EUSIPCO 2000
Tampere, Finland,
Sep. 5–8, 2000

PUBLIC KEY WATERMARKING BY EIGENVECTORS OF LINEAR TRANSFORMS

Joachim J. Eggers, Jonathan K. Su
Telecommunications Laboratory
University of Erlangen-Nuremberg
Cauerstrasse 7/NT, 91058 Erlangen, Germany
{eggers,su}@LNT.de

Bernd Girod
Information Systems Laboratory
Stanford University
Stanford, CA 94305-9510, USA
girod@ee.stanford.edu

ABSTRACT

Digital watermarks are signals embedded in multimedia data to allow copyright enforcement. In most watermarking schemes the embedded signal must be known for watermark detection, which leads to severe security risks. Van Schyndel et al. proposed a public watermark detection principle that works without explicit reference to the embedded signal. In this paper, extensions of this scheme are considered, and the applicability in practice is investigated. The new approaches are significantly less complex than the previously proposed scheme. Further, they are more robust against attacks via exhaustive search for the embedded watermark and against attacks that intend to confuse the public watermark detector. However, one drawback of all discussed schemes is the large signal length that is necessary for robust detection.

1 Introduction

Unauthorized copying and distribution of digital data is a severe problem for intellectual property rights. The embedding of digital watermarks into multimedia content was proposed to tackle this problem, and many different schemes have been presented in the last years. However, almost all of them are symmetric, meaning the key used for the watermark embedding must be available at the watermark detector and gives enough information to enable removal of the watermark to be detected. This leads to a security problem if the detectors are implemented in consumer devices that are spread all over the world. Therefore, the development of an asymmetric scheme becomes important. In such a scheme the detector only needs to know a public key, which does not give enough information to destroy the embedded watermark.

Fig. 1 depicts the scenario investigated in this paper. With aid of a private and a public key, the watermark is embedded into the host signal $\vec{x}$. The public signal $\vec{s} = \vec{x} + \vec{w}$ is subject to attacks on the watermark signal $\vec{w}$, and the attacker has access to the public key. Finally, the existence of the watermark should be detectable from the received signal $\vec{r} = \vec{s} + \vec{v}$, where $\vec{v}$ denotes the modifications introduced by the attacks. We consider only the basic modulation scheme, thus the detector simply indicates whether a watermark is found ("1") or not ("0").

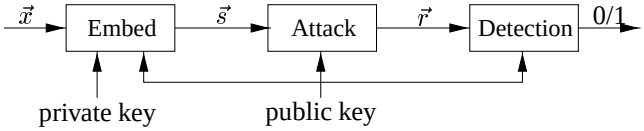

Figure 1: General asymmetric watermarking scheme.

In this paper we discuss a public key watermark detection scheme which is a modification of a proposal by van Schyndel et al. [3]. We briefly review the scheme proposed in [3] and then discuss a modification called *eigenvector watermarking*. A rough estimation for the necessary signal length enabling robust public detection is given, which is valid for van Schyndel's and our new scheme. Further, we discuss the security and robustness in the presence of malicious attacks. Finally, two specific realizations using the discrete Fourier transform (DFT) or permutation matrices are investigated and compared.

2 Van Schyndel's Key Independent Watermark Detection

Van Schyndel et al. [3] proposed to embed a (generalized) Legendre sequence [2, 3] as watermark into the host signal $\vec{x}$. Length-N Legendre sequences $\vec{a}$ have an optimal periodic autocorrelation function $\varphi_{aa}[n]$, meaning

$$\varphi_{aa}[n] = \begin{cases} N-1 & n = mN,\ m \in \mathbb{Z}; \\ 0 & \text{else.} \end{cases} \tag{1}$$

Further, Legendre sequences have a simple relationship to their DFT which we call "Fourier invariance", namely

$$\mathbf{G}_{\mathcal{DFT}}\vec{a} = \vec{A} = A_1\ \vec{a}^*, \tag{2}$$

where the scalar A_1 can be complex and $\vec{a}^*$ denotes the conjugate Legendre sequence. $\mathbf{G}_{\mathcal{DFT}}$ describes the DFT in matrix notation, where the unitary definition of the DFT is used here. Large letters, e.g. $\vec{A}$, denote frequency domain values. Van Schyndel et al. proposed to add the Legendre sequence $\vec{a}$ to the host signal $\vec{x}$ to get the public signal $\vec{s} = \vec{x} + \vec{a}$. The existence of the Legendre watermark can be detected by correlating the public signal $\vec{s}$ with its conjugate Fourier transform $(\mathbf{G}_{\mathcal{DFT}}\vec{s})^* = \vec{S}^*$. It is assumed that the average host signal $\vec{x}$ does not have the property of Fourier invariance.

The embedded Legendre sequence $\vec{a}$ need not be known explicitly for this detection process. It is sufficient to know the sequence length N that determines $\mathbf{G}_{\mathcal{DFT}}$ uniquely.

One shortcoming is that only $N-2$ different, non-degenerate Legendre sequences of length N exist. Therefore, an attacker might be able to determine the embedded Legendre sequence by exhaustive search. Another disadvantage is that Legendre sequences exist only for prime length N. Nevertheless, the basic idea indicates how a public key watermarking scheme might work.

3 Eigenvector Watermarking

The key idea of the Legendre watermaking scheme is that the DFT maps a Legendre sequences back to itself, except for conjugation and a scale factor. This idea can be extended to other pairs of a transform and a corresponding sequence that have similar properties. Since the watermark is added to the host signal, the transform should be linear.

3.1 Basic Concept

The *eigenvector watermarking* scheme is based on an $N \times N$ transform matrix $\mathbf{G}$ and a watermark vector $\vec{w}$ with the property

$$\mathbf{G}\vec{w} = \lambda_0 \vec{w}, \tag{3}$$

thus $\vec{w}$ is an eigenvector of $\mathbf{G}$.

The watermark $\vec{w}$ is embedded into the host signal $\vec{x}$ by addition, so the public signal is $\vec{s} = \vec{x} + \vec{w}$. The power of the watermark sequence $\vec{w}$ must be so small that $\vec{x}$ and $\vec{s}$ are perceptually equivalent. We characterize the embedding strength by the ratio $\gamma = ||\vec{x}|| \,/\, ||\vec{w}||$, where $||\cdot||$ is the Euclidean norm. Equivalently, we use the document-to-watermark power ratio $\mathrm{DWR} = 20 \log_{10} \gamma$.

The described embedding scheme is analogous to that for common spread spectrum watermarking. However, (3) must hold to enable watermark detection without explicit knowledge of $\vec{w}$. With help of $\mathbf{G}$, the embedded watermark $\vec{w}$ can be detected by measuring the correlation between the received signal $\vec{r}$ and its transformed signal $\mathbf{G}\vec{r}$. Thus, public detection (indicated by the subscript p) is based on

$$c_p = \frac{\vec{r}^H \mathbf{G} \vec{r}}{N}, \tag{4}$$

where $\vec{r}^H$ is the conjugate transpose of $\vec{r}$. Fig. 2 depicts this public detector.

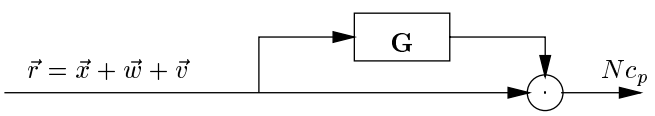

Figure 2: Public detection principle using the transform $\mathbf{G}$. The operator "$\cdot$" denotes the dot product of two vectors.

Without interfering host signal and without any attack, meaning $\vec{r} = \vec{s}$ and $\gamma = 0$, the value $c_{0p} = \vec{w}^H \mathbf{G} \vec{w}/N = \lambda_0 ||\vec{w}||^2 / N$ is returned. In practice, when $\gamma \gg 0$ and $\vec{v} \neq 0$, the detection process can be formulated as hypothesis test, where a measured value c_p being close to c_{0p} indicates that the hypothesis "the watermark is embedded" is true. $\mathbf{G}$ and $\vec{w}$ must be chosen such that the probability of a false decision is minimized. Assumptions on the statistics of the host signal, the watermark, and the attack are necessary to formulate the hypothesis test more precisely. In this paper we focus on the general applicability of the proposed scheme and thus, do not give a more precise description.

The choice of an appropriate matrix $\mathbf{G}$ is crucial when designing an eigenvector watermarking scheme. Two important goals are robust detection even for large values of γ and robustness against malicious attacks. Further, efficient computation of (4) and compact representation of $\mathbf{G}$ are desirable in practical systems. These restrictions will be discussed further in the following subsections.

3.2 Interference from the Host Signal $\vec{x}$

First of all, we analyze the influence of the host signal on the detection performance. We assume that no attack occurred, so that $\vec{r} = \vec{s}$. Expanding (4) yields

$$\begin{aligned} c_p &= \frac{1}{N} \vec{s}^H \mathbf{G} \vec{s} \\ &= \frac{1}{N} \left(\vec{x}^H \mathbf{G} \vec{x} + \vec{w}^H \mathbf{G} \vec{x} \right) + \lambda_0 c_b. \end{aligned} \tag{5}$$

The term $c_b = (\vec{x}^H \vec{w} + ||\vec{w}||^2)/N$ (subscript b for blind) equals the detection value for "common" blind spread spectrum watermark detection, where the watermark $\vec{w}$ is correlated with the public signal $\vec{s}$. It is difficult to analyze the influence of $\vec{x}$ on detection robustness for all possible $\vec{x}$, $\vec{w}$ and $\mathbf{G}$. Therefore, the interference from $\vec{x}$ is only roughly investigated in this article. We assume that $\vec{x}$ is a zero-mean, white, stationary signal. We mainly try to compare the detection robustness for the public scheme with that of common blind detection using a private key.

It is well known that the variance of c_b decreases for increasing N proportional to $\sigma_{\mathbf{x}}^2/N$, where $\sigma_{\mathbf{x}}^2$ denotes the variance of the host signal samples. The proposed public detection principle suffers from additional interference by the terms $\vec{w}^H \mathbf{G} \vec{x}/N$ and $\vec{x}^H \mathbf{G} \vec{x}/N$. The vector $\vec{w}^H \mathbf{G}$ is a deterministic signal that is independent from $\vec{x}$ and thus the variance of $\vec{w}^H \mathbf{G} \vec{x}/N$ will also decrease proportional to $\sigma_{\mathbf{x}}^2/N$. The influence of $\vec{x}^H \mathbf{G} \vec{x}/N$ is much more important. To achieve a working public watermarking scheme, the matrix $\mathbf{G}$ should be chosen such that $\mathrm{E}\left\{\vec{x}^H \mathbf{G} \vec{x}\right\}/N \approx 0$ and $\mathrm{Var}\left\{\vec{x}^H \mathbf{G} \vec{x}\right\}/N$ decreases for increasing signal length N. This can be achieved by a matrix $\mathbf{G}$ that produces elements of the vector $\mathbf{G}\vec{x}$ that are uncorrelated with the corresponding elements of $\vec{x}$. In this case the expectation $\mathrm{E}\left\{\vec{x}^H \mathbf{G} \vec{x}/N\right\}$ is zero and the variance is $\mathrm{Var}\left\{\vec{x}^H \mathbf{G} \vec{x}/N\right\} = \sigma_{\mathbf{x}}^4/N$. The variance is proportional to $\sigma_{\mathbf{x}}^4/N$ even for more general matrices $\mathbf{G}$.

The detection robustness can be measured by the ratio $\mathrm{E}\{c|\vec{r}\text{ is watermarked}\}/\mathrm{STD}\{c\}$, where larger values indicate higher robustness ($\mathrm{STD}\{\cdot\}$ is the standard deviation). For public and blind detection we expect

$\mathrm{E}\{c|\vec{r}\text{ is watermarked}\} = ||\vec{w}||^2/N = \sigma_w^2$. Thus the robustness of blind detection is given by $\sqrt{N}/\gamma$. The discussion above shows that the robustness of the proposed public detection scheme is proportional to $\sqrt{N}/\gamma^2$. Therefore, the signal length in the public scheme has to be increased by a factor of γ^2 to achieve approximately the same detection performance as blind detection. This is a very demanding requirement, since $\gamma \gg 1$ in practical applications.

Fig. 3 shows some experimental results where real-valued Legendre sequences are embedded as watermarks. Following van Schyndel's proposal, the DFT matrix $\mathbf{G}_{\mathcal{DFT}}$ is used for the public detection. The example is for DWR $= 20$ dB, or equivalently $\gamma = 10$. We tried to detect the watermark from 20 random received signals $\vec{r}$, where only the first one contains the watermark. The measured detection values c_b (common blind detection) and c_p (public detection) show that a signal length of about 10^6 samples is necessary to make public detection as reliable as blind detection for a signal length of about 10^4 samples.

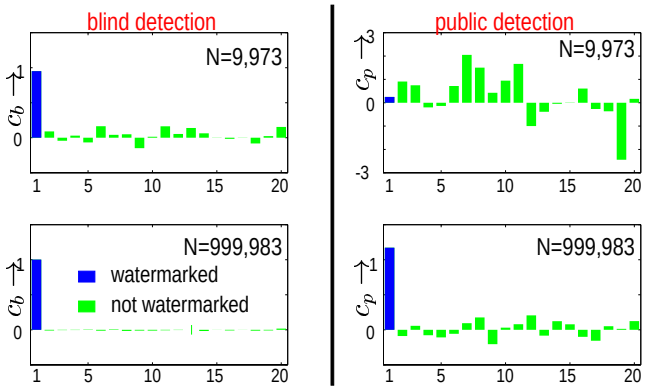

Figure 3: Detecting Legendre sequences of length N=9,973 and N=999,983 for 20 different received signals. Left side: Blind detection using the watermark sequence. Right side: Public detection based on van Schyndel's scheme.

3.3 Security against Exhaustive Search for $\vec{w}$

The matrix $\mathbf{G}$ and the desired detection value c_{0p} must be public. Thus, an attacker might try to search for $\vec{w}$ by exploiting the knowledge that $\vec{w}$ has to fulfill (3). One promising attempt is to compute the eigenvalues λ_i of $\mathbf{G}$ and search for corresponding eigenvectors $\vec{e}_i$ with $\lambda_i ||\vec{e}_i||^2 = c_{0p}$. Suppose the geometric multiplicity M_g of the eigenvalue λ_0 is equal to one. In this case the corresponding eigenvetor $\vec{e}_0 = \vec{w}$ is uniquely defined and can be computed easily. Then, the attacker computes the normalized correlation $\alpha = (\vec{s}^H\vec{e}_0)/||\vec{e}_0||^2$ and removes the watermark by $\vec{s} - \alpha\vec{e}_0$. To avoid such an attack, the eigenvalue λ_0 belonging to the eigenvector $\vec{w}$ should have a geometric multiplicity $M_g = N - \mathrm{rank}(\mathbf{G} - \lambda_0\mathbf{I}) \gg 1$. In this case, the eigenvectors for λ_0 are no longer uniquely defined and the attacker has to search for the embedded sequence in a sub-space of M_g dimensions. The complexity of such a search increases exponentially with the number of dimensions, and thus an attack via exhaustive search becomes unfeasible. We conclude that a large geometric multiplicity M_g of the eigenvalue λ_0 enhances the security of the proposed eigenvector watermarking scheme.

3.4 Confusing the Watermark Detector

The eigenvector watermark must be detectable using (4). When an exhaustive search for the embedded watermark $\vec{w}$ is too complex, an attacker can try to confuse the public detector by adding a properly scaled sequence $\vec{z}$ which satisfies

$$\mathbf{G}\vec{z} = -\beta\lambda_0\vec{z} \quad \text{with} \quad \beta > 0, \tag{6}$$

where $-\beta\lambda_0$ is an eigenvalue of $\mathbf{G}$ different from the eigenvalue λ_0 of the embedded watermark. Thus, $\vec{z}$ will be orthogonal to $\vec{w}$. We assume that $\vec{z}$ is normalized such that $||\vec{z}||=||\vec{w}||$. The influence of this attack can be described in terms of the parameter β. The public detector is confused by the added sequence $\vec{z}/\sqrt{\beta}$, since the detector measures for the composite signal $\vec{w} + \vec{z}/\sqrt{\beta}$ the value

$$\left(\vec{w} + \frac{\vec{z}}{\sqrt{\beta}}\right)^H \mathbf{G}\left(\vec{w} + \frac{\vec{z}}{\sqrt{\beta}}\right) = \vec{w}^H\mathbf{G}\vec{w} + \frac{\vec{z}^H\mathbf{G}\vec{z}}{\beta} = 0. \tag{7}$$

In general, the quality of the attacked signal will be reduced by the added attack sequence $\vec{z}$. We assume that the signal quality can be measured by the mean squared error of $\vec{y} = \vec{s} + \vec{z}/\beta = \vec{x} + \vec{w} + \vec{z}/\beta$ relative to the host signal $\vec{x}$. If we relate this to the embedding distortion, we find the distortion penalty

$$\frac{D_y}{D_s} = \frac{\left|\left|\vec{w} + \vec{z}/\sqrt{\beta}\right|\right|^2}{||\vec{w}||^2} = 1 + \frac{1}{\beta} \tag{8}$$

for a successful confusion attack. We exploited that $\vec{w}$ is orthogonal to $\vec{z}$ and that both vectors have the same norm. Whether or not an attack sequence $\vec{z}$ with the property (6) exists and which value of β results depends strongly on the set of all eigenvalues of $\mathbf{G}$.

4 Public Detection Using the DFT

In this section, we consider an eigenvector watermarking scheme where the detection matrix is equal to the DFT matrix ($\mathbf{G}$=$\mathbf{G}_{\mathcal{DFT}}$). The advantages of using $\mathbf{G}_{\mathcal{DFT}}$ are that almost no overhead for transmitting the detection matrix is necessary (only the transformation length must be transmitted), and that fast algorithms for computing the transform are known (nevertheless, the computational complexity for long sequences is still very demanding). For real signals, eigenvector watermarking using $\mathbf{G}_{\mathcal{DFT}}$ is almost the same as van Schyndel's approach using Legendre sequences. Only the conjugation involved in the detection process for Legendre watermarks is missing in the eigenvector watermarking scheme.

The eigenvalues and eigenvectors of $\mathbf{G}_{\mathcal{DFT}}$ have been analyzed in [1]. For $N > 4$, $\mathbf{G}_{\mathcal{DFT}}$ has the eigenvalues $\lambda \in \{\pm 1, \pm j\}$ with different multiplicity. It is not difficult

to construct eigenvectors for all eigenvalues, and thus we can easily construct a "random" watermark sequence for all values of N.

Note that the constraint on the required signal length (Section 3.2) is valid for the eigenvector and Legendre sequence watermarking scheme. We assume in the following discussion that this constraint can be met in some watermarking applications. The Legendre watermarking scheme suffers from the small number of Legendre sequences. This weakness does not exist for the eigenvector watermarking scheme using $\mathbf{G}_{\mathcal{DFT}}$. For every sequence length N the subspace of eigenvectors for one certain eigenvalue has about $N/4$ dimensions. Thus, the scheme works for all sufficiently large signal lengths N and is secure against exhaustive search attacks. However, the confusion attack will always work, since the negative of each eigenvalue of $\mathbf{G}_{\mathcal{DFT}}$ is again an eigenvalue. Therefore, the attacker can always find a sequence $\vec{z}$, where $\beta = 1$ leads to a perfect confusion attack. Thus the distortion penalty for the eigenvector watermarking scheme using the DFT is $D_y/D_s = 2 \equiv 3$ dB. Note that this distortion penalty is only an upper bound due to the additional interference from the host signal. It can be shown that a confusion attack also works for the Legendre watermarking scheme, where the same distortion penalty is valid.

5 Public Detection Using Permutation Matrices

A very attractive type of transformation matrices $\mathbf{G}$ are permutation matrices $\mathbf{G}_{\text{perm}}$. First of all, these transformations allow for a fast detection algorithm since only re-indexing operations are involved in the transformation step. Further, for a signal length N, $\mathbf{G}_{\text{perm}}$ can be described uniquely by at most $N-1$ integer values. Using some sophisticated algorithm for the design of permuation matrices, $\mathbf{G}_{\text{perm}}$ can be described with even fewer values.

Requirements for a useful permutation matrix $\mathbf{G}_{\text{perm}}$ in the context of the eigenvector watermarking scheme can be derived from the corresponding eigenvalues. The eigenvalues of $\mathbf{G}_{\text{perm}}$ can be determined by the cycles of $\mathbf{G}_{\text{perm}}$. Here, a cycle of a permutation matrix is a sub-matrix that maps a set of sample positions uniquely onto itself. A fixed point of a permutation matrix is equal to a cycle of length 1. The matrix $[0\ 1; 1\ 0]$ has a cycle of length 2 and $[0\ 0\ 1; 1\ 0\ 0; 0\ 1\ 0]$ has a cycle of length 3. However, $[0\ 0\ 1; 0\ 1\ 0; 1\ 0\ 0]$ has two cycles, one of length 2 and another one of length 1. Note that a permutation matrix may have P different cycles of length R_p, with $\sum_{p=1}^{P} R_p = N$. The set of eigenvalues of a permutation matrix is equal to the set of the Pth complex roots of 1. Thus, a cycle of length 1 has the eigenvalue 1, a cycle of length 2 has the eigenvalues ± 1 and a cycle of length 3 has the eigenvalues $\{1, -0.5 \pm j0.5\sqrt{3}\}$. Since every cycle has at least the eigenvalue 1, we conclude that the multiplicity of the eigenvalue 1 is equal to the number of cycles in the permutation matrix. The multiplicity of all other eigenvalues of a permutation matrix will always be smaller or equal to the number of cycles. Thus, an eigenvector belonging to the eigenvalue 1 should be chosen as watermark, and the number of cycles in $\mathbf{G}_{\text{perm}}$ should be large to resist attacks via exhaustive search for the embedded watermark.

Knowing $\mathbf{G}_{\text{perm}}$, the attacker would like to apply the confusion attack with the attack sequence $\vec{z}$ being an eigenvector to the eigenvalue -1, meaning $\beta = 1$. From the discussion above, it is clear that such an eigenvalue always exists if $\mathbf{G}_{\text{perm}}$ has cycles of even length. To increase the distortion penalty for the confusion attack, one might choose $\mathbf{G}_{\text{perm}}$ such that only cycles of length 3 are present. In this case, the attacker has to construct $\vec{z}$ by the sum of an eigenvector for $-0.5 + 0.5\sqrt{3}$ and an eigenvector for $-0.5 - 0.5\sqrt{3}$. Public detection is impossible if the resulting vector is embedded with $\beta = 0.5$. Thus, a distortion penalty of $D_y/D_s = 3 \equiv 4.771$ dB for the confusion attack is achieved for this scheme. Whether this distortion penalty is large enough in practical applications is not known yet. However, this distortion penalty is the largest value of all public watermarking schemes that we have analyzed so far.

6 Conclusion

A public key watermarking scheme based on Legendre sequences was reviewed. The basic idea of this scheme was extended to general linear transforms, where the eigenvectors of these transforms are embedded as watermarks. We showed that for both schemes very long signals are needed. Using the DFT for the eigenvector watermarking scheme, we could achieve the same robustness as for the scheme based on Legendre watermarks. However, higher security against exhaustive search attacks and more flexibility in the signal length could be achieved. Even more promising is the usage of permutation matrices having only cycles of length 3. In this case, the complexity of the detector is significantly decreased, and high security against exhaustive search attacks is achieved. Unfortunately, the distortion penalty for a confusion attack is modest (4.77 dB in our best scheme) thus being the Achilles' Heel of such a public key watermarking scheme for most applications.

References

[1] J. H. McClellan and T. W. Parks. Eigenvalue and Eigenvector Decomposition of the Discrete Fourier Transform. *IEEE Transaction on Audio and Electroacoustics*, 20(1):66–74, March 1972.

[2] M. R. Schroeder. *Number Theory in Science and Communication*. Springer Verlag, New York, Heidelberg, Berlin, Tokyo, 2rd edition, 1986.

[3] R. G. van Schyndel, A. Z. Tirkel, and I. D. Svalbe. Key independent watermark detection. In *Proceedings of the IEEE Intl. Conference on Multimedia Computing and Systems*, volume 1, Florence, Italy, June 1999.