High-rate quantization data hiding robust to arbitrary linear filtering attacks *[†]

Fernando Pérez-González[1], Carlos Mosquera[1], Marcos Alvarez[1] and Reginald Lagendijk[2]

[1] Dept. Teoria de la Señal y Comunicaciones. ETSI Telecom., Universidad de Vigo, 36311 Vigo, Spain

[2] Department of Mediamatics, Faculty of EEMCS, Delft University of Technology. 4, Mekelweg, 2628 CD DELFT. The Netherlands

ABSTRACT

Rational Dither Modulation (RDM) is a high-rate data hiding method invariant to gain attacks. We propose an extension of RDM to construct a scheme that is robust to arbitrary linear time-invariant filtering attacks, as opposed to standard Dither Modulation (DM) which we show to be extremely sensitive to those attacks. The novel algorithm, named Discrete Fourier Transform RDM (DFT-RDM) basically works in the DFT domain, applying the RDM core on each frequency channel. We illustrate the feasibility of DFT-RDM by passing the watermarked signal through an implementation of a graphic equalizer: the average error probability is small enough to justify the feasibility of adding a coding with interleaving layer to DFT-RDM. Two easily implementable improvements are discussed: windowing and spreading. In particular, the latter is shown to lead to very large gains.

1. INTRODUCTION

Until very recently, it was thought that one of the main drawbacks of quantization-based schemes, as opposed to spread-spectrum ones, was their fragility when subject to gain (also known as valumetric) attacks. Motivated by this problem, several researchers have proposed different ways of estimating this gain at the decoder, either by embedding an auxiliary pilot signal known to the decoder,[1] or through blind estimation,[2] ,[3] i.e., exploiting certain statistical properties of the received signal due to the periodic structure of the watermarked signal. In[4] we took a radically different approach and demonstrated how to construct a simple, yet effective, gain-invariant method which retains most desirable properties of Dither Modulation (DM), which in turn is the most popular algorithm in the family of side-informed ones. Our method is based on the use of a gain-invariant adaptive quantization step-size which is used by the embedder and which holds the desirable feature of allowing its robust estimation at the decoder. This adaptive step-size is constructed as a nonlinear function of past watermarked samples. The memory of such function, i.e., the number of past samples used at a certain instant to compute the step-size is a crucial design parameter: a small memory is useful when the target is to cope with varying gains; on the other hand, the larger this memory is, the more robust the step-size estimation at the decoder will be, therefore improving performance. In fact, our method, named Rational Dither Modulation (RDM), has been proven to asymptotically achieve the performance of DM. Full details on RDM and its performance can be found in.[5] Since its inception, several other works have shown the feasibility of combining RDM with channel/source coding[6, 7] or applied it to image data hiding with Watson-like perceptual distances.[8]

*This work was partially funded by *Xunta de Galicia* under projects PGIDT04 TIC322013PR and PGIDT04 PXIC32202PM; MEC project DIPSTICK, reference TEC2004-02551/TCM; FIS project IM3, reference G03/185, and European Comission through the IST Programme under Contract IST-2002-507932 ECRYPT.

[†] ECRYPT disclaimer: The information in this paper is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Further author information: (Corresponding author: FPG)
E-mail: {fperez, mosquera, malvarez}@gts.tsc.uvigo.es, Telephone: +34 986 812124
R.L.Lagendijk@ewi.tudelft.nl, Telephone: +31 15 278 3731

Security, Steganography, and Watermarking of Multimedia Contents VIII, edited by Edward J. Delp III, Ping Wah Wong,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6072, 60720C, © 2006 SPIE-IS&T · 0277-786X/06/$15

Now that the gain attack can be considered somewhat solved, it is interesting to mention that there is a much less known but almost as simple attack to quantization based schemes: linear time-invariant (LTI) filtering. LTI filtering is a recurrent operation with multimedia signals which in many instances can be considered as an unintentional attack. For instance, for audio signals the filter can take the form of a graphical equalizer where the user can play with the weights assigned to each spectral band.

Unfortunately, existing side-informed methods have not been designed to survive such attacks, which therefore pose considerable concerns upon their alleged superiority over other alternatives. To see how devastating those attacks can be, consider the outcome of the following experiment on binary DM when the host signal is white with power 25 dB above that of the watermark, and the watermarked signal is attacked by a lowpass filter of variable cutoff frequency. When this frequency is as high as 0.99π rad, the bit error rate (BER) is 0.5; when the cutoff frequency is 0.999π rad, the BER is still 0.11. Note that given the tiny fraction of bandwidth that is thrown away in both cases, the attacked signals are very likely to go perceptually unnoticed. The reason for this failure lies in the distortion error due to filtering that moves the watermarked signal away from the correct quantization centroids. This distortion error increases with the the host signal power; in other words, the larger the host to watermark ratio, the worse the performance. This is clearly in contrast to the performance of quantization-based schemes subject to additive white Gaussian noise (AWGN) attacks, where the host signal power is virtually irrelevant for most cases of practical interest.

In this paper, we propose an extension of RDM that allows us to construct a high-rate scheme that is robust to LTI filtering attacks. Our solution does not assume any prior knowledge of the filter neither at the embedder nor at the decoder; however, should this knowledge be available, it could be appropriately exploited to improve performance. The idea is to perform RDM over parallel channels in the DFT domain, taking advantage of the orthogonality of this transform and the fact that it converts a filtering operation in the time domain into an approximate multiplication on each channel. For this reason, the new scheme is termed DFT-RDM. As we will see, the inevitable distortions incurred during the filtering process, cause the solution not to be truly LTI-filtering-invariant; however, by carefully choosing the different design parameters, it is possible to achieve any desired degree of resilience against this attack, at the expense of increased complexity or reduced payload.

2. NOTATION AND REVIEW OF RDM

For simplicity, we assume one-dimensional real-valued hosts arranged in a vector $\mathbf{x}$; extension to the 2-D case is possible but it has not been pursued here. For convenience, and without loss of generality, we will regard the one-dimensional host as being given in the time-domain. Let x_k denote the kth element of $\mathbf{x}$. We will also use $\{x_k\}$ when referring to the sequence whose samples are x_k. In data-hiding applications a message is embedded by modifying $\mathbf{x}$ to a vector $\mathbf{y}$ which we will call *watermarked signal*. The difference $\mathbf{w} \triangleq \mathbf{y} - \mathbf{x}$ is termed *watermark*. The kth elements of $\mathbf{y}$ and $\mathbf{w}$ are denoted by y_k and x_k, respectively.

For analytical purposes, it will be convenient to model the different sequences involved in the embedding and decoding operations as realizations of random processes. We will reserve uppercase letters to denote random variables and lowercase letters to denote specific values. Then, X_k is a random variable modeling the kth sample of the host sequence, and $\{X_k\}$ is the random process modeling the whole sequence $\{x_k\}$. In particular, throughout this paper we will assume that $\{X_k\}$ is a white process with variance σ_x^2. As it will become apparent later, the case of colored hosts can be taken into account in our analysis by combining the coloring filter with the channel filter. Boldface letters will be used to denote vectors: uppercase for random vectors, and lowercase for realizations. The vector containing the available samples at the decoder, also called *observations* throughout this paper, will be denoted by $\mathbf{z}$. Those observations z_k are obtained after passing the watermarked signal $\{y_k\}$ through the attacking channel.

We will also need to define the *embedding distortion* D_w as the average power of the watermark, i.e., $D_w \triangleq \frac{1}{N}\text{E}\left[||\mathbf{Y}-\mathbf{X}||^2\right] = \frac{1}{N}\text{E}\left[||\mathbf{W}||^2\right]$, where E$[\cdot]$ denotes statistical expectation and $||\cdot||$ stands for Euclidean (i.e., ℓ_2) norm. The *Document to Watermark Ratio* (DWR), often expressed in decibels, is given by σ_x^2/D_w. We will assume that this DWR is large, as it occurs in the vast majority of practical applications, due to percpetual reasons. Values larger than 20 dB are customary.

Although the procedures here described can be extended to the case of multilevel (even lattice-block) quantizers,[5] the underlying principles of both RDM and DFT-RDM are fully illustrated by considering the simpler binary case (i.e., every embedding operation is meant to conceal one bit), to which our discussion will be confined.

Let us start by recalling the basic principles of binary DM data-hiding. Define the shifted lattices

$$\Lambda_b \triangleq 2\Delta\mathbb{Z} - b\Delta/2, \quad b = -1, 1 \tag{1}$$

which describe the centroids for the respective quantizers $Q_{-1}(\cdot)$ and $Q_1(\cdot)$, here assumed to be based on Euclidean distances. Given the kth information symbol $b_k \in \{-1, 1\}$, embedding in the kth sample is performed using the rule $y_k = Q_{b_k}(x_k)$, $k = 1, \cdots, N$. After observing the kth sample z_k, the decoded binary symbol $\hat{b}_k$ is decided according to a minimum Euclidean distance rule

$$\hat{b}_k = \arg \min_{-1,1} |z_k - Q_{b_k}(z_k)|, \; k = 1, \cdots, N, \tag{2}$$

which can be also seen to be equivalent to quantizing z_k with a quantizer with step-size Δ.

On the other hand, RDM is based on the following embedding and decoding equations[5]

$$\begin{aligned} y_k &= g(\mathbf{y}_{k-1}) Q_{b_k}\left(\frac{x_k}{g(\mathbf{y}_{k-1})}\right), & (3) \\ \hat{b}_k &= \arg \min_{-1,1} \left| \frac{z_k}{g(\mathbf{z}_{k-1})} - Q_{b_k}\left(\frac{z_k}{g(\mathbf{z}_{k-1})}\right) \right|. & (4) \end{aligned}$$

where $\mathbf{y}_{k-1} \triangleq (y_{k-1}, y_{k-2}, \cdots, y_{k-L})^T$, $\mathbf{z}_{k-1} \triangleq (z_{k-1}, z_{k-2}, \cdots, z_{k-L})^T$, with L a designer parameter that determines the *memory* of the system. The function $g : \mathbb{R}^L \to \mathbb{R}$ is designed in such a way that $g(\rho\mathbf{y}) = \rho g(\mathbf{y})$, for any $\rho > 0$ and any $\mathbf{y} \in \mathbb{R}^L$. One convenient choice for the g function is any of the Hölder or ℓ_p vector-norms:

$$g(\mathbf{y}_{k-1}) = \left(\frac{1}{L}\sum_{m=k-L}^{k-1} |y_m|^p\right)^{1/p}, \; p \geq 1. \tag{5}$$

Knowing that best performance is obtained when tuning the p parameter to the shape parameter of the host distribution (when $\mathbf{X}$ follows a Generalized Gaussian), we will assume from now on $p = 2$, which suits a Gaussian host distribution and leads to simplified expressions. To understand why RDM is gain-invariant, let us consider a fixed amplitude scaling attack with no noise; hence, the observed vector $\mathbf{z}$ becomes $\mathbf{z} = \rho\mathbf{y}$, with ρ the channel gain unknown to both embedder and detector. Then, substituting $z_k = \rho y_k$ into (4) and taking into account the fact that $g(\mathbf{z}_{k-1}) = \rho g(\mathbf{y}_{k-1})$, it is immediate to see that the ρ in the numerator and the denominator cancel out and the decision $\hat{b}_k$ is identical, irrespective of the value of ρ.

3. LTI-FILTERING ATTACKS

In the remainder of this paper we will consider attacks of the type $z_k = y_k * h_k$, where $*$ denotes convolution and h_k is the impulse response of a real-valued LTI filter. We will assume that the cursor of the filter (i.e., the largest magnitude coefficient) is located at the origin to facilitate the analysis. Note that in practice this requires that some synchronization measures are effected to allign the watermarked signal and the observations. Also note that in the considered channel model there is no explicit noise component, this is because the filtered-host interference (FHI) will be the dominant impairment. However, in the event such noise were present, it could be easily incorporated in the subsequent developments.

To see the source of this FHI, suppose that DM is used to hide one bit per sample using the rule given above. The observed sequence can be written as

$$z_k = y_k * h_k = h_0 y_k + \sum_{i \neq k} h_i y_{k-i} \approx h_0 y_k + \sum_{i \neq k} h_i x_{k-i} \tag{6}$$

where the given approximation follows from the large DWR assumption. The second term in the right hand side of (6) corresponds to the FHI. For i.i.d. zero-mean Gaussian hosts, this FHI will be also zero-mean Gaussian, independent of x_k and with variance σ_f^2 given by

$$\sigma_f^2 = \sigma_X^2 \sum_{i \neq k} |h_i|^2 = \frac{\sigma_X^2}{2\pi} \int_{-\pi}^{\pi} |h_0 - H(e^{j\omega})|^2 d\omega, \tag{7}$$

where $H(e^{j\omega})$ denotes the Fourier Transform of h_k. Now, assuming that $h_0 = 1$, it is possible to directly apply the results in[9] to derive the bit error probability for DM under LTI attacks:

$$P_{e,DM}\left(\frac{\Delta}{\sigma_f}\right) = 2\sum_{k=0}^{\infty}\left\{ Q\left(\frac{(4k+1)\Delta}{2\sigma_f}\right) - Q\left(\frac{(4k+3)\Delta}{2\sigma_f}\right)\right\} \tag{8}$$

(See also[9] for several useful approximations). In any case, it is worth noting that while for large DWRs the performance of DM in an AWGN channel is solely governed by the so-called Watermark to Noise Ratio (WNR), here the role of the noise is played by the FHI, so performance will depend in exactly the same way on the Watermark to Interference Ratio (WIR), defined as WIR $\triangleq D_w/\sigma_f^2$. The problem is that the WIR is inversely proportional to the DWR, so if the DWR is increased in, say, 10 dB, so as to make the watermark less perceptible, the WIR will be correspondingly decreased in 10 dB. This reveals the strong dependence of the performance of DM upon the operating DWR, which typically imposes hard to meet constraints, as illustrated by the following example.

EXAMPLE 1. *From Eq.* (8) *it is immediate to see that a WIR of 10 dB will guarantee a bit error probability of* $6 \cdot 10^{-3}$*, which may be reasonable for many applications. Suppose that* h_k *is an ideal low-pass filter with cutoff frequency* $\omega = \omega_c$ *and such that* $h_0 = 1$*. Then, from* (7) *we have*

$$\sigma_f^2 = \frac{\sigma_x^2}{\pi}\int_{\omega_c}^{\pi} d\omega = \sigma_x^2\left(1 - \frac{\omega_c}{\pi}\right) \tag{9}$$

From here, it is possible to calculate, for a given DWR, the cutoff frequency that yields a WIR of 10 dB. For instance, when DWR=25 dB, we have that $\omega_c = \pi(1 - 3.16 \cdot 10^{-4})$ *rad, while for DWR=15 dB,* $\omega_c = \pi(1 - 3.16 \cdot 10^{-3})$ *rad. Observe how narrow the stop-band is for both cases.*

It is also important to remark that if $h_0 \neq 1$ then, besides the FHI problem, DM would be facing a scaling attack which, as we have noted above, is not able to combat in an effective manner. On the other hand, RDM would be free of the scaling attack problem but would not be exempt of FHI either, so its performance in this case will be also quite poor.

A possible countermeasure would be to "invert" (equalize) the filtering operation, but this would not only be extremely complex, but also would hardly work, because the residual errors due to an imperfect equalizer estimate would likely overrun the watermark. Another potential solution would be to spectrally shape the watermark in such a way that the filtering takes no effect on it, but this clearly presumes too much knowledge about the attack at the embedder, and so it is feasible only for a very limited set of unintentional attacks. In the next section we discuss a third possibility that consists in applying RDM in the Fourier Transform domain.

4. DISCRETE FOURIER TRANSFORM RDM

The key idea for constructing a quantization-based data-hiding scheme robust to LTI filtering is the gain-invariance property of RDM combined with the fact that linear convolutions can be approximated by multiplications in the Discrete Fourier Transform (DFT) domain[‡]. Moreover, orthogonality between DFT coefficients guarantees a small amount of cross-interference between signal components.

[‡] Exact multiplications would be achieved with circular convolutions.

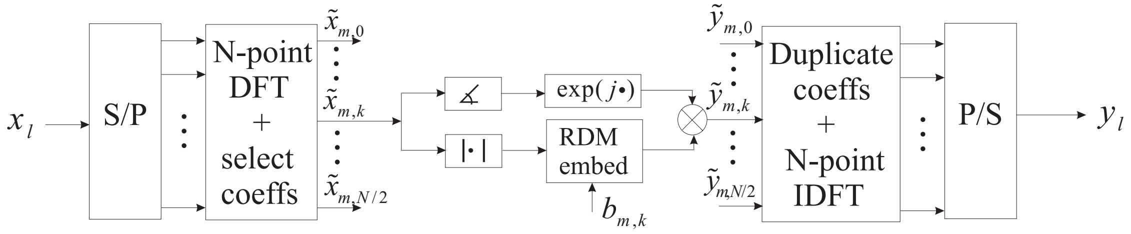

Figure 1. Embedding in DFT-RDM, detailing the operation in the kth channel.

Some additional notation is required to describe the operation in the DFT domain. We will consider a non-overlapping block-wise operation, with block-size N. Let $\mathbf{x}_m$ denote the mth block of samples of the host signal, then $\tilde{x}_{m,k}$ will denote the kth coefficient of the length-N DFT of such block, i.e.,

$$\tilde{x}_{m,k} = \sum_{l=0}^{N-1} x_{m,l} e^{-j(2\pi/N)lk} = \sum_{l=mN}^{mN+N-1} x_l e^{-j(2\pi/N)lk} \tag{10}$$

Even though in principle it would be possible to independently watermark the real and the imaginary parts of every DFT coefficient, a problem arises when multiplying such coefficient by a complex factor, as we will soon see happening, because both parts mix in a way that depends on the phase of that complex factor. For this reason, it is preferrable to watermark only the magnitude of each DFT coefficient. Then, for the DFT domain implementation of RDM (DFT-RDM in the sequel), the modulus of the kth coefficient of the mth block of the watermarked signal, $\tilde{y}_{m,k}$, is obtained as

$$|\tilde{y}_{m,k}| = g(\tilde{\mathbf{y}}_{m-1,k}) Q_{b_{m,k}} \left(\frac{|\tilde{x}_{m,k}|}{g(\tilde{\mathbf{y}}_{m-1,k})} \right), \tag{11}$$

where $b_{m,k}$ denotes the mth ($m > 0$) bit sent through the kth channel ($0 < k < N/2$), see Fig. 1. The phase of $\tilde{y}_{m,k}$ is simply set to that of $\tilde{x}_{m,k}$ to minimize the embedding distortion. This in turn implies that the watermark $\tilde{w}_{m,k}$ is collinear with $\tilde{x}_{m,k}$. The remaining watermarked samples $\tilde{y}_{m,k}$, $N/2 + 1 < k < N - 1$, are chosen to preserve the symmetry in such a way that the resulting sequence in the time domain $\{y_k\}$ is real. This is accomplished by making $\tilde{y}_{m,k} = \tilde{y}^*_{m,N-k}$, where superscript $*$ denotes conjugate.

The time-domain watermarked sequence $\{y_l\}$ is obtained as the inverse DFT (IDFT) of the watermarked DFT coefficients in a non-overlapping block-by-block basis:

$$y_l = \frac{1}{N} \sum_{k=N}^{N+N-1} \tilde{y}_{m,k} e^{j(2\pi/N)lk}, \quad mN \leq l \leq mN + N - 1 \tag{12}$$

The function $g(\tilde{\mathbf{y}}_{m-1,k})$ operates on vector $\tilde{\mathbf{y}}_{m-1,k} \triangleq (\tilde{y}_{m-1,k}, \tilde{y}_{m-2,k}, \cdots, \tilde{y}_{m-L,k})^T$. Thus, the proposed scheme essentially constructs an RDM-like scheme for each DFT channel. On the other hand, given the kth coefficient of the mth block of the observed signal $\tilde{z}_{m,k}$, decoding uses the same rule as RDM in (4) with $|\tilde{z}_{m,k}|$ being the input to the decoder. See Fig. 2. The proposed scheme allows to transmit $N/2+1$ bits per DFT-block of size N, so it achieves an approximate rate of $1/2$ bits per host sample, that is, half of that attainable with time-domain RDM. This is the price to be paid for only watermarking the magnitude in the DFT domain.

It is interesting to point out one important implementation issue due to embedding in just the magnitude of the DFT coefficients. For purposes of discussion only, assume that symbol $b_{m,k}$ is embedded in a DFT coefficient $\tilde{x}_{m,k}$ such that $\tilde{x}_{m,k} = \Delta/2$, and that $g(\tilde{\mathbf{y}}_{m-1,k}) = 1$. Then, if the in each DFT channel are given by the shifted lattices in (1), it is easy to verify that no matter which value $b_{m,k}$ takes, $\tilde{y}_{m,k}$ will always become $\Delta/2$, and, of course, a bit error will occur with probability $1/2$. This defect is attributable to the phase-preserving

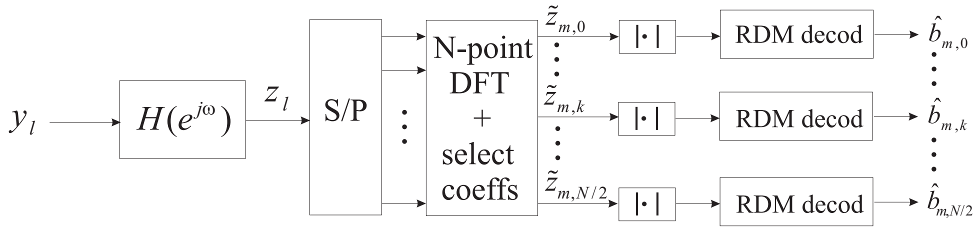

Figure 2. Decoding in DFT-RDM

operation, since it does not allow for sign changes, as standard dither modulation requires. One simple loophole is to admit changes of π radians on the phase of $\tilde{x}_{m,k}$; however, for certain signals, such as audio, this phase changes may produce annoying clicking artifacts. For this reason, we have chosen a different alternative that for large DWRs yields exactly the same performance and which consists in employing the following centroids $\Lambda_b \triangleq 2\Delta\mathbb{Z} + (1-b)\Delta/2, \quad b = -1, 1$, that is, a $\Delta/2$ shift of the sets described by (1). It is straightforward to check that with this definition, sign changes are avoided.

To get a rough understanding of why the DFT-RDM scheme proposed above is robust to LTI filtering attacks, given by their frequency response $H(e^{j\omega})$, note that if N is sufficiently large (so that the bandwith of the DFT window is small enough), then $\tilde{z}_{m,k} \approx \tilde{h}_k \tilde{y}_{m,k}$, with $\tilde{h}_k = H(e^{j2\pi k/N})$. On the other hand, $g(\tilde{\mathbf{z}}_{m-1,k}) = g(\tilde{h}_k \tilde{\mathbf{y}}_{m-1,k}) = |\tilde{h}_k| g(\tilde{\mathbf{y}}_{m-1,k})$, where the last equality follows from the properties of the function g. From here, one can easily conclude the approximate invariance of the proposed method against LTI filtering attacks. Note that this approximation is better for larger values of N and for smoother frequency responses of the LTI filter.

By assessing the variance of both the host and the watermark signals in the DFT domain, it is possible to find the relation between the DWR and the step-size used in the quantizers of DFT-RDM. This relation turns out to be functionally identical to that of time-domain RDM, i.e.,

$$\mathrm{DWR} = \frac{\sigma_x^2}{\mathrm{Var}\{W_{m,l}\}} = \frac{3}{\Delta^2_{\text{DFT-RDM}}} \tag{13}$$

This means that using the same Δ in each DFT channel as for the standard time-domain RDM will guarantee the desired DWR.

5. PERFORMANCE ANALYSIS

The performance of DFT-RDM will be measured by the bit error probability. In general, this probability will be different for each DFT channel, being of prime importance to derive analytical approximations as well as bounds to such probabilities. As one would expect, performance will strongly depend on the filter h_l. The derivation of an exact expression of the bit error probability as well as upper and lower bounds is a little involved, so due to space restrictions we will simply present here the main results, leaving the proofs for a later publication.

The random variable representing the observed signal in the DFT-domain $\tilde{Z}_{m,k}$ is modeled as $\tilde{Z}_{m,k} = \tilde{h}_k(\tilde{X}_{m,k} + \tilde{W}_{m,k}) + \tilde{N}_{m,k}$, where $\tilde{N}_{m,k}$ is a zero-mean complex Gaussian random variable uncorrelated with both $\tilde{X}_{m,k}$ and $\tilde{W}_{m,k}$, and characterized by the following joint second-order statistics

$$\sigma^2_{R,k} \triangleq \mathrm{Var}[\mathrm{Re}(\tilde{N}_{m,k})] = \sigma_x^2 \sum_l [\mathrm{Re}(f_{l,k})]^2 \quad ; \quad \sigma^2_{I,k} \triangleq \mathrm{Var}[\mathrm{Im}(\tilde{N}_{m,k})] = \sigma_x^2 \sum_l [\mathrm{Im}(f_{l,k})]^2$$

$$\mu_{RI,k} = \mathrm{E}[\mathrm{Re}(\tilde{N}_{m,k})\mathrm{Im}(\tilde{N}_{m,k})] \quad = \quad \sigma_x^2 \sum_l \mathrm{Re}(f_{l,k})\mathrm{Im}(f_{l,k}) \tag{14}$$

where $f_{l,k}$ is given by

$$f_{l,k} \triangleq (\delta_l - h_l/\tilde{h}_k) * \phi^*_{l,k}, \; k = 0, \cdots, N-1 \tag{15}$$

δ_l is the Kronecker delta function, and $\phi_{l,k}$ denotes the impulse response of the kth DFT basis function, i.e., $\phi_{l,k} \triangleq \exp(-j(2\pi/N)lk)$, for $l, k = 0, \cdots N-1$, and is zero otherwise.

Note that for the particular cases of $k = 0$ and $k = N/2$, the DFT coefficients of the host, the watermark and the watermarked signal are zero, and so are $\sigma^2_{I,k}$ and $\mu_{RI,k}$. From the quantities in (15) it is possible to obtain the noise principal components in the kth DFT channel as

$$\begin{aligned}\sigma^2_{1,k} &= \frac{1}{2}\left(\sigma^2_{I,k} + \sigma^2_{R,k} + \sqrt{4\mu^2_{RI,k} + (\sigma^2_{R,k} - \sigma^2_{I,k})^2}\right)\\ \sigma^2_{2,k} &= \frac{1}{2}\left(\sigma^2_{I,k} + \sigma^2_{R,k} + \sqrt{4\mu^2_{RI,k} - (\sigma^2_{R,k} - \sigma^2_{I,k})^2}\right), \quad k = 1, \cdots, N/2 - 1 \end{aligned} \tag{16}$$

Let $P_{e,k}$ denote the bit error probability corresponding to the kth channel of the DFT. Then,

$$P_{e,k} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{0}^{\infty} f_{g(\mathbf{Z}_k)}(s) P_{\mathrm{DM}}\left(\frac{\Delta s}{\sigma_{\tilde{N},k}(\theta)}\right) ds d\theta \tag{17}$$

with

$$\sigma_{\tilde{N},k}(\theta) \triangleq \sqrt{\sigma^2_{1,k}\cos^2\theta + \sigma^2_{2,k}\sin^2\theta} \tag{18}$$

and[5]

$$f_{g(\mathbf{Z}_k)}(s) \approx \frac{2s}{\sqrt{2\pi}\sigma_r}\exp\left(-\frac{(s^2 - N\sigma_x^2)^2}{2\sigma_r^2}\right), \quad s \geq 0, \tag{19}$$

and $\sigma_r \triangleq \sqrt{2}N\sigma_x^2/\sqrt{L}$. Interestingly, when $L \to \infty$, we have that $f_{g(\mathbf{Z}_k)}(s) \to \delta(s - \sqrt{N}\sigma_x)$, and $P_{e,k} \to \frac{1}{2\pi}\int_{-\pi}^{\pi} P_{e,\mathrm{DM}}\left(\Delta\sqrt{N}\sigma_x/\sigma_{\tilde{N},k}(\theta)\right) d\theta$.

While the computation of the per-channel bit error probability in Eq. (18) requires numerically evaluating a double integral, some useful bounds result after using the fact that the error probability of DM monotonically increases with the noise variance and is convex w.r.t. σ^2 if $\Delta/\sigma > \sqrt{12}$. Let σ_k^2 denote the variance of the per-channel FHI. This variance is $\sigma_k^2 = \sigma^2_{1,k} + \sigma^2_{2,k} = \sigma^2_{R,k} + \sigma^2_{I,k} = \sigma_x^2 \sum_l |f_{l,k}|^2$. Then, we can write the following bounds:

$$P_{e,k} \leq \int_0^{\infty} f_{g(\mathbf{Z}_k)}(s) P_{e,\mathrm{DM}}\left(\frac{\Delta s}{\sigma_k}\right) ds = P_{e,\mathrm{RDM}}\left(\frac{\Delta\sigma_x\sqrt{N}}{\sigma_k}\right) \tag{20}$$

$$P_{e,k} \geq P_{e,\mathrm{DM}}\left(\frac{\sqrt{2}\Delta\sigma_x\sqrt{N}}{\sigma_k}\right), \quad \frac{\Delta\sigma_x\sqrt{N}}{\sigma_k} > \sqrt{12}, \tag{21}$$

From the performance plot of DM in[5] we can readily see that for the range of validity of Eq. (22) $P_{e,\mathrm{DM}} \lesssim 1.5{\cdot}10^{-2}$, so at least for probabilities smaller than $1.5 \cdot 10^{-2}$, it is guaranteed that the right hand side of Eq. (22) acts as a lower bound on $P_{e,k}$. As we will see, in practice this range is about one order of magnitude larger.

6. IMPROVEMENTS

Two easily implementable improvements to the basic DFT-RDM scheme are discussed next.

6.1. Windowing

Here we concentrate on ways to reduce the variance of the per-channel FHI σ_k^2. A valuable explanation arises from applying Parseval's relation to the expression of σ_k^2 above, and taking the Fourier Transform of $f_{l,k}$ in (16), which yields

$$\sigma_k^2 = \sigma_x^2 \int_{-\pi}^{\pi} \left|\Phi_0(e^{j\omega})\right|^2 \cdot \left|1 - \frac{H(e^{j(\omega + 2\pi k/N)})}{H(e^{j2\pi k/N})}\right|^2 d\omega \tag{22}$$

where $\Phi_0(e^{j\omega})$ is the Fourier Transform of the rectangular window, which has a squared-sinc aspect. This squared-sinc function weights the error due to the unevenness of the frequency response of the LTI filter. Observe that for a flat frequency response of $H(e^{j\omega})$, the noise variance in (23) is null. This fact points at the advantages of *equalizing* the channel response at the decoder's front end, which requires that the frequency response of the LTI filter be properly estimated. From (23) it is possible to see that if the frequency response of the filter is sufficiently smooth so as to consider it approximately constant within the main lobe of the weighting function, then reducing the energy of the sidelobes will be advantageous. In the following we show how to apply this window and which modifications should be made to the developments presented thus far to take windowing into account.

Let $\mathbf{v} = (v_0, v_1, \cdots, v_{N-1})^T$ be the window, where N is the DFT-size. For reasons that will become apparent shortly, we will constrain $\mathbf{v}$ to the class of windows for which $(v_l)^{-1}$ exists for all $l = 0, \cdots, N-1$. Note that this rules out some popular windows such as the Hanning, Blackman or Bartlett windows, which are null at the borders. Embedding now operates on the windowed host signal

$$\tilde{x}_{m,k} = \sum_{l=0}^{N-1} v_l x_{m,l} e^{-j(2\pi/N)lk} = \sum_{l=mN}^{mN+N-1} v_{l-mN} x_l e^{-j(2\pi/N)lk} \tag{23}$$

and proceeds in the same way as in Section 4 to produce a set of watermarked blocks in the DFT domain. Now, given the mth block of the watermarked signal $\tilde{y}_{m,k}$, the time domain samples are obtained as

$$y_l = v_l^{(i)} \frac{1}{N} \sum_{k=0}^{N-1} \tilde{y}_{m,k} e^{j(2\pi/N)lk}, \quad mN \leq l \leq mN+N-1 \tag{24}$$

where $\mathbf{v}^{(i)} = (v_0^{(i)}, v_1^{(i)}, \cdots, v_{N-1}^{(i)})^T$, with $v_l^{(i)} \triangleq (v_l)^{-1}$, for all $l = 0, \cdots, N-1$. The reason for post-multiplying by the "inverse" window in (25) can be easily understood by considering the case where the DWR is infinity (i.e., $\Delta = 0$) because in this case, combining (25) and (24), one obtains $y_l = x_l$, as desired. At the receiver, the decoding rule in (4) is applied to the mth block of the observed signal in the DFT domain $\tilde{z}_{m,k}$, which is computed in the same way as $\tilde{x}_{m,k}$ in (24), that is, applying the DFT on the windowed mth block of z_l.

The relation between the DWR and the step-size in the case of a windowed DFT-RDM can be shown to be

$$\text{DWR} = \frac{\sigma_x^2}{\text{Var}\{W_{m,l}\}} = \frac{3N^2}{\Delta^2_{\text{WDFT-RDM}} \cdot \|\mathbf{v}\|^2 \cdot \|\mathbf{v}^{(i)}\|^2} \tag{25}$$

Concerning the performance analysis, the derivations in Sect. 5 extend to the case of windowing, after simply changing the definition of $\phi_{l,k}$ in the previous section to $\phi_{l,k} = v_l \exp(-j(2\pi/N)lk)$, for $l = 0, \cdots, N-1$.

It is important to remark that windowing will increase the peak embedding distortion. This means that, depending on the application, a particular window may be rendered unacceptable even though the embedding distortion in a mean-squared sense meets the target.

6.2. Spreading

From the performance analysis in Section 5 it is apparent that increasing the size of the DFT entails a gain in the effective SNR which is proportional to that size. Unfortunately, from an implementation perspective, increasing the DFT size implies a large computational burden, even if Fast Fourier Transforms (FFT) algorithms are employed. There is however a well-known simple technique with a small computational complexity that allows to calculate only N evenly spaced coefficients out of a DFT of size $M \cdot N$. This technique amounts to adding (i.e., *aliasing*) M length-N signal blocks in the time-domain and then computing the size-N DFT. The obvious drawback of this approach is that we reduce the payload by a factor of M, as we are collapsing M blocks down to a single one.

Assuming that no windowing (or better to say, that a rectangular windowing) is applied, the kth coefficient of the pth block of the length-N DFT of host signal is now

$$\tilde{x}_{p,k} = \sum_{l=0}^{N-1} \left(\sum_{m=pMN}^{pNM+MN-1} x_{m,l} \right) e^{-j(2\pi/N)lk} = \sum_{l=0}^{N-1} \left(\sum_{m=pM}^{pM+M-1} x_{mN+l} \right) e^{-j(2\pi/N)lk} \tag{26}$$

where the parentheses have been inserted to stress the fact that the aliasing stage should be applied first to reduce the required number of arithmetic operations.

In order to perform the embedding a slight modification to the block-diagram of Sect. 4 is necessary to cope with the contingency that the pth watermarked block in the DFT domain only consists of the N samples $\tilde{y}_{p,k}$, which correspond to $M \cdot N$ time-domain samples. Surely, the aliasing operation cannot be losslessly reverted, but we are only interested in a watermark such that the desired result is achieved when aliasing is performed again at the receiver. Then, as $\tilde{w}_{p,k} = \tilde{y}_{p,k} - \tilde{x}_{p,k}$, the time domain watermark is obtained by sharing the total watermark among the M involved blocks:

$$w_l = \frac{1}{MN} \sum_{k=0}^{N-1} \tilde{w}_{p,k} e^{j(2\pi/N)lk}, \quad pMN \leq l \leq pMN + MN - 1 \tag{27}$$

Finally, the embedder constructs the watermarked samples as $y_l = x_l + w_l$, for all l. As we have mentioned, decoding works in an identical way to (4), constructing the pth block of the observation in the DFT domain as in (27). Interestingly, the aliasing operation can be seen as a particular instance of spreading, this being the name coined by Chen and Wornell to refer to a technique that consists in projecting onto a lower dimensional space where embedding and decoding takes place. Here, through aliasing we are transforming the original $M \cdot N$-dimensional block into a N-dimensional block, using a length-M projection vector $\mathbf{s} = \mathbf{1}$. Of course, nothing precludes us from using other projection sequences, with those having coefficients in $\{\pm 1\}$ being the preferred choice for implementation simplicity.

From the results reported in the literature the spreading operation should be expected to afford a gain of $10 \log_{10} M$ in the effective SNR. Remarkably, the gain is higher, amounting to $10 \log_{10} M^2$, so spreading is a recommended technique for considerably improving the SNR at the expense of reducing the data rate. To understand where the additional gain comes from, we need to look at the step-size for a given DWR, which in this case, labeled as SDFT-RDM, are related in the following way

$$\text{DWR} = \frac{\sigma_x^2}{\text{Var}\{W_l\}} = \frac{3M}{\Delta^2_{\text{SDFT-RDM}}} \tag{28}$$

so the step-size can be made $\sqrt{M}$ times larger to achieve the same DWR. This additional gain implies an actual dependence on $M\sqrt{N}$ of the arguments of the bounds on $P_{e,k}$.

7. EXPERIMENTAL RESULTS

Several experiments were conducted to validate the goodness of our theoretical analysis and the proposed bounds and to evaluate the actual performance of the proposed method in practical scenarios. In all cases, the host signal is white, the DWR was set to 25 dB and the memory L of the g function is 100. No other impairment than LTI filtering has been introduced. For each reported experiment, 10 Monte Carlo runs with $N \times 10^3$ transmitted bits each were averaged, with N the DFT-size. Needless to say, in all cases reported, the bit error probability of both DM and time-domain RDM is nearly 0.5, thus rendering the retrieval of the hidden information completely unreliable. Fortunately, DFT-RDM manages to get a BER small enough to be driven to any desired error probability by means of error correcting codes. Additional experiments, not discussed here, show the benefits of combining DFT-RDM with Reed-Solomon coding. Figure 3 illustrates the behavior of DFT-RDM with $N = 256$ when the attack consists in a low-pass filter with cut-off frequency $\omega_c = 0.8\pi$ rad. Notice the excellent match between our analytical results and the experimental ones. Also notice that the presented lower bound is valid for bit error probabilities of 10^{-1} and smaller; the lower bound labeled as RDM-lower bound is based on using the error probability of RDM instead of that of DM in (22), and is generally tighter.

Figure 4 corresponds to the watermarked signal being fed through a 10-band graphic audio equalizer. This corresponds to an adaptation of the EQU equalizer for the XMMS multimedia player for Unix platforms, which is modeled after the popular *Winamp* player for the Windows operating system. Experiments have been performed when the equalizer default presets are chosen, yielding the frequency response depicted in Fig. 5. A Hamming window has been used following the procedure outlined in Section 6.1. Note that the lower bounds in Fig. 4 are

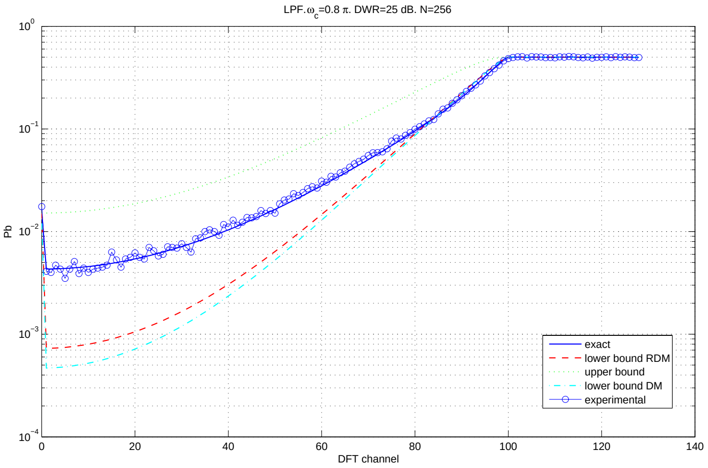

Figure 3. BER vs. DFT channel for a low-pass filter with $\omega_c = 0.8\pi$.

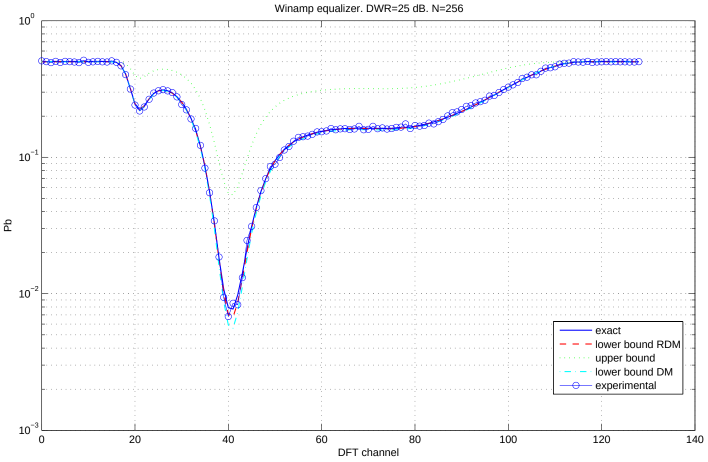

Figure 4. BER vs. DFT channel for the 10-band equalizer.

remarkably tight. This is due to the fact that the noise $\tilde{N}_{m,k}$ in our model is circularly symmetric, which is not the case for the low-pass filtering attack.

A third example corresponds to a simple desynchronization attack, which amounts to resampling the watermarked signal at instants $t = (n + \xi)T$, where T is the original sampling period and $\xi \in (-1/2, 1/2]$ determines the time shift. This resampling typically yields a negligible distortion, but its effects can be devastating: it is easy to see that even for very small values of $|\xi|$ the performance of DM is totally impaired (i.e., BER=0.5). On the other hand, the worst case can be shown to be $\xi = 1/2$ for which the watermarked signal is resampled with a half-period shift. The performance of DFT-RDM with $N = 256$ and a Hamming window for such worst case is plotted in Figure 6. The average BER among all channels is 0.086. Notice how performance worsens for high frequencies. This is explained by the so-called Gibbs phenomenon, which is reduced as the length of the interpolating filter increases. In the preceding example, this length is 401 samples.

Finally, Fig. 7 represents the theoretical bit error probability as a function of the discrete frequency for different combinations of the design parameters when the attack consists in the 10-band equalizer described above.

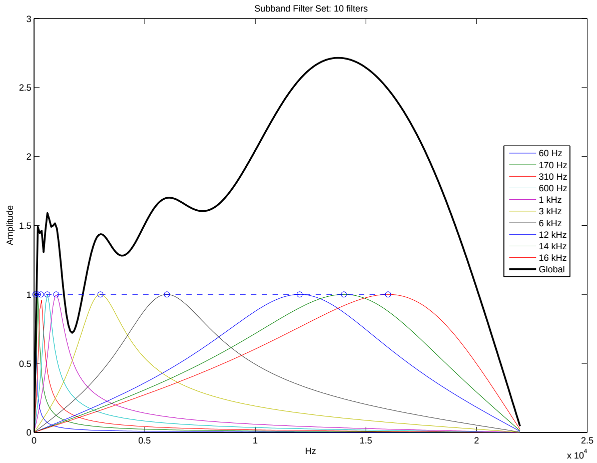

Figure 5. Magnitude (solid bold line) and subband filters response of the 10-band XMMS audio equalizer used in the experiments.

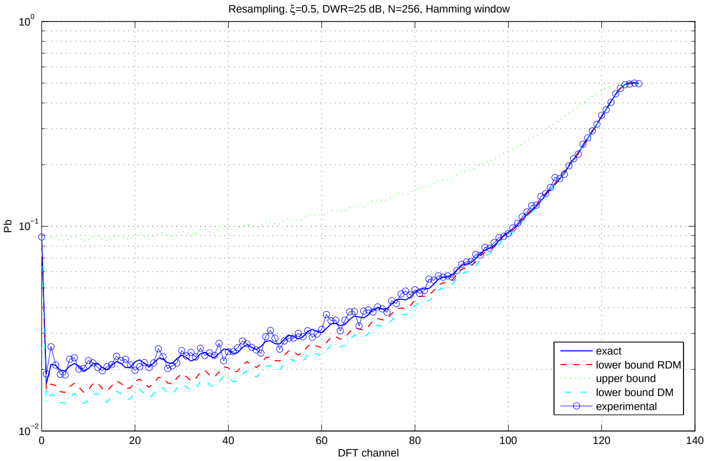

Figure 6. BER vs. DFT channel for a resampling attack with $\xi = 0.5$.

The benefits of windowing can be clearly seen, as well as the advantages of increasing the DFT-length. Recall that they come at the price of an increased peak distortion and complexity, respectively. Another advantageous choice, this time at the expense of a reduced payload, is the use of spreading. Note that with a spreading factor of $M = 4$ the performance of a 128-point DFT scheme is considerably better than that achieved with 1024-point DFTs. This is a consequence of the effective SNR being increased as $10 \log_{10} M^2$.

8. CONCLUSIONS

We have presented a novel side-informed algorithm that is robust to arbitary LTI attacks thanks to the combination of a RDM core and a DFT-based operation that essentially converts each DFT channel into a multiplicative one. Two easily implementable and non-exclusive improvements, namely windowing and spreading, have been shown to provide significant performance gains. The algorithm here presented can be extended to take into account distortion compensation, multidimensional lattices and channel coding.

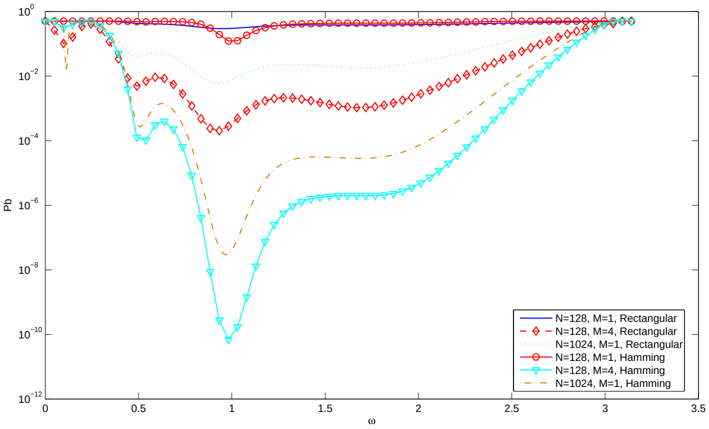

Figure 7. BER vs. discrete frequency for the 10-band equalizer attack and several design choices.

Ongoing research includes a systematic window design, a DFT-window positioning algorithm and the adaptation to colored hosts. In this regard, a frequency-based operation like that of DFT-RDM is well-suited for applying water-pouring concepts.

REFERENCES

1. J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Trans. on Signal Processing*, vol. 4, pp. 1003–1019, April 2003.
2. I. Shterev, R. Lagendijk, and R. Heusdens, "Statistical amplitude scale estimation for quantization-based watermarking," in *Security, Steganography, and Watermarking of Multimedia Contents VI, Proc. SPIE Vol. 5306* (P. W. Wong and E. J. Delp, eds.), (San Jose, CA, USA), January 2004.
3. F. Balado, K. Whelan, G. Silvestre, and N. Hurley, "Joint iterative decoding and estimation for side-informed data hiding," *IEEE Trans. on Signal Processing*, vol. 53, pp. 4006–4019, October 2005.
4. F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo, "Ensuring gain-invariance in high-rate data-hiding," in *Security, Steganography, and Watermarking of Multimedia Contents VII, Proc. SPIE Vol. 5681* (P. W. Wong and E. J. Delp, eds.), (San Jose, CA, USA), January 2005.
5. F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: a high-rate data-hiding method robust to gain attacks," *IEEE Trans. on Signal Processing*, vol. 53, pp. 3960–3975, October 2005.
6. A. Abrardo, M. Barni, F. Pérez-González, and C. Mosquera, "Trellis-coded rational dither modulation for digital watermarking," in *Proceedings of the 4th International Workshop on Digital Watermarking*, (Siena, Italy), September 2005.
7. K. Whelan, G. Silvestre, and N. Hurley, "Iterative decoding of scale invariant image data-hiding," in *IEEE International Conference on Image Processing*, (Genoa, Italy), September 2005.
8. Q. Li and I. Cox, "Rational dither modulation watermarking using a perceptual model," in *Proc. IEEE Work. on Multimedia signal Processing, MMSP'05*, (Shanghai, China), October 2005.
9. F. Pérez-González, F. Balado, and J. Hernandez, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. on Signal Processing*, vol. 4, pp. 960–980, April 2003.