Journal of Information Hiding and Multimedia Signal Processing ©2012 ISSN 2073-4212
Ubiquitous International Volume **3**, Number **2**, April **2012**

Performance Analysis of Dither Modulation against Composite Attacks

Xinshan Zhu and Yanming Chen

School of Information Engineering
Yangzhou University
196 Huayang West Road, Yangzhou 225127, China
xszhu_hm@hotmail.com

Received May 2011; revised April 2012

Abstract. *In this paper, we analyze the performance of dither modulation (DM) against the composite attacks including valumetric scaling, additive noise and constant change. The analyses are developed under the assumptions that the host vector and noise vector are mutually independent and both of them have independently and identically distributed components. We derive the general expressions of the probability density functions of several concerned signals and the decoding error probability. The specific analytical results are presented for the case of generalized Gaussian host signal. Numerical simulations confirm the validity of the given theoretical analyses.*

Keywords: Digital watermarking, quantization index modulation, composite attacks, valumetric scaling, constant change

1. **Introduction.** Since Cox et al. pointed out that digital watermarking problem could be viewed as a communication problem with side information [1], much attention has been paid to the quantization-based watermarking for cancelling the host signal interference [2, 3]. One of the most important methods proposed so far is quantization index modulation (QIM) [4]. Chen et al. [4] presented the basic QIM algorithm called dither modulation (DM) and several variants of it, i.e., distortion compensated dither modulation (DC-DM) and spread transform dither modulation (STDM) [4]. The theoretical performance of QIM methods has been extensively investigated in [5, 6, 7].

In [4], Chen et al. considered the simple case where the watermark is transmitted in an additive white Gaussian noise (AWGN) channel. They gave a relatively crude approximation to the error probability of the minimal distance detector. Eggers et al. [5] proposed the scalar Costa scheme (SCS), approximately equivalent to DC-QIM, and analyzed the decoding performance of it under the AWGN attack. The careful performance analyses were done by Gonzàlez et al. [6] for a large class of QIM methods. They assumed that the watermark is impaired by an additive attacker and considered the following two cases: the channel noise follows a uniform and Gaussian distributions. Bartolini et al. [8] concentrated on analyzing the performance of the STDM algorithm at a practical level. By assuming the host signal is normally distributed, they derived the theoretical error probabilities in closed form for the gain attack plus noise addition, and the quantization attack. Boyer et al. [9] theoretically evaluated the performance of scalar DC-QIM against AWGN from the detection viewpoint. In [10], the authors proposed an improved DM scheme to resist linear-time-invariant filtering and provided a thorough analysis of it, resulting in both accurate predictions and bounds on the error probability. Recently, a new logarithmic QIM (LQIM) was presented in [11] and its analytical performance was obtained in the presence of AWGN.

The objective of this paper is to analyze the performance of DM against composite attacks, which is lacking in the literature. Obviously, in watermarking applications, it is more often that the watermark undergoes multiple attacks. Specifically, the combination of valumetric scaling, additive noise and constant change will be considered. On the other hand, most of previous analyses are restricted to the Gaussian host, even sometime regardless of the distribution of the host signal, which we will try to overcome. The generalized Gaussian distribution (GGD) is adopted to model the host signal in our analysis. Since the GGD is a parametric family of distributions, we will observe how the choice of distribution model affects the performance of DM.

The paper is organized as follows. Section 2 reviews the original DM and describes the problem to be solved. Next, Section 3 accurately derives the general PDF models concerned with our analysis. In Section 4, the performance of DM under the composite attacks is mathematically analyzed by the derived PDFs. The decoding error probability is given in closed form. Then, in Section 5, the theoretical results are confirmed by numerical simulations. Finally, Section 6 concludes.

Notation: In the remainder of this article, we use boldface lower-case letters to denote column vectors, e.g. $\boldsymbol{x}$, and scalar variables are denoted by italicized lower-case letters, e.g. x. The probability distribution function (PDF) of a random variable (r.v.) x is denoted by $p_X(x)$, whereas if x is discrete its probability mass function (PMF) is designated by $P_X(x)$. We write $x \sim p_X(x)$ to indicate that a r.v. x is distributed as $p_X(x)$. $p_{X|Y}(x|y)$ means the conditional probability of x given y. And the subscripts of the distribution functions will be dropped wherever it is clear the random variable they refer to. Finally, the mathematical expectation and standard deviation are respectively represented by μ_x and σ_x for a r.v. x.

2. **Review of DM and Problem.**

We will concentrate our attention on DM in this study. The uncoded binary DM can be summarized as follows.

Let $\boldsymbol{x} \in \mathbb{R}^N$ be a host signal vector in which we wish to embed the watermark message m. First, the message m is represented by a vector $\boldsymbol{b}$ with NR_m binary antipodal components, i.e., $b_j = \pm 1, j = 1, \cdots, NR_m$, where R_m denotes the bit rate. The host signal $\boldsymbol{x}$ is then decomposed into NR_m subvectors (blocks) of length $L = \lfloor 1/R_m \rfloor$, denoted by $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{NR_m}$. In the binary DM, two L-dimensional uniform quantizers $Q_{-1}(\cdot)$ and $Q_{+1}(\cdot)$ are constructed, whose centroids are given by the lattices $\Lambda_{-1} = 2\Delta\mathbb{Z}^L + \boldsymbol{d}$ and $\Lambda_{+1} = 2\Delta\mathbb{Z}^L + \boldsymbol{d} + \Delta\mathsf{a}$ with $\boldsymbol{d} \in \mathbb{R}^L$ a key-dependent dithering vector and $\mathsf{a} = (1, \cdots, 1)^T$. Each message bit b_j is hidden by using $Q_{b_j}(\cdot)$ on $\boldsymbol{x}_j$, resulting in the watermarked signal $\boldsymbol{y} \in \mathbb{R}^N$ as

$$\boldsymbol{y}_j = Q_{b_j}(\boldsymbol{x}_j), \quad j = 1, \cdots, NR_m \tag{1}$$

The watermark detector receives a distorted, watermarked signal, $\boldsymbol{z}$, and decodes a message $\widehat{m}$ using the minimal distance decoder

$$\widehat{b}_j = \arg\min_{-1,1} \|Q_{b_j}(\boldsymbol{z}_j) - \boldsymbol{z}_j\|, \quad j = 1, \cdots, NR_m \tag{2}$$

where $\|\cdot\|$ stands for Euclidean (i.e., ℓ_2) norm.

In practical watermarking applications, the watermarked signal might undergo composite attacks. It is well known that quantization-based watermarking is vulnerable to valumetric scaling attack. While the vector at the input of the decoder is scaled by ρ_j, i.e., $\boldsymbol{z}_j = \rho_j \boldsymbol{y}_j$, the quantization bins at the decoder are not scaled accordingly, thus producing a mismatch between embedder and decoder that dramatically affects performance. Also, the original DM is not robust to constant change distortion, i.e., $\boldsymbol{z}_j = \boldsymbol{y}_j + c_j\mathsf{a}$ with c_j a constant value. No decoding error is made for $|c_j| < \Delta/2$, however, the bit error rate (BER) is equal to 1 for $\Delta/2 < |c_j| < 3\Delta/2$. In this work, the two attacks are considered together with additive noise $\boldsymbol{\nu}_j$, yielding the attacked signal as

$$\boldsymbol{z}_j = \rho_j \boldsymbol{y}_j + \boldsymbol{\nu}_j + c_j\mathsf{a}. \tag{3}$$

We will analyze the performance of DM in the case of (3). In the analysis, $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ and $\boldsymbol{\nu}$ are all regarded as random vectors. And we assume that both $\boldsymbol{x}$ and $\boldsymbol{\nu}$ have independently and identically distributed (i.i.d.) components and $\boldsymbol{\nu}$ is independent from $\boldsymbol{y}$. Since the mean value of additive noise can be counted by the third term in (3), it is reasonable to assumed that $\mu_\nu = 0$.

3. **PDF Models.** Define the extracted vector $\boldsymbol{r}$, $\boldsymbol{r} \overset{\triangle}{=} Q_b(\boldsymbol{z}) - \boldsymbol{z}$. Obviously, a crucial aspect when performing a rigorous analysis lies in computing the PDF of $\boldsymbol{r}$. Let us begin with the issue.

A. PDF Model of the watermarked signal

We use a lower-case letter to indicate any element of the vector denoted by the boldface one. The previously used index j is dropped for no specific values (or subverctors) are concerned. Given $x \sim p_X(x)$, from the relation (1), we get

$$p_Y(y|b) = \sum_{k=-\infty}^{\infty} \delta(y - y_k) \int_{y_k-\Delta}^{y_k+\Delta} p_X(x)dx, \tag{4}$$

where the variable y_k is defined as

$$y_k = 2k\Delta + (b+1)\Delta/2 + d$$

and $\delta(\cdot)$ denotes the delta function.

A few observations are in order about the PDF of y. First, for different d, the PDF $p_Y(y|b)$ is different. That means each element of $\boldsymbol{y}$ obeys different distributions by randomly selecting $\boldsymbol{d}$ during embedding. However, due to the fact

$$P_Y(y_k + 2\Delta|b) = P_Y(y_{k+1}|b) \tag{5}$$

exists, it is sufficient for us to consider the case $d \in [-\Delta, \Delta)$. Further, if the PDF $p_X(x)$ is symmetric, i.e., $p_X(x) = p_X(-x)$, and $d = -\Delta/2$, from (4), it is easily derived that

$$p_Y(y|b=-1) = p_Y(-y|b=+1). \tag{6}$$

But for $d = 0$, the PDF $p_Y(y)$ satisfies

$$p_Y(y|b) = p_Y(-y|b). \tag{7}$$

These two properties of $p_Y(y)$ are exhibited in Fig. 1 and Fig. 2 respectively.

B. PDF Model of the attacked signal

Taking the equation (3) into account and using the fact that for any $\rho > 0$ $p_{\rho Y}(y) = \frac{1}{\rho} p_Y(\frac{y}{\rho})$ holds, the PDF of $\boldsymbol{z}$ can be obtained by convolution [12]

$$p_Z(z|b) = \sum_{k=-\infty}^{\infty} P_Y(y_k|b) p_\nu(z - \rho y_k - c), \tag{8}$$

where the convolution follows from the independence between $\boldsymbol{y}$ and $\boldsymbol{\nu}$. In (8), if the effect of different d on $P_Y(y)$ is ignored (this generally holds when the embedding distortion is acceptable), $p_Z(z|b, d \neq 0)$ can be approximately viewed as the translate of $p_Z(z|b, d = 0)$, that is,

$$p_Z(z + \rho d|b, d \neq 0) \approx p_Z(z|b, d = 0). \tag{9}$$

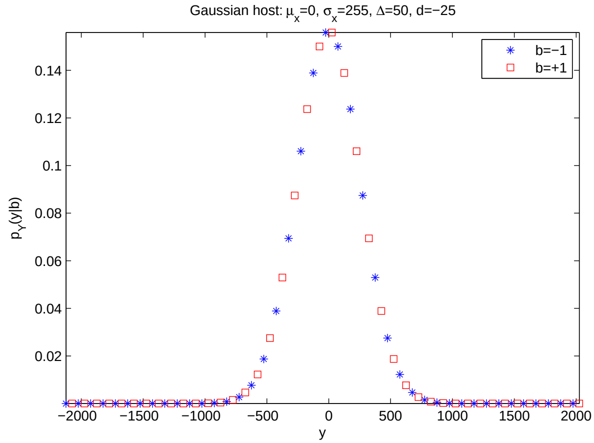

FIGURE 1. The PDF curves of y for different values of b.

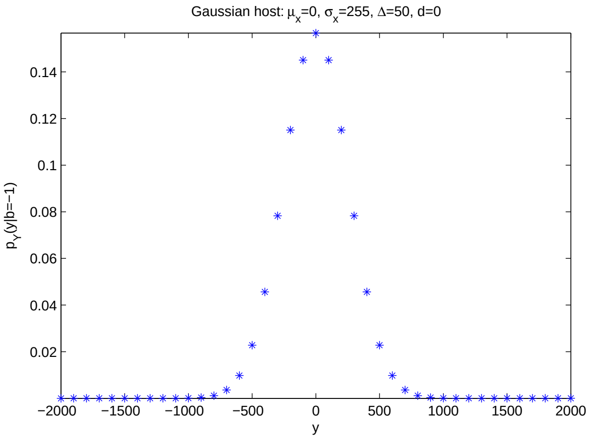

FIGURE 2. The PDF curve of y with $b = -1$ and $d = 0$.

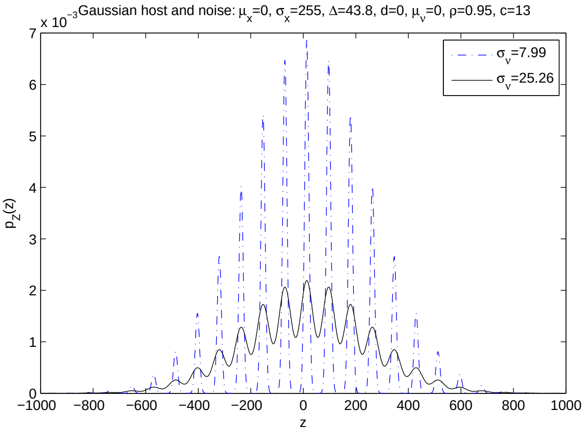

FIGURE 3. The PDF curves of z.

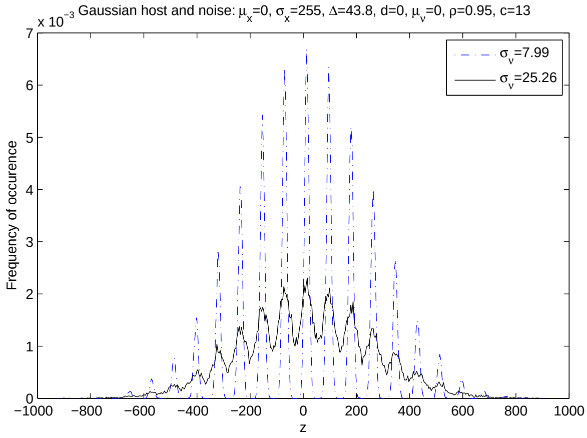

FIGURE 4. The empirical distribution curves of z.

Now, let us analyze the properties of $p_Z(z)$ when both y and ν are distributed symmetrically around the origin. In the case $d = -\Delta/2$, combining (6) and (8), we derive

$$p_Z(z + 2c|b = 0) = p_Z(-z|b = 1). \tag{10}$$

But with $d = 0$, it results

$$p_Z(z + 2c|b) = p_Z(-z|b). \tag{11}$$

Fig. 3 depicts qualitatively the PDFs of z in case of Gaussian host and noise. It can be seen that there is a bell curve present around each discrete value of y due to the existence of Gaussion noise, and the two adjacent ones even overlap for large noise strength. Meanwhile, the distance between two discrete points of y is scaled by the scaling factor ρ and $p_Z(z)$ is translated by constant value c. The corresponding empirical density curves of Z are plotted in Fig. 4. We found the analytical PDF of z fits well with empirical observations.

C. PDF Model of the extracted signal

With the given definition of r previously, the PDF of it is written as

$$p_R(r|b) = \begin{cases} \sum\limits_{j=-\infty}^{\infty} p_Z(z_j - r|b, d), & r \in [-\Delta, \Delta) \\ 0, & \text{else} \end{cases} \tag{12}$$

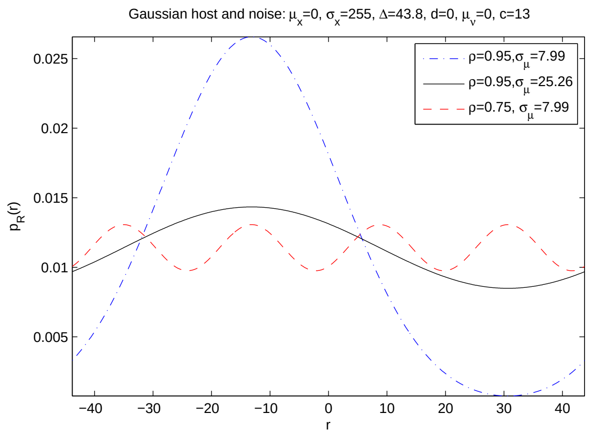

FIGURE 5. The PDF curves of r.

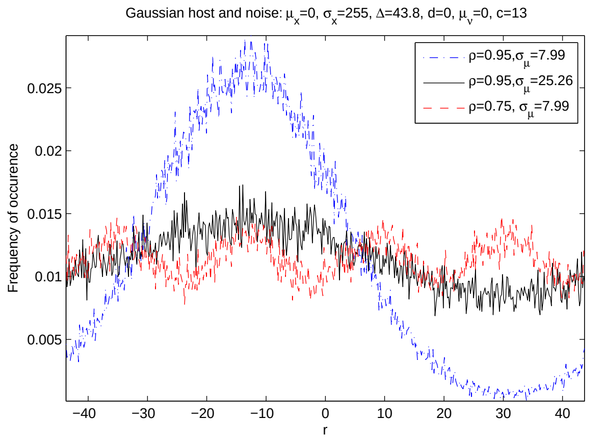

FIGURE 6. The empirical distribution curves of r.

where z_j has the similar definition with y_k. Substituting (8) into (12) yields

$$p_R(r|b) = \begin{cases} \sum_j \sum_k P(y_k|b) p_\nu(\mu_{jk} - r), & r \in [-\Delta, \Delta) \\ 0, & \text{else} \end{cases} \tag{13}$$

with $\mu_{jk} = z_j - \rho y_k - c$. Obviously, once the PDFs of x and ν are given, the PDF of r is easily computed by (13). If ignoring the effect of d on $P_Y(y)$, in view of (13), we derive

$$p_R(r - \epsilon d|b, d \neq 0) \approx p_R(r|b, d = 0) \tag{14}$$

with $\epsilon = \rho - 1$. This shows that for the case $d \neq 0$ $p_R(r|b)$ can be approximately obtained by translating $p_R(r|b, d = 0)$. Moreover, while $|\epsilon|$ is small enough for neglecting the term ϵd, (14) reduces to

$$p_R(r|b, d \neq 0) \approx p_R(r|b, d = 0). \tag{15}$$

Thus, despite the choice of d, $p_R(r)$ approximately remains unchanged for small $|\epsilon|$.

By assuming $p_X(x)$ and $p_\nu(\nu)$ are symmetrical, for $p_R(r)$, we have the following properties

$$p_R(r - 2c|b) = p_R(-r|b), \quad for\, d = 0 \tag{16}$$

$$p_R(r - 2c|b = 0) = p_R(-r|b = 1), \quad for\, d = -\frac{\Delta}{2} \tag{17}$$

which is similar to $p_Z(z)$. In particular, combining (10), (11) and (13), we derive

$$p_R(r|b, \epsilon) = p_R(r|b, -\epsilon) \quad for\, d = 0 \tag{18}$$

$$p_R(r|b = 0, \epsilon) = p_R(r|b = 1, -\epsilon) \quad for\, d = -\frac{\Delta}{2}. \tag{19}$$

These properties of $p_R(r)$ are helpful for us to analyze the performance of DM.

Fig. 5 and Fig. 6 respectively plot the probability density curves of r and the corresponding empirical ones with Gaussian host and noise in use. It is shown that the values of r are distributed around zero with higher probability for weak attacks, and the density curve of r becomes smooth as attacks become stronger, resulting in the increase of BER. By the scale factor ρ, the distribution curve of r is either dilated or compressed. Simultaneously, it is approximately translated by $\epsilon d + c$. Comparison of Fig. 5 and Fig. 6 reveals the analytical PDF of r fits perfectly with its empirical distribution.

4. **Performance Analysis.** As the previous literatures, the decoding bit error probability P_e is used as the final performance measurement. Applying the definition of $\boldsymbol{r}$, it is straightforward to write P_e as

$$P_e = P(\|\boldsymbol{r}\| > \|\Delta \mathsf{a} - |\boldsymbol{r}|\|\,|b). \tag{20}$$

where $|\boldsymbol{r}|$ denotes the vector of absolute values of components of $\boldsymbol{r}$. Defining $s \overset{\triangle}{=} |\boldsymbol{r}|^T \mathsf{a}$, the above expression is equivalent to

$$P_e = \int_{L\Delta/2}^{L\Delta} p_S(s|b)ds. \tag{21}$$

To compute P_e, we need know the PDF $p_S(s)$ of s. The exact solution for $p_S(s)$ may be achieved by several means. One of the standard procedures is by performing multifold integral operation, which is feasible for a small L. However, it becomes impractical as L increases. To solve the problem, it is nature to use mathematically tractable approximations. Let us assume that all components of $\boldsymbol{d}$ are equal, so that the vector $\boldsymbol{r}$ has i.i.d components. At this point, by the well known central limit theorem (CLT), s thus can be approximated by a Gaussian random variable, whose mean and variance are $L\mu_{|r|}$ and $L\sigma^2_{|r|}$. Using the derived PDF in (13), $\mu_{|r|}$ and $\sigma^2_{|r|}$ are represented as

$$\mu_{|r|} = \sum_j \sum_k P(y_k|b) \int_{-\Delta}^{\Delta} |r| p_\nu(\mu_{jk} - r)dr \tag{22}$$

$$\sigma^2_{|r|} = \sum_j \sum_k P(y_k|b) \int_{-\Delta}^{\Delta} r^2 p_\nu(\mu_{jk} - r)dr - \mu^2_{|r|}. \tag{23}$$

Then, the probability P_e is computed as

$$P_e \approx \Phi(\frac{\sqrt{L}(\Delta - \mu_{|r|})}{\sigma_{|r|}}) - \Phi(\frac{\sqrt{L}(\Delta/2 - \mu_{|r|})}{\sigma_{|r|}}), \tag{24}$$

where $\Phi(\cdot)$ stands for the cumulative distribution function (CDF) of the standard Gaussian distribution. It should be pointed out the CLT approximation to P_e is only valid for very large L. In reality, the condition is generally met in order to improve the watermarking robustness.

Now, we can observe several useful properties of P_e from the previous analysis. If $|\epsilon|$ is small enough, by the property (15), it is easily understood that P_e approximately equals under the situations $d \neq 0$ and $d = 0$. Therefore, without loss of generality, d is set to 0. Furthermore, if both $p_X(x)$ and $p_\nu(\nu)$ are symmetrical, in use of (18), P_e is symmetrical with respect to the point $\beta = 1$ for $d = 0$. As a result, the symmetry of P_e also holds for $d \neq 0$ approximately.

Theoretically, P_e can be estimated only if the PDFs $p_X(x)$ and $p_\nu(\nu)$ are given. For the following analysis we consider a specific case where the host signal is statistically modeled by the GGD. The GGD model is used because it includes a family of distributions and suitable for many practical applications. The PDF $p(t)$ of the GGD is

$$p(t) = \frac{\kappa\beta}{2\Gamma(\beta^{-1})} e^{-|\kappa(t-\mu)|^\beta}, \tag{25}$$

where $\kappa = \frac{1}{\sigma}\sqrt{\Gamma(3\beta^{-1})/\Gamma(\beta^{-1})}$, and $\Gamma(u) = \int_0^\infty t^{u-1}e^{-t}dt$ is the Gamma function. Thus, the distribution is completely specified by the mean μ, the standard deviation σ and the shape parameter β, and is denoted as $GGD(\beta; \mu, \sigma)$. Note that Gaussian and Laplacian distributions are just two special cases of GGD with $\beta = 2$ and $\beta = 1$, respectively.

First, the PMF $P_Y(y)$ is calculated according to the distribution model of x. Given $p_X(x) \sim GGD(\beta_x;\mu_x,\sigma_x)$, in view of (4), we immediately write

$$P_Y(y_k|b) = \Psi_x(y_k+\Delta) - \Psi_x(y_k-\Delta), \tag{26}$$

where the CDF $\Psi_x(t)$ is defined as

$$\Psi_x(t) = \frac{1}{2} + sgn(t-\mu_x)\frac{\gamma(\beta_x^{-1}, |\kappa_x(t-\mu_x)|^{\beta_x})}{2\Gamma(\beta_x^{-1})} \tag{27}$$

[13], $\gamma(s,u) = \int_0^u t^{s-1}e^{-t}dt$ is the lower incomplete gamma function, and $sgn(t)$ denotes the sign function.Then, the integration terms in (22) and (23) are derived by the PDF $p_\nu(\nu)$. As an example, the additive Gaussian noise is considered, i.e., $p_\nu(\nu) \sim \mathcal{N}(0,\sigma_\nu^2)$. This leads to

$$\begin{aligned}\int_{-\Delta}^{\Delta}|r|p_\nu(t-r)dr = &\int_t^{t+\Delta} rp_\nu(r)dr - \int_{t-\Delta}^{t} rp_\nu(r)dr \\ &+t(\int_{t-\Delta}^{t} p_\nu(r)dr - \int_t^{t+\Delta} p_\nu(r)dr) = \sum_{i=1}^{3} f_i(t)\end{aligned} \tag{28}$$

and

$$\int_{-\Delta}^{\Delta} r^2 p_\nu(t-r)dr = \int_{t-\Delta}^{t+\Delta}(t^2-2tr+r^2)p_\nu(r)dr = \sum_{i=4}^{6} f_i(t), \tag{29}$$

where

$$\begin{aligned}
f_1(t) &= \frac{\sigma_\nu}{\sqrt{2\pi}}(e^{-\frac{t^2}{2\sigma_\nu^2}} - e^{-\frac{(t+\Delta)^2}{2\sigma_\nu^2}}), \quad f_2(t) = f_1(-t), \\
f_3(t) &= t(2\Phi(\frac{t}{\sigma_\nu}) - \Phi(\frac{t-\Delta}{\sigma_\nu}) - \Phi(\frac{t+\Delta}{\sigma_\nu})) \\
f_4(t) &= t^2(\Phi(\frac{t+\Delta}{\sigma_\nu}) - \Phi(\frac{t-\Delta}{\sigma_\nu})), \quad f_5(t) = 2t(f_1(-t) - f_1(t)) \\
f_6(t) &= \frac{\sigma_\nu(t-\Delta)}{\sqrt{2\pi}}e^{-\frac{(t-\Delta)^2}{2\sigma_\nu^2}} - \frac{\sigma_\nu(t+\Delta)}{\sqrt{2\pi}}e^{-\frac{(t+\Delta)^2}{2\sigma_\nu^2}} + \sigma_\nu^2(\Phi(\frac{t+\Delta}{\sigma_\nu}) - \Phi(\frac{t-\Delta}{\sigma_\nu}))
\end{aligned}$$

Using the above results, the mean and variance of $|r|$ become

$$\mu_{|r|} = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{i=1}^{3} P(y_k|b)f_i(\mu_{jk}) \tag{30}$$

$$\sigma_{|r|}^2 = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{i=4}^{6} P(y_k|b)f_i(\mu_{jk}) - \mu_{|r|}^2. \tag{31}$$

Therefore, the approximated P_e is obtained for large L by computing (26), (30), and (31), then putting them into (24). Since the calculation of $p_Y(y)$ is relatively simple in (4), the above analysis can be easily extended for other host distributions. However, the derivation of the integration terms in (22) and (23) might become very complex for the noise ν with other PDFs. Thus, they are computed numerically in the case.

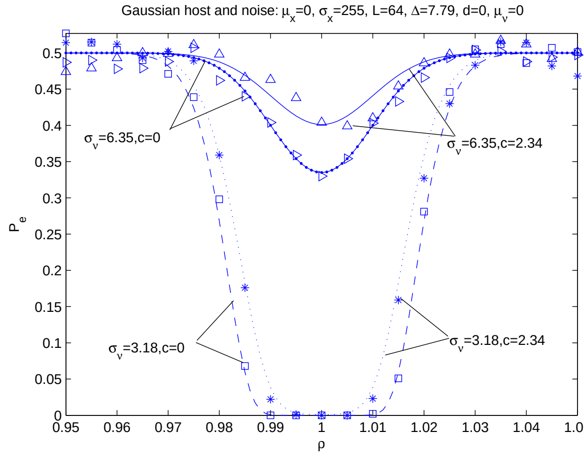

FIGURE 7. Bit error probability versus ρ while fixing σ_ν and c. Lines and symbols stand for theoretical values and empirical data, respectively.

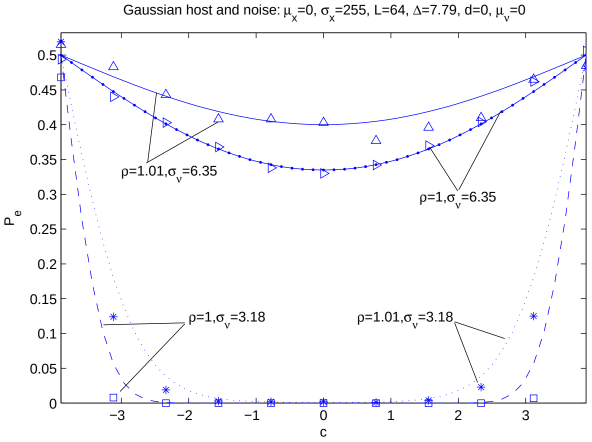

FIGURE 8. Bit error probability versus c while fixing ρ and σ_ν. Lines and symbols stand for theoretical values and empirical data, respectively.

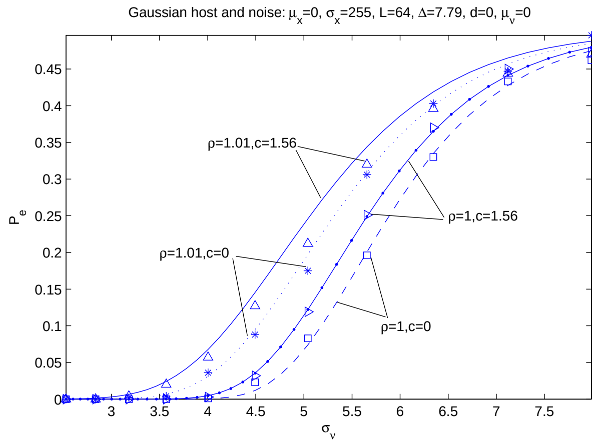

FIGURE 9. Bit error probability versus σ_ν while fixing ρ and c. Lines and symbols stand for theoretical values and empirical data, respectively.

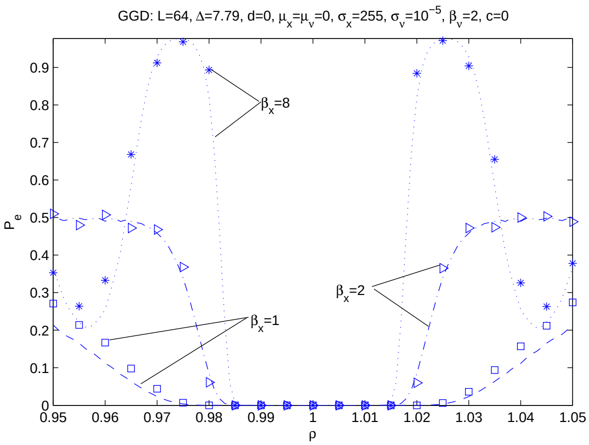

FIGURE 10. Bit error probability versus ρ for different values of β_x. Lines and symbols stand for theoretical values and empirical data, respectively.

5. **Experimental Results.** In order to verify the obtained theoretical results, we first conduct experiments on artificial signals. A set of 64000 random data, independently drawn from the GGD, are used as the host signal. It is decomposed into 1000 subvectors, and each one conceals one bit information. Under the composite attacks, we obtain the empirical BER. And the theoretical values of P_e are calculated in the same cases.

Fig.7 illustrates the plots of the P_e's versus the scaling factor ρ for several values of σ_ν and c. It is shown that DM is definitely very sensitive to the scaling attack. The probability of error is unacceptably high when ρ movies beyond the range $[0.98, 1.02]$. The existence of noise and constant change causes the increase of P_e further. And the effect of constant change becomes relatively distinct for strong noise. The theoretical approximation of P_e agrees almost perfectly with the empirical results particularly in the case of weak attacks. Fig. 7 also demonstrates that P_e is symmetrical around the point $\rho = 1$. Fig. 8 shows the performance of DM against constant change attack while fixing

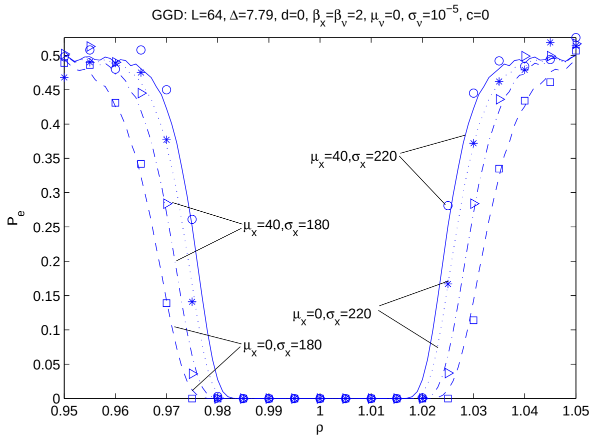

FIGURE 11. Bit error probability versus ρ for different values of μ_x and σ_x. Lines and symbols stand for theoretical values and empirical data, respectively.

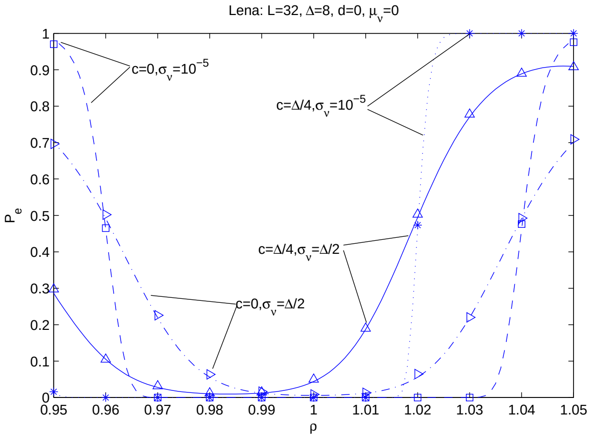

FIGURE 12. Bit error probability versus ρ for different values of c and σ_ν. Lines and symbols stand for theoretical values and empirical data, respectively.

ρ and σ_ν. As can be seen, the probability of error starts to grow rapidly as long as the absolute value of c approaches to $\Delta/2$. The given bound is decreased due to other two attacks. The estimated P_e's approximately equal to the empirical ones, which proves that our analysis is valid, but the accuracy of estimation gets worse for large c. At the same time, Fig. 8 shows that P_e is symmetrical around $c = 0$. The performance of DM against Gaussian noise attack is exhibited in Fig. 9 while fixing ρ and c. It is clear that the the probability of error becomes larger as σ_ν and ρ increases. The negative effect of the constant change on P_e is observed remarkably for large σ_ν. Similar to the previous tests, the theoretical P_e's fit the empirical ones very well and the maximal difference between them is lower than 0.02.

Our analysis also leads to insights about how statistical properties of the host signal affect the performance of DM. In Fig. 10, the theoretical performance values of DM are shown for different values of β_x together with the empirical data. In terms of scaling attack, impressively DM manifests better performance while the host signal obeys the GGD with a smaller value of β_x. The slope of the P_e versus ρ curves become less steep and lower values of P_e are attainable for large channel distortions as β_x decreases. This behavior can be explained as follows. For the GGD, the smaller β_x is, the more impulsive the shape, and the heavier the tails, so the lower the probabilities that the bigger values of x present over the range of interests. As a result, the introduced distortion $\epsilon\boldsymbol{y}$ by scaling attack degrades for the same value of ρ, and thus, P_e becomes smaller. We also observe that the theoretical approximation agrees almost perfectly with the empirical results for the cases $\beta_x = 2, 8$, but does worse for $\beta_x = 1$. This is because the CLT approximation to P_e may underestimate the importance of the tails of $p_X(x)$ with $\beta_x = 1$ and gives the smaller results than the true P_e [6]. In Fig. 11, the theoretical and empirical performance values of DM are shown for different values of σ_x and μ_x. As can be seen, for the same quantization step, the performance of DM goes down against valumetric scaling as σ_x and $|\mu_x|$ increase. The reason is apparent. Due to the increase of σ_x and $|\mu_x|$ the large distortions are introduced by scaling operation. However, for both constant change and noise cases, the performance of DM is insensitive to statistical properties of the host signal, because the two operations are independent from the watermarked signal. Hence, herein we just provide the results for scaling attack.

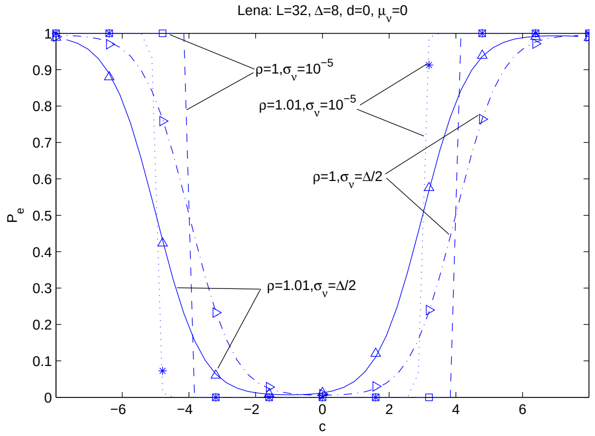

FIGURE 13. Bit error probability versus c for different values of ρ and σ_ν. Lines and symbols stand for theoretical values and empirical data, respectively.

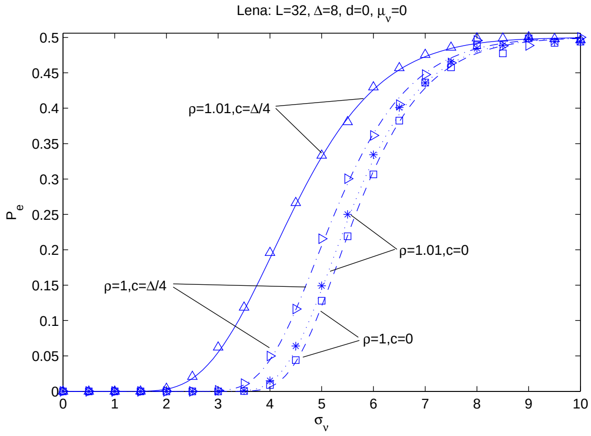

FIGURE 14. Bit error probability versus σ_ν for different values of ρ and c. Lines and symbols stand for theoretical values and empirical data, respectively.

Subsequently, we conduct experiments on real images. The watermark embedding is performed in the spatial domain, which allows us to measure the performance of DM without the impact of transform operations. Specifically, all pixels of one image are pseudorandomized and arranged in a vector as the host signal. Each 32 pixels conceal one bit information. The composite attacks with different values of ρ, σ_ν and c are carried out on the watermarked image to obtain the empirical BER. The computation of P_e need estimate the distribution parameters of image pixels. We employ the maximum likelihood (ML) estimator for the GGD [14].

The experimental results on Lena image are shown in Fig. 12-14. Since the PDF of image pixels are not symmetrical, the results are slightly different from ones given by simulation on artificial signals. Fig.12 depicts the plots of the P_e's versus the scaling factor ρ for several values of σ_ν and c. As can be seen, while $c \neq 0$, P_e is not symmetrical around $\rho = 1$ yet. Constant change distortions transmit the performance curves of DM. The same behavior can be also observed in Fig. 13, which illustrates the sensitivity of DM to the addition/subtraction of a constant luminance value while fixing ρ and σ_ν. In addition, Fig.12 shows the additive noise attack increases BER for small scaling distortions, but reduces BER as the scaling distortions become serious enough. The performance of DM against Gaussian noise attack is exhibited in Fig. 14 while fixing ρ and c. From all the tests, it is observed that the analytical prediction and empirical results for DM are sufficiently close, which verifies the validity of the given theoretical results.

Since the performance of DM depends on the distribution of host signal, different test results will be obtained on different images. Moreover, it is well known that there are no statistical distributions suitable for modeling the luminance component of common images in the spatial domain [15]. This makes that the prediction error of BER could be very large sometime. We tested the performance of DM on 'Crowd' and 'Mandrill' images for comparison purpose. The two images and Lena respectively represent three types of images: the PDF shape is very impulse for Crowd, a little flat for Mandrill and very flat for Lena. The estimated mean, variance and shape parameter of each image are displayed in Table. 1. The quantization step Δ for embedding was set to 8 in all tests. The robustness to amplitude scaling is shown in Fig. 15. We observe that DM achieves the best performance on Crowd, which is consistent with the results in Fig. 10. On the

TABLE 1. Distribution parameters.

Image	μ_x	σ_x	β_x
Crowd	85.2	50.9	1.5
Mandrill	129.1	42.4	3.3
Lena	99.1	47.9	10.6

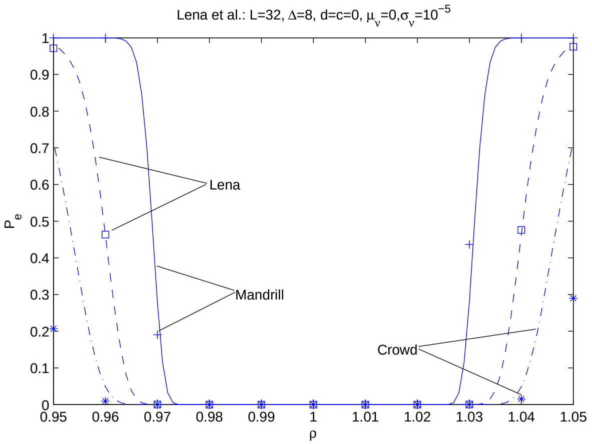

FIGURE 15. Bit error probability versus ρ for different images. Lines and symbols stand for theoretical values and empirical data, respectively.

contrary, DM performs worse on Mandrill than Lena. The behavior can be explained according to the obtained conclusions on artificial signals. Mandrill has smaller shape parameter and variance than Lena, which helps reduce BER of DM. However, due to the large mean luminance of Mandrill, not only the performance gain is canceled out but also the BER grows up. We also note that the analytical P_e departs from the empirical data too much on Crowd. That is mainly due to the fact that GGD is a poor model for the Crowd image. With respect to constant change and additive noise, it has been pointed out that the performance of DM is insensitive to the statistical properties of host signal. We may get similar results on other images as shown in Fig. 13 and Fig. 14. Thus, they are not provided here.

6. **Conclusion.**

Throughout this paper, we have theoretically evaluated the performance of DM facing the combination of valumetric scaling, additive noise and constant change. The analyses were developed under the assumption that both the host vector and noise vector have i.i.d components and the two vectors are independent. We accurately derived the general forms of the PDFs of the watermarked signal, the attacked signal and the extracted signal. By the derived PDFs, the decoding error probability was generally expressed in closed form. The specific analytical results were presented for the case of generalized Gaussian host and Gaussian noise. Moreover, the theoretical results can be easily extended by modeling the host and noise signals with other distributions. Simulations on artificial and real data show us the robustness of DM against the composite attacks and the analytical error probability agrees with the empirical one very well. Our analyses also discovered that DM is more robust to valumetric scaling for the impulsive PDF shape of host signal, but performs worse with the increase of the absolute mean value and variance of host signal. However, with respect to constant change and additive noise distortions, DM is insensitive to the statistical properties of host signal. These can ultimately guide the design of efficient watermarking algorithms based on DM.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (Grant No. 60803122, 61103018), by the Natural Science Foundation of Jiangsu Province (Grant No. BK2011442), by the Innovative Foundation of Yangzhou University (Grant No. 2011CXJ023), by the Opening Project of State Key Laboratory of Digital Publishing Technology, and by the Opening Project of State Key Laboratory of Software Development Environment (Grant No. SKLSDE-2011KF-08). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] I. J. Cox, M. L. Miller, and A. L. McKellips, Watermarking as communications with side information, *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127-1141, 1999.

[2] C. C. Lin and P. F. Shiu, High capacity data hiding scheme for dct-based images, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 3, pp. 220-240, 2010.

[3] C. H. Lin and C. Y. Yang, Multipurpose watermarking based on blind vector quantization (BVQ), *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 3, pp. 236-246, 2011.

[4] B. Chen and G. W. Wornell, Quantization index modulation: a class of provably good methods fordigital watermarking and information embedding, *Proc. of IEEE Symposium on Information Theory*, vol. 47, no. 4, pp. 1423-1443, 2001.

[5] J. J. Eggers, R. B auml, R. Tzschoppe, and B. Girod, Scalar costa scheme for information embedding, *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1003-1019, 2003.

[6] F. Pérez-Gonzàlez, F. Balado, and J. R. H. Martin, Performance analysis of existing and new methods for data hiding with known-host information in additive channels, *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 960-980, 2003.

[7] B. Li, J. He, J. Huang, and Y. Q. Shi, A survey on image steganography and steganalysis, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142-172, 2011.

[8] F. Bartolini, M. Barni, and A. Piva, Performance analysis of st-dm watermarking in presence of nonadditive attacks, *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2965-2974, 2004.

[9] J. P. Boyer, P. Duhamel, and J. Blanc-Talon, Performance analysis of scalar dc-qim for zero-bit watermarking, *IEEE Trans. Information Forensics and Security*, vol. 2, no. 2, pp. 283-289, 2007.

[10] F. Pérez-Gonzàlez and C. Mosquera, Quantization-based data hiding robust to linear-time-invariant filtering, *IEEE Trans. Information Forensics and Security*, vol. 3, no. 2, pp. 137-152, 2008.

[11] N. K. Kalantari and S. M. Ahadi, A logarithmic quantization index modulation for perceptually better data hiding, *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1504-1517, 2010.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.

[13] N. Saralees, A generalized normal distribution, *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685-694, 2005.

[14] M. N. Do and M. Vetterli, Wavelet-based texture retrieval using generalized gaussian density and kullbackcleibler distance,, *IEEE Trans. Image Processing*, vol. 11, no. 2, pp. 146-158, 2002.

[15] A. N. Netravali and B. G. Haskell, *Digital Pictures. Representation, Compression and Standards*, Plenum, New York, 1995.