

# Rational Dither Modulation: A Novel Data-Hiding Method Robust to Value-metric Scaling Attacks

Fernando Pérez-González\*, Mauro Barni†

\* Signal Theory and Communications Dept.  
University of Vigo, E-36200 Vigo, Spain  
e-mail: {fperez,mosquera}@gts.tsc.uvigo.es

Andrea Abrardo†, Carlos Mosquera\*

†Dept. of Information Engineering  
Univ. of Siena, Via Roma 56, 53100 - Siena, Italy  
e-mail: barni@dii.unisi.it

**Abstract**—A novel quantization-based data-hiding method, named Rational Dither Modulation (RDM), is presented. This method amounts to simple modifications of the well-known Dither Modulation (DM) scheme, which is largely vulnerable to scaling attacks. With such modifications, RDM becomes invariant to those attacks. Since RDM does not work by trying to estimate the step-size of the quantizers, it does not need any pilot-sequence. Moreover, RDM is suitable for a scalar operation, thus avoiding the cumbersome constructions of spherical codes. It is also shown that RDM approaches the performance of DM asymptotically with the size of the memory needed for the method to operate. Simulation results show the accuracy of our theoretical analysis and the superiority of RDM compared to the Improved Spread Spectrum method.

## I. INTRODUCTION

Since their very inception, the Achilles' heel of informed embedding methods has been their sensitivity to value-metric scaling attacks. These attacks tend to have a minimal impact on distortion as perceived by a human observer, and yet they produce an unacceptable degradation in the performance of existing implementations of the Quantization Index Modulation paradigm, proposed by Chen and Wornell [1], which are generally based on structured quantizers. This is clearly a drawback compared to (binary) spread-spectrum methods, which are intrinsically robust to those type of attacks. Not surprisingly, there has recently been an increasing research interest in finding new ways of mitigating value-metric scaling attacks [2],[3], in most cases, by keeping a standard embedding and decoding mechanism (e.g., dither modulation, DM) and proposing new ways of estimating the quantization step-size. Unfortunately, a very accurate estimate is necessary in order to guarantee reasonable operation, due to the large sensitivity of the bit error rate with respect to the estimation error. Another important class of methods aim at constructing a codebook whose codewords are somewhat evenly distributed on the surface of a hypersphere, and then perform decoding based

\*This work was partially funded by *Xunta de Galicia* under projects PGIDT02 PXIC32205PN and PGIDT04 PXIC32202PM; CYCIT project AMULET, reference TIC2001-3697-C03-01; FIS project G03/185, and European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT.

ECRYPT disclaimer: the information in this document reflects only the author's views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

on angle measurements, so they become invariant to value-metric scaling attacks. Unfortunately, both embedding and detection become rather involved and also have the drawback of reducing data-hiding payload [4].

Here we follow a radically different approach, inspired by differential phase modulations used in communications, and which is inherently robust to value-metric scaling attacks. The proposed method amounts to a simple modification of the standard DM scheme, and moreover approaches asymptotically the performance of the latter. Our novel scheme relies on "dithering" the ratio of two quantities, so it is termed Rational Dither Modulation (RDM).

As customary, we write  $\mathbf{y} = \mathbf{x} + \mathbf{w}$ , where  $\mathbf{w}$  and  $\mathbf{y}$  denote the respective vectors for the watermark and the watermarked image. The **invariant value-metric scaling** (IVS) attack consists in a constant scaling of the amplitudes of the watermarked image coefficients. We will assume that additive zero-mean white Gaussian noise is also added by the attacker. Let  $\rho > 0$  denote the scaling parameter, then the attacked image  $\mathbf{z}$  can be written as  $\mathbf{z} = \rho(\mathbf{y} + \mathbf{n})$ , where  $\mathbf{n}$  denotes the noise vector with zero-mean i.i.d. components. In the sequel, we will use uppercase letters to denote random variables and lowercase to denote specific values. Vectors are written in boldface.

Let  $D_w$  denote the embedding distortion measured in a MSE-sense;  $\sigma_n^2$  denotes the noise variance. The attacking distortion  $D_c$  will be measured as if it were only produced by  $\mathbf{n}$ , irrespective of the value of  $\rho$ . This stresses the fact that scaling alone does not produce a perceptually noticeable effect, and thus  $D_c = 0$  if  $\mathbf{n} = \mathbf{0}$ . The host image will be statistically modeled as samples drawn from a pdf  $f_X(x)$ , with zero-mean and variance  $\sigma_x^2$ . Finally, we define the Document to Watermark Ratio (DWR) as  $\sigma_x^2/D_w$  and the Watermark to Noise Ratio (WNR) as  $D_w/D_c$ .

For the purposes of illustrating the main concepts in RDM, we do not pursue distortion compensation (DC) and moreover concentrate on scalar schemes. However, the ideas presented here can be extended to account for DC mechanisms and multidimensional codes. For the standard binary scalar DM, the well-known embedding procedure uses the lattices:

$$\Lambda_{b_k} = 2\Delta\mathbb{Z} + b_k\Delta/2 \quad (1)$$

where  $b_k \in \{-1, 1\}$  is the information symbol to be embedded in the  $k$ -th host sample. These lattices give the centroids for

the respective quantizers  $Q_{-1}(\cdot)$  and  $Q_1(\cdot)$ , so embedding is simply performed as  $y_k = Q_{b_k}(x_k) = x_k + w_k$ .

Given the attacked sample  $z_k$ , decoding is performed by using a minimum Euclidean distance rule, i.e.,

$$\hat{b}_k = \arg \min_{-1,1} |z_k - Q_{b_k}(z_k)|^2. \quad (2)$$

Now consider an IVS attack with no additive noise, so the vector at the input of the decoder can be written as  $\mathbf{z} = \rho\mathbf{y}$ , which is equivalent to scaling the output of the embedder by  $\rho$ . Unfortunately, the quantization bins at the decoder are not scaled accordingly, thus producing a mismatch between embedder and decoder which dramatically affects performance, even in the absence of attacking noise [6]. In fact, substituting  $z_k = \rho y_k$  into (2) it is easy to see that the decoded bit depends on  $\rho$ . For the typical ranges of WNR and DWR, and deviations of the scaling factor as small as 10% (i.e.,  $\rho = 1.1$ ), the bit error probability  $P_b$  is already driven to 0.5, thus rendering the DM method useless.

The remainder of this paper is organized as follows: RDM is introduced in Section II while the stationary probability density function (pdf) of the watermarked image is obtained in Section III. The analytical derivation of the bit error rate (BER) for RDM with a large memory size is discussed in Section IV. Section V is devoted to providing numerical results and to comparing RDM with the Improved Spread Spectrum (ISS) method. Finally, in Section VI we give our main conclusions.

## II. RATIONAL DITHER MODULATION

In this section we propose a crucial modification of the basic DM idea that will lead to a fully robust solution against IVS attacks. Let  $\mathbf{y}_k$  denote the vector containing  $L$  past samples of vector  $\mathbf{y}$  taken at instant  $k$ , i.e.,  $\mathbf{y}_{k-1} = (y_{k-1}, y_{k-2}, \dots, y_{k-L})^T$ . A similar definition holds for  $\mathbf{z}_{k-1}$ . It is important to recognize that the construction of vectors  $\mathbf{y}_{k-1}$  and  $\mathbf{z}_{k-1}$  are strictly causal.

We will consider the set  $\mathcal{G}$  of functions  $g : \mathcal{Y}^L \rightarrow \mathbb{R}$  having the property that for any  $\rho > 0$ ,  $g(\rho\mathbf{y}_k) = \rho g(\mathbf{y}_k)$ . Then, given the  $k$ -th bit  $b_k$ , the embedding rule becomes

$$y_k = g(\mathbf{y}_{k-1}) Q_{b_k} \left( \frac{x_k}{g(\mathbf{y}_{k-1})} \right) \quad (3)$$

where the quantizer  $Q_{b_k}(\cdot)$  is induced by lattice (1). Given  $z_k$ , decoding is now done by following a similar rule to (2), that is,

$$\hat{b} = \arg \min_{-1,1} \left| \frac{z_k}{g(\mathbf{z}_{k-1})} - Q_{b_k} \left( \frac{z_k}{g(\mathbf{z}_{k-1})} \right) \right|^2. \quad (4)$$

From this equation, and the properties of the function  $g$ , it is immediate to see that the resulting method is completely insensitive to IVS attacks. It is also interesting to notice that the implementation of the generic RDM amounts to small modifications to the DM method; in embedding, it is necessary to divide  $x_k$  by  $g(\mathbf{y}_{k-1})$  prior to performing the quantization, while in decoding, the divisor becomes  $g(\mathbf{z}_{k-1})$ . This is due to the unavailability of  $\mathbf{y}_{k-1}$  at decoding, so  $\mathbf{z}_{k-1}$  becomes an estimate.

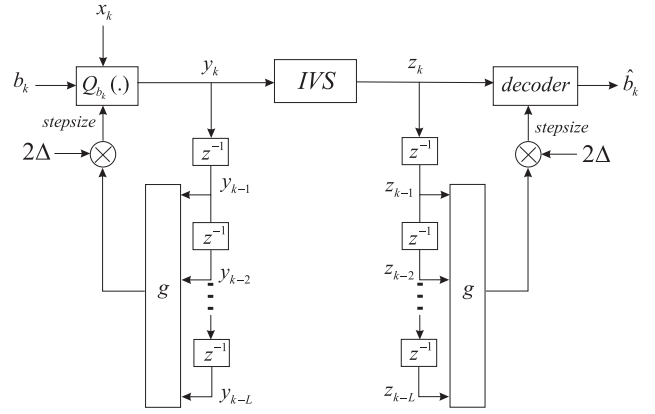


Fig. 1. Block-diagram of RDM.

Figure 1 summarizes the operation of RDM. The set  $\mathcal{G}$  includes, but is not limited to, the  $l_p$  vector-norms, given by

$$g(\mathbf{y}_k) = \left( \frac{1}{L} \sum_{m=k-L}^{k-1} |y_m|^p \right)^{1/p} \quad (5)$$

## III. STATIONARY PDF

One relevant question regarding RDM is what is the distribution of the watermarked signal. To this end, consider first the case  $L = 1$  for which the value of  $p$  in (5) becomes irrelevant. In this case,  $g(y_{k-1}) = |y_{k-1}|$ . We are interested in determining the stationary probability density function of  $Y_k$  provided it exists.

From 3, for embedding symbol  $b_k$  at the  $k$ -th sample, the set of possible centroids is  $2\Delta|y_{k-1}|Z + b_k\Delta|y_{k-1}|/2$ , so it is clear that the BER will depend on the pdf of  $Y_{k-1}$ . In fact, since we are quantizing to a discrete set of centroids, it turns out that the pdf  $f(y_k|y_{k-1}, b_k)$  of  $Y_k$  conditioned on a transmitted symbol  $b_k$  and on  $Y_{k-1} = y_{k-1}$ , is discrete and has the form

$$\begin{aligned} & \sum_m p_m(y_{k-1}, b_k) \delta(y_k - \Delta|y_{k-1}|(2m + b_k/2)) \\ &= \sum_m \frac{p_m(y_{k-1}, b_k)}{|2m + b_k/2|\Delta} \delta \left( |y_{k-1}| - \frac{y_k}{(2m + b_k/2)\Delta} \right) \end{aligned} \quad (6)$$

where  $\delta$  denotes Dirac's delta,  $p_m(y_{k-1}, b_k)$  and is the probability that  $X_k$  takes its value in the interval  $[\Delta|y_{k-1}|(2m - 1 + b_k/2), \Delta|y_{k-1}|(2m + 1 + b_k/2))$ .

If the conditions for the convergence of  $Y_k$  are met, then there will exist  $Y = \lim_{k \rightarrow \infty} \{Y_k\}$ , with pdf  $f_Y(y)$ . Next, we derive an equilibrium equation for this pdf. Assuming that  $b_k$  takes the values  $\pm 1$  with equal probability, it follows that

$$\begin{aligned} f(y_k) &= \frac{1}{2} \int_{-\infty}^{\infty} f(y_k|y_{k-1}, b_k = -1) f(y_{k-1}) dy_{k-1} \\ &+ \frac{1}{2} \int_{-\infty}^{\infty} f(y_k|y_{k-1}, b_k = +1) f(y_{k-1}) dy_{k-1} \end{aligned} \quad (7)$$

If the pdf of the host  $f_X(x)$  is symmetric about the origin, then it is not difficult to show that  $p_m(y, +1) = p_{-m}(y, -1) \triangleq p_m(y)$ , with

$$p_m(y) = \int_{\frac{|y|(4m-1)}{|4m+1|}}^{\frac{|y|(4m+3)}{|4m+1|}} f_X(x) dx, \quad (8)$$

and that  $p_m(y) = p_m(-y)$ .

Consider now the first integral in (7), that we will denote by  $I_1$ . Substituting (6) into (7) and performing the integration for the case  $y_k \geq 0$ , we obtain

$$\begin{aligned} I_1 &= \frac{1}{2} \sum_{m=1}^{\infty} \frac{1}{|2m-1/2|\Delta} p_m \left( \frac{y_k}{(2m-1/2)\Delta}, -1 \right) \\ &\cdot f_{Y_{k-1}} \left( \frac{y_k}{(2m-1/2)\Delta} \right) \\ &+ \frac{1}{2} \sum_{m=1}^{\infty} \frac{1}{|2m+1/2|\Delta} p_m \left( \frac{y_k}{(2m+1/2)\Delta}, +1 \right) \\ &\cdot f_{Y_{k-1}} \left( \frac{y_k}{(2m+1/2)\Delta} \right), \quad y_k \geq 0. \end{aligned} \quad (9)$$

For  $y_k < 0$  the result is identical to (9) with the sums ranging from  $m = -\infty$  to  $m = 0$ . The second integral in (7) is solved in a similar way.

Combining the results of the two integrals in (7) and the properties above, and after some straightforward algebra, we arrive at one expression relating  $f_{Y_k}(y_k)$  and  $f_{Y_{k-1}}(y_{k-1})$ , from which it is possible to write the equilibrium equation that the stationary distribution must satisfy, by simply replacing  $f_{Y_k}(\cdot)$  and  $f_{Y_{k-1}}(\cdot)$  by  $f_Y(\cdot)$ . This can be written in a compact form as

$$\begin{aligned} f_Y(y) &= \sum_m \frac{1}{|2m+1/2|\Delta} p_m \left( \frac{y}{(2m+1/2)\Delta} \right) \\ &\cdot f_Y \left( \frac{|y|}{|2m+1/2|\Delta} \right). \end{aligned} \quad (10)$$

Noticing that for all  $m$ ,  $p_m \geq 0$ , and that for at least one  $m = m_0$ ,  $p_{m_0} \neq 0$ , it is possible to conclude from (10) that  $f_Y(0) = 0$ .

Equation (10) obviously specializes for any particular host distribution. For a Gaussian host pdf it is possible to show that the stationary pdf of  $Y_k$  can be well-approximated by the following mixture

$$\begin{aligned} f_Y(y) &\approx \frac{4|y|}{\pi\sigma_x^2\Delta} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} \\ &\cdot \exp \left( -\frac{2y^2}{\sigma_x^2\Delta^2(2m+1)^2} - \frac{y^2}{2\sigma_x^2} \right). \end{aligned} \quad (11)$$

For analyzing the general order (i.e.,  $L \geq 1$ ), let  $\mathbf{Y}_k = (Y_k, \dots, Y_{k-L+1})^T$  and  $\tilde{\mathbf{Y}} = \lim_{k \rightarrow \infty} \{\mathbf{Y}_k\}$ ,  $\tilde{\mathbf{Y}} \in \mathbb{R}^L$ , provided that this limit exists. Let also  $f_{g(\tilde{\mathbf{Y}})}(s)$  denote the pdf of  $g(\tilde{\mathbf{Y}})$ . It can be shown that the implicit relation that defines the distribution of  $Y$  is similar to (10), with  $f_{g(\tilde{\mathbf{Y}})}$

instead of  $f_Y$  in the right hand side. For large  $L$ ,  $f_{g(\tilde{\mathbf{Y}})}(s)$  can be approximated using the central limit theorem as follows

$$f_{g(\tilde{\mathbf{Y}})}(s) \approx \frac{ps^{p-1}}{\sqrt{2\pi\sigma_r}} \exp \left( -\frac{(s^p - M_{yp})^2}{2\sigma_r^2} \right), \quad s \geq 0. \quad (12)$$

where  $M_{yp}$  is the  $p$ -th absolute moment of  $Y$ .

We have stepped here the issue of the existence of the stationary pdf's. It can be shown, using Markov chains theory, that a necessary condition for the convergence of  $\{Y_k\}$  and  $\{\mathbf{Y}_k\}$ , is that  $\Delta < 2$ .

#### IV. ANALYTICAL DERIVATION OF THE BER

We assume that the watermarked signal  $\mathbf{Y}$  is sent through an IVS channel, producing a vector  $\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N})$ . As noted previously, RDM is invariant to gain attacks, so the bit error rate (BER) analysis can be carried out by setting  $\rho = 1$ .

As decoding errors will occur at the same rate for  $b_k = 1$  and  $b_k = -1$ , we will compute the probability of decoding  $\hat{b}_k = -1$  when  $b_k = 1$ . Let us define  $P_e[s, y]$  as the probability of  $Z = Y + N$  falling in the set of intervals  $\bigcup_{l=-\infty}^{\infty} [(2l+1)\Delta s, (2l+2)\Delta s)$ , when  $Y = y$ :

$$P_e[s, y] \triangleq \sum_{l=-\infty}^{\infty} \int_{(2l+1)\Delta s}^{(2l+2)\Delta s} f_N(z-y) dz \quad (13)$$

with  $f_N(n)$  the pdf of the additive noise. As opposed to DM, RDM does not use a fixed discrete grid due to the variable step-size<sup>1</sup>, and the evaluation of the probability of error is more involved. In order to determine the bit error probability  $P_e$ , we must take the expectation of  $P_e[s, y]$  with respect to the joint pdf of  $g(\tilde{\mathbf{Z}})$  and  $Y$ , which is equivalent to

$$P_e = \int_0^{\infty} f_{g(\tilde{\mathbf{Z}})}(s) \int_{-\infty}^{\infty} P_e[s, y] f_{Y|g(\tilde{\mathbf{Z}})}(y|s) dy ds. \quad (14)$$

For large  $L$ , and due to the low variance of  $g(\tilde{\mathbf{Y}})$ , and correspondingly of  $g(\tilde{\mathbf{Z}})$ ,  $f_{Y|g(\tilde{\mathbf{Z}})}(y|s)$  can be safely approximated by  $f_{Y|g(\tilde{\mathbf{Y}})}(y|s)$ , i.e., the difference in the quantization steps at the embedder and the decoder will have a small impact in the final result; in such case, the probability of  $Y$  conditioned on  $g(\tilde{\mathbf{Z}})$  has the form of an impulse train. From here, it is possible to show that

$$\begin{aligned} P_e &= \int_0^{\infty} f_{g(\tilde{\mathbf{Z}})}(s) \sum_{m=-\infty}^{\infty} p_m \left( \frac{4m+1}{2} \Delta s \right) \\ &\cdot P_e \left[ s, \frac{4m+1}{2} \Delta s \right] ds, \quad L \gg 1, \end{aligned} \quad (15)$$

with  $f_{g(\tilde{\mathbf{Z}})}(s) \approx f_{g(\tilde{\mathbf{Y}})}(s)$ , where the latter is given by (12).

Recognizing that the summation in (15) is nothing but the probability of error of Dither Modulation for a step-size  $2\Delta s$ , here denoted by  $P_{DM}(2\Delta s)$ ,  $P_e$  can be rewritten in a more compact form as follows

$$P_e = \int_0^{\infty} f_{g(\tilde{\mathbf{Z}})}(s) P_{DM}(2\Delta s) ds. \quad (16)$$

<sup>1</sup>In the limit, for  $L = \infty$ , the embedding quantization step is fixed and equal to  $2\Delta M_{yp}^{1/p}$ .

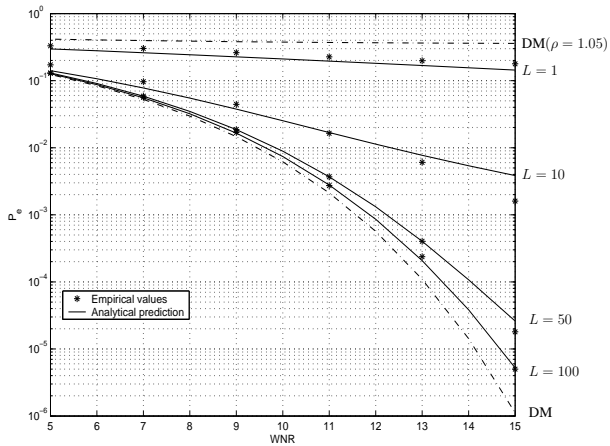


Fig. 2. Empirical and analytical values of the probability of error for different values of the memory size  $L$ . Gaussian host, DWR = 25 dB,  $c = 2$ ,  $p = 2$ .

where the probability of error of DM is supplied in [5]. Therefore, from (16) it is possible to conclude that, for large  $L$ , the performance of RDM is equivalent to averaging that of Dither Modulation for all the possible values of the step-size. In particular, if  $L \rightarrow \infty$ , we have that  $f_{g(\tilde{\mathbf{Z}})}(s) \rightarrow \delta\left(s - M_{yp}^{1/p}\right)$ , and  $P_e \rightarrow P_{DM}\left(2\Delta M_{yp}^{1/p}\right)$ .

Good approximations can also be obtained for the case  $L = 1$ , following a different strategy.

## V. NUMERICAL RESULTS

Figure 2 shows the empirical and analytical values of the BER for RDM in the Gaussian case ( $c = 2$ ), if we use the  $l_2$  norm in the definition of normalization function  $g(\cdot)$ . The performance of Dither Modulation in the ideal case, and assuming a gain attack of 5%, i.e.,  $\mathbf{Z} = 1.05(\mathbf{Y} + \mathbf{N})$ , is plotted as a reference. The measured probability of error was established after averaging a number of simulations of 50,000 bits each, such that the total number of bits in error was higher than 50 for all cases. Analytical expressions for the case  $L = 1$  (not given here) and the large  $L$  (Eq. (16)) setting have been plotted as well. It is important to see that we can reduce the performance loss with respect to Dither Modulation as much as desired for a sufficiently high order  $L$ . This is especially relevant if we consider that Dither Modulation is useless beyond a small degree of gain attack, unless some other measures are taken into account, as shown in the figure for a channel gain of  $\rho = 1.05$ .

Next we compare RDM with the recently proposed Improved Spread Spectrum (ISS) [7]. Since ISS needs some amount of spreading for attaining a satisfactory performance, it is interesting to plot the spreading factor  $N$  that is required to achieve *the same* probability of bit error as RDM, for different values of the memory size  $L$ . This is plotted in Figure 3 for a DWR of 25 dB, where it can be seen that for  $L = 10$ , and depending on the WNR, a spreading factor between 300 and 3000 is necessary. This simply means that for a WNR of 15

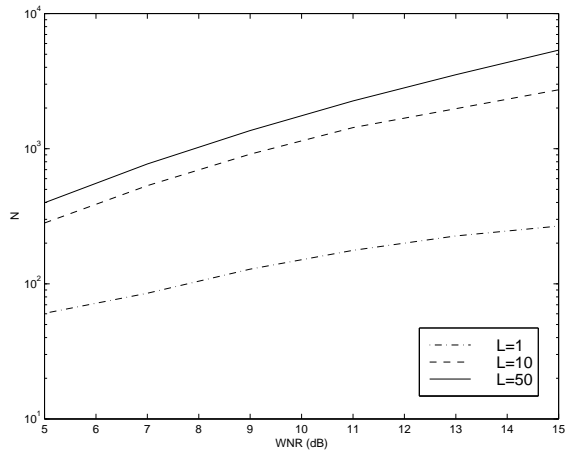


Fig. 3. Spreading factor  $N$  needed in ISS for achieving the same BER as in RDM, for different values of  $L$ . Gaussian host, DWR = 25 dB,  $p = 2$ .

dB, ISS reduces the payload by a factor of 3000 compared to RDM to achieve the same bit error rate.

## VI. CONCLUSIONS

We have introduced here a novel data hiding scheme (RDM), which can be proven to be invariant to IVS attacks. RDM has the advantage of not needing any pilot signal, can work in a scalar-basis (as opposed to spherical codes, which need many more dimensions) and can approach DM asymptotically with the memory size  $L$ . RDM allows a much higher capacity than spread-spectrum methods (including ISS), which are also invariant to IVS attacks. RDM can benefit also from the gains afforded by distortion compensation and channel coding. Finally, RDM is even moderately robust to varying value-metric scalings.

## REFERENCES

- [1] 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*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [2] J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.
- [3] K. Lee, D.S. Kim, T. Kim, and K.A. Moon, "Blind EM estimation of the scale factor for quantization-based audio watermarking," in *2nd International Workshop on Digital Watermarking*, Seoul, Korea, October 2003.
- [4] A. Abrardo and M. Barni, "Orthogonal dirty paper coding for informed 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.
- [5] Fernando Pérez-González, Félix Balado, and Juan R. Hernández, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 960–980, April 2003.
- [6] F. Bartolini, M. Barni, and A. Piva, "Performance analysis of spread transform dither modulation in presence of non-additive attacks," *IEEE Trans. on Signal Processing*, 2003, To appear.
- [7] H.S. Malvar and D.A. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. on Signal Processing*, vol. 4, no. 51, pp. 898–905, April 2003.