International Journal of Image and Graphics
Vol. 5, No. 1 (2005) 191–207

COVER NOISE INTERFERENCE SUPPRESSION IN MULTIMEDIA DATA HIDING

LITAO GANG
Infodesk Corp, 660 Whiteplain Road, Tarrytown, NY 10592, USA
lgang@idsk.com

ALI N. AKANSU
ECE Department, New Jersey Institute of Technology,
University Heights, Newark, NJ 07102, USA

Received 6 June 2003
Revised 16 November 2003
Accepted 9 December 2003

In this paper, we investigate the general problem of data hiding and propose an approach for effective cover noise interference rejection in oblivious applications. We first evaluate the performance in the commonly used direct sequence modulation approach where a low-power signal is embedded into the original cover signal. The optimal detection is derived and its performance is analyzed. Second, we study a novel approach in oblivious data hiding and evaluate its performance and compare it with existing algorithms. Both simulation studies and empirical data hiding results validate its efficiency in the multimedia oblivious applications.

Keywords: Data hiding; watermarking; ML detection; data security.

1. Introduction

In the Internet era, copyright and data integrity protection has raised a great concern. Digital watermarking and information embedding technologies provide one of the potential solutions to these problems. Research in these areas has gained substantial attention and recent years have seen a flurry of activities on this subject reported in the literature. The purpose of the data hiding is to embed information into an original content cover signal without perceptual artifacts. The information can be extracted to solve ownership disputes, track piracy, or identify malicious tampering. Watermarking can be regarded as a special case of data hiding. In the following discussions, these two terms are used interchangeably. And the algorithms discussed in data hiding are readily applicable to watermarking. *Robust and fragile* watermarking techniques have different contexts in practice. Secure and resilient to attacks, the former is often applied in copyright management. While the latter is usually employed for content tampering detection. In this paper, our discussion is concentrated on robust watermarking techniques.

Embedded message extraction has different requirements. In *escrow* applications, the embedded message is retrieved with the assistance of the original cover signal. In *oblivious* cases, the message is decoded without reference to the original signal. In most applications, the latter is more meaningful because the original cover is usually unavailable at the decoder. This poses a big challenge in message decoding. The cover signal is unknown and acts as a kind of noise, called *cover noise*. Its rejection is a big concern for reliable message extraction. Costa[10] proved at least in theory the oblivious data hiding can achieve the same capacity as an escrow case. To gain this goal, the embedding involves design of an optimal code book by taking advantage of the statistical property of the cover signal. Some recent research has explored the problem in length.[11,18] However, the optimal code book design is pretty demanding, if not impossible at all.[9] It is often more essential to investigate embedding and decoding performance in the deployment than the theoretical hiding capacity calculation. Our focus in this paper is on the empirical cover noise suppression techniques and multimedia applications.

Among the proposed message embedding schemes, the direct sequence (DS) modulation approaches are extensively studied and widely employed. The algorithms embed a key-generated vector in the cover signal. Perceptual models are exercised to reduce the artifacts. Although originally proposed for escrow applications, the DS schemes have also been used in oblivious data hiding cases, in images,[20] video.[12,13,22] and audio.[3,15,17,23] The advantages in this kind of schemes are easy distortion control and resilience to additive noise attacks. Complexity of the schemes includes computation cost and rigid re-synchronization requirement. Because of the imperceptible requirement, the watermark signal is of limited energy, which dramatically degrades performance in oblivious applications. In the first half of the paper, we derive the optimal detectors and analyze the detection bit error rate (BER) assuming Gaussian distributed cover signal, Both theoretical analysis and simulation studies highlight inefficiency in the cover noise suppression. In the second half, a simple nonlinear data hiding algorithm, *set partitioning*, is suggested. Distortion is calculated and suboptimal detectors are worked out. Some encouraging experiment results with audio and image content are presented subsequently.

This paper is organized as follows. In Sec. 2, the performance of a widely used DS modulation is analyzed. The ubiquitous correlator type decoder is not optimal. We suggested a simple fix-up, and the maximum likelihood (ML) detector is reached. Our assumption of Gaussian distributed cover signal is a trade-off between ease of analysis and good statistical approximation. The inferior host noise rejection stems from low signal noise ratio (SNR). Instead of linearly superimposing a watermark signal into a cover signal, we come forward with a simple nonlinear hiding scheme. Its signaling and distortion calculation is conducted in Sec. 3. The optimal and suboptimal detection in the additive white Gaussian noise (AWGN) channel is visited in Sec. 4. Our multimedia experiment results are presented in Sec. 5. And further discussion and comparison with existing algorithms is addressed in Sec. 6. Some conclusions and future work are summarized in the last section.

2. Cover Noise Interference in Direct Sequence Modulation

2.1. *Multiplicative embedding and detection*

Direct sequence modulation is one of the earliest data hiding schemes. The basic scheme modulates a small value hiding signature vector $\mathbf{w}$ into a cover signal vector $\mathbf{c}$,

$$\mathbf{x} = \mathbf{c} + b\mathbf{w}\,, \tag{1}$$

where b is the antipodal information bit taking the value of either $+1$ or -1 to represent bit value 1 or 0. And $\mathbf{x}$ is composed of selected coefficients in a data hiding domain, while $\mathbf{w}$ is independent of $\mathbf{x}$ and can be Gaussian or uniformly distributed. In order to keep the embedding imperceptible, some adaptation is needed to constrain distortion strength. One revised version is,

$$x_i = c_i + bw_i \cdot \alpha c_i\,, \tag{2}$$

where α is the gain factor for artifact control. The justification is the human perceptual systems are more sensitive to relative, rather than absolute distortion. The hiding signal is scaled by α and proportional to the cover signal. This can keep the distortion under JND (Just Noticeable Distortion) to meet transparency requirement. On the other hand, it is desirable to maximize embedding energy to enhance the robustness. In the following discussions, we study the deep embedding algorithm where $\mathbf{w}$ is an antipodal random sequence, i.e. w_i taking the value of either $+1$ or -1.

Given the cover signal $\mathbf{c}$ the information bit can be extracted via correlation. In an oblivious application scenario where $\mathbf{c}$ is unavailable, two fix-ups can be exercised. One keeps the correlation type detector but altering the embedding rule; The other derives a new detector while keeping the embedding intact. We will address the problem following these two paths.

In the linear oblivious data hiding, the challenge is cover noise rejection. For simplicity, we neglect the noisy channel effect. Therefore the received the signal at the decoder side is assumed just the data embedding output, $\mathbf{r} = \mathbf{x}$.

2.2. *A revised direct sequence embedding*

A quick fix-up for (2) introduces the absolute value operation:

$$x_i = c_i + bw_i \cdot \alpha|c_i|\,, \tag{3}$$

and a correlator-alike detector is

$$q = \mathbf{w}^T \cdot \mathbf{r} = \mathbf{w}^T\mathbf{c} + b\alpha\mathbf{w}^T|\mathbf{c}| \approx \alpha b \sum_{i=0}^{N-1} w_i|c_i|\,. \tag{4}$$

The above approximation holds as long as c_i is a zero-mean random variable, and $\mathbf{c}$ is independent of $\mathbf{w}$. Although the detection works its performance is inferior. In fact, the optimum ML detector can be derived as follows. Considering the simple hypothesis testing:

$$H1\colon\ r_i = c_i + k_i|c_i| \quad \text{versus} \quad H0\colon\ r_i = c_i - k_i|c_i|\,, \tag{5}$$

where $\mathbf{k} = \alpha\mathbf{w}$, k_i is either $+\alpha$ or $-\alpha$.

The ML ratio is calculated as

$$R = \frac{P(H1|\mathbf{r})}{P(H0|\mathbf{r})}\,. \tag{6}$$

For simplicity, we assume that $\mathbf{c}$ is composed of N components which are i.i.d. Gaussian distributed, $c_i \sim N(0, \sigma^2)$. The conditional pdf can be expressed as

$$f(r_i|H1) = \frac{1}{\sqrt{2\pi}\sigma(1 + s(r_i)\cdot k_i)} \cdot \exp\left[\frac{-r_i^2}{2\sigma^2} \cdot \frac{1}{(1 + s(r_i)\cdot k_i)^2}\right]\,, \tag{7}$$

where $s(\cdot)$ is the sign function

$$s(x) = \begin{cases} +1\,, & x > 0 \\ -1\,, & x < 0 \\ \ \ 0\,, & x = 0 \end{cases}\,. \tag{8}$$

Similarly, $f(r_i|H0)$ can be obtained. If we assume $H1$ and $H0$ have equal *a priori* probabilities, $P(H0) = P(H1)$, the ML ratio on r_i can be expressed as

$$R_i = \frac{P(r_i|H1)}{P(r_i|H0)} = \left(\frac{1 - s(r_i)\cdot k_i}{1 + s(r_i)\cdot k_i}\right) \cdot \exp[-\beta r_i^2 \cdot s(r_i)\cdot s(k_i)]\,, \tag{9}$$

where $\beta = \gamma\frac{1}{\sigma^2}$ and $\gamma = \frac{1}{2(1+\alpha)^2} - \frac{1}{2(1-\alpha)^2}$.

The ML ratio (6) is finally obtained as

$$R = \prod_{i=0}^{N-1}\left(\frac{1 - k_i}{1 + k_i}\right)^{s(r_i)} \cdot \exp\left[\sum_{i=0}^{N-1} -s(r_i)\cdot s(k_i)\cdot r_i^2\beta\right]. \tag{10}$$

The above calculation is quite tedious. One straightforward observation is that for sufficiently large sequence length N, $\mathbf{w}$ has roughly the same count of -1's and $+1$'s,

$$\prod_{i=0}^{N-1}\left(\frac{1 - k_i}{1 + k_i}\right)^{s(r_i)} \approx 1\,. \tag{11}$$

Under this approximation, a computation-efficient suboptimal detection statistic results as,

$$R = \sum_{i=0}^{N-1} -\gamma r_i^2 \cdot s(r_i)\cdot s(k_i)\,. \tag{12}$$

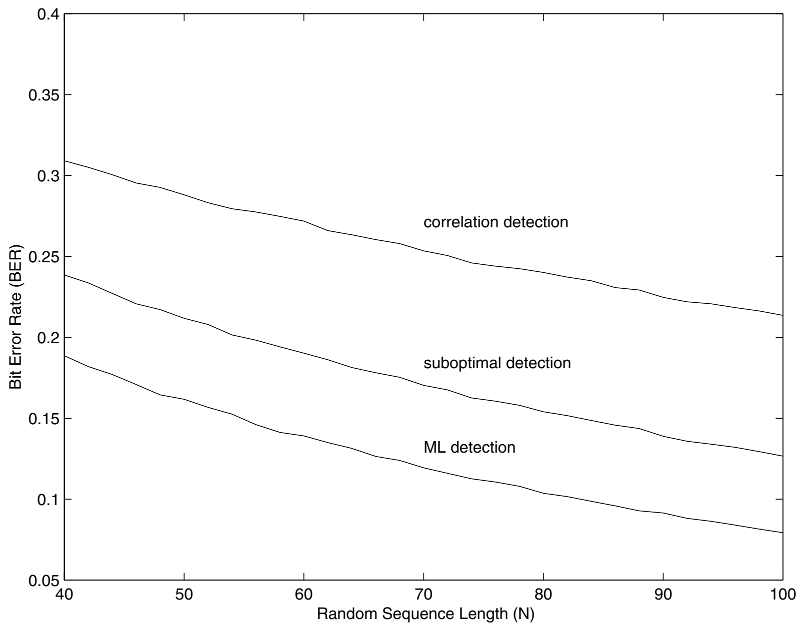


Fig. 1. Detection performance comparison.

The decision threshold can be conveniently selected as $R = 0$. The sub-optimal detector (12) has comparable computation complexity as (4). Nevertheless it outperforms the latter as depicted in Fig. 1. But it is inferior to the optimum detector (10) due to the approximation (11). In our simulation, the threshold ratio value is chosen as $\alpha = 0.1$, and the original coefficient is Gaussian distributed with variance $\sigma^2 = 50^2$.

Any data hiding scheme alters some statistical properties of the original signal. From the embedding operation it is obvious that the message embedding impact is the change of variance in c_i. Intuitively speaking, the ML detector based on the distinction of variance outperforms the correlator-type detector based on the mean value.

2.3. *ML detection in DS embedding*

Alternatively, we try to derive the ML detector for the embedding rule (1). The hypotheses testing at the decoder side is

$$H1\colon\ r_i = c_i + k_i c_i \quad \text{versus} \quad H0\colon\ r_i = c_i - k_i c_i\,. \tag{13}$$

After embedding, the variance of the modified coefficients is equal to $\sigma_1^2 = (1+\alpha)^2\sigma^2$ or $\sigma_0^2 = (1-\alpha)^2\sigma^2$.

Similar to the analysis aforementioned, the ML ratio on $\mathbf{r}$ can be calculated as

$$R = \frac{P(\mathbf{r}|H1)}{P(\mathbf{r}|H0)} = \prod_{i=0}^{N-1} \frac{1-k_i}{1+k_i} \cdot \exp\left[\sum_{i=0}^{N-1} -\gamma r_i^2 \cdot s(k_i)\right]. \tag{14}$$

For an even value of N, suppose the random sequence is generated as $\mathbf{w} = [\mathbf{p}, -\mathbf{p}]$ where $\mathbf{p}$ is an $N/2$ length random vector, the above can be further reduced to a neat result,

$$R = \sum_{i=0}^{N-1} s(k_i) \cdot r_i^2 \gamma\,, \tag{15}$$

considering $\prod_{i=0}^{N-1}(\frac{1-k_i}{1+k_i}) = 1$. If $R > 0$, the bit value 1 is decided, 0 otherwise.

The BER performance can be analyzed as follows. Divide c_i into two sets 0 and 1 based on the value of k_i, $c_i \in$ set 0 if $k_i > 0$ and $c_i \in$ set 1 otherwise. The test statistic (15) is rewritten as

$$R = \sum_{\{r_i \in \text{set } 0\}} r_i^2 \gamma - \sum_{\{r_i \in \text{set } 1\}} r_i^2 \gamma\,. \tag{16}$$

Denote variable

$$t_j = \sum_{r_i \in \text{set } j} r_i^2\,, \quad (j = 0, 1)\,. \tag{17}$$

It can be proved that t_j ($j = 0$, 1) is $M = N/2$ degree of freedom Γ distributed with pdf expressed as[14]

$$f(t_j) = \frac{t_j^{M/2-1} \cdot e^{-\frac{t_j}{2\sigma_j^2}}}{\sigma_j^M \cdot 2^{M/2} \cdot \Gamma(M/2)} = A_j \cdot t_j^{n-1} e^{-C_j t_j}\,, \quad (j = 0, 1)\,, \tag{18}$$

where $A_j = \frac{1}{\sigma_j^M \cdot 2^{M/2} \cdot \Gamma(M/2)}$, $C_j = \frac{1}{2\sigma_j^2}$ and $n = M/2 = N/4$.

Suppose the bit value 1 is embedded, BER turns out to be

$$\begin{aligned} \text{BER} = P(t_1 < t_0) &= \int_0^{+\infty} f(t_0)dt_0 \cdot \int_0^{t_0} f(t_1)dt_1 \\ &= \int_0^{+\infty} f_0(t_0) \int_0^{t_0} A_1 t_1^{n-1} e^{-C_1 t_1} dt_1 dt_0\,. \end{aligned} \tag{19}$$

For an integer n, using the formula

$$\int x^n e^{-ax} dx = -\frac{e^{-ax}}{a^{n+1}} \cdot [(ax)^n + n(ax)^{n-1} + n(n-1)(ax)^{n-2} + \cdots + n!]\,, \tag{20}$$

and

$$\int_0^{+\infty} s^n e^{-as} ds = \frac{n!}{a^{n+1}}\,, \tag{21}$$

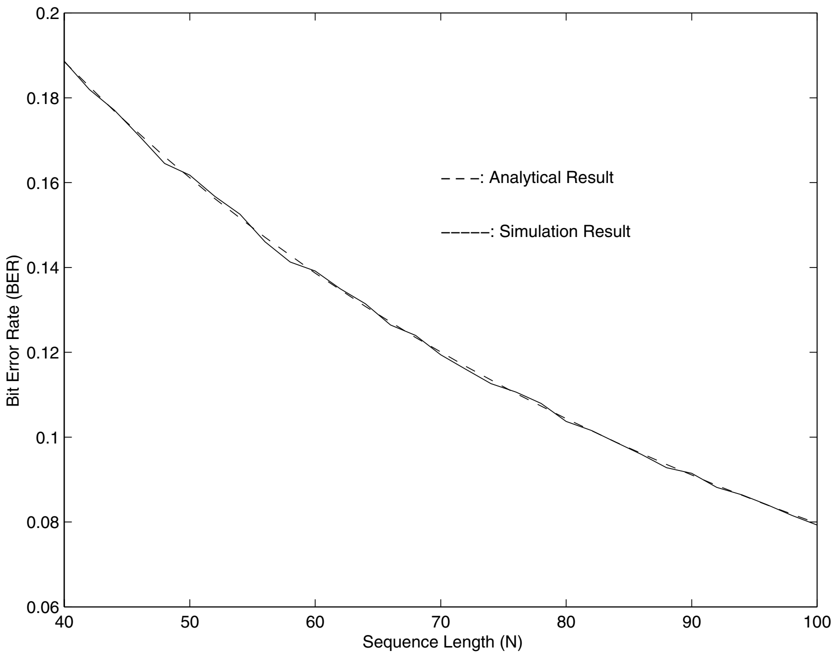


Fig. 2. Performance in the linear modulation.

after some algebraic steps, BER yields as

$$\mathrm{BER} = \left[\left(1 + \frac{C_0}{C_1}\right)(2n-2)! + \sum_{i=2}^{n} \frac{(n-1)!}{(n-i)!}\left(1 + \frac{C_0}{C_1}\right)^i\right] \cdot \frac{-A_0 A_1}{C_0 + C_1^{2n}} + \frac{A_0 A_1[(n-1)!]^2}{(C_0 C_1)^n}. \tag{22}$$

Figure 2 demonstrates that the theoretical result (22) is a perfect match for the simulation output. Again α is set at 0.1 and $\sigma^2 = 50^2$. Further simulation studies show substantial improvements over the revised DS scheme (3) with similar computation complexity. Still the scheme is not quite successful in oblivious applications. With $\alpha = 0.1$ and $N = 1000$, (22) yields $\mathrm{BER} = 3.91 \cdot 10^{-6}$. To achieve performance $\mathrm{BER} \leq 10^{-9}$, sequence length must be $N > 1800$. This is the inherent limitations in this class of DS schemes.

Our Gaussian distribution assumption for the cover signal may not be accurate. The coefficient c_i is usually in a transform domain in practice. Its statistics can be modeled as generalized Gaussian distribution (GGD) or Laplacian distribution.[1] Recently, Cheng *et al.*[7,8] conducted more vigorous mathematical analysis based on the GGD statistical model. Our clean result (22) can still be used as good performance prediction in the DS schemes. Both our analysis and GGD distribution model reveal the limitations in the DS embedding approaches.

3. Cover Noise Rejection and Set Partitioning

The shortcoming of the DS schemes lies in the inefficiency in the cover noise suppression. The hiding signal energy is much lower than that of the original cover noise, resulting in low SNR. As shown in the above discussion, provided embedded bit value 0 and 1 have equal *a priori* probabilities, the optimal detector always faces the following hypothesis problem:

$$H1:\ \text{bit value 1 is embedded} \quad \text{versus} \quad H0:\ \text{bit value 0 is embedded}\,, \qquad (23)$$

and an optimum detection is based on the ML ratio.

In a noise free scenario where $r = x$, how can the ML decoder make a reliable decision on a given r? Answer is simple and straightforward, just make $H0$ and $H1$ have *no* element in common. Since the conditional probability $P(H0|x) = 0$ or $P(H1|x) = 0$, correct decision is always expected. In order to increase robustness in a noisy environment, we can simply keep the elements in $H0$ and $H1$ some distance apart. This simple idea thus extends to *set partitioning* approach. Two separate sets are constructed on the real axis (Fig. 3). The coefficient after embedding should be kept in a set according to the hidden bit value. To embed bit value 1, the coefficient x should be kept in set 1. If the value of the original coefficient c is already in set 1, no alteration needed. Otherwise it is replaced by the nearest element in set 1 to minimize distortion. Similarly the value of x is kept in set 0 to embed bit value 0.

We may define signal patterns to represent bit values embedded into a vector. For example, to embed 1 bit into a vector $\mathbf{c}$, an antipodal signaling pattern can be defined as follows,

$$\begin{cases} \text{Pattern A (bit 1)}: & [\text{set } 1,\ \text{set } 0,\ \text{set } 1,\ \text{set } 0,\ \text{set } 1] \\ \text{Pattern } -\text{A (bit 0)}: & [\text{set } 0,\ \text{set } 1,\ \text{set } 0,\ \text{set } 1,\ \text{set } 0] \end{cases}. \qquad (24)$$

The resulting vector $\mathbf{x}$ should comply with Pattern A for bit value 1, or Pattern −A for 0. For instance, the resulting sequence should be $x_0 \in$ set 1, $x_1 \in$ set 0, $x_2 \in$ set 1, $x_3 \in$ set 0 and $x_4 \in$ set 1 in order to embed bit value 1.

To calculate the distortion introduced, c is assumed uniformly distributed in the limited range $(-a, a)$. It should be mentioned that this assumption is not very

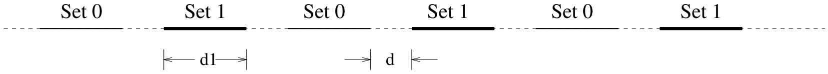


Fig. 3. Set partitioning scheme.

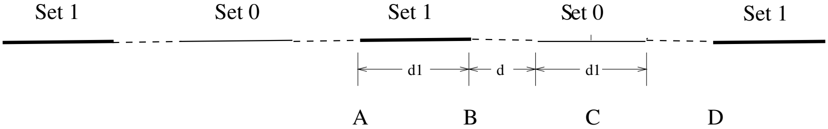


Fig. 4. Average distortion calculation.

accurate in many transform domains. However it is a good compromise between accuracy and ease of analytical work. This assumption is justified in our experiments. As depicted in Fig. 4, suppose the bit value 1 is to be embedded, and denote the error introduced in embedding as $e = x - c$, consider the typical region AD:

If c is in the range AB, no modification is needed, thus $e = 0$. If c is in the range BD, e is uniformly distributed in the range $(-d - d1/2, d + d1/2)$. With the conditional probabilities $P(c \in \text{AB}|c \in \text{AD}) = \frac{d1}{2d1+2d}$ and $P(c \in \text{BD}|c \in \text{AD}) = \frac{2d+d1}{2d1+2d}$, the average distortion yields as

$$D = \frac{(2d + d1)}{(2d1 + 2d)} \cdot \frac{(2d + d1)^2}{12} = \frac{1}{12} \frac{(2d + d1)^3}{(2d + 2d1)}\,. \tag{25}$$

The result holds if a bit value 0 is embedded instead.

4. Set Partitioning Detection and Performance

4.1. *ML ratio calculation and suboptimal detection*

Given the received coefficient r_i after the AWGN channel transmission, the ML ratio is obtained as[16]

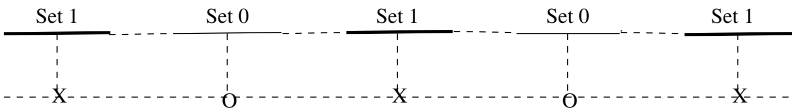
$$R = \frac{P(x_i \in \text{ set } 1|r_i)}{P(x_i \in \text{ set } 0|r_i)} = \frac{\sum_{x_i \in \text{set } 1} P(x_i) f(r_i|x_i)}{\sum_{x_i \in \text{set } 0} P(x_i) f(r_i|x_i)}\,, \tag{26}$$

where $f(r_i|x_i)$ is the Gaussian noise conditional probability function,

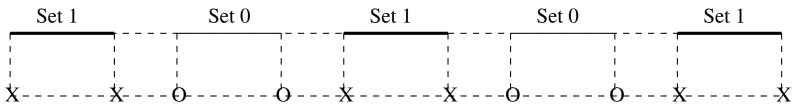
$$f(r_i|x_i) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[\frac{-(r_i - x_i)^2}{2\sigma^2}\right]. \tag{27}$$

Further mathematical analysis does not lead to a closed-form result because (26) involves infinite elements in the two sets. This greatly increases the computational cost and no practical detector can be obtained. In the following discussions, we explore some heuristic detectors employable in practice.

In a sequence embedding, the simplest detection rule is majority vote. This is a hard decision based on individual coefficients. Given a received coefficient r_i, if it has closer minimum distance to the set 1 than to the set 0, it is assumed that transmitted signal x_i originates from set 1. With the embedding (24), if a received sequence pattern is obtained as [set 0, set 0, set 1, set 0, set 0], which is more similar to the pattern A (2 coefficient difference) than the pattern $-$A (3 coefficient difference), the decision is made in favor of the bit value 1. More elaborate approaches make some heuristic simplifications in (26). Our first approximation assumes merely those signals at the centers of the set segments as the transmitted signals. Our second approximation assumes that the end points in those two sets are the only signal candidates because they have much higher transmission probabilities (any coefficient not in the desired set is replace by those values). Hence we have two signal patterns xoxo and xxoo as depicted in Fig. 5.



(a) Suboptimal Detection 1



(b) Suboptimal Detection 2

Fig. 5. Sub-optimal detection in set partitioning.

Given a received coefficient r_i, we can merely consider the "leader" u_i and v_i, of the signal candidates in those two sets, the ML ratio is reduced to

$$R \approx \frac{P(r_i|x_i = u_i)}{P(r_i|x_i = v_i)}\,. \tag{28}$$

In the case where one bit is embedded into a vector, a minimum distance detector can be deployed. With embedding signaling (24), suppose a 5-coefficient sequence $\mathbf{r}$ is received, and the nearest x and o points to r_i are denoted as u_i (in set 1) and v_i (in set 0). Two sequence candidates are constructed as follows,

$$\begin{cases} \text{Pattern A type}: & \mathbf{a} = [u_0, v_1, u_2, v_3, u_4] \\ \text{Pattern } -\text{A type}: & \mathbf{b} = [v_0, u_1, v_2, u_3, v_4] \end{cases}\,. \tag{29}$$

If $\|\mathbf{r} - \mathbf{a}\| < \|\mathbf{r} - \mathbf{b}\|$, $\mathbf{r}$ is more similar to the Pattern A, bit value 1 is decided, or bit value 0 otherwise.

4.2. *BER-DNR performance*

To evaluate the performance, BER is measured versus SNR in AWGN environment and the BER-SNR curve is used as a distortion-robustness benchmark. As data hiding signal energy is equivalent to the distortion introduced, we replace SNR with distortion noise ratio (DNR) which is defined as the ratio of distortion energy D to noise variance σ^2, i.e. DNR $= \frac{D}{\sigma^2}$.

Our simulation studies take the following steps. A Laplacian random sequence $\mathbf{c}$ is generated which is composed of N i.i.d components with zero means and variance value $\sigma_c^2 = 50^2$. A bit value b is embedded into $\mathbf{c}$, generating an embedding result $\mathbf{x}$, then an AWGN vector $\mathbf{n}$ is added $\mathbf{x}$ to simulate channel transmission. The bit is decoded given a received vector $\mathbf{r} = \mathbf{x}+\mathbf{n}$. To validate the algorithms, this procedure is repeated for different values of sequence length N, signaling parameters d, $d1$, and noise variance σ^2.

Figure 6 displays the simulation result for the three different decoders. The sequence length $N = 11$, and signaling $d/d1 = 1$ are chosen. It is evident that

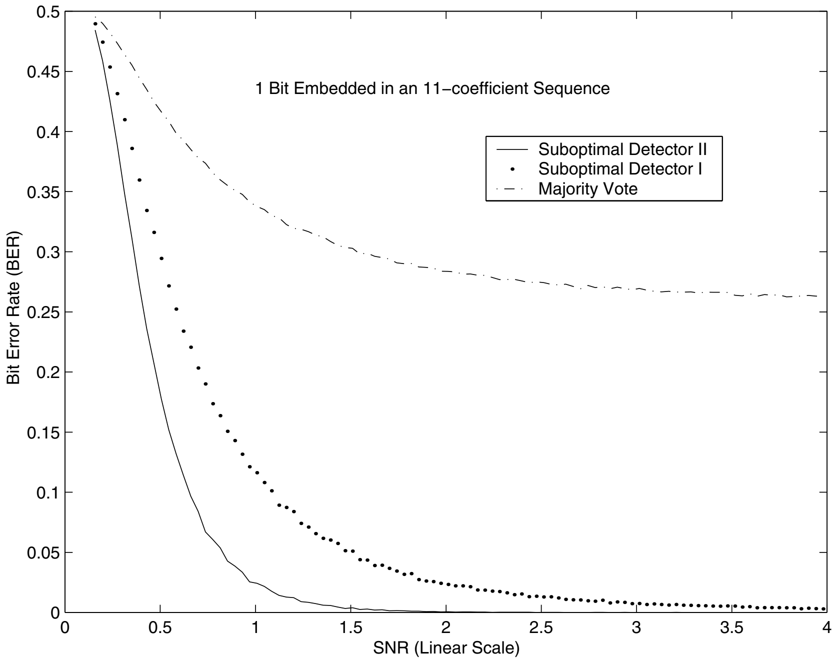


Fig. 6. Detection performance comparison.

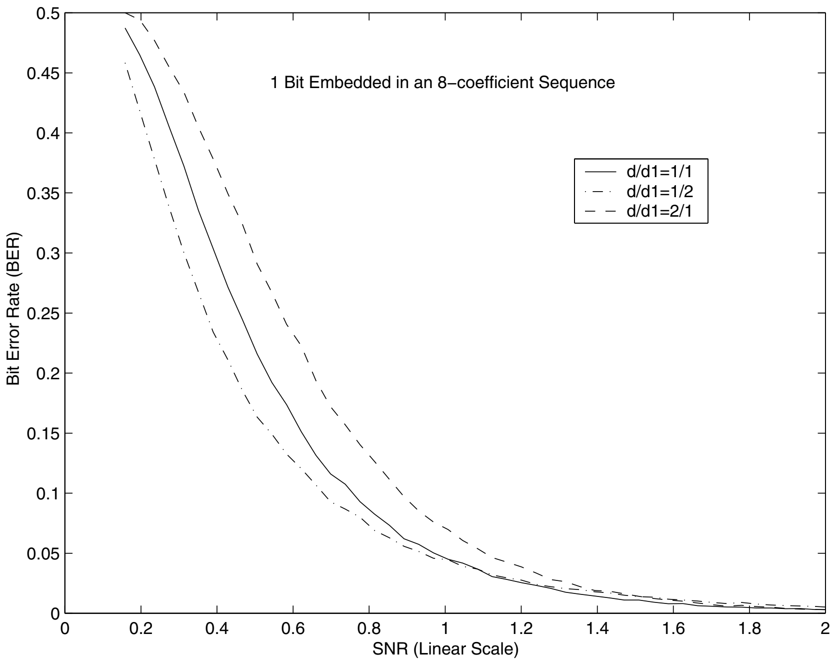


Fig. 7. BER-DNR at different $d/d1$.

the suboptimal detectors outperforms the majority vote detector. This is intuitive because of the superiority of the soft decision decoder. In addition, the results demonstrate that the suboptimal detector II offers remarkable decoding improvements over Detector I. It should be emphasized though, this does not mean the detector II always outperforms detector I. The latter may have better performance for some $d/d1$ signaling schemes. In the following discussions, the detector II is used as the default detector.

In our simulations, it is also established that the BER-DNR curve is only related with the ratio of $d/d1$, not the individual values of d and $d1$. Figure 7 depicts the result of embedding one single bit into an 8-coefficient sequence. It shows that the smaller $d/d1$ performs better at lower DNR. However at higher SNR, larger $d/d1$ is more advantageous. Because data hiding distortion is not expected to be more than moderate or severe compression distortion in practice, resulting in a lower DNR, a smaller $d/d1$ signaling is advisable.

5. Multimedia Application Experiments

Because of the efficient cover noise rejection property, the set partitioning approach can be employed in oblivious multimedia data hiding in place of the DS schemes. In our image data embedding experiments, information bits are embedded into the Discrete Fourier Transform (DFT) amplitude domain. Since the DFT amplitude is assumed less significant perceptually[21] thus more embedding energy is tolerable. Information is embedded into the DFT domain by modifying a set of the DFT amplitude coefficient in an image. In our experiments, the 256×256 Lena image is first divided into 64 sub-images, and DFT transform is taken for each sub image. Second, the bits are embedded into some selected medium-frequency DFT amplitude coefficients. The embedding is applied to the first half of the DFT coefficients and a mirror operation is applied to the second half to keep the DFT symmetric property. Our experiments demonstrates successful extraction and robustness to the JPEG compression, (Fig. 8).

(a) Original Lena

(b) Marked Lena

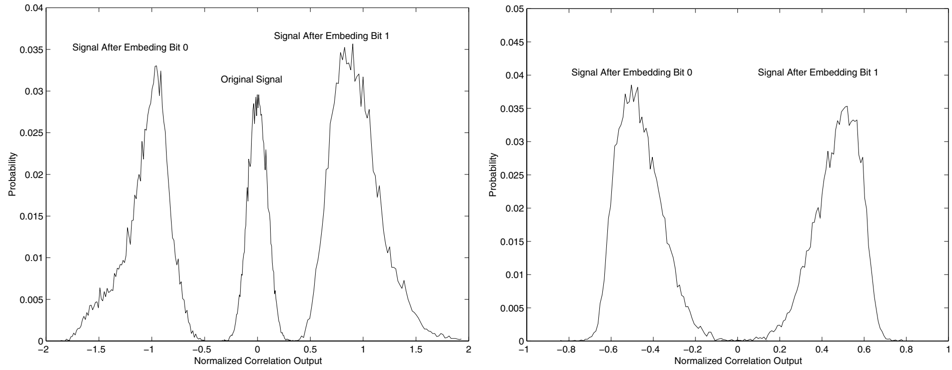
Fig. 8. Lena image before and after embedding.

Though the above algorithm is resistant to the additive noise, it is vulnerable to the scaling attacks. Data embedding in the phase domain can enhance robustness against these attacks. Another advantage is that the DFT phase is perceptually more significant, implying potential resistance again malicious attacks. Our following discussion is on DFT phase embedding in audio signals and the approach can be easily extended to images.

Bender *et al.*[2] proposed a scheme to convey messages in the DFT phase. An audio signal is divided into frames and DFT is applied to each frame. The DFT phases are modified from the second frame but keep the phase difference between frames intact. The underlying justification is human auditory system is more sensitive to the relative phase rather than the absolute phase. In the above scheme the frame continuity is destroyed when the next bit is to be embedded. This may result in a beat pattern. Moreover, the abrupt phase also may alter the signal spectrum. Informal listening tests show that small modifications in DFT phase are inaudible. This property is exploited for data embedding.

The set partitioning scheme can be applied in the DFT phase domain, provided an predefined signal pattern, the original DFT phase value θ_i at one frequency bin is replaced by the nearest element in the set 1 or set 0 accordingly. It is obvious that the DFT phase noise tends to have larger variance if the corresponding amplitude is smaller. After all, large phase noise has little perceptual effect if the amplitude at this frequency bin is sufficiently small. Our heuristic fix-up is a weighted minimum distance detector. Denote the received DFT amplitude and phase as r_i and ϕ_i respectively, employing suboptimal detector I or II, two phase signal candidates can be constructed as $\mathbf{a}$ and $\mathbf{b}$. This distance detector statistic becomes $q = \sum_{i=0}^{N-1} r_i[(\phi_i - a_i)^2 - (\phi_i - b_i)^2]$.

We use a 512-point DFT in our experiments. The phase modification is applied to the frequency bands ranging from 2 kHz to 8 kHz and the modification range is chosen from $\pi/12$ to $\pi/4$. To measure the embedding performance, a normalized



(a) Before 64 k MP3 compression (b) After 64 k MP3 compression.

Fig. 9. Normalized output distribution in phase modulation.

correlation output is defined as

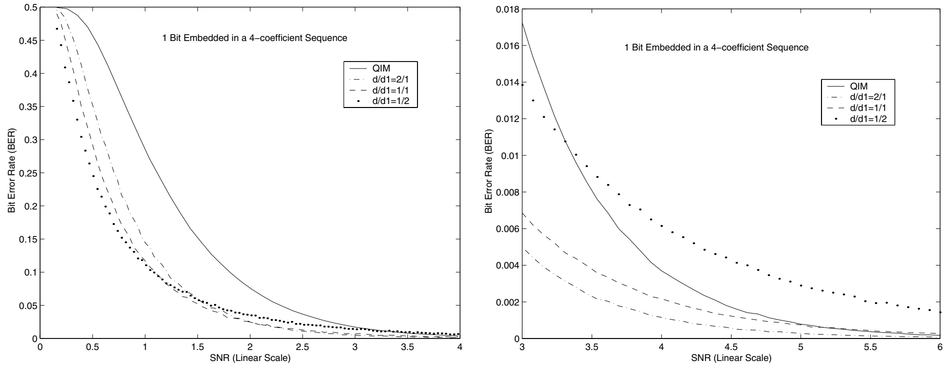
$$q = \frac{\sum_{i=0}^{N-1} r_i[(\phi_i - a_i)^2 - (\phi_i - b_i)^2]}{\|\mathbf{r}\|} . \tag{30}$$

The statistical distribution of q is shown in Fig. 9. q has different statistical properties after embedding bit value 1 or 0. Experimental results indicate that this scheme is effective in oblivious applications. The audio quality can preserved at lower SNR. Our experiments demonstrate robustness to MP3 compression and other attacks.

6. Comparison with Existing Schemes

A very effective oblivious scheme, Quantization Index Modulation (QIM)[4] is a special case of the set partitioning with $d1 = 0$. The resulting coefficient x is a quantized value in that scheme. In contrast, the set partitioning scheme provides us the flexibility to choose different values of $d/d1$. In most applications where DNR is low, signaling with $d/d1 = \infty$ (QIM) is not a very good choice.

In Fig. 10(a), one single bit is embedded into a 4-coefficient sequence. Several $d/d1$ selections demonstrates substantial improvements over the QIM scheme. The performance gain is remarkable at lower DNR. At higher DNR, the QIM scheme only performs slightly better than the signaling scheme $d/d1 = 1/1$ as shown in Fig. 10(b). The set partitioning scheme offers the designer an improvement over the QIM scheme by selecting an appropriate $d/d1$ ratio. The reason to select a smaller $d/d1$ ratio is twofold; first, data hiding operates at lower DNR at practice; second, this selection guarantees a fair performance at severe compressions or tampering attacks. In comparison, QIM approach does not survive very noisy channel. This conclusion rules out QIM in applications. It should be noted that given the same distortion energy, the maximum error e in $d/d1 = 1$ signaling is larger than that in



(a) BER-DNR at lower DNR

(b) BER-DNR at higher DNR

Fig. 10. BER-DNR at different DNR.

the QIM scheme. However, even under the same maximum error constraints, which implies less distortion energy in $d/d1 = 1$ signaling, other signaling selections still demonstrate significant improvements over the QIM scheme at lower DNR.

Chen *et al.*[5,6] suggested post-quantization processing to improve the achievable distortion-robustness trade-offs in the QIM scheme. The compensation term, on the one hand, increases the noise disturbance; on the other hand, reduces the distortion injected. A properly chosen scaling factor α can achieve the optimal distortion-robustness trade-off. Borrowing the distortion compensation idea, a linear post processing can be similarly applied in set partitioning $\hat{x} = x + (1 - \alpha)(c - x)$ where x is the embedding output before post-processing, and $\hat{x}$ is the result after distortion compensation.

We have found in our simulations that with a proper value of α, the set partitioning performance can be further boosted. $d/d1 = \infty$ is often a good choice with the post processing. Some other signaling schemes achieve as well as the DC-QIM. As a general case of DC-QIM approach, the set partitioning scheme has some room to improve performance further.

The challenge in the set partitioning scheme (QIM included) is that accurate mathematical analysis is quite elusive. Up to now, there is not available any accurate results to predict the performance in practice. Some recent theory research[6,18,19] are based on oversimplified assumptions. Much theoretical work addresses the the hiding capacity. Though important in theory, it does not shed much light on the practical data hiding applications. Perez-Gonzalez *et al.*[19] have conducted rigorous analysis on the QIM and DC-QIM schemes. Still more serious work needs to be done, especially the performance analysis where DNR is low. We have observed the performance gap between QIM-alike schemes and the regular signal constellation modulation techniques in practice. The inherent periodic nature in the former technique involves infinite signal points in BER calculation while the latter deals with limited signal points. Therefore the analysis results[4] often deviate from the simulation output. Our data embedding and detection is very heuristic, and performance analysis is crude and needs further refinement. More elaborate detection algorithm is one line of our future research.

7. Future Work and Conclusions

In this paper, the problem of cover noise rejection in oblivious data hiding is addressed in detail. At the beginning, we analyzed the performance in a commonly used DS embedding. Then we derived optimal detectors to boost performance. Both the analytical and simulation results expose the inherent insufficiency in the cover noise suppression. To facilitate the ML ratio decoding, a nonlinear algorithm, set partitioning is proposed. And several heuristic detectors are proposed. Though lack of rigorous mathematical work, the practical approach proposed has been shown promising in the cover noise rejection and offers encouraging results in the oblivious multimedia data hiding applications. Our future research will focus on the more serious performance analysis, improved detection, and appropriate post processing. Another line of future work is integration of DS scheme with set partitioning.

References

1. M. Barni, F. Bartolini and F. Rigacci, "Statistical modelling of full frame dct coefficients," *Proc. EUSIPCO 98, 9th European Signal Processing Conference* **6**(8–11), 1512–1516 (September 1998).
2. W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Technique for data hiding," *IBM System Journal* **35**(3–4), 313–336 (1996).
3. L. Boney, A. H. Tewfik and K. N. Hamdy, "Digital watermarks for audio signals," *Proc. IEEE International Conference on Multimedia Computing and Systems*, pp. 473–480 (June 1996).
4. B. Chen and G. W. Wornell, "Dither modulation: A new approach to digital watermarking and information embedding," *Proc. of SPIE: Security and Watermarking of Multimedia Contents* **3657**, 344–353 (January 1999).
5. B. Chen and G. W. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking," *Proc. of SPIE: Security and Watermarking of Multimedia Contents II* **3971**, 344–353 (January 2000).
6. B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory* **47**, 1423–1443 (May 2001).
7. Q. Cheng and T. S. Huang, "Optimal detection and decoding of multiplicative watermarks in DFT domain," *Proc. Int. Conf. Acoust. Speech, Signal Process* (May 2002).
8. Q. Cheng and T. S. Huang, "Robust optimum detection of transform domain multiplicative watermarks," *IEEE Trans. on Signal Processing* **51** (April 2003).
9. J. Chou, K. Ramchandran and S. Pradham, "Turbo coded trellis-based constructions for data hiding," *Proc. SPIE: Security Watermarking Multimedia Content*, IV, January 2002.
10. M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory* **29**(3), 439–441 (May 1983).
11. J. J. Eggers, R. Bauml, R. Tzschoppe and B. Girod, "Scalar costa scheme for information embedding," *IEEE Trans. on Signal Processing* **51**, 1003–1019 (April 2003).
12. F. Hartung, P. Eisert and B. Girod, "Digital watermarking of MPEG-4 facial animation parameters," *Computers & Graphics* **22**(4), 425–435 (August 1998).
13. F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing* **66**(3), 283–301 (May 1998).
14. C. W. Helstrom, "Probability and stochastic processes for engineers," *Macmillan, New York* (1991).
15. M. Ikeda, K. Takeda and F. Itakura, "Audio data hiding by use of band-limited random sequences," *Proc. ICASSP'99* **4**, 2315–2318 (1999).
16. S. M. Kay, "Fundamentals of statistical signal processing," *Volume 2*, Prentice-Hall PTR (1993).
17. A. N. Lemma, J. Aprea, W. Oomen and L. van de Kerklof, "A temporal domain audio watermarking technique," *IEEE Trans. on Signal Processing* **51**, 1088–1097 (April 2003).
18. P. Moulin and M. Mihcak, "A framework of evaluating the data hiding capacity of image sources," *IEEE Trans. Image Processing* **11**, 1029–1042 (September 2002).

19. F. Perez-Gonzalez, F. Balado and J. R. H. Martin, "Performance analysis of existing and new methods for data hiding with known-host information in additive channel," *IEEE Trans. on Signal Processing* **51**, 960–980 (April 2003).
20. C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Trans. Journal on Selected Area in Communications* **16**(4), 525–539 (May 1998).
21. J. J. K. Ruanaidh, W. J. Dowling and F. M. Boland, "Phase watermarking of digital images," *Proc. ICIP'96*, pp. 239–242 (September 1996).
22. M. D. Swanson, B. Zhu and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE Trans. Journal on Selected Area in Communications* **16**(4), 540–550 (May 1998).
23. I.-K. Yeo and H. J. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," *IEEE Trans. on Speech and Audio Processing* **11**, 381–386 (July 2003).

Litao Gang received his PhD and MS degrees, both in Electrical Engineering, from the New Jersey Institute of Technology, NJ, USA and the Beijing Institute of Technology, Beijing, respectively.

He is currently a system engineer in InfoDesk Corp. Tarrytown, New York, USA. His research interests include multimedia signal processing, multimedia copyright protectio management, multimedia data security, and software/hardware implementations of multimedia algorithms.

Ali N. Akansu received the MS and PhD degrees from the Technical University of Istanbul and the Polytechnic University, in 1983 and 1987, respectively, both in Electrical Engineering. Since 1987, he has been with the New Jersey Institute of Technology, where he is a professor of electrical and computer engineering. He was the Founding Director of the New Jersey Center for Multimedia Research (NJCMR) between 1996–2000.

He was the Vice President of Research and Development of IDT Corporation between June 2000 and September 2001. He was the founding President and CEO of AVWay.com Inc. His current research interests include signal theory, linear transforms and algorithms, image-video compression, signal processing for digital communications, Internet multimedia, and information security.

Dr. Akansu has served as an associate editor of IEEE Transactions on Signal Processing. He was the Lead Guest Editor of two special issues of IEEE Transactions on Signal Processing. He authored, co-authored or co-edited 4 books and over 200 refereed journal and conference papers on his research contributions.