IMPROVING THE DETECTION RELIABILITY OF CORRELATION-BASED WATERMARKING TECHNIQUES

A. Giannoula, A. Tefas, N. Nikolaidis and I. Pitas

Department of Informatics, Aristotle University of Thessaloniki,
Box 451, Thessaloniki 540 06, GREECE
e-mail: alexia@comm.utoronto.ca,{tefas,nikolaid,pitas}@zeus.csd.auth.gr

ABSTRACT

The performance of watermarking schemes based on correlation detection is closely related to the frequency characteristics of the watermark sequence. In order to improve both detection reliability and robustness against attacks, embedding of watermarks with high-frequency spectrum, in the low frequencies of the DFT domain, is introduced in this paper and theoretical analysis of correlation-based watermarking techniques with multiplicative embedding is performed. The proposed watermarking framework is successfully applied to audio signals, demonstrating its superiority with respect to both robustness and inaudibility. Experiments are conducted, in order to verify the validity of the theoretical analysis results.

1. INTRODUCTION

Watermarking technology has recently emerged as a means of copyright enforcement and content verification of multimedia data and thus, a promising tool against digital piracy. Watermarking techniques based on correlation detectors have been widely popular in the watermarking community. There has been limited effort, though, in theoretically evaluating the performance of such techniques with respect to detection reliability. In [1, 2] the statistical properties of the correlation detectors when pseudorandom watermark signals are used are explored. Research conducted so far, establishes the dependence of the system detection reliability on the frequency characteristics of the watermark signal. A theoretical performance analysis of additive embedding, correlation-based watermarking schemes using chaotic sequences is provided in [3], where the superiority of highpass skew tent chaotic watermarks against both white and lowpass watermarks, in the case of distortion-free host signals is demonstrated. High-frequency watermarks, however, are severely affected when lowpass attacks, such as compression and lowpass filtering, are imposed on the host signal.

In this paper, a watermark with highpass spectral characteristics, embedded in the low frequencies of the DFT domain is proposed. This approach guarantees that the system becomes robust to lowpass attacks while preserving its correlation properties and thus, leading to an enhancement of the detector reliability. For this purpose, piecewise-linear Markov chaotic watermarks, and in particular skew tent maps, will be employed. Their major advantage is their easily controllable spectral/correlation properties, a fact that makes them a good alternative to the widely used pseudorandom signals [3, 4]. The benefits of high-frequency watermark embedding in the low-frequency subbands of the DFT domain is established by theoretical evaluation of the statistic properties of the correlator. Theoretical performance analysis of correlation-based techniques using chaotic sequences and obeying a multiplicative rule of embedding, is also undertaken throughout this paper. It is worth noting that the proposed method is generic and can be applied on various modalities, such as images, audio and video, in combination with appropriate countermeasures for other attacks, e.g. geometric distortions. In this paper, we present a simple application of the investigated technique in watermarking of audio data.

2. DESCRIPTION OF THE WATERMARKING MODEL

Let x and X be the original signal and its Discrete Fourier Transform (DFT) coefficients, correspondingly, both of length N_s. Watermark embedding is performed by modifying the magnitude $F = |X|$ of the DFT coefficients, which will be considered hereafter as the host signal. Since we require that the watermark signal affects a specific low frequency subband of the host signal, it can be represented by the following formula:

$$W(n) = \begin{cases} W_o(i), & if\ aN_s \le n \le bN_s, 0 \le i < N-1 \\ W'_o(i), & if\ (1-b)N_s \le n \le (1-a)N_s, \\ & 0 \le i < N-1, \\ 0, & otherwise \end{cases} \quad (1)$$

where $n = 0, 1, \ldots, N_s - 1$ and coefficients a, b $(0 < a < b \le 0.5)$ control the frequency terms that will be modified. The watermark signal W_o that is used for the construction of W consists of N samples, where $N = \lceil (b-a)N_s \rceil$, and it is generated through an appropriately selected function, $W_o = g(K, N)$, where K denotes the secret key, accessible only to the copyright owner or authorized users. W_o is embedded in the low frequency components around coefficient 0, according to a multiplicative superposition rule. Due to the symmetry of the DFT magnitude, a reflected version of the signal $W'_o(i) = W_o(N-i-1)$ is also embedded in the low frequency components around coefficient $N_s - 1$. Multiplicative embedding is used since it incorporates a simple perceptual masking effect by modifying coefficients proportionally to their magnitude: $F^{'} = F + pWF$, where $F^{'}$ is the watermarked signal and p is a constant that controls the watermark embedding power.

The correlation detector will be employed in this paper to examine whether the signal under test F_t contains a watermark W_d or not under a binary-decision hypothesis test framework, where the following hypotheses are considered:

- H_0: The test signal F_t under investigation contains the watermark W_d, i.e., $F_t = F_o + pW_dF_o$, F_o being the host signal (DFT magnitude).

- H_1: The test signal F_t under investigation does not contain the watermark W_d.

Event H_1 can be analyzed to the events H_{1a}, representing the case where the test signal is not watermarked, i.e., $F_t = F_o$ and H_{1b}, representing the case where the test signal is watermarked with a different watermark $W_d^{'} \neq W_d$, i.e., $F_t = F_o + pW_d^{'}F_o$. The three events H_o, H_{1a}, H_{1b} mentioned above, can be combined in:

$$F_t = F_o + pW_eF_o \tag{2}$$

where the watermark W_d is indeed embedded in the signal if $p \neq 0$ and $W_e = W_d$ (event H_0), and it is not embedded in the signal if $p = 0$ (no watermark is present, event H_{1a}) or $p \neq 0$ and $W_e = W_d^{'} \neq W_d$ (wrong watermark presence, event H_{1b}).

The correlation between the signal under investigation and the watermark sequence is given by:

$$c = \frac{1}{N}\sum_{n=0}^{N-1}\left(F_o(n)W_d(n) + pW_e(n)F_o(n)W_d(n)\right) \tag{3}$$

In order to decide on the valid hypothesis, c is compared against a suitably selected threshold T. The performance of such a correlation-based technique can be measured in terms of the probability of false alarm $P_{fa}(T)$ (probability of erroneously detecting the existence of a specific watermark in a signal that is not watermarked or that is watermarked with a different watermark) and the probability of false rejection $P_{fr}(T)$ (probability of erroneously rejecting the existence of a specific watermark in a signal that is indeed watermarked) and can be graphically represented by the receiver operating characteristic (ROC) curve (plot of P_{fa} versus P_{fr}). For the interested reader, alternative correlation-based detection schemes can be found in [5].

3. THEORETICAL PERFORMANCE ANALYSIS

To proceed with the extraction of the correlation detector statistics, we first need to make certain assumptions about the host signal. We assume that the host signal is wide-sense stationary and has a first order exponential autocorrelation function [2]:

$$R_{F_o}[k] = \mu_{F_o}^2 + \sigma_{F_o}^2\beta^k, \quad k \geq 0, \;\; |\beta| \leq 1 \tag{4}$$

where β is the parameter of the autocorrelation function and $\sigma_{F_o}^2$ is the host signal variance. Despite the rather simplistic assumptions, the validation of the theoretical results using numerical experiments on audio signals (Section 4), demonstrates that these assumptions hold in a great extent. In the rest of this paper, we will denote by $R_g[\boldsymbol{k}]$ the r-th order correlation statistic of a wide-sense stationary signal g:

$$R_g[k_1, k_2, \ldots, k_r] = E\left[g[n]g[n+k_1]g[n+k_2]\ldots g[n+k_r]\right] \tag{5}$$

We, also, adopt the procedure of subtracting the mean value $E[F_t]$ from the test signal F_t prior to detection, for additional improvement in the detection reliability [3]. By subtracting $E[F_t]$ from the test signal we obtain the signal F_t':

$$F_t' = F_t - E[F_t] = F_o' + pW_eF_o \tag{6}$$

where $F_o' = F_o - \mu_{F_o}$. Obviously $\mu_{F_o'} = 0$.

For the watermark sequences explored in this paper, the Central Limit Theorem for random variables with small dependency [6] can be used, in order to establish that the involved correlator output pdfs under the two hypotheses, $f_{c|H_0}$, $f_{c|H_1}$, attain a Gaussian distribution. Therefore, these pdfs can be described by their mean $\mu_{c|H_0}$, $\mu_{c|H_1}$, and variance values $\sigma_{c|H_0}^2$, $\sigma_{c|H_1}^2$. Using expression (3), the mean value and the variance of the correlation detector c, under the above assumptions, can be evaluated:

$$\begin{aligned}\mu_c \;\; &= E[c] = \frac{1}{N}\left(\sum_{n=0}^{N-1}E[F_o'(n)]E[W_d(n)] + \sum_{n=0}^{N-1}pE\left[F_o(n)\right]E\left[W_e(n)W_d(n)\right]\right) \\ &= \frac{1}{N}\sum_{n=0}^{N-1}pE\left[F_o(n)\right]E\left[W_e(n)W_d(n)\right]\end{aligned} \tag{7}$$

$$\begin{aligned}\sigma_c^2 \;\; = \;\; & \frac{1}{N^2}\left[\sum_{n=0}^{N-1}(E[F_o^{'2}(n)]E[W_d^2(n)] \right. \\ & +p^2E[F_o^2(n)]E[W_d^2(n)W_e^2(n)] \\ & +2pE[F_o(n)F_o'(n)]E[W_e(n)W_d^2(n)]) \\ & +\sum_{n=0}^{N-1}\sum_{m=0,m\neq n}^{N-1}\big(E[F_o'(n)F_o'(m)]E[W_d(n)W_d(m)] \\ & +pE[F_o'(n)F_o(m)]E[W_e(m)W_d(n)W_d(m)] \\ & +pE[F_o(n)F_o'(m)]E[W_e(n)W_d(n)W_d(m)] \\ & \left.+\; p^2E[F_o(n)F_o(m)]E[W_e(n)W_e(m)W_d(n)W_d(m)]\big)\right] \\ & -\mu_c^2\end{aligned} \tag{8}$$

The above formulas are general and can be applied to all three events, H_0, H_{1a} and H_{1b}. The statistical independence between the host signal F_o and both watermarks W_e and W_d, has been taken into account to derive these formulas.

The performance of the correlation-based detector depends, as we will show, on the autocorrelation and cross-correlation functions, or equivalently the power spectrum of the constructed watermarks. Therefore, functions that generate watermarks attaining desirable spectral/correlation properties are required. The controllable spectrum of certain chaotic signals has lately motivated us to use chaotic maps in watermarking schemes [3, 7], as an efficient alternative to pseudorandom watermarks. The class of the eventually expanding piecewise-linear Markov maps $\mathcal{M} : [0,1] \rightarrow [0,1]$, is particularly amenable to mathematical analysis [8]. The Markov sequences that will be employed hereafter, attain an exponential autocorrelation function of the form (4), where β is an eigenvalue of the corresponding Frobenius-Perron (FP) matrix [9], which is involved in the statistics evaluation process. Zero mean chaotic watermark sequences W_o can be constructed by subtracting the sequence mean value. Bearing in mind that $W_d[n] = W_e[n+k]$ according to [3] the mean value and the variance of the correlation c are formulated as follows:

$$\mu_c = p\,\mu_{F_o}\left(R_x[k] - \mu_x^2\right) \tag{9}$$

$$\sigma_c^2 \;\; = \;\; \frac{1}{N}\left(R_x[0,k,k] - 2\mu_xR_x[0,k] + 2\mu_x^2R_x[0] - 2\mu_xR_x[k,k]\right.$$

$$
\begin{aligned}
&+ \quad 4\mu_x^2 R_x[k] - 3\mu_x^4\big)\, p^2 R_{F_o}[0] + \frac{1}{N}\,(2pR_x[k,k] \\
&- \quad 4p\mu_x R_x[k] + (1-2p\mu_x)R_x[0] + 4p\mu_x^3 - \mu_x^2\big)\,\sigma_{F_o}^2 \\
&+ \quad \frac{2}{N^2}\left[\sum_{m=1}^{N-1}(N-m)(R_{F_o}[m]-\mu_{F_o}^2)R_x[m]\right. \\
&+ \quad (4p\mu_x^3-\mu_x^2)\sum_{m=1}^{N-1}(N-m)(R_{F_o}[m]-\mu_{F_o}^2) \\
&- \quad p\mu_x\sum_{m=1}^{N-1}(N-m)(R_{F_o}[m]-\mu_{F_o}^2)(2R_x[m]+2R_x[k] \\
&\quad +R_x[m+k]+R_x[k-m]) \\
&+ \quad p\sum_{m=1}^{N-1}(N-m)(R_{F_o}[m]-\mu_{F_o}^2)(R_x[k,k-m] \\
&+ \quad R_x[k,m+k]) + \sum_{m=1}^{N-1}(N-m)R_{F_o}[m] \\
&\quad \big\{p^2 R_x[m,k,m+k] - p^2\mu_x(R_x[m,k]+R_x[m,m+k] \\
&+ \quad R_x[k,k-m]+R_x[k,m+k]) + p^2\mu_x^2(2R_x[m]+2R_x[k] \\
&+ \quad R_x[k-m]+R_x[m+k])\} \\
&- \quad \left. 3p^2\mu_x^4\sum_{m=1}^{N-1}(N-m)R_{f_o}[m]\right] - \mu_c^2
\end{aligned} \tag{10}
$$

where $R[\boldsymbol{k}]$ is given by (5). The above formulas involve several moments of the watermark x. Furthermore, by setting the watermark shift k equal to zero and the embedding factor $p > 0$ expressions for the event H_o can be derived. The event H_{1a} can be described by an embedding factor p of zero value and the event H_{1b} by setting $p > 0$ and $k > 0$.

A close examination of equation (9), leads to the conclusion that the mean of the correlation detector obtains the same value for all watermarks of the same variance applied on the same host signal with a particular embedding power p. Therefore, the system performance is determined only by the variance of the correlation detector, which in turn is affected by the watermark spectrum. Thus, one can improve the system performance by selecting watermarks of desirable spectrum properties. High-frequency watermarks ($\beta \to -1$), which, as demonstrated in [3], lead to reduced correlation variance and, thus, improved performance when no attacks are inflicted, are indeed a good choice.

If furthermore, we restrict our attention to the class of *skew tent* Markov maps

$$
\mathcal{T}(x) = \begin{cases} \frac{1}{\lambda}\,x & ,\ 0 \le x \le \lambda \\ \frac{1}{\lambda-1}\,x + \frac{1}{1-\lambda} & ,\ \lambda < x \le 1 \end{cases} \quad \lambda \in (0,1) \tag{11}
$$

whose first, second and third order correlation statistics have been derived in [3], equations (9) and (10), can be modified so that they describe the two events H_0 and H_{1a} in a more compact form:

$$
\mu_c = \begin{cases} 0 & ,\ p = 0\ (H_{1a}) \\ \frac{p\mu_{F_o}}{12} & ,\ k = 0, p \neq 0\ (H_0) \end{cases} \tag{12}
$$

$$
\sigma_c^2 = \begin{cases}
\dfrac{\sigma_{F_o}^2}{12N^2}\,\dfrac{N - 2\beta e_2 - N\beta^2 e_2^2 + 2(\beta e_2)^{N+1}}{(1-\beta e_2)^2},\ p = 0\ (H_{1a}) \\
\dfrac{\sigma_{F_o}^2}{N}(\dfrac{1}{12}+\dfrac{p^2}{80}) + \dfrac{p^2\mu_{F_o}^2}{180N} \\
+\dfrac{2\sigma_{F_o}^2}{N^2}\left[\dfrac{3a-2-ap}{12(3a-2)}\,\dfrac{(\beta e_2)^{N+1} - N\beta^2 e_2^2 + N\beta e_2 - \beta e_2}{(1-\beta e_2)^2}\right. \\
+\dfrac{p^2}{144}\,\dfrac{\beta^{N+1} - N\beta^2 + N\beta - \beta}{(1-\beta)^2} \\
\left.+\ \dfrac{15ap + 3ap^2 - 2p^2}{180(3a-2)}\,\dfrac{(\beta e_1)^{N+1} - N\beta^2 e_1^2 + N\beta e_1 - \beta e_1}{(1-\beta e_1)^2}\right] \\
+\dfrac{p^2\mu_{F_o}^2}{90N^2}\,\dfrac{e_1^{N+1} - Ne_1^2 + Ne_1 - e_1}{(1-e_1)^2},\ k = 0, p \neq 0\ (H_0)
\end{cases} \tag{13}
$$

where e_1, e_2 are eigenvalues of the FP matrix $\boldsymbol{P}_3$ and β is the parameter of the host signal autocorrelation function given by (4).

Using the derived statistics for the correlator output, theoretical expressions for the ROC curves can be constructed.

4. EXPERIMENTAL RESULTS

The first set of experiments aimed at verifying the validity of the theoretical results derived throughout this paper. The proposed watermarking model was incorporated in a simple audio watermarking scheme. A music mono audio signal, of approximately 5.94 sec duration, sampled at 44.1 KHz with 16 bits per sample ($N = 262144$) was fed to the system. The host signal (DFT coefficients) was assumed to comply with the signal model of (4). The parameter β for the specific audio signal was estimated equal to 0.95, using Mean Square Error minimization between the model autocorrelation function and the test signal autocorrelation function. The skew tent map parameter λ was chosen to be equal to 0.3, thus leading to watermarks of highpass spectral characteristics. All sets of experiments were conducted using a total number of 10000 keys. The watermarks were embedded in the low frequency subband defined by $a = 0.01,\ b = 0.11$ and an embedding factor $p = 0.27$ was used, which produced a watermarked signal with SNR=23.1 db. ROC curve evaluation for this specific experiment, was performed for the event H_{1a}, corresponding to the signal not being watermarked ($p = 0$). Figure 1 demonstrates the close resemblance of the theoretical ROC curve evaluated using (12),(13) with the one evaluated experimentally. The small divergence that can be observed, can be attributed to the fact that the real audio signal is discrete-valued (quantized), as well as to the assumption made for its autocorrelation function (4). Multiple experiments involving different audio signals also verified the validity of the theoretical results.

A large number of experiments were performed to investigate the robustness of the proposed watermarking scheme to lowpass attacks and compare its performance with the performance of two alternative techniques: a) a scheme involving white pseudorandom watermark sequences ($w(i) \in \{-1, 1\}$) multiplicatively embedded in the same low frequency subband ($a = 0.01,\ b = 0.11$) of the DFT domain, producing watermarked signals with $SNR = 23$ db; b) a scheme based on the time-domain audio watermarking technique presented in [10], which involves watermark modulation according to the amplitudes of the original audio samples and filtering with a lowpass Hamming filter of 25th-order with cut-off frequency of 2205 Hz, in order to improve imperceptibility and robustness to lowpass attacks. The SNR value of the watermarked

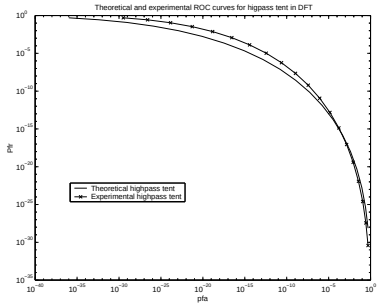

Figure 1: Theoretical and experimental ROC curves for embedding of highpass tent watermarks in the DFT domain.

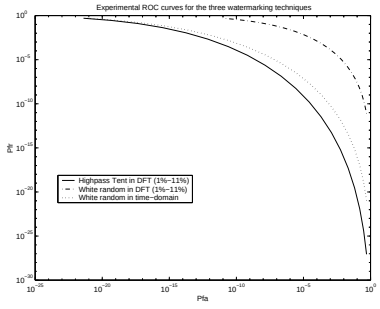

Figure 2: ROC curves for the three watermarking schemes without inflicted attack.

signal in this case was equal to 22 db. The parameters used in all three schemes were chosen so that all watermarked signals were just below the audibility threshold in order to ensure a fair comparison. The ROC curve evaluation has been performed under the worst case assumption (the signal was watermarked with a different watermark, event H_{1b}). Figure 2 indicates the superiority of our technique against both schemes described above, when no distortions are inflicted. White pseudorandom watermarks embedded in the DFT domain exhibited the worst performance. Figure 3 illustrates the superior performance of highpass tent watermarks embedded over the low DFT frequencies, against the alternative techniques described above in the case of MPEG-I layer III encoding at 64 kbps. Similar results were obtained for other signal distortions (mean/median lowpass filtering, resampling, requantization and cropping), but are not presented here due to lack of space.

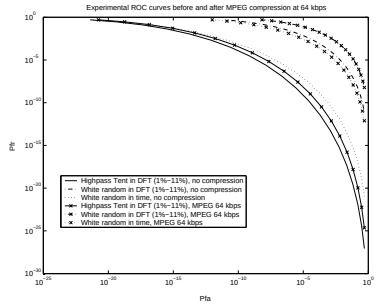

Figure 3: ROC curves for the three watermarking schemes after MPEG compression at 64 kbps.

5. CONCLUSIONS

The use of high-frequency spectrum watermarks generated by chaotic sequences and embedded in the low-frequency subband of the DFT coefficients is proposed in this paper. The proposed technique guarantees improved system detection reliability, imperceptibility and robustness to lowpass attacks. Theoretical analysis of the correlation detector is being performed and closed-form expressions are derived for the correlation detector statistics of skew tent chaotic watermarks. Various sets of experiments, on a simple audio watermarking application, verify the validity of the theoretical results and demonstrate the robustness of the proposed scheme to common lowpass attacks, such as MPEG compression and mean or median filtering.

6. REFERENCES

[1] J.R. Hernandez and F. Perez-Gonzalez, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1142–1166, July 1999.

[2] J.-P. Linnartz, T. Kalker, and G. Depovere, "Modeling the false alarm and missed detection rate for electronic watermarks," in *Proc. of 2nd Information Hiding Workshop*, Oregon, USA, April 1998, pp. 329–343.

[3] A. Tefas, A. Nikolaidis, N. Nikolaidis, V. Solachidis, S. Tsekeridou, and I. Pitas, "Statistical analysis of markov chaotic sequences for watermarking applications," in *2001 IEEE International Symposium on Circuits and Systems (ISCAS 2001)*, Sydney, Australia, May 6-9 2001.

[4] S. Tsekeridou, V. Solachidis, N. Nikolaidis, A. Nikolaidis, and I. Pitas, "Statistical analysis of a watermarking system based on bernoulli chaotic sequences," *Elsevier Signal Processing, Sp. Issue on Information Theoretic Issues in Digital Watermarking*, vol. 81, no. 6, pp. 1273–1293, 2001.

[5] T. Furon, B. Macq, N. Hurley, and G. Silvestre, "Janis: just another n-order side-informed watermarking scheme," in *Proc. of ICIP 2002*, Rochester, New York, Sept. 22-25 2002, pp. 153–156.

[6] Patrick Billingsley, *Probability and Measure*, Wiley, 1995.

[7] G. Voyatzis and I. Pitas, "Chaotic watermarks for embedding in the spatial digital image domain," in *Proc. of ICIP'98*, Chicago, USA, 4-7 October 1998, vol. II, pp. 432–436.

[8] S.H. Isabelle and G.W. Wornell, "Statistical analysis and spectral estimation techniques for one-dimensional chaotic signals," *IEEE Trans. on Signal Processing*, vol. 45, no. 6, pp. 1495–1506, June 1997.

[9] T. Kohda, H Fujisaki, and S. Ideue, "On distributions of correlation values of spreading sequences based on markov information sources," in *Proc. of ISCAS'00*, Geneva, Switzerland, 28-31 May 2000, vol. V, pp. 225–228.

[10] N. Nikolaidis P. Bassia, I. Pitas, "Robust audio watermarking in the time domain," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 232–241, June 2001.