidden bits. *Cover image escrow* methods need the original image for extracting the hidden bits. On the other hand, *oblivious* detection methods extract the hidden bits without any knowledge of the original image;

$$\tilde{\mathbf{b}} = \begin{cases} D(\tilde{I}, K, I) & \text{escrow} \\ D(\tilde{I}, K) & \text{oblivious} \end{cases} \tag{2}$$

In this paper, we consider issues for *optimal oblivious* data hiding.

In most data hiding methods, the bit sequence to be embedded, *viz.* $\mathbf{b}$, is converted to a form *suitable for embedding* in the cover image. In Eq (3), $\mathcal{S}$ converts the bit sequence $\mathbf{b}$ to a *signature sequence* $\mathbf{s}$. The signature sequence is embedded in the cover image by the embedding function $\mathcal{E}$ to obtain the stego-image $\hat{\mathbf{I}}$. The overall embedding and detection operations therefore take the following form:

$$\begin{array}{ll} \mathbf{s} = \mathcal{S}(\mathbf{b}) & \hat{I} = \mathcal{E}(I, \mathbf{s}) \\ \tilde{I} = \hat{I} + N & \\ \tilde{\mathbf{s}} = \mathcal{D}(\tilde{I}) & \tilde{\mathbf{b}} = \mathcal{S}^{-1}(\tilde{\mathbf{s}}) \end{array} \tag{3}$$

From a signal processing perspective, data hiding methods can be classified into two categories, depending on the type of embedding and detecting operators. In the first category lies methods where the $\mathcal{E}$ adds the signature sequence *linearly* to I, and $\mathcal{D}$ detects $\tilde{\mathbf{s}}$ from $\tilde{I}$ by *correlative processing*. In the second category $\mathcal{E}$ and $\mathcal{D}$ are *non-linear*. One of the important characteristics of the non-linear methods is their ability to suppress the noise due to the original image (or self-noise), even though the original image is not available at the receiver.

In the next section, we present a unified viewpoint of the linear data hiding methods. In Section 3 we present some examples of non-linear data hiding methods, and explain how they achieve self-noise suppression. In Section 4 we

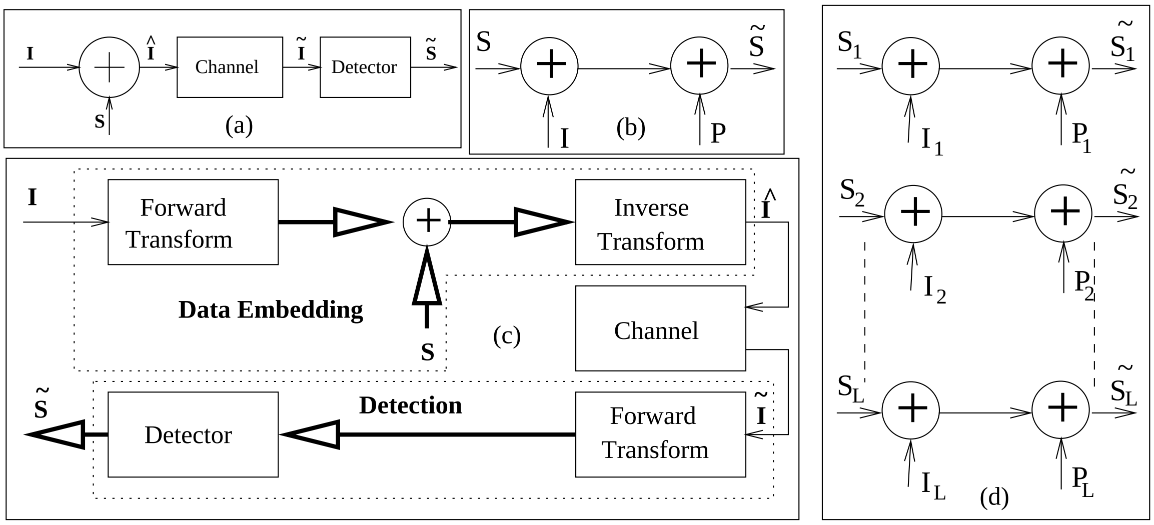


Figure 1. (a) Block Diagram of Data Hiding. (b) The Data Hiding Channel. (c) Block Diagram of Second Generation Data Hiding Schemes (d) Decomposition of the Data Hiding Channel.

take a communications theory perspective of data hiding. Oblivious data hiding is seen as a sophisticated *signaling* method. The main difference between "conventional" signaling methods and those used for (oblivious) data hiding is the *lack of reference* (of the signal constellation) for the latter. We then look at the relationship between data hiding and data compression to get a better insight into how a typical source of channel noise affects data hiding.

The rest of the paper is an analysis of optimal solutions for *self-noise suppression*, which is used to *obtain the reference* or the *origin of the signal constellation*. We propose a new algorithm for self-noise suppression, and evaluate its performance in an additive noise scenario. Exact analysis of the proposed algorithm is carried out and compared with numerical simulations. Optimal design strategies are then presented for a given *additive noise variance* and *permitted distortion*.

2. LINEAR DATA HIDING

Early data hiding methods,[2,3] primarily introduced the hidden information or signature directly in the spatial domain. Later methods, however, introduced the signature in some transform domain. DCT, DFT or wavelet transforms were the most used transforms for embedding the signature (compression transforms like DCT and Wavelets were obvious choices due to the intricate relationship between data hiding and data compression). However, as mentioned earlier, in all the linear schemes, the signature sequence was always embedded by *adding* a random sequence, and detected by correlative processing.

Figure 1 (a) shows the block diagram of a data hiding. Figure 1 (b) is an interpretation of the data hiding as a channel for communication. The *data hiding channel*[4–6] has two sources of noise - $\mathbf{I}$, the noise due to the original image, and $\mathbf{P}$, the noise due to processing. $\mathbf{S}$ is the input to the channel.

A *generic block diagram* of linear data hiding is depicted in Figure 1 (c). The forward transform block decomposes the image $\mathbf{I}$ into its coefficients of L bands. A component of the signature / message signal is added to each band. The inverse transform block reconstructs the modified image $\hat{\mathbf{I}}$. The image with the signature (stego-image) may undergo some modification *en route* to the receiver, (for example, lossy compression). The independent signature components in each band are detected after the forward transform of the received image $\tilde{I}$. Figure 1 (d) is an interpretation of the data hiding schematic of Figure 1 (c), as a communication channel. The forward and inverse transform blocks decompose the data hiding channel of Figure 1 (b) into *independent, multiple channels* (corresponding to the "bands" of the decomposition).

For most decompositions, the low frequency bands (channels) have a large component of *image noise* or *self-noise* and a small component of *processing noise*. The high frequency channels on the other hand have low image

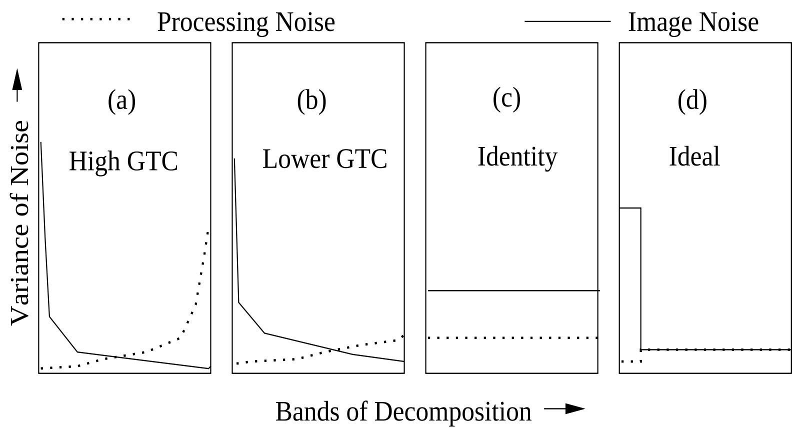


Figure 2. Typical Distribution of Image and Processing Noise for Different Decompositions

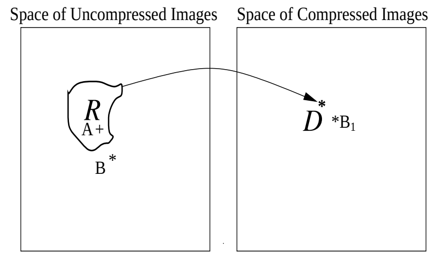


Figure 3. A lossy compression - decompression sequence maps all points in the range R to a single point in the domain D

noise and high processing noise. For cover image escrow methods, image noise is not an issue as it can be canceled by subtracting the original image (which is available at the detector). Therefore such methods[7] could utilize the robust low frequency bands. On the other hand, (linear) blind image (oblivious) methods cannot use the robust low frequency bands due to the high amount of image noise in those bands. Therefore, blind image data hiding methods[8,9] add the signature to the *mid-frequency* coefficients.

The *purpose* of the decomposition (forward and inverse transform blocks) is to obtain a *favorable distribution* of the image and processing noise in the different bands. Most existing methods use DCT or wavelet transforms for data hiding. These transforms are good choices for escrow data hiding, as the low frequency bands of these transforms suffer the least amount of noise. However, the choice of transforms is not so obvious for oblivious data hiding methods. Ramkumar *et. al*[6] discuss the relationship between the transform coding gain (GTC)[10] of the transform employed and the robustness (or achievable capacity) of the data hiding, for different processing noise scenarios. While high GTC transforms like DCT and wavelet transforms are better suited for low processing noise scenarios, lower GTC transforms like Hadamard and Hartley transforms are better suited for higher robustness requirements. When the processing noise is low, the high GTC transforms yield many channels with very low image noise. As processing noise increases, those channels become very noisy. However, the high frequency channels (bands) of lower GTC transforms are not *as vulnerable to processing noise.* This makes them a better choice for high processing noise scenarios.

For a given processing scenario, the ideal decomposition[4] would be the one which results in image noise variances (σ_i^2) close to the processing noise variances (σ_p^2) in a *maximum number of bands.* Typically, for high GTC decompositions, (Figure 2 (a)) $\sigma_i >> \sigma_p$ in the low frequency bands and $\sigma_p >> \sigma_i$ in the high frequency bands. For lower GTC transforms, the discrepancy is reduced (Figure 2 (b)). On the other hand, for the identity transform $\sigma_i >> \sigma_p$ in the single band (Figure2 (c)). Therefore, for the ideal decomposition, the image and processing noise variances should be distributed as shown in Figure 2 (d).

3. NON LINEAR DATA HIDING

The non-linear methods are capable of utilizing the robust low frequency bands even though the original image is not available at the detector. In Ramkumar *et. al*[11] the signature is introduced in 8 low frequency DCT coefficients (of each 8×8 block). The vector $\mathbf{x}$ of the low-frequency DCT coefficients is scrambled by means of an (invertible) cyclic all-pass filter $\mathcal{F}$ with pseudo random coefficients. Let $\mathbf{y} = \mathcal{F}(\mathbf{x})$ The signature is added and detected in the scrambled 'domain' $\mathbf{y}$. To embed the bit we modify the signs of many small amplitude coefficients of $\mathbf{y}$ so that the resulting sequence has more positive than negative coefficients. Coefficients with large amplitudes in the scrambled domain $\mathbf{y}$ are untouched. Altering only the small magnitude coefficients guarantees that the distortion introduced is tolerable. The modified sequence $\hat{\mathbf{y}}$ is unscrambled to obtain the modified (DCT) coefficients $\hat{\mathbf{x}} = \mathcal{F}^{-1}(\hat{\mathbf{y}})$. For

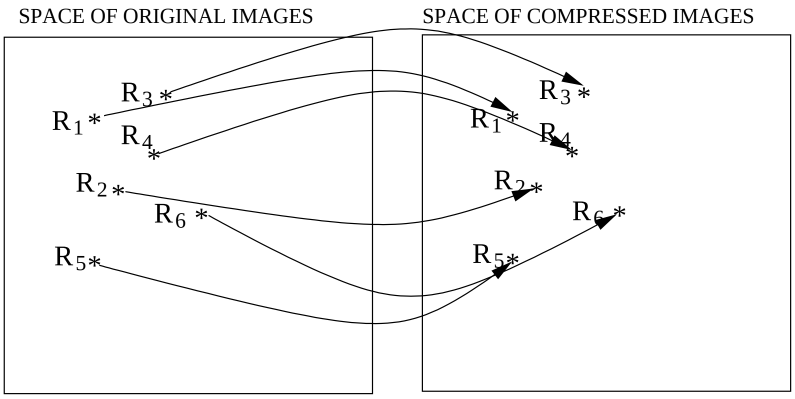


Figure 4. Data Hiding when Compression Method to be Employed is Known. The signal constellation should be chosen from points in the space of compressed images (which are known *a priori*).

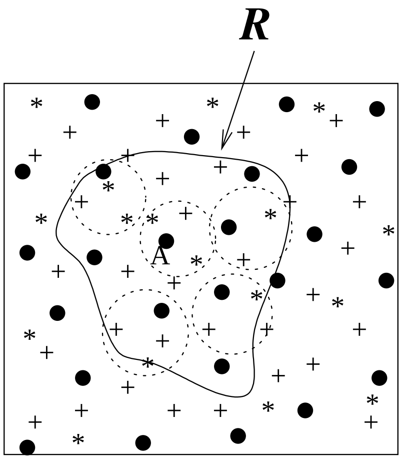


Figure 5. Data Hiding with Simultaneous Robustness to Multiple Compression Schemes

detecting the buried bit, the received vector $\tilde{\mathbf{x}}$ is scrambled by the filter $\mathcal{F}$ to obtain $\tilde{\mathbf{y}}$. The excess number of positive coefficients is counted. Note that by treating both high and low magnitude coefficients of $\tilde{\mathbf{y}}$ with equal weight (only the sign of the coefficient is considered), suppression of image noise is achieved. Unlike linear detection methods using correlative processing (which would attach more significance to the high amplitude coefficients), in this case, large magnitude coefficients affect the result of the detection process in the same way as the small magnitude coefficients.

In the data hiding scheme by Wang *et. al.*,[12] the significant wavelet coefficients are altered. The coefficients are modified so that they *quantize* to an even or odd value depending on the bit to be embedded. In[13] Wu *et. al.* introduce a similar scheme based on JPEG quantizers. The signature is introduced in the DCT domain. Chen *et. al*[14] provide a more formal treatment of data hiding techniques, that use the quantization index to embed bits (methods which force the quantized *indices* to take a desired value depending on the information signal to be embedded).

4. DATA HIDING AS A SIGNALING TECHNIQUE

Consider a (metric) space $\mathcal{I}$ of $M \times N$ images, b bits per pixel (2^{MNb} possible images). Alternately, every point in $\mathcal{I}$ is an $M \times N$ image. Let $\mathbf{I}$ represent the original (cover) image. To embed a bit sequence $\mathbf{b}$ of length n_b, we should be able to define a constellation with a minimum of 2^{n_b} points. The problem now is the choice of a signaling set or a signal constellation, such that *any* valid image (any point in $\mathcal{I}$) can be "moved" to a point in the constellation corresponding to the arbitrary bit sequence to be hidden, *without perceptual distortion* of the original image. The new point to which the image is moved is then the stego image, $\hat{\mathbf{I}}$. Obviously, the space of images should be *tiled* by the constellation to achieve this with a reasonably low amount of distortion. As we also need the hidden bits to survive some distortion that $\hat{\mathbf{I}}$ is expected to undergo before it reaches the detector, we need the points of the constellation to be well separated.

4.1. Data Hiding and Ideal Compression

One of the most common sources of distortion introduced in $\hat{\mathbf{I}}$ is lossy compression. As the image is represented by fewer bits in the compressed domain, many original image points are mapped by the compressor to one image point after (lossy) compression and decompression. As an example, in Figure 3, all points in the range R are mapped to a single point D. Consider an image A (represented by $+$) in the region R. Let us say we want to hide one bit of information in the image A that would survive compression. The space $\mathcal{I}$ is tiled by two regions that represent 0 or 1. For example, if the image A is located in a region representing 0, it could be left intact if the bit to be hidden is 0. To hide a bit 1 however, A has to be moved to a point B (represented by $*$) which simultaneously belongs to region 1 and lies *outside the range* R, so that after compression (and decompression), the image is mapped to a different point B_1. To hide n_b bits in an image which can survive compression, the image has to be distorted such that after decompression the image is mapped to any of 2^{n_b} possible points. In other words, the space of images has to be tiled by 2^{n_b} regions.

Now it is easy to see that no data hiding would be possible with an *ideal* compressor. If δ_t is the visual distortion permitted (δ_t *may not* be a measure of the mean square error), then there exists a finite number of points to which the original image may be "moved". However, an ideal compressor with the same threshold $\delta_c = \delta_t$ would map all such points to a *single point* in the space of decompressed images! So unless we employ different standards (a measure of δ) for the quality of the image after data hiding and that for the decompressed image, (or unless $\delta_c > \delta_t$), *no data hiding would be possible with ideal compressors.*

4.2. Data Hiding With Known Compression

When the compression method the image is likely to undergo is known in advance, it is very easy to design efficient data hiding techniques. For example, if it is known in advance that the images will only undergo DCT based JPEG compression with the *default quantization matrix*, and that image is not expected to undergo compression more severe than quality factor 50%, it is possible to achieve data hiding capacities of the order of 6000 to 16000 bits per 256×256 images![15]

Again, consider the space $\mathcal{I}$ of original images. When the compression method is known, we make use of the fact that points (or "states") R_1 to R_n are mapped to the same points R_1 to R_n in the space of decompressed images. Therefore, the number of valid "states" of the compression technique that lie within an envelope of "unnoticeable visual distortion" yields a direct measure of the number of bits that can be hidden in an image. The problem of designing data hiding schemes becomes more difficult if the data hiding scheme has to survive *different types* of compression schemes

The problem becomes more complicated if the hidden data has to survive *multiple* compression methods. To see how the requirement of robustness to different compression schemes (simultaneously) can drastically reduce the data hiding capacity, consider 3 compression schemes $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$. In Figure 5 the '+'s denote points in $\mathcal{I}$ which are permissible $\mathcal{C}_1$-compressed (and decompressed) images. Similarly filled 'o's and '*'s stand for $\mathcal{C}_2$ and $\mathcal{C}_3$ compressed images. Let A be the original image $\mathbf{R}$ an envelope of the possible points A could be moved to, without noticeable visual distortion. If the data hiding scheme has to survive only one of the 3 compression schemes, one can see that there are roughly 9 points to which the image can be moved in each case. However, if the hidden data has to survive any compression scheme, then the number of possible states (2^p, where p is the number of bits that can be hidden) is limited to the number of non-intersecting regions (marked by dotted circles) where at least one of the valid points of different compression schemes can be found.

If the exact effect of compression is not known (the valid states are not known *a priori*), the job of designing efficient data hiding techniques warrants a totally different approach. As one has no idea of the "valid" compression points, the centers of the non-intersecting regions have to be considerably well separated to ensure that *at least one valid compression point of all compression methods* lie in each hyper-sphere. However, this might imply considerable distortion of the image in the mean-square sense, to embed the information bits. In Ramkumar *et. al.*[15] we point out different possible ways that would enable us introduce *significant* distortion in the mean-square sense without affecting the visual fidelity of the image.

5. SIGNALING FOR DATA HIDING

The next question that arises is

- Given a sequence of bits $\mathbf{b}$ of length K (typically, $K << MN$), how do we *map* the bit sequence to the new "state" to which the image should be moved?

Let $\mathcal{E}_0$ and $\mathcal{D}_0$ be such that,

$$\hat{\mathbf{I}} = \mathcal{E}_0(\mathbf{b}, \mathbf{I}) \quad \mathbf{b} = \mathcal{D}_0(\hat{\mathbf{I}}) \tag{4}$$

$\mathcal{E}_0$ and $\mathcal{D}_0$ together define the *signaling scheme* used for the steganographic communication, or data hiding.

As before, let the received image, $\tilde{\mathbf{I}} = \hat{\mathbf{I}} + \mathbf{n}$. Therefore,

$$\tilde{\mathbf{b}} = \mathcal{D}_0(\tilde{\mathbf{I}}) = \mathcal{D}_0(\hat{\mathbf{I}} + \mathbf{n}) \tag{5}$$

It is usually a better idea to embed the information bits in some transform domain. Let $\mathcal{T}$ be some invertible transform such that $\mathbf{C} = \mathcal{T}(\mathbf{I})$. The embedding and detecting operations now take the form

$$\hat{\mathbf{C}} = \mathcal{E}_0(\mathbf{b}, \mathbf{C}) \quad \mathbf{b} = \mathcal{D}_0(\hat{\mathbf{C}}) \ . \tag{6}$$

Let the transform coefficients of the received image be $\tilde{\mathbf{C}} = \hat{\mathbf{C}} + \nu$. Where ν is the additive noise introduced by the channel.

$$\tilde{\mathbf{b}} = \mathcal{D}_0(\tilde{\mathbf{C}}) = \mathcal{D}_0(\hat{\mathbf{C}} + \nu) \tag{7}$$

Obviously, we would like to minimize the channel noise ν. We know that most of the noise would be concentrated in the high frequency components of the image. A compression scheme like JPEG quantizes the high frequency coefficients very coarsely. Therefore a significant portion of the noise can be *eliminated* if the data is embedded in the transform domain, and high frequency coefficients are *ignored* (not used for data hiding). We could use a subset (low-to-medium frequencies) $\mathbf{c} \in \Re^D$ of the coefficients $\mathbf{C} \in \Re^{MN}$ for data hiding.

We can now consider any image as a point in D dimensional metric space (of D-dimensional vectors $\mathbf{c}$). We need embedding and detecting functions

$$\hat{\mathbf{c}} = \mathcal{E}_0(\mathbf{b}, \mathbf{c}) \quad \mathbf{b} = \mathcal{D}_0(\hat{\mathbf{c}}) \ . \tag{8}$$

A difficulty with the above formulation of $\mathcal{E}_0$ and $\mathcal{D}_0$ is that while $\mathbf{b}$ is K-dimensional, and $\mathbf{c}$ and $\hat{\mathbf{c}}$ are D-dimensional. Therefore, we introduce another mapping $\mathcal{S}$ such that $\mathbf{s} = \mathcal{S}(\mathbf{b})$ is D-dimensional. As an example, $\mathcal{S}$ may be an error protection scheme that introduces redundancy to map the K-dimensional bit sequence $\mathbf{b}$ to a D-dimensional bit sequence $\mathbf{s}$. Alternately, $\mathbf{s}$ may be a real valued sequence. Therefore, the over-all embedding and detecting sequences take the form

$$\begin{array}{ll} \mathbf{s} = \mathcal{S}(\mathbf{b}) & \hat{\mathbf{c}} = \mathcal{E}(\mathbf{s}, \mathbf{c}) \\ \tilde{\mathbf{s}} = \mathcal{D}(\tilde{\mathbf{c}}) & \tilde{\mathbf{b}} = \mathcal{S}^{-1}(\tilde{\mathbf{s}}) \end{array} . \tag{9}$$

Note that $\mathcal{S}$ is *image independent*. Therefore, the embedding and detecting operations have been split into two parts - an image dependent part ($\mathcal{E}$ and $\mathcal{D}$) and an image independent part $\mathcal{S}$ and $\mathcal{S}^{-1}$. Now, $\mathbf{s}$ represents a point in a signal constellation with *known origin*. The advantages of this approach are

- $\mathcal{E}$ and $\mathcal{D}$ can be implemented as simple *periodic functions* (for example, linear quantizer error).
- A wealth of knowledge exists for the options for the mapping $\mathcal{S}$, which is a "conventional" signaling method in communications.

5.1. Self-Noise Suppression

Figure 6 is an illustration of the function of $\mathcal{E}$ and $\mathcal{D}$. In the figure, $D = 2$. A bit sequence $\mathbf{b}$ is mapped by $\mathcal{S}$ to a point $\mathbf{s}$ in the *bold rectangular region near the origin*. The filled box represents the position of $\mathbf{s}$ in D-dimensional space. The filled circles represent the position of $\mathbf{c}_1$ and $\mathbf{c}_2$ (transform coefficients of 2 images). $\mathcal{E}$ maps $\mathbf{c}_1$ to the point $\hat{\mathbf{c}}_1$ and $\mathbf{c}_2$ to $\hat{\mathbf{c}}_2$. $\mathcal{D}$, on the other hand, would map both $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ to $\mathbf{s}$. We call the pair ($\mathcal{E}$, $\mathcal{D}$) as the *self-noise suppression* (SNS) method.

For linear data hiding techniques, we saw that for the purpose of detection of the hidden bits in an image, the image itself (or its transform coefficients) is noise. The SNS scheme suppresses the image component in $\tilde{\mathbf{c}}$ and extracts the component $\tilde{\mathbf{s}}$ which is needed for extraction of $\mathbf{b}$. The SNS scheme, which obtains the origin of the signal constellation, is characterized by step sizes $\Delta_i, i = 1 \cdots D$ corresponding to each of the D dimensions.

Before we explore specific SNS techniques, consider the linear cover image escrow data hiding method of Figure 8 (a). Let $\nu \sim [f_\nu(\nu), \sigma_\nu^2]$ be additive noise in the channel.

$$\begin{array}{ll} \hat{\mathbf{c}} = \mathbf{c} + \mathbf{s} & \tilde{\mathbf{c}} = \hat{\mathbf{c}} + \nu \\ \tilde{\mathbf{s}} = \tilde{\mathbf{c}} - \mathbf{c} & \tilde{\mathbf{s}} = \mathbf{s} + \nu \end{array} \tag{10}$$

Let the signature be a binary sequence ($s(k) = \pm t_k, k = 1 \cdots D$). For simplicity we further assume that $t_k = t \,\forall\, k$. This is equivalent to the scenario in Figure 8 (b), of transmitting $\mathbf{s}$ over a channel with additive noise variance σ_ν^2.

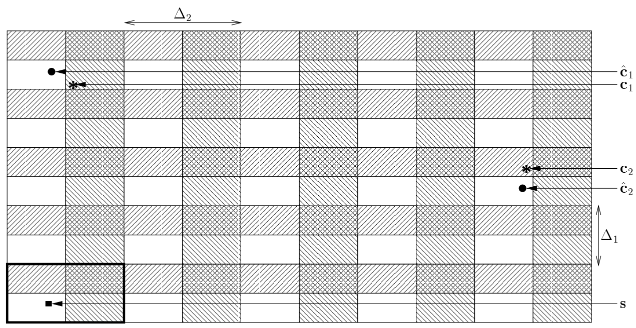


Figure 6. The SNS operators $\mathcal{E}$ and $\mathcal{D}$

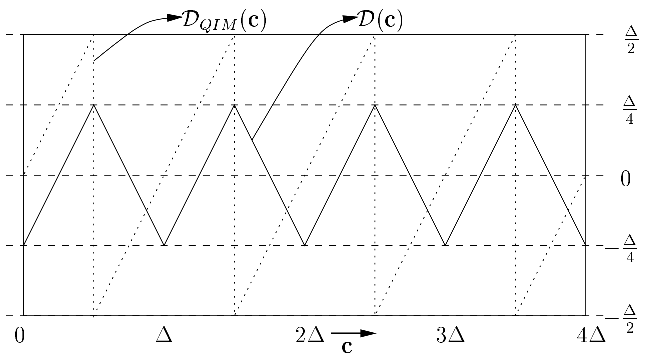


Figure 7. The mapping $\tilde{\mathbf{s}} = \mathcal{D}(\tilde{\mathbf{c}})$ for the proposed scheme (CM - continuous line) and QIM (Dither Modulation - dotted line)

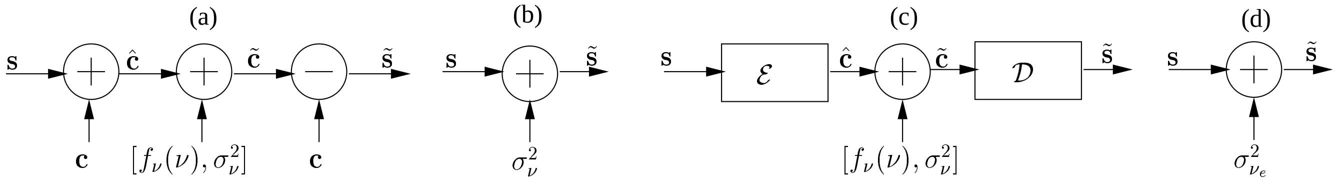


Figure 8. (a) Linear Cover Image Escrow Data Hiding. (b) Equivalent Additive Noise Channel. (c) Non-linear Oblivious Detection Data Hiding (d) Equivalent Additive Noise Channel.

In such a scenario, the normalized inner product of the $\mathbf{s}$ and $\tilde{\mathbf{s}}$ (for sufficiently large D) can be written as

$$\rho = \frac{\mathbf{s}^T \tilde{\mathbf{s}}}{|\mathbf{s}||\tilde{\mathbf{s}}|} = \frac{\int_{-\infty}^{\infty} t(t+\nu) f_\nu(\nu) \mathrm{d}\nu}{\sqrt{\int_{-\infty}^{\infty} t^2 (t+\nu)^2 f_\nu(\nu) \mathrm{d}\nu}}. \tag{11}$$

If the pdf $f_\nu(\nu)$ is *even*, then it can be easily seen that

$$\rho^2 = \frac{t^2}{t^2+\sigma_\nu^2} \quad \text{or} \quad \sigma_\nu^2 = \frac{t^2(1-\rho^2)}{\rho^2} \tag{12}$$

On the other hand for the data hiding scenario in Figure 8 (c), the situation is very different. More specifically, $\sigma_\nu^2 \neq \frac{t^2(1-\rho^2)}{\rho^2}$ where ρ is once again obtained as in Eq (11). Typically,

$$\frac{t^2(1-\rho^2)}{\rho^2} = \sigma_{\nu_e}^2 > \sigma_\nu^2, \tag{13}$$

We may consider $\sigma_{\nu_e}^2$ as the variance of the *equivalent* additive noise. This is a penalty paid for having to "guess" the origin of the signal constellation. The channel of Figure 8 (c) can now be replaced by the channel of Figure 8 (d). We shall see later that for the proposed SNS technique, analytical evaluation of ρ is possible (similar to Eq (11)). From the value of ρ, the equivalent additive noise variance ($\sigma_{\nu_e}^2$) can be evaluated.

5.2. Dither Modulation

Chen *et. al.*[14] presented *dither modulation* technique as a special case of *quantization index modulation* (QIM) for self-noise suppression. In this method,

$$\hat{c}(k) = Q_k(c(k) + s(k)) - s(k), k = 1 \cdots D. \tag{14}$$

where Q_k is a uniform quantizer with step size Δ_k. Detection of the signature can then be performed as

$$\tilde{s}(k) = Q_k(\tilde{c}(k)) - \tilde{c}(k). \tag{15}$$

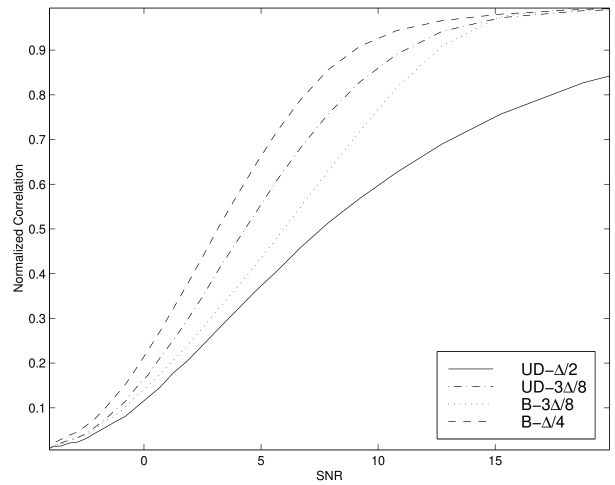


Figure 9. Performance of Dither Modulation for Uniformly Distributed and Binary Signature Sequences

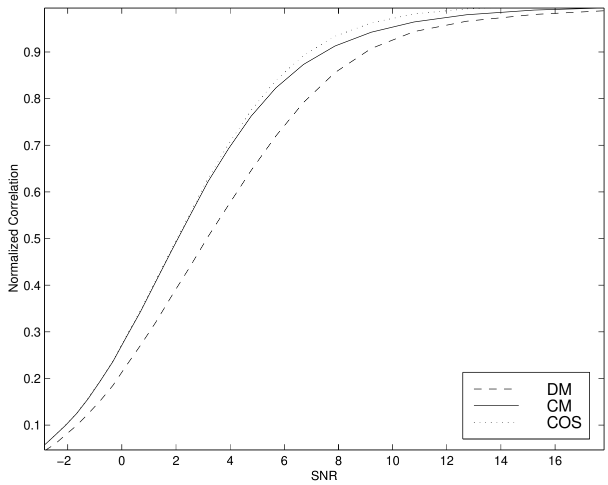


Figure 10. Comparison of DM (QIM), CM-SNS and Cosine modulated SNS

The detector can be represented by the dotted line in Figure 7. Figure 9 illustrates the simulated performance of this SNS technique for $s(k)$

1. uniformly distributed between $-\frac{\Delta}{2}$ and $\frac{\Delta}{2}$,
2. uniformly distributed between $-3\frac{\Delta}{8}$ and $3\frac{\Delta}{8}$.
3. binary $\pm 3\frac{\Delta}{8}$, and
4. binary $\pm\frac{\Delta}{4}$.

The simulations were obtained for Gaussian sequences $\mathbf{c}$ ($\sigma_c = 200$) of length 4096 for $\Delta = 30$. The normalized correlation ρ was obtained by averaging over many realizations of additive Gaussian noise ν.

Note that embedding any signature sequence $\mathbf{s}$ (even a sequence of zeroes!) results in a mean square distortion of $\frac{\Delta^2}{12}$. The SNR in the X-axis therefore represents the ratio of the power of the distortion introduced to embed the signature, viz, $\frac{\Delta^2}{12}$, to the variance of the additive noise σ_ν^2.

$$\text{SNR} \ = 10\log_{10}\frac{\Delta^2}{12\sigma_\nu^2} \tag{16}$$

It is clear that the best performance is obtained for binary $\pm\frac{\Delta}{4}$ sequences. This is due to the fact that as long as $-\Delta_k/4 \leq s(k) \leq \Delta_k/4$, corresponding points in neighboring quantization cells are *maximally separated.*

In Ramkumar *et. al*[16] we suggested an alternate *continuous* modulation (CM) technique (bold line in Figure 7). Let $\mathbf{c}, \mathbf{s} \in \Re^D$. The algorithm for $\mathcal{D}(\tilde{\mathbf{c}})$ of the CM method is as follows:

$$\begin{aligned} q(k) &= \text{rem}(\frac{|\tilde{c}(k)|}{\Delta_k}), \quad k = 1\cdots D \\ \tilde{s}(k) &= (q(k) \geq \frac{\Delta_k}{2}) \ \ ? \ \ (\frac{3\Delta_k}{4} - q(k)) \ : \ (q(k) - \frac{\Delta_k}{4}) \end{aligned} \tag{17}$$

In the above equation $x = (\text{Condition}) \quad ? \quad x_1 \quad : \quad x_2$ stands for "If Condition is true $x = x_1$, else, $x = x_2$".

Let $\mathbf{p} = \mathcal{D}(\mathbf{c})$. To introduce the signature $\mathbf{s}$, we need to modify $\mathbf{c}$ to obtain $\hat{\mathbf{c}}$ such that $\mathbf{s} = \mathcal{D}(\hat{\mathbf{c}})$. To achieve this, the distortion $e(k)$ introduced in coefficient $c(k), k = 1\cdots D$ is equal to

$$|e(k)| = |\hat{c}(k) - c(k)| = |s(k) - p(k)|. \tag{18}$$

The algorithm for embedding the sequence $\mathbf{s}$ in $\mathbf{c}$ is as follows

$$
\begin{aligned}
e(k) &= s(k) - p(k) \\
e(k) &= (p(k) > \Delta_k/2) \quad ? \quad -e(k) \quad : \quad e(k) \\
\hat{c}(k) &= (c(k) \geq 0) \quad ? \quad c(k) + e(k) \quad : \quad c(k) - e(k)
\end{aligned}
$$

Figure 10 compares the performance of the proposed CM-SNS technique (bold line in Figure 7) with that of the dither modulation (DM) technique for $s(k) = \pm\frac{\Delta}{4}$. The better performance of the proposed technique (CM) is not surprising, considering the periodic function used by CM is continuous, as opposed to the DM method (the figure also illustrates the performance of another continuous periodic function - a cosine function (COS) which performs even better than CM. However, due to reasons of analytical tractability, in this paper, we restrict ourselves to CM-SNS).

5.3. Analysis of CM-SNS

We shall now analytically evaluate the equivalent noise for the CM-SNS scheme, when the additive noise in the channel is ν (Figure 8 (c)). Let $\nu \sim [f_\nu(\nu), \sigma_\nu^2]$, and $s(k) \pm \frac{\Delta}{4}$. The expected value of the normalized correlation between $\mathbf{s}$ and $\tilde{\mathbf{s}}$ can be obtained as

$$
\rho_n = \frac{2(\int_0^{\frac{\Delta}{2}} (\frac{\Delta}{4} - \nu) f_\nu(\nu) \mathrm{d}\nu + \int_{\frac{\Delta}{2}}^{\Delta} (\nu - 3\frac{\Delta}{4}) \mathrm{f}_\nu(\nu) \mathrm{d}\nu + \int_{\Delta}^{\frac{3\Delta}{2}} (5\frac{\Delta}{4} - \nu) \mathrm{f}_\nu(\nu) \mathrm{d}\nu + \cdots)}{\sqrt{2(\int_0^{\frac{\Delta}{2}} (\frac{\Delta}{4} - \nu)^2 f_\nu(\nu) \mathrm{d}\nu + \int_{\frac{\Delta}{2}}^{\Delta} (\nu - 3\frac{\Delta}{4})^2 \mathrm{f}_\nu(\nu) \mathrm{d}\nu + \int_{\Delta}^{\frac{3\Delta}{2}} (5\frac{\Delta}{4} - \nu)^2 \mathrm{f}_\nu(\nu) \mathrm{d}\nu + \cdots)}} \tag{19}
$$

For Gaussian $f_\nu(\nu)$, the above integral can be solved and is expressed in terms of the Gaussian error function,

$$
\mathrm{erf}(t) = \frac{2}{\pi} \int_0^t \mathrm{e}^{\frac{-y^2}{2}} \mathrm{dy} \tag{20}
$$

The variance of the equivalent additive noise $\sigma_{\nu_e}^2$ can then be obtained as

$$
\sigma_{\nu_e}^2 = \frac{\frac{\Delta^2}{12}(1 - \rho_n^2)}{\rho_n^2} \tag{21}
$$

Note that we do not use the signature energy $(\frac{\Delta^2}{16})$ in Eq (21) as we did in Eq (13). We use the energy of the *distortion* introduced for embedding the signature instead. Figure 11 is a plot of the normalized correlation *vs* the standard deviation of additive noise for various values of the quantizer step size Δ, obtained from simulations. The stars $(*)$ represent the corresponding values calculated by solving Eq (19). Figure 12 is a plot of $\sigma_{\nu_e}^2$ *vs* σ_ν^2 for various values of Δ.

As mentioned earlier, the choice of Δ dictates the distortion introduced by the embedding function $\mathcal{D}$. If the permitted distortion has a variance γ^2, then $\Delta = \sqrt{(}12\gamma^2)$. This implies that Δ is chosen without any consideration of the expected noise variance σ_ν^2! Obviously, this can not be an optimal solution. This problem can be overcome by introducing *thresholding* in the SNS method.

6. CM-SNS WITH THRESHOLDING

Let γ^2 be the variance of permitted distortion due to data embedding. Let $\Delta_0^2 = 12\gamma^2$. The question we are faced with now is that given γ and some additive noise σ_ν^2, what is the optimal choice of Δ for the SNS method?

We define a modified embedding function $\mathcal{E}_t$ with the same detecting function $\mathcal{D}$. Let $\mathbf{p} = \mathcal{D}(\mathbf{c})$. In the modified embedding method, the distortion $|e(k)|$ introduced in coefficient $c(k)$, is *hard limited* to $-\frac{\beta}{2} < e(k) < \frac{\beta}{2}$, where $\beta < \Delta_0 < \Delta$. The algorithm for embedding the sequence $\mathbf{s}$ in $\mathbf{c}$ is therefore

$$
\begin{aligned}
e(k) &= s(k) - p(k) \\
e(k) &= (e(k) > \frac{\beta}{2}) \quad ? \quad \mathrm{sign}(e(k))\frac{\beta}{2} \quad : \quad e(k) \\
e(k) &= (p(k) > \Delta/2) \quad ? \quad -e(k) \quad : \quad e(k) \\
\hat{c}(k) &= (c(k) \geq 0) \quad ? \quad c(k) + e(k) \quad : \quad c(k) - e(k)
\end{aligned}
$$

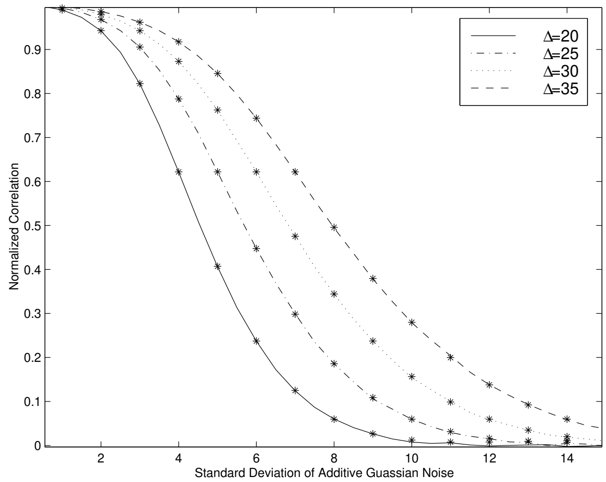


Figure 11. Effect of Additive Gaussian Noise. The lines represent values obtained from simulations. The $*$s represent the values calculated from Eq (19)

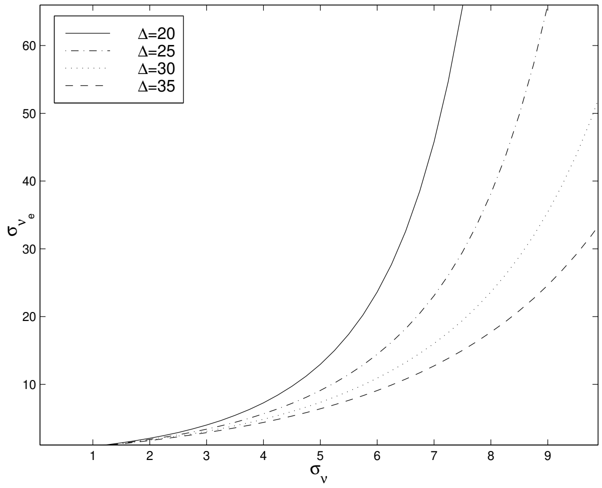


Figure 12. Plot of $\sigma^2_{\nu_e}$ *vs* σ_ν for $\Delta = 20, 25, 30$, and 35.

The distortion $\mathbf{e}$ introduced by the modified embedding function $\mathcal{E}_t$ has a probability distribution and variance given by

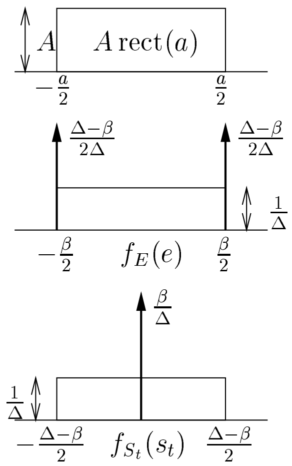
$$\begin{aligned} f_E(e) &= \frac{1}{\Delta}\mathrm{rect}(\beta) + \frac{\Delta-\beta}{2\Delta}\left(\delta(e-\frac{\beta}{2}) + \delta(e+\frac{\beta}{2})\right) \\ \sigma_e^2 &= \frac{\beta^2}{12\Delta}(3\Delta - 2\beta) \end{aligned} \tag{22}$$

Therefore, we can choose $\Delta > \Delta_0$, and $\beta < \Delta_0$, such that the distortion introduced is equal to $\gamma^2 = \Delta_0^2/12$ if

$$\gamma^2 = \Delta_0^2/12 = \frac{\beta^2}{12\Delta}(3\Delta - 2\beta) \tag{23}$$

Note that, with the modified embedding function, if $\hat{\mathbf{c}} = \mathcal{E}_t(\mathbf{c}, \mathbf{s})$, then $\mathcal{D}(\hat{\mathbf{c}}) \neq \mathbf{s}$. The difference $\mathbf{s}_t = \mathbf{s} - \mathcal{D}(\hat{\mathbf{c}})$ has a probability distribution and variance given by

$$\begin{aligned} f_{S_t}(s_t) &= \frac{\beta}{\Delta}\delta(s_t) + \frac{1}{\Delta}\mathrm{rect}(\Delta - \beta) \\ \sigma_{s_t}^2 &= \frac{(\Delta-\beta)^3}{12\Delta} \end{aligned} \tag{24}$$



Alternately, we could assume that a distortion of variance $\Delta^2/12$ (corresponding to $\mathbf{s}$) was introduced in $\mathbf{c}$ by the embedding scheme, *along* with a noise of variance $\sigma^2_{s_t}$, given by Eq (24). Once again, the *equivalent additive noise* due to thresholding can be obtained by a measure of correlation. Let

$$\rho_t = \frac{\frac{\beta}{\Delta}\frac{\Delta}{4} + \frac{2}{\Delta}\int_0^{\frac{\Delta-\beta}{2}}(\frac{\Delta}{4} - x)\mathrm{dx}}{\sqrt{\frac{\beta}{\Delta}\frac{\Delta^2}{16} + \frac{2}{\Delta}\int_0^{\frac{\Delta-\beta}{2}}(\frac{\Delta}{4} - x)^2\mathrm{dx}}} = \frac{\sqrt{3}\beta(2\Delta-\beta)}{\Delta\sqrt{6\beta^3 - \frac{4\beta^3}{\Delta} + \Delta^3}}. \tag{25}$$

The equivalent additive noise, is therefore

$$\sigma^2_{s_{t_e}} = \frac{\Delta^2(1-\rho_t^2)}{12\rho_t^2} \tag{26}$$

The plot of σ_{s_t} *vs* $\sigma_{s_{t_e}}$ for different values of Δ is shown in Figure 13.

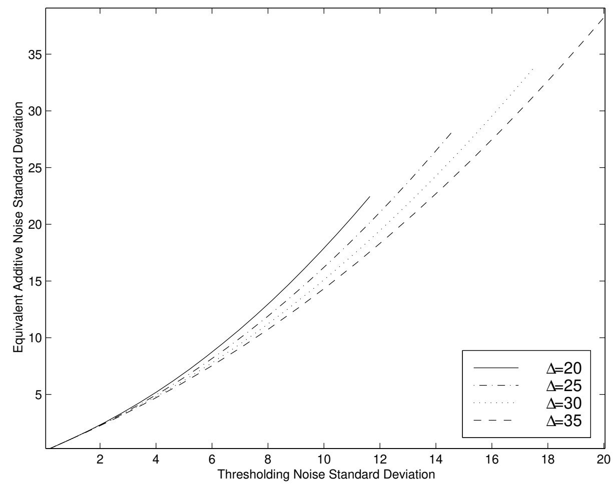


Figure 13. Plot of standard deviation of thresholding noise (σ_{s_t}) *vs* standard deviation of *equivalent noise* due to thresholding, ($\sigma_{s_{t_e}}$).

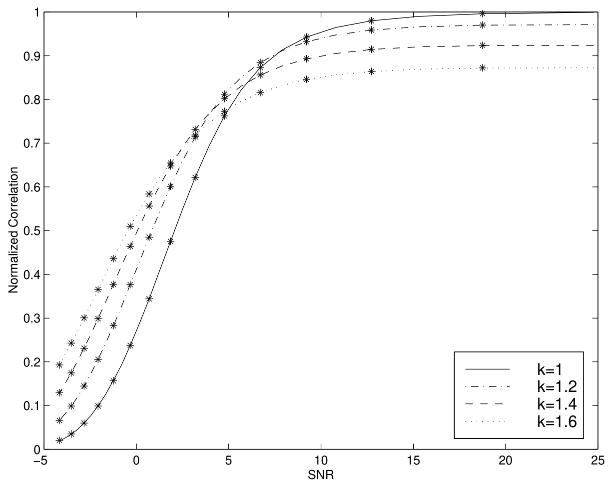


Figure 14. Plot of Correlation *vs* SNR for $k = 1, 1.2, 1.4$ and 1.6

6.1. Combined Effect of Channel Noise and Thresholding Noise

Let the additive noise in the channel is Gaussian with variance σ_ν^2. The thresholding noise has a probability distribution given by Eq (24). The probability distribution of the combined noise, $\mathbf{z} = \nu + \mathbf{s}_t$, *viz* $f_Z(z)$ can be shown to be

$$f_Z(z) = \int_{-\infty}^{\infty} f_\nu(x) f_{S_t}(z-x)\mathrm{d}x = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \int_{z-\frac{\Delta-\beta}{2}}^{z+\frac{\Delta-\beta}{2}} \mathrm{e}^{\frac{-x^2}{2\sigma_\nu^2}} \mathrm{d}x + \frac{\beta}{\Delta}\frac{1}{\sqrt{2\pi\sigma_\nu^2}} \mathrm{e}^{\frac{-z^2}{2\sigma_\nu^2}} \tag{27}$$

$$= \frac{\beta}{\Delta}\frac{1}{\sqrt{2\pi\sigma_\nu^2}} \mathrm{e}^{\frac{-z^2}{2\sigma_\nu^2}} + \frac{1}{2\Delta}\mathrm{erf}\left(\frac{z+\frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_\nu}\right) - \frac{1}{2\Delta}\mathrm{erf}\left(\frac{z-\frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_\nu}\right) \tag{28}$$

The normalized correlation ρ_{nt} and hence the equivalent additive noise can then be obtained by solving

$$\rho_{nt} = \frac{2(\int_0^{\frac{\Delta}{2}}(\frac{\Delta}{4}-z)f_Z(z)\mathrm{d}z + \int_{\frac{\Delta}{2}}^{\Delta}(\mathrm{z}-3\frac{\Delta}{4})\mathrm{f_Z(z)dz} + \int_{\Delta}^{\frac{3\Delta}{2}}(5\frac{\Delta}{4}-\mathrm{z})\mathrm{f_Z(z)dz} + \cdots)}{\sqrt{2(\int_0^{\frac{\Delta}{2}}(\frac{\Delta}{4}-z)^2 f_Z(z)\mathrm{d}z + \int_{\frac{\Delta}{2}}^{\Delta}(\mathrm{z}-3\frac{\Delta}{4})^2\mathrm{f_Z(z)dz} + \int_{\Delta}^{\frac{3\Delta}{2}}(5\frac{\Delta}{4}-\mathrm{z})^2\mathrm{f_Z(z)dz} + \cdots)}} \tag{29}$$

Once again, the solution for the above integral can be obtained in terms of the Gaussian error function, and the equivalent noise variance σ_{nt}^2 is obtained from

$$\sigma_{nt}^2 = \frac{\Delta^2(1-\rho_{nt}^2)}{12\rho_{nt}^2}. \tag{30}$$

Figure 14 is a plot of the normalized correlation ρ_{nt} versus the SNR for values of $k = \Delta/\Delta_0$ ranging from 1 to 1.6. The $k = 1$ case corresponds to no thresholding (or $\Delta_0 = \Delta = \beta$). For all four plots, $\Delta_0 = 30$. This implies that the distortion introduced to embed the signature is the same for all the four cases. The plots have been obtained from simulations. The $*$'s represent the corresponding values obtained from calculating the normalized correlation from Eq (29).

The steps to obtain the optimal parameters for CM-SNS, for a given permitted distortion γ^2 and additive noise variance σ_ν^2, is as follows:

- Obtain $\Delta_0^2 = 16\gamma^2$.
- Let $k > 1$ such that $\Delta = k\Delta_0$.

- Evaluate β under the constraint of Eq (23).
- Choose k to maximize ρ_{nt} (Eq (29)).

7. CONCLUSIONS

We have presented a new self-noise suppression technique suitable for blind steganography. Exact analysis of the proposed technique is carried out and is found to agree extremely well with simulations. Issues regarding optimal choices of the parameters for the SNS technique are also analyzed. Note that the proposed method is equally applicable for signals like video, music / speech, though we have assumed the cover object to be an image for the purpose of illustration. However, the signature sequence has been assumed to be binary, in all the presented analysis. This is because the choice of $\pm\frac{\Delta}{4}$ is optimal for the SNS scheme. There may be some situations,[15] however, where the optimality of the conventional signaling part $\mathcal{S}$, (which follows SNS) may demand the use of non-binary signatures. Therefore joint optimization of $\mathcal{S}$ and $\mathcal{E}$ (or $\mathcal{E}_t$) may be called for. We also noted from Figure 10 that other modulation techniques (like the Cosine modulation) are possible, which may yield better performance. Extension of this work for non binary signatures and other continuous periodic functions for SNS, is underway.

REFERENCES

1. R.J. Anderson, F.A. Petitcolas, "On the Limits of Steganography", IEEE Journal on Selected Areas of Communications Vol **16**, No 4, pp 474-481, May 1998.
2. R.G. van Schyndel, A.Z. Tirkel, C.F. Osborne, "A Digital Watermark", IEEE International Conference on Image Processing, Vol **2**, pp 86-90, 1994.
3. G. Caronni, "Assuring Ownership Rights for Digital Images", Proceedings of Reliable IT Systems, VIS-95, Vieweg Publishing Company, 1995.
4. M. Ramkumar, A.N. Akansu, "Information Theoretic Bounds for Data Hiding in Compressed Images", IEEE Second Workshop on Multimedia Signal Processing, Redondo Beach, California, USA, pp 267-272, Dec 1998.
5. M.Ramkumar, A.N. Akansu, A. Alatan, "On the Choice of Transforms for Data Hiding in Compressed Video", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol **VI**, pp 3049-3052, March 1999. ICASSP-99.
6. M.Ramkumar, A.N Akansu, "Capacity Estimates for Data Hiding in Compressed Images", Submitted to IEEE Trans. on Image Processing, 1998.
7. I.J. Cox, J. Kilian, F.T. Leighton, and T.G. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Transactions on Image Processing, **6** (12) pp 1673-1687, 1997.
8. W. Zeng, B. Liu, "On Resolving Rightful Ownerships of Digital Images by Invisible Watermarks", Proceedings of ICIP, vol **1**, pp 552-555, 1997.
9. A.G. Bors, I. Pitas, "Image Watermarking Using DCT Domain Constraints", IEEE International Conference on Image Processing, Lausanne, Switzerland, **3**, Sep 1996.
10. A. N. Akansu, R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands and Wavelets.* Academic Press Inc., 1992.
11. M.Ramkumar, A.N. Akansu, "A Robust Scheme for Oblivious Detection of Watermarks/ Data Hiding in Still Images", Proceedings of SPIE, Multimedia Systems and Applications, Boston, MA, **3528**, pp 474 - 481, Nov 1998.
12. H.-J.M. Wang, P.-C. Su, C.-C.J. Kuo, "Wavelet Based Digital Image Watermarking", Optics Express, **3**, No 12, pp 491 - 496, Dec 1998.
13. M.Wu, B. Liu, "Watermarking for Image Authentication", Proceedings of IEEE International Conference on Image Processing, October 4-7, 1998, Chicago, Illinois, USA, Vol. **2**, pp 437 - 441.
14. B. Chen, G.W. Wornell, "Digital Watermarking and Information Embedding Using Dither Modulation", IEEE Second Workshop on Multimedia Signal Processing, Redondo Beach, California, USA, pp 273-278, Dec 1998.
15. M.Ramkumar, A.N Akansu, "On the Design of Robust Data Hiding Systems", to be presented in the 33rd ASILOMAR Conference on Signals, Systems and Computers, Pacific Grove, CA, Oct 1999.
16. M.Ramkumar, A.N. Akansu, A.A.Alatan, "A Robust Data Hiding Schemes for Images Using DFT", to be presented in the 1999 International Conference on Image Processing (ICIP-99), Kobe, Japan, Oct. 1999.