CAPACITY ANALYSIS OF PERIODICAL WATERMARKING

E. Topak, S. Voloshynovskiy, O. Koval, T. Pun

University of Geneva - CUI, 24 rue General Dufour, CH 1211, Geneva 4, Switzerland

ABSTRACT

Periodical watermark embedding has been especially proposed to cope with geometrical attacks. Using a diversity approach, this method allows to decrease the probability of error in the case of additive attacks. It is usually admitted that the worst additive attack consists in the addition of additive white Gaussian noise (AWGN). However, our theoretical capacity analysis of periodical watermarking demonstrates that periodical AWGN in the optimal attacking strategy leads to more significant drop of the capacity than AWGN test channel from rate distortion.

1. INTRODUCTION

Watermarking has emerged as a means of addressing intellectual property and security issues in the context of digital media dissemination. For this purpose, a trade-off between watermark invisibility and robustness to intentional/unintentional attacks should be resolved assuming that sufficient information-theoretic requirements are satisfied. A security-capacity analysis should also be performed to examine possible security leakages.

Geometrical attacks, including translation, cropping, rotation, scaling, change of aspect ratio, shearing or general affine transforms, are of particular importance for practical robust watermarking. Without the need for signal removal, this type of distortions leads to a change of the channel state by signal de-synchronization.

There are several methods proposed for watermark recovery under geometrical attacks. They can be classified in three groups:

- methods that are performed in a fully transform invariant domain. In these methods, the Fourier-Mellin transform is applied to the magnitude of the host image spectrum with a log-log or log-polar coordinate mapping [1];
- methods that use extra synchronization templates. In these methods, template points are estimated in the FFT domain and they are then removed by using local interpolation [2];
- methods that are based on the self-reference principle. In these methods, the autocorrelation function (ACF) is used for watermark recovering [3].

The first two of these methods have the following drawbacks. Regarding the first one, the quality of the stego image is not high enough due to the embedding performed into perceptually important frequency components. Moreover, significant problems of watermark detection appear when rotation is applied simultaneously with change of aspect ratio. For the second one, efficient removal attack exists, allowing to destroy synchronization pattern [2]. Hence, there only remains the third class of methods to show good performances against geometrical attacks.

Self-reference based methods utilize the following property of the ACF of periodical signals [3]: it is presented as a regular grid of periodically placed local maximas (peaks). So, comparing the ACF of the geometrically distorted watermark with that of original watermark allows to perform its successful recovery under geometrical attacks.

Additionally, developers can benefit from periodical watermarking (Figure 1) for the resistance to additive attacks (for example, independent identically distributed (i.i.d.) AWGN attack) due to diversity decoding.

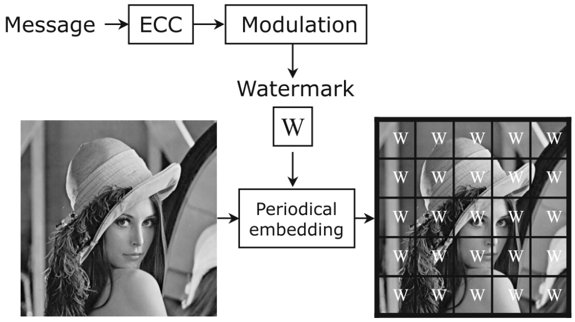

Figure 1: Periodical watermarking.

Most of the research up to date has focused on the development of practical methods that enable recovery under geometrical transforms. However, there is a lack of investigation regarding the information-theoretic properties of these techniques including the capacity analysis under optimal attacking strategies.

In this paper, we first investigate the capacity of periodical watermarking under an additive Gaussian attack when different levels of correlation are assumed. Our second goal is to show how the attacker could use the information given by the algorithm structure to decrease the rate of reliable communications.

The paper is organized as follows. In Section 2, an information-theoretic analysis of periodical watermarking is presented. In Section 3, a possible attacking senario is given when the structure of data-hiding algorithm is known, in some details, by the attacker. Finally Section 4 concludes the paper.

Notations. We use capital letters to denote scalar random variables X, bold capital letters to denote vector

Further author information: (Send correspondence to S.Voloshynovskiy): E-mail: svolos@cui.unige.ch, http://sip.unige.ch

random variables $\mathbf{X}$, corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively. The superscript N is used to designate length-N vectors $\mathbf{x} = x^N = [x_1, x_2, ..., x_N]^T$ with ith element x_i. The variance of X is denoted by σ_X^2. The covariance matrix of $\mathbf{X}$ is denoted by $\mathbf{C_{XX}}$. λ_i is the eigenvalue of $\mathbf{C_{XX}}$. We use $X \sim \mathcal{N}\left(0, \sigma_X^2\right)$ to indicate that a random variable X is Gaussian. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. ρ_{ij} is the correlation coefficient between X_i and X_j. The forward and inverse Fourier Transform of $\mathbf{W}$ is denoted by $\mathcal{F}\left\{\mathbf{W}\right\}$ and by $\mathcal{F}^{-1}\left\{\mathbf{W}\right\}$. log means $\log_2$ everywhere. $R_{\mathbf{WW}}(x,y)$ stands for the autocorrelation function of $\mathbf{W}$.

2. INFORMATION THEORETIC MODELLING OF PERIODICAL EMBEDDING

Periodical embedding of the watermark can be modelled in terms of parallel Gaussian channels [4] as in Figure 2. In this approach, each embedding block is interpreted as a channel through which the watermark information $\mathbf{W}$ is sent and each channel is attacked with corresponding noise $\mathbf{Z}_i$. Assume that $\mathbf{W} \sim \mathcal{N}\left(\mathbf{0}, \sigma_W^2 \mathbf{I}_n\right)$ is the watermark, where n is the embedding block linear size, and $\mathbf{Z}_i$ are the noise vectors such that $(Z_{1_i}, Z_{2_i}, \ldots, Z_{N_i}) \sim p(z_1, z_2, \ldots, z_N) = \mathcal{N}(\mathbf{0}, \mathbf{C_{ZZ}})$;

$$\mathbf{C_{ZZ}} = \sigma_Z^2 \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1N} \\ \rho_{21} & 1 & \ldots & \rho_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{N1} & \rho_{N2} & \ldots & 1 \end{bmatrix}. \quad (1)$$

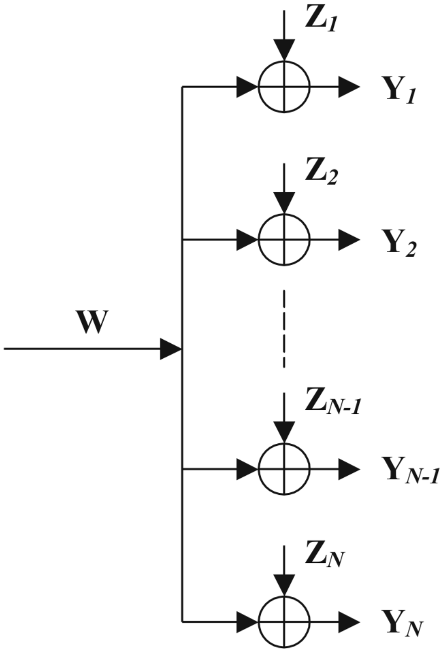

Figure 2: Parallel Gaussian channels model of the periodical embedding.

The maximum value of $\mathrm{I}\left(\mathbf{W}; \mathbf{Y}_1, ..., \mathbf{Y}_N\right)$ over the pdf of $\mathbf{W}$ will give the capacity of this embedding. Therefore, by comparing this value for different attacks, we can decide which one is the worst. Due to the i.i.d. distribution of both watermark and noise vectors, $\mathrm{I}\left(\mathbf{W}; \mathbf{Y}_1, ..., \mathbf{Y}_N\right)$ is equal to:

$$\mathrm{I}(\mathbf{W}; \mathbf{Y}_1, ..., \mathbf{Y}_N) = n\mathrm{I}(W; Y_1, ..., Y_N). \quad (2)$$

$\mathrm{I}(W; Y_1, ..., Y_N)$ can be written as

$$\begin{aligned} \mathrm{I}(W; Y_1, ..., Y_N) &= h(Y_1, ..., Y_N) - \\ &- h(Y_1, ..., Y_N | W). \end{aligned} \quad (3)$$

Since $Y_i = W + Z_i$ for each $i = 1, 2, ..., N$ and Z_i's are independent from W, equation (3) becomes:

$$\begin{aligned} \mathrm{I}(W; Y_1, ..., Y_N) &= h(Y_1, ..., Y_N) - h(Z_1, ..., Z_N) = \\ &= \frac{1}{2}\log\left((2\pi e)^N \left|\mathbf{C_{ZZ}} + \sigma_W^2 \mathbf{1}^T\mathbf{1}\right|\right) - \\ &- \frac{1}{2}\log\left((2\pi e)^N \left|\mathbf{C_{ZZ}}\right|\right) = \\ &= \frac{1}{2}\log\left(\frac{\left|\mathbf{C_{ZZ}} + \sigma_W^2 \mathbf{1}^T\mathbf{1}\right|}{\left|\mathbf{C_{ZZ}}\right|}\right), \end{aligned} \quad (4)$$

where $\mathbf{1} = [111\ldots1]$ and $|\mathbf{C_{ZZ}}|$ is the determinant of $\mathbf{C_{ZZ}}$. Since covariance matrices are symmetric matrices, they can be diagonalized by an orthogonal matrix. Thus, $\left|\mathbf{C_{ZZ}} + \sigma_W^2 \mathbf{1}^T\mathbf{1}\right|$ can be expanded to:

$$\begin{aligned} \left|\mathbf{C_{ZZ}} + \sigma_W^2 \mathbf{1}^T\mathbf{1}\right| &= \left|\mathbf{Q\Lambda Q}^T + \sigma_W^2 \mathbf{1}^T\mathbf{1}\right| = \\ &= \left|\mathbf{\Lambda} + \mathbf{Q}^T\sigma_W^2 \mathbf{1}^T\mathbf{1Q}\right| = \\ &= \left|\mathbf{\Lambda} + \sigma_W^2 \mathbf{Q}^T\mathbf{1}^T\mathbf{1Q}\right|, \end{aligned} \quad (5)$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose non-zero elements are equal to the eigenvalues of $\mathbf{C_{ZZ}}$ and $\mathbf{Q}$ is a matrix whose columns are eigenvectors of $\mathbf{C_{ZZ}}$. According to [5], we have:

$$\begin{aligned} \left|\mathbf{\Lambda} + \sigma_W^2 \mathbf{Q}^T\mathbf{1}^T\mathbf{1Q}\right| &= \left(1 + \sigma_W^2 \mathbf{1Q\Lambda}^{-1}\mathbf{Q}^T\mathbf{1}^T\right)\left|\mathbf{\Lambda}\right| = \\ &= \left(1 + \sigma_W^2 \mathbf{1C}_{\mathbf{ZZ}}^{-1}\mathbf{1}^T\right)\left|\mathbf{\Lambda}\right|, \end{aligned} \quad (6)$$

where $\mathbf{Q\Lambda}^{-1}\mathbf{Q}^T = \mathbf{C}_{\mathbf{ZZ}}^{-1}$. Similarly, $|\mathbf{C_{ZZ}}|$ can be expanded as:

$$\begin{aligned} \left|\mathbf{C_{ZZ}}\right| &= \left|\mathbf{Q\Lambda Q}^T\right| = \\ &= \left|\mathbf{Q}^T\mathbf{Q}\right|\left|\mathbf{Q\Lambda Q}^T\right| = \\ &= \left|\mathbf{Q}^T\right|\left|\mathbf{Q}\right|\left|\mathbf{Q\Lambda Q}^T\right| = \\ &= \left|\mathbf{Q}^T\right|\left|\mathbf{Q\Lambda Q}^T\right|\left|\mathbf{Q}\right| = \\ &= \left|\mathbf{Q}^T\mathbf{Q\Lambda Q}^T\mathbf{Q}\right| = \\ &= \left|\mathbf{\Lambda}\right|. \end{aligned} \quad (7)$$

Therefore, inserting the results of equations (6)-(7) in equation (4), one obtains:

$$\mathrm{I}(W; Y_1, ..., Y_N) = \frac{1}{2}\log\left(1 + \sigma_W^2 \mathbf{1C}_{\mathbf{ZZ}}^{-1}\mathbf{1}^T\right). \quad (8)$$

In the next sections we will show how the structure of $\mathbf{C_{ZZ}}$ will influence the system performance.

2.1 AWGN attack

In the case of an AWGN attack, $\mathbf{C_{ZZ}}$ will be a diagonal matrix (i.e. $\rho_{ij}=0$ if $i \neq j$)

$$\mathbf{C_{ZZ}} = \sigma_Z^2 \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}. \tag{9}$$

Therefore, the inverse of $\mathbf{C_{ZZ}}$ will be equal to $\frac{1}{\sigma_Z^2}\mathbf{I}_N$ and equation (8) can be rewritten as follows:

$$\begin{aligned} \mathrm{I}(W; Y_1, ..., Y_N) &= \frac{1}{2}\log\left(1 + \sigma_W^2 \mathbf{1}\frac{1}{\sigma_Z^2}\mathbf{I}\mathbf{1}^T\right) = \\ &= \frac{1}{2}\log\left(1 + \frac{\sigma_W^2}{\sigma_Z^2}\mathbf{1}\mathbf{1}^T\right) = \\ &= \frac{1}{2}\log\left(1 + N\frac{\sigma_W^2}{\sigma_Z^2}\right) \end{aligned} \tag{10}$$

that is the maximum embedding rate under an AWGN attack.

2.2 Periodical noise attack

If we apply the same noise in all channels (i.e. $Z_1 = Z_2 = \ldots = Z_N$), $\mathbf{C_{ZZ}}$ will have the following form:

$$\mathbf{C_{ZZ}} = \sigma_Z^2 \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix}. \tag{11}$$

However, since $|\mathbf{C_{ZZ}}| = 0$, $\mathbf{C_{ZZ}^{-1}}$ does not exist. Nevertheless, we can write [5]

$$|\mathbf{C_{ZZ}}| = \prod_{i=1}^{N} \lambda_i. \tag{12}$$

For $\mathbf{C_{ZZ}}$, $\lambda_1 = N\sigma_Z^2$ and $\lambda_{i,i\neq 1} = 0$. Similarly, for $\mathbf{C_{ZZ}} + \sigma_W^2 \mathbf{1}^T\mathbf{1}$, $\lambda_1 = N\left(\sigma_Z^2 + \sigma_W^2\right)$ and $\lambda_{i,i\neq 1} = 0$. Putting these values into equation (12) and afterwards into equation (4), we will have

$$\begin{aligned} \mathrm{I}(W; Y_1, ..., Y_N) &= \frac{1}{2}\log\left(\frac{N\left(\sigma_Z^2 + \sigma_W^2\right)}{N\sigma_Z^2}\right) = \\ &= \frac{1}{2}\log\left(1 + \frac{\sigma_W^2}{\sigma_Z^2}\right) \end{aligned} \tag{13}$$

That is the maximum embedding rate under periodical noise attack. If the embedding distortions are bounded in the L_2-norm sense, it was recently shown [6] that the worst attack consists in a MMSE estimation of the host image and AWGN test channel from rate distortion theory. Comparing (10) and (13), it becomes clear that performance of periodical watermarking will be decreased more when AWGN in the worst attack will be replaced by the periodical noise.

3. PERIODICAL EMBEDDING: SECURITY LEAKAGE

In Section 2, it was shown that the capacity reduction of periodical watermarking is more severe in the case of periodical noise attack than for the case of AWGN attack. This feature can be effectively used by the attacker to reduce communication rate of practical watermarking systems. The rest of this section is dedicated to the main aspects of the periodical attacking of periodical watermarking.

3.1 Estimation of the watermark

Assuming additive watermarking, embedding can be modelled as:

$$\mathbf{S} = \mathbf{W} + \mathbf{X}, \tag{14}$$

where $\mathbf{S}$ is the stego image, $\mathbf{W} \sim \mathcal{N}\left(\mathbf{0}, \sigma_W^2\mathbf{I}_n\right)$ is the watermark message and $\mathbf{X} \sim \mathcal{N}\left(\overline{\mathbf{x}}, \mathbf{C_{XX}}\right)$ is the original image. A Maximum A Posteriori (MAP) estimation can be used to estimate $\mathbf{W}$ [7] as

$$\begin{aligned} \widehat{\mathbf{W}} &= argmax_{\mathbf{W}\in\mathbf{R}^N} P_{\mathbf{S}|\mathbf{W}}(\mathbf{s}|\mathbf{w})P_{\mathbf{W}}(\mathbf{w}) = \\ &= \sigma_W^2\mathbf{I}_n\left(\sigma_W^2\mathbf{I}_n + \mathbf{C_{XX}}\right)^{-1}\mathbf{S}. \end{aligned} \tag{15}$$

3.2 Calculation of the autocorrelation function of the watermark

The autocorrelation function of the estimated watermark (Figure 3) can be efficiently calculated in the Fourier transform domain by [8]

$$R_{\mathbf{WW}}(x,y) = \mathcal{F}^{-1}\left\{\left|\mathcal{F}\left\{\widehat{W}(m,n)\right\}\right|^2\right\}, \tag{16}$$

where $\mathcal{F}\left\{\widehat{W}(m,n)\right\}$ is the 2-D Fourier transform of $\widehat{\mathbf{W}}$.

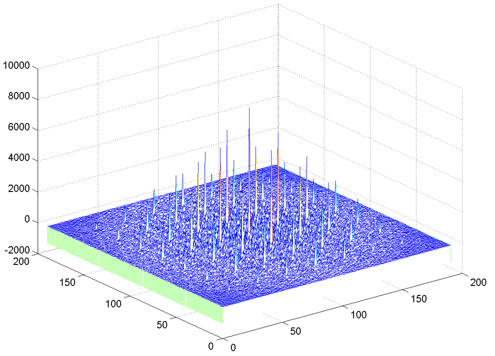

Figure 3: Autocorrelation function of a locally Gaussian watermark with embedding block size is 24x24 and image size is 96x96.

Having the estimate of the watermark ACF, the attacker is able to estimate the block size (Figure 4). Therefore, he/she should generate an i.i.d. AWGN with the dimensionality of the embedding block and insert this noise in all blocks. This attacking strategy will constrain the information-theroretical limits of periodical watermarking by equation (13). It should be pointed out that when adding i.i.d. AWGN to the whole image, the performance is bounded by (10), which corresponds to a higher capacity.

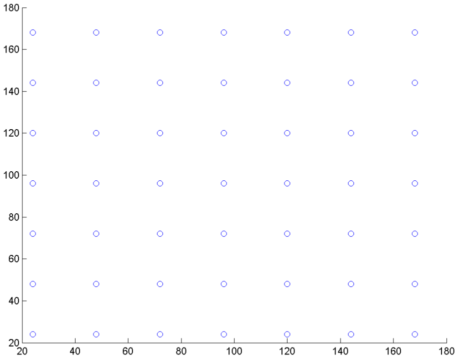

Figure 4: Orientation of the peaks of Figure 3 on the X&Y axes

4. CONCLUSION

Periodical embedding has been developed against geometrical attacks, and has also been used in various watermarking schemes in a diversity framework. For the evaluation of these schemes, additive white Gaussian noise has so far been considered the worse attack. We however demonstrate in this paper that for periodical embedding of the watermark, periodical noise is worse than additive white Gaussian noise. The only property of the periodical noise that matters is its embedding block size. Since this information is also leaked by the autocorrelation function of the embedded periodical watermark, it is easy to design a proper attacking noise. Periodical watermarking schemes should therefore be tested against periodical noise instead of additive white Gaussian noise.

5. ACKNOWLEDGMENTS

This paper was partially supported by the Swiss SNF grant No 21-064837.01, SNF Professorship grant No PP002-68653/1 and Interactive Multimodal Information Management (IM2) projects. The authors are thankful to Jose Vila, Frederic Deguillaume and Yuriy Rytsar for many helpful and interesting discussions. S. Voloshynovskiy is also thankful to K. Mihcak for interesting discussions during his stay at Microsoft Research, Redmond, USA.

REFERENCES

[1] J. J. K. . Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking", *In Proceedings of ICIP 97, IEEE International Conference on Image Processing*, pp. 536-539, Santa Barbara, CA, October 1997

[2] S. Voloshynovskiy, S. Pereira, V. Iquise and T. Pun, "Attack modelling: Towards a second generation benchmark", *Signal Processing*, 81, 6, pp. 1177-1214, June 2001. Special Issue: Information Theoretic Issues in Digital Watermarking, 2001

[3] M. Kutter, "Watermarking resisting to translation, rotation and scaling," *Proceedings of SPIE*, November 1998

[4] T. M. Cover, J. A. Thomas *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.

[5] http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/ *Matrix Reference Manual*.

[6] P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *IEEE Trans. on Information Theory*, Vol. 49, No. 3, pp. 563-593, March 2003

[7] S. Voloshynovskiy, S. Pereira, A. Herrigel, N. Baumgärtner and T. Pun, "Generalized watermark attack based on watermark estimation and perceptual remodulation" *IS&T/SPIE's 12th Annual Symposium, Electronic Imaging 2000: Security and Watermarking of Multimedia Content II*, Vol. 3971 of SPIE Proceedings, San Jose, California USA, 23-28January 2000

[8] A. L. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989