# ESTIMATION OF ATTACKER'S SCALE AND NOISE VARIANCE FOR QIM-DC WATERMARK EMBEDDING

*R. (Inald) L. Lagendijk and Ivo D. Shterev*

Information and Communication Theory Group
Faculty of EEMCS, Delft University of Technology, The Netherlands
{R.L.Lagendijk,I.Shterev}@ewi.tudelft.nl

**ABSTRACT**

Quantization-based watermarking schemes are vulnerable to amplitude scaling. Therefore, the scaling factor needs to be estimated at the decoder side, such that the received (attacked) watermarked image can be inversely scaled prior to detection of embedded message bits. In this paper we propose a maximum likelihood (ML) approach to the estimation of the amplitude scaling factor and the variance of the noise in the attack channel. We model the probability density function (PDF) of the received (attacked) watermarked image amplitudes in case quantization index modulation with distortion compensation (QIM-DC) is used. Using this PDF, the ML estimator can be formulated. Our approach also handles the case that (subtractive) dithered quantization is employed. The behavior of the likelihood as a function of the scale and noise variance is such that efficient gradient-based optimization is unlikely to be successful. Hence, alternative optimization approaches need to be considered in future work.

## 1. INTRODUCTION

Watermark embedding schemes based on quantization theory are known to perform close to the information theoretical bounds [1]. A major drawback of these embedding schemes is, however, their vulnerability to common (non-)linear amplitude scaling. Two approaches have been proposed in literature to combat the amplitude scaling attack. These approaches are, firstly, the usage of scaling-robust codes such as modified trellis codes [2], and secondly, the estimation and inversion of the (non-)linear amplitude scaling [3, 4, 5]. In this paper we propose a maximum likelihood (ML) approach to the estimation of a linear scaling factor and the variance of the attacker's additive noise. The approach is based on the ML scale estimation procedure for audio watermark embedding we introduced in [5]. We extend this procedure to include dithered quantization as well as estimation of the attacker's noise variance. We apply the resulting estimation procedure on attacked watermarked images.

In order to mathematically formulate the ML-estimators of scale and variance, in Section 2 we consider the PDF of the watermarked image data $X(i,j)$[1]. The watermark embedding is based on dithered quantization-index modulation with distortion compensation (Figure 1) [6]. The model we use for the attack on the watermarked data is given by:

$$Y(i,j) = \beta X(i,j) + N'(i,j) \quad (1)$$
$$= \beta(X(i,j) + N(i,j)), \quad (2)$$

where $Y(i,j)$ is the received watermarked and attacked image, $N(i,j)$ is the attackers iid (zero-mean Gaussian) noise with variance $\sigma_n^2$, and $\beta$ is the amplitude scaling factor. The model in Eq. (2) assumes that scaling is applied after adding the attacker's noise $N(i,j)$. This is slightly different from the model commonly used (Eq. (1)), where scaling is applied *before* the attacker's noise is added [5]. Since, in our

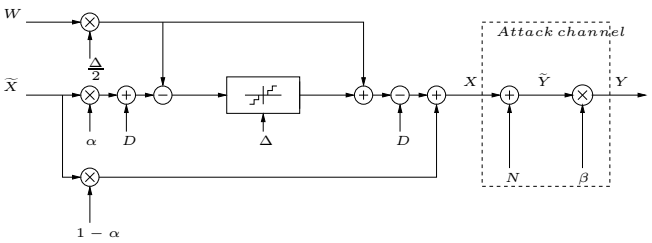


**Fig. 1**. *Block diagram of dithered QIM-DC.*

estimation procedure we estimate *both* the scaling factor $\hat{\beta}$ and the noise variance $\hat{\sigma}_n^2$, we can interchange noise addition and scaling without loss of generality. Clearly, we have $\text{var}(N) = \frac{1}{\beta^2}\text{var}(N')$. As we will see later on in this paper, such model has significant modeling and computational advantages. In Section 3 we describe the ML-estimation procedure, and graphically illustrate the behavior of the likeli-

This research is supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs..

[1]Throughout this paper we assume that $E[X(i,j)] = 0$.

hood function. Section 4 gives several experimental results, and Section 5 concludes the paper.

## 2. PDF MODEL OF THE WATERMARKED DATA

The marginal PDF of the watermarked data $X(i,j)$ consists of two conditional PDFs, one for each of the message bits embedded in the data, weighted by the probabilities of occurrence of the messages:

$$f_X(x) = f_{X|W=0}(x)P(W=0) + f_{X|W=1}(x)P(W=1),$$

where $f_{X|W=b}(x)$ is the conditional PDF[2] of the watermarked data if message bit $b$ is embedded, and $P(W=b)$ is the *a priori* probability of message bit $b \in \{0,1\}$. Throughout this paper we assume that the message bits are equally probable.

In [5, 7] we have derived expressions for the conditional PDFs $f_{X|W=b}(x)$ in the absence of dithering. The resulting expressions are given by:

$$f_{X|W=0}(x) = \sum_{k=-\infty}^{\infty} \frac{1}{1-\alpha} f_{\widetilde{X}}\Big(\frac{x-k\Delta}{1-\alpha}\Big) I_{A_{k|W=0}}(x)$$

$$f_{X|W=1}(x) = \sum_{k=-\infty}^{\infty} \frac{1}{1-\alpha} f_{\widetilde{X}}\Big(\frac{x-\frac{2k+1}{2}\Delta}{1-\alpha}\Big) I_{A_{k|W=1}}(x)$$

where $f_{\widetilde{X}}(x)$ is the PDF of the host data, $\Delta$ is the step size of the watermark embedding quantizer, $\alpha = \frac{\sigma_\Delta^2}{\sigma_\Delta^2+\sigma_n^2}$, and $\sigma_\Delta^2 = \frac{\Delta^2}{12}$. The indicator functions $I_{A_{k|W=b}}(x)$ are defined as:

$$I_{A_{k|W=b}}(x) = \begin{cases} 1 & \text{if } x \in A_{k|W=b} \\ 0 & \text{if } x \notin A_{k|W=b} \end{cases}$$

and

$$A_{k|W=0} = \Big\{\frac{\Delta}{\alpha}(k - \frac{1-\alpha}{2}) < X < \frac{\Delta}{\alpha}(k + \frac{1-\alpha}{2})\Big\}$$

$$A_{k|W=1} = \Big\{\frac{\Delta}{\alpha}(k + \frac{\alpha}{2}) < X < \frac{\Delta}{\alpha}(k + \frac{2-\alpha}{2})\Big\}$$

A graphical illustration of the resulting PDF of $X(i,j)$ is given in Figure 2 (left-hand side figure). The particular structure of $f_X(x)$ observed is solely due to the embedding of watermark bits using a scalar quantizer. The estimation approach of Eggers *et al.* [3, 5] explicitly makes use of the periodic structure of $f_X(x)$. The ML approach developed in this paper explicitly utilizes knowledge of the PDF structure, but does not rely on the periodic structure of the PDF $f_X(x)$. In the above model we have not yet included (subtractive) dithering of the quantizers. In this paper, we refrain from explicitly modeling this effect, but take the dithering of the quantizers into account directly in the ML-estimation procedure. Alternatively, the PDF of the dithered watermark can be modeled, as we discuss in [7]. The watermarked data is corrupted by the attacker's noise.

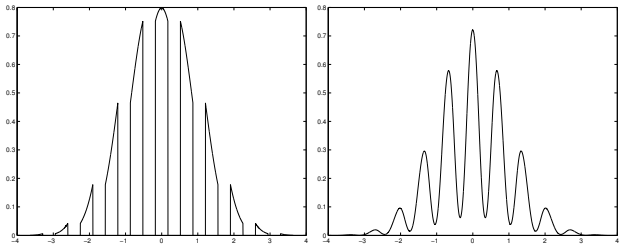

**Fig. 2**. *Illustration of the PDF of $X(i,j)$ (left) and $\widetilde{Y}(i,j)$ (right)*

The resulting PDF of $\widetilde{Y}(i,j) = X(i,j) + N(i,j)$ is given by

$$f_{\widetilde{Y}}(y;\sigma_n^2) = f_X(y) * f_N(y;\sigma_n^2),$$

where $f_N(y;\sigma_n^2)$ is the PDF of the attackers noise with variance $\sigma_n^2$. The effect of the convolution with the PDF $f_N(y;\sigma_n^2)$ is also illustrated in Figure 2 (right-hand side figure) for Gaussian noise.

Finally, the PDF of the scaled version $Y(i,j) = \beta\widetilde{Y}(i,j)$ is given by

$$f_Y(y;\beta,\sigma_n^2) = \frac{1}{\beta} f_{\widetilde{Y}}(\frac{y}{\beta};\sigma_n^2).$$

Note that in the above PDFs, we explicitly indicate the dependency on the attacker's parameters, namely the amount of additive noise $\sigma_n^2$ and the amplitude scaling $\beta$. In this paper we assume that the host data $\widetilde{X}(i,j)$ and attacker noise $f_N(y,\sigma_n^2)$ can be regarded as i.i.d. processes. Hence the joint PDF of the image $\mathbf{Y} = \{Y(i,j), 0 \le i \le N-1, 0 \le j \le M-1\}$ is equal to the product of the marginal PDFs:

$$f_{\mathbf{Y}}(\mathbf{y};\beta,\sigma_n^2) = \prod_{i,j} \frac{1}{\beta} f_{\widetilde{Y}(i,j)}(\frac{y}{\beta};\beta,\sigma_n^2). \tag{3}$$

## 3. MAXIMUM LIKELIHOOD ESTIMATOR

Using the PDF $f_{\mathbf{Y}}(\mathbf{y};\beta,\sigma_n^2)$ in Eq. (3) we can formulate the maximum likelihood estimator of the unknown parameters $\sigma_n^2$ and $\beta$ as:

$$\begin{aligned}(\hat{\beta},\hat{\sigma}_n^2) &= \arg\max_{\beta,\sigma_n^2} L(\beta,\sigma_n^2) \\ &= \arg\max_{\beta,\sigma_n^2} \log f_{\mathbf{Y}}(\mathbf{y};\beta,\sigma_n^2) \\ &= \arg\max_{\beta,\sigma_n^2} MN\log(\frac{1}{\beta}) + \sum_{i,j} \log f_{\widetilde{Y}(i,j)}(\frac{y}{\beta};\sigma_n^2)\end{aligned} \tag{4}$$

[2] We drop the index $(i,j)$ in the PDFs for notational convenience where possible.

All terms in Eq. (4) have been derived in Section 2, hence the likelihood function $L(\beta, \sigma_n^2)$ can be evaluated for a given combination $(\beta, \sigma_n^2)$.

We remark that the actual evaluation of $\log f_{\widetilde{Y}}(\frac{y}{\beta}; \sigma_n^2)$ requires the PDF $f_{\widetilde{Y}}(y; \sigma_n^2)$, which does *not* depend on $\beta$. In fact, into this PDF we substitute the *inversely scaled* (using the current estimate $\hat{\beta}$) amplitudes of the attacked watermarked image $Y(i,j)$. The efficient evaluation of the (rather complex expression of the) likelihood function is possible thanks to the model Eq. (2). If the model Eq. (1) had been used, the expression for the likelihood function would be dependent on $\beta$ in a more elaborate way, making efficient evaluation of $f_{\widetilde{Y}}(y; \sigma_n^2)$ far more difficult [7].

Clearly, the sophistication of the optimization method needed to maximize the likelihood function depends greatly on the behavior $L(\beta, \sigma_n^2)$. Figure 3 illustrates the behavior of the likelihood function for $\beta \in [0.5, 1.5]$ and $\sigma_n^2 \in [0.1, 10.0]$. In this case we have assumed that the host image $\widetilde{X}(i,j)$ is Gaussian distributed with $\sigma^2_{\widetilde{X}} = 900$, the signal-to-watermark ratio SWR = 30 dB (yielding $\Delta = 3.29$), the signal-to-noise ratio SNR = 33 dB (yielding a watermark-to-noise ratio WNR=3 dB, and $\alpha = 0.67$), and $\beta = 1.21$.

The optimum of the likelihood value can be found close to the actual attacker's parameters, but we also observe that $L(\beta, \sigma_n^2)$ consists of ridges with deep valleys in between. In fact, in case $f_{\widetilde{Y}}(y; \sigma_n^2) = 0$ for certain amplitudes, the likelihood function may become equal to negative infinity if the (inversely scaled) amplitudes of $Y(i,j)$ fall in these zero regions of the PDF of $\widetilde{Y}$. Such zero regions are more likely to occur for larger watermark-to-noise ratios (WNR); hence for larger WNR the behavior of $L(\beta, \sigma_n^2)$ becomes more irregular and efficient numerical optimization procedures for Eq. (4) (such as gradient-based optimization) become less likely to be successful.

In the above approach we have excluded dithering of the quantization process. Dithered quantization is necessary for security purposes, because without dither an attacker can easily recover the embedded bits. To include the dither sequence $D(i,j)$ (See Figure 1) in the estimation of $\beta$ and $\sigma_n^2$, we observe that:

- the input to the quantizer has a PDF that is different from the case that dither is not used. However, if the variance of the dither sequence is small compared to the variance of the host data, we can approximately ignore the effect of the dither on the PDF of $X(i,j)$. In our QIM-DC scheme, the dither is uniformly distributed over $[-\Delta, \Delta]$. The amount of dither has been chosen to be as small as possible without creating a security leak [7].

- the subtracted dither $D(i,j)$ after the quantizer in the watermark embedding scheme can be compensated for in the parameter estimation process by simply re-adding $D(i,j)$ to the inversely scaled (using the current estimate of $\beta$) attacked watermarked image $Y(i,j)$. Hence, in Eq. (4) we simply replace the argument $\frac{y}{\beta} = \frac{y(i,j)}{\beta}$ by $(\frac{y(i,j)}{\beta} - D(i,j))$. Again, this makes possible an efficient evaluation of the likelihood function.

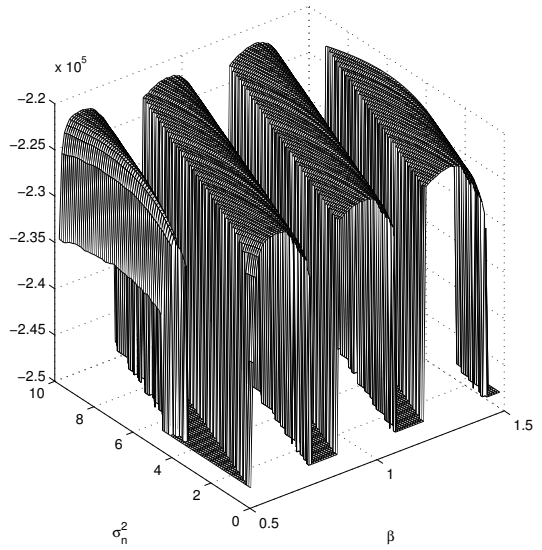


**Fig. 3**. *Illustration of $L(\beta, \sigma_n^2)$ without dithered quantization.*

Figure 4 illustrates the behavior of the likelihood function under the same conditions as those in Figure 3, but now taking into account dithered quantization. The maximum of the likelihood function can still be found in approximately the same location, and the behavior of the likelihood function itself has changed marginally. This confirms the validity of the assumption that we can safely ignore the effect of the dither on the PDF of $X(i,j)$ in the maximum likelihood parameter estimation.

An effect that we see in both Figures 3 and 4 is that the (correct) optimum of the likelihood function is relatively insensitive to the variance of the attacker's noise. This suggests that in a practical context we can limit the search of a proper value of $\sigma_n^2$ to a limited set of values.

## 4. EXPERIMENTAL RESULTS

We have applied the proposed scale and variance estimation procedure on synthetic Gaussian distributed images of size 256x256. The numbers obtained in this way give a performance ceiling, since the PDF of real images can obviously be modeled less accurately. Various embedding set-

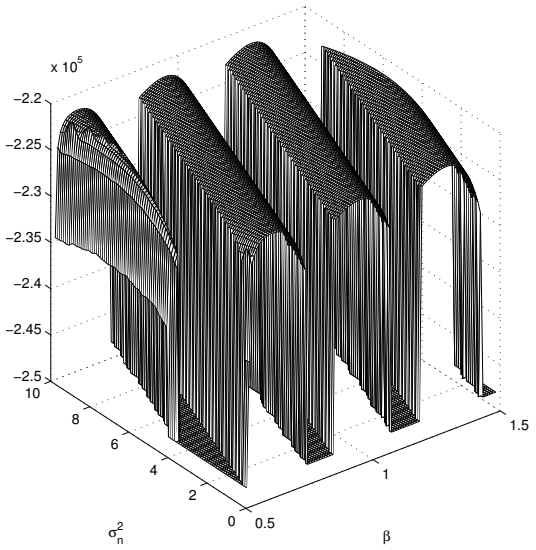


**Fig. 4**. *Illustration of $L(\beta, \sigma_n^2)$ including dithered quantization.*

tings have been used to benchmark the estimation procedure. The following table list our current results based on synthetic (Gaussian) data, with $\beta = 0.91$, watermark-to-signal ratio = 30 dB, and watermark-to-noise ratio (WNR)= 0, -10, and -20 dB.

**Table 1**. Experimental results using synthetic data.

| WNR | 0dB | -3dB | -10 dB |
|---|---|---|---|
| $\beta$ | 0.91 | 0.91 | 0.91 |
| $\sigma_n^2$ | 0.9 | 1.8 | 9.0 |
| $\hat{\beta}$ | 0.91 | 0.91 | 0.90 |
| $\beta$ search resolution | 0.01 | 0.01 | 0.01 |
| variance $\hat{\beta}$ | 0.00 | 0.01 | 0.05 |
| $\hat{\sigma}_n^2$ | 1.4 | 1.4 | 2.1 |
| $\sigma_n^2$ search resolution | 0.1 | 0.1 | 0.1 |
| variance $\hat{\sigma}_n^2$ | 0.3 | 0.7 | 5.0 |

Similar results are obtained for other values of $\beta$ and watermark-to-signal ratio. Our results show that the value of $\beta$ can be estimated much more reliably than $\sigma_n^2$. However, as we already remarked, the value of $\sigma_n^2$ seems not be important in finding the correct scaling factor $\beta$. Estimating $\beta$ below a WNR of -10 to -20 dB is useless, as the attacker's noise will effectively make the extraction of message bit very difficult (bit error rates approaching 0.5.

## 5. CONCLUSIONS

We have presented a maximum likelihood approach to the estimation of the amplitude scaling factor and variance of the noise of the attack channel. The estimation procedure is not dependent on whether dithered quantization or non-dither quantization is applied. Compared to our earlier work [5], our current approach is computationally very efficient. The optimum of the likelihood function is found around the correct values of the parameters $\beta$ and $\sigma_n^2$ for a wide range of watermark-to-noise ratios.

A major disadvantage of the ML approach is that the likelihood function shows a very irregular behavior for varying $\beta$ and $\sigma_n^2$. For that reason, in this paper we have optimized $L(\beta, \sigma_n^2)$ using a full search of the parameter space. In our future work we will focus on finding more efficient optimization procedures, but it is unlikely that these will be gradient-based.

Finally, our current amplitude scaling model includes only a linear scaling factor; clearly attackers may choose to use far more complicated non-linear amplitude scaling functions, for which our current estimation approach does not offer a solution.

## 6. REFERENCES

[1] P.smalllin, and A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *IEEE Trans. Information Theory*, vol. 49 (2), pp. 563-593, 2003.

[2] M.L. Miller, G. J. Doerr and J. Cox, "Dirty-Paper Trellis Codes For Watermarking", *Proc. IEEE Int. Conf. Image Processing*, pp. 129-132, 2002.

[3] J.J. Eggers, R. Bauml and B. Girod, "Estimation of Amplitude Modifications before SCS Watermark Detection," *Proc. IS&T/SPIE 16th Symposium on Electronics Imaging*, pp. 387-398, 2002.

[4] B. Bradley, "Improvement to CDF Grounded Lattice Codes," *Proc. IS&T/SPIE 16th Symposium on Electronics Imaging*, 2004.

[5] I.D. Shterev, R. L. Lagendijk and R. Heusdens, "Statistical Amplitude Scale Estimation for Quantization-based Watermarking," *Proc. IS&T/SPIE 16th Symposium on Electronics Imaging*, 2004.

[6] B. Chen and G. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding", *IEEE Trans. on Information Theory*, vol. 47, pp. 1423-1443, 2001.

[7] I.D. Shterev, R. L. Lagendijk, "Amplitude Scale Estimation for Quantization-based Watermarking", *IEEE Trans. on Signal Processing*, submitted June 2004.