Sicherheit in
Mediendaten
Berlin, Germany,
Sep. 19, 2000

Asymmetric Watermarking Schemes

Joachim J. Eggers[1], Jonathan K. Su[1], and Bernd Girod[2]

[1] Telecommunications Laboratory, University of Erlangen-Nuremberg
Cauerstr. 7/NT, 91058 Erlangen, Germany
{eggers,su}@LNT.de
[2] Information Systems Laboratory, Stanford University
Stanford, CA 94305-9510, USA
girod@ee.stanford.edu

Abstract. Unauthorized copying and distribution of digital data is a severe problem in protecting intellectual property rights. The embedding of digital watermarks into multimedia content has been 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 watermark embedding must be available at the watermark detector. 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 asymmetric schemes becomes important. In such a scheme the detector only needs to know a public key, which does not give enough information to make watermark removal possible. In this paper, we review recent proposals for asymmetric watermarking and analyze their performance.

1 Introduction

The digital representation of audio signals, images, and video has become popular due to the ease of transmitting digital data and copying without loss of quality. However, the problem arises that unauthorized copying and distribution of digital data is simplified, too. For this reason, researcher have started looking for techniques that allow copy control of digital multimedia data and enable copyright enforcement. It was realized that common cryptographic means are not sufficient since the data is without any protection as soon it is used, e.g., decrypted and displayed in the case of image or video data. A potential aid in solving this problem is digital watermarking. Digital watermarking is the imperceptible embedding of information into multimedia data, where the information remains detectable as long the quality of the content itself is not rendered useless. It is commonly assumed that digital watermarking is only one of several measures that have to be combined to build a good copy protection mechanism [8].

One particular problem with state-of-the-art watermarking schemes is that they are symmetric. The keys necessary for watermark embedding and detection are identical. Thus, the watermark detector knows all critical parameter of the watermarking scheme that also allow efficient removal of the embedded watermark. We will discuss such methods in more detail in Section 3. Using watermark technology for copy protection, the watermark detector needs to be implemented in many cheap consumer devices all over the world. A symmetric watermarking scheme presents a security risk, since the detector has to know the required private key. However, cheap tamper-proof devices are

hardly produceable [8], and thus, pirates can obtain the private key from such devices and use them to outwit the copy protection mechanism. For this reason, we would like to develop a watermarking scheme where detection of the watermark is possible with a public key that does not give enough information to impair the embedded watermark. Such a scheme is called asymmetric. The intention of this paper is to give an overview of proposals for such a mechanism and discuss their pros and cons.

First, we will explain our notation and describe a general point of view on watermarking schemes in Section 2. For a better understanding of the differences between symmetric and asymmetric schemes, both methods are described. In Section 3, we will discuss two symmetric watermarking schemes and possible attacks if the private key can be accessed by an attacker. Several proposals of asymmetric watermarking schemes are discussed in Section 4. Finally, an assessment of the state-of-the-art is given, and future research directions are proposed.

2 Digital Watermarking: A Communications Problem

We view digital watermarking as a communications problem, where the watermark information $b \in \mathcal{B}$, with $\mathcal{B}$ denoting the finite set of all possible watermark messages, is transmitted over an hostile channel. The host signal $\underline{x}$ serves as the carrier for the watermark information. In this paper, we adopt vector notation for signals, that is $\underline{x} = [x[0], x[1], \ldots, x[N-1]]^T$ with $x[i]$ being the ith signal sample. We do not focus on a specific data type. $\underline{x}$ can denote audio, image or video data, or any transform domain representation of such multimedia data. In practice, watermarking schemes have to be optimized for the specific features of different host signals. Here, our intention is to compare basic concepts without considering details that are strongly dependent on the specific multimedia data.

Any modification of the host signal $\underline{x}$ does affect its quality, thus an assessment of watermarking schemes is not possible without defining a quality measurement. Good quality measurements are again strongly dependent on the data at hand. However, as a rough approximation, the *mean squared error* (MSE) between the original host signal and any modified signal can be used as a quality measurement.

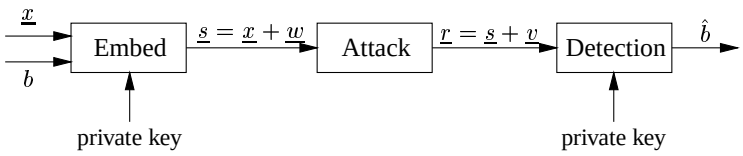

Fig. 1. General blind symmetric watermarking scheme.

Fig. 1 depicts a general blind symmetric watermarking scheme. The term "blind" indicates that the host signal $\underline{x}$ is not known at the watermark detector. The watermark information b is embedded into the host signal $\underline{x}$ dependent on a private key. All modifications introduced by the embedding process are denoted by the watermark signal $\underline{w}$, so that the public signal $\underline{s}$ can be expressed as $\underline{s} = \underline{x} + \underline{w}$. The distortion introduced

by the embedding of the watermark is given by $D_E = \mathrm{E}\left\{(s-x)^2\right\} = \mathrm{E}\left\{w^2\right\}$. Here, $\mathrm{E}\{\cdot\}$ denotes expectation.

The public signal $\underline{s}$ is subject to a variety of different *attacks*. We use the term *attack* for any signal processing that, intentionally or not, reduces the reliability of watermark detection. The modifications introduced by the attack(s) can be summarized by the additive, but not necessarily independent, signal $\underline{v}$. Of course, an attack is useless if the attacked signal $\underline{r} = \underline{s} + \underline{v}$ has such poor quality that its value is lost. Thus, the quality of the attacked signal must be sufficiently good. Many watermarking schemes can be successfully attacked by desynchronizing the embedded watermark relative to the watermark signal the detector is looking for. We do not consider desynchronization attacks formally, but point out where synchronization is a particularly difficult problem. Assuming synchronization, the quality of the attacked signal $\underline{r}$ is measured relative to the original host signal $\underline{x}$. We measure the distortion of an attacked signal by $D_A = \mathrm{E}\left\{(r-x)^2\right\}$.

Finally, the detector computes an estimate $\hat{b}$ of the transmitted watermark information b, depending on the private key and the received signal $\underline{r}$. The probability $\Pr(\hat{b} \neq b)$ of false detection should be as small as possible.

The constraints on the qualities D_E and D_A are strongly dependent on the given data and the application in mind. However, it is reasonable to assume that the allowable D_A is at least at the order of D_E, and in many cases even much larger. We use the ratio $D_{A,\mathtt{min}}/D_E$ as a robustness criteria, with $D_{A,\mathtt{min}}$ being the minimal distortion for a successful attack. Chen and Wornell [2] introduced the term "distortion penalty" for $D_{A,\mathtt{min}}/D_E$.

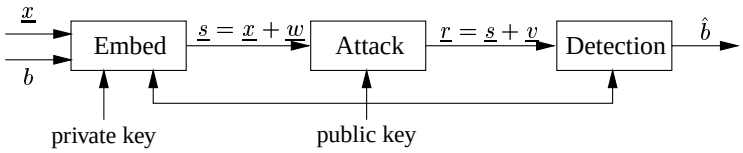

Fig. 2. General asymmetric watermarking scheme.

Fig. 2 depicts a general asymmetric watermarking scheme. With aid of a private and a public key, the watermark is embedded into the host signal $\underline{x}$. The significant difference to the symmetric scheme depicted in Fig. 1 is that all entities, embedding, attack and detection, have access to the public key necessary for watermark detection. Obviously, an attacker can try to use the knowledge of the public key to destroy the embedded watermark information.

3 Symmetric Watermarking and Public Detection

In this section, two techniques for blind symmetric watermarking will be briefly reviewed and their security risks in combination with public detectors will be discussed.

3.1 Spread-Spectrum Watermarking

Spread-spectrum watermarking is one of the first methods used for blind symmetric watermarking (e.g., [3, 11]) and is still the most popular one. Many modifications are possible, depending on the characteristics of the host signal and the application in mind. Here, it is sufficient to focus on the basic approach, which can be described as follows:

1. A random signal $\underline{z}$ is defined. For instance, the samples $z[i]$ can be drawn independently and equiprobably from the binary alphabet $\{-1, +1\}$. $\underline{z}$ serves as the private key of the watermarking scheme. Watermark embedding is implemented by simple addition of the watermark signal $\underline{w} = b\alpha\underline{z}$. The scale factor α determines the power of the watermark signal and must be chosen such that the watermark is imperceptible, but sufficient reliably detectable. The factor $b \in \mathcal{B}$ depends on the watermark information to be transmitted. For instance, unipolar transmission is obtained for $\mathcal{B} = \{0, 1\}$, and for bipolar transmission $\mathcal{B} = \{-1, +1\}$.
2. For detection, the correlation between the received signal $\underline{r} = \underline{x} + \underline{w} + \underline{v}$ and the private-key signal $\underline{z}$ is measured. Spread-spectrum watermarking relies on the assumption that the key signal $\underline{z}$ is statistically independent from the host signal $\underline{x}$ and the distortion $\underline{v}$, which leads to $\mathrm{E}\{zr\} = \mathrm{E}\{z(x + v)\} + b\alpha\mathrm{E}\{z^2\} = 0 + b\alpha\mathrm{E}\{z^2\}$. For finite-length signals, the correlation can be measured only with a certain level of accuracy. The estimated watermark information $\hat{b}$ can be obtained from a hypothesis test on the measured correlation $c = (1/N)\sum_{i=0}^{N-1} z[i]r[i]$. Here, the power of the interfering signals $\underline{x}$ and $\underline{v}$ becomes important and the detection performance increases with the signal length N and embedding strength α. For $\mathcal{B} = \{-1, +1\}$, the estimated watermark information can be obtained easily from the sign of the measured correlation: $\hat{b} = \mathtt{sign}(c)$.

Without knowing $\underline{z}$, it is difficult to attack the embedded watermark. However, knowing the private key, the watermark can be removed easily. First, the watermark information b is detected and the watermark signal $\underline{w} = b\alpha\underline{z}$ is reconstructed. In the most simple case, when the attacker has the purely watermarked public signal $\underline{s}$ without any further distortion, he can subtract $\underline{w}$ to obtain the non-distorted host signal $\underline{x}$. Having an already distorted public signal $\underline{r} = \underline{s} + \underline{v}$, the attacker maximizes the signal quality and minimized the ability of watermark detection by cancelling the signal components that are correlated with $\underline{w}$. This can be achieved by $\underline{r} - (\underline{w}^T\underline{r}/\,||\underline{w}||^2)\underline{w}$. Thus, keeping the key $\underline{z}$ private is crucial to the security of spread-spectrum watermarking.

3.2 Quantization Index Modulation (QIM)

Chen and Wornell [1, 2] proposed a blind watermarking technique where the host signal $\underline{x}$ is quantized differently depending on the watermark information to be embedded. The general scheme is called *quantization index modulation* (QIM).

A quantizer can be uniquely described by a set of reconstruction points $\mathcal{Q}$ in an L-dimensional space and a rule for assigning a length-L input signal to one of the points defined in $\mathcal{Q}$. Here, we will always use the minimum-distance rule for selecting the appropriate points and characterize different quantizers solely by their reconstruction points $\mathcal{Q}$.

The basic principle of QIM can be described as follows:

1. A set of different quantizers $\{\mathcal{Q}_0, \mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_{B-1}\}$ is defined. The index set $\mathcal{B}$=$\{0, 1, 2, \ldots, B-1\}$ denotes the B considered watermark messages.
2. For embedding the watermark information $b \in \mathcal{B}$, the host signal $\underline{x}$ is quantized using the quantizer $\mathcal{Q}_b$ to obtain the public signal $\underline{s}$. Thus, the expected embedding distortion D_E is equal to the introduced quantization noise.
3. The watermark detector quantizes the received signal $\underline{r}$ by the union of all quantizers $\{\mathcal{Q}_0, \mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_{B-1}\}$. The detector determines the index of the quantizer containing the reconstruction point closest to the received signal. This index corresponds to the received watermark information $\hat{b}$.

QIM does not suffer from host signal interference like blind spread spectrum watermarking does. Thus, QIM offers high watermark rates when the distortion introduced by attacks is small. Since quantization always includes a loss of detail ("many-to-one mapping"), it is not possible to reconstruct the host signal $\underline{x}$ perfectly from a given public signal $\underline{s}$, even if the watermark information b and the watermarking scheme is completely known. Nevertheless, QIM must be considered a symmetric watermarking technique, since encoder and decoder have to have the same knowledge; in particular the involved quantizer sets must be known.

Chen and Wornell argued in [2] that QIM could be used as a public-key watermarking scheme, since successful removal of the watermark information is not possible without introducing additional distortion. Although this is true in theory, we have found that the distortion introduced by successful attacks is smaller than the embedding distortion and therefore might be acceptable in many circumstances. Here, we will discuss two example QIM schemes. First, binary dither modulation using a shifted uniform scalar quantizer [1, 2], and second, dithered hexagonal lattice quantization.

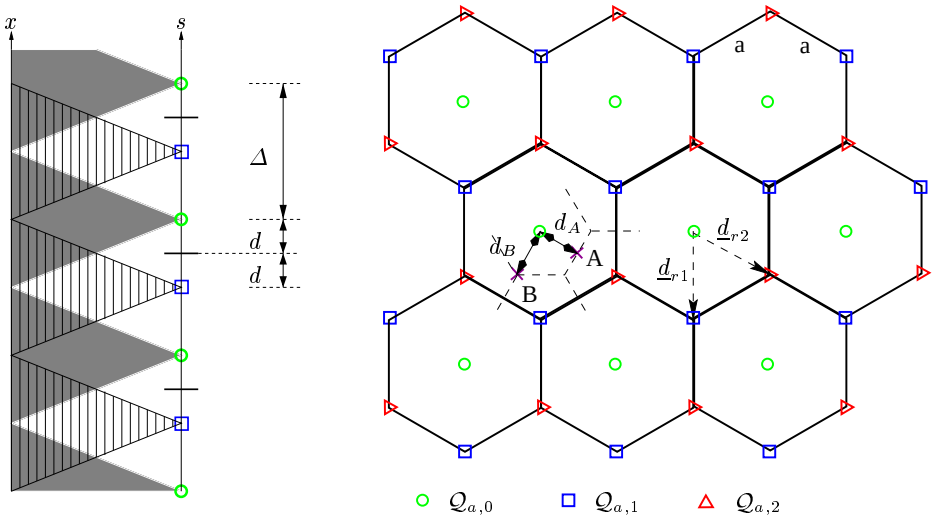

Fig. 3. Example QIM schemes based on dithered uniform scalar quantization (left) and dithered hexagonal lattice quantization (right)

For binary dither modulation, the mapping of the range of host signal values x onto a public signal value s is depicted in the left diagram in Fig. 3. The set $\mathcal{Q}_0$ (circles) is defined by an uniform scalar quantizer with step size Δ. The set $\mathcal{Q}_1$ (squares) is another uniform scalar quantizer, however, with an offset of $\Delta/2$. Switching between both quantizers can be considered dithered quantization. Since decoding the watermark is based on the closest distance rule, a detection error can occur after modifying the public signal sample s by at least $\pm d$.

If the detection process is publicly known, an attacker can perturb the public signal $\underline{s}$ in such a way that the attacked signal $\underline{r}$ exactly lies on the decision boundary between different quantizer points. For binary dither modulation these points are depicted by the short lines in Fig. 3 (left). After such an attack, the decoder can only randomly guess whether the received signal sample was originally quantized by $\mathcal{Q}_0$ or $\mathcal{Q}_1$. Thus, the watermark information is completely lost. Note that no channel coding can help to recover information from the such attacked signal.

Of course, Chen and Wornell's statement that QIM can never be attacked without introducing additional distortion still holds. However, the distortion penalty for the described attack and binary dither modulation is only $D_{A,\mathtt{min}}/D_E = 1.75 \equiv 2.43$ dB.

We also investigated a two-dimensional QIM scheme using hexagonal lattice quantization. Fig. 3 (right) depicts the considered quantizer $\mathcal{Q}_0$, $\mathcal{Q}_1$ and $\mathcal{Q}_2$. We depicted also the decision boundaries for $\mathcal{Q}_0$. They form a hexagonal lattice. The quantizer $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are obtained from $\mathcal{Q}_0$ by dithering with the dither vectors $\underline{d}_{r1}$ and $\underline{d}_{r2}$. Making the sets $\mathcal{Q}_0$, $\mathcal{Q}_1$ and $\mathcal{Q}_2$ public again enables the attack of putting the public signal samples $\underline{s}$ on the decision boundary between the three possible quantizers. Two special versions of this attack are considered. First, a point in the middle of two quantizers $\mathcal{Q}_i$ and $\mathcal{Q}_j$, with $i \neq j$, is chosen . We marked such a point in Fig. 3 (right) by "A". Another option is to choose a point in the middle of all three quantizers. An example point is marked in Fig. 3 (right) by "B". Attack A gives a smaller attack distortion since the distance $d_A < d_B$. On the other hand, some watermark information remains since the detector can exclude one of the three possible quantizers. The watermark information is completely erased in case of attack B. For the attack "A", a distortion penalty of $D_{A,\mathtt{min}}/D_E = 1.6 \equiv 2.04$ dB results, whereas for the attack "B" a distortion penalty of $D_{A,\mathtt{min}}/D_E = 1.8 \equiv 2.55$ dB is achieved. These values are very similar to the distortion penalty found for binary dither modulation.

We conclude that for the investigated QIM schemes, public-key watermarking is possible, but the distortion penalty for successfull attacks is too low for many practical applications.

4 Properties of Proposed Asymmetric Watermarking Schemes

In the last section it was illustrated that some symmetric watermarking schemes are no longer robust when the private key for watermark detection is made public. In this section we review proposals for asymmetric watermarking and discuss their robustness against attacks.

4.1 Spread Spectrum Watermarking with Partly Known Key

Hartung and Girod [10] discussed a public-key watermarking approach that is a simple modification of spread-spectrum watermarking. Recall from Section 3.1 that the private key $\underline{z}$ must be known for spread-spectrum watermark detection. However, if long watermark signals can be obtained, it is possible to detect a spread-spectrum watermark even if some samples of $\underline{z}$ are modified. Based on this idea, every recipient of the watermarked data gets a different "public key" $\underline{z}_k$, where only a subset of the samples of $\underline{z}_k$ match those in $\underline{z}$. The rest of $\underline{z}_k$ is chosen randomly. Using this approach, client k cannot modify the part of the watermark detected by $\underline{z}_j$ (with $j \neq k$) and client j cannot modify the part of the watermark detected by $\underline{z}_k$.

However, it was mentioned already in [10] that recipient k, knowing the public key sequence $\underline{z}_k$ can easily make detection based on this public key impossible. For this, the correlation between $\underline{z}_k$ and the received signal $\underline{r}$ must be removed, which can be achieved by $\underline{r} - (\underline{z}_k^T \underline{r} / ||\underline{z}_k||^2)\underline{z}_k$. Using this attack, the quality of the attacked signal might be even better than that of the watermarked signal.

It is obvious that such a public watermarking scheme cannot solve the copy-protection problem described in Section 1. Thus, the term "public watermark" is misleading. Instead, the scheme has applications in multiple watermarking of one document, where several spread-spectrum watermarks are combined in an elegant way.

4.2 Asymmetric Watermarking based on One-Way Signal Processing

Furon and Duhamel [7] concluded from comparisons to public-key cryptosystems that a one-way signal-processing function is needed to build an asymmetric watermarking scheme. They identified the power density spectrum (PDS) of a signal as a candidate of such an one-way function. The PDS of a signal describes the signal to some extent, but in general does not allow perfect reconstruction due to the loss of the signal phase. Furon and Duhamel implemented and tested their approach for image and audio signals [7–9]. We briefly describe the basic principle of the proposed scheme. After that an effective attack is discussed.

First, the host signal is randomly permuted. For brevity, we denote the permuted host signal by $\underline{x}$ and its power by P_x. The main purpose of the permutation is to break statistical dependencies between adjacent signal samples so that $\underline{x}$ has a flat PDS. As in spread-spectrum watermarking, an independent watermark signal $\underline{w}$ is added to the permuted host signal to obtain the permuted public signal $\underline{s} = \underline{x} + \underline{w}$. However, in this scheme, the watermark signal is colored noise, which can be obtained by filtering a white noise signal $\underline{z}$ of power P_z. Let $H(\Omega)$ denote the frequency response of the selected filter. The PDS of the watermark signal is given by $\Phi_{ww}(\Omega) = P_z \left|H(\Omega)\right|^2$. Since the watermark signal $\underline{w}$ is independent from the host signal $\underline{x}$, the PDS of the public signal $\underline{s}$ is straightforwardly derived as $\Phi_{ss}(\Omega) = P_x + P_z \left|H(\Omega)\right|^2$.

The public detection process is based on the specific shape of the PDS of the public signal $\underline{s}$. Furon and Duhamel describe an hypothesis test that, given a permuted received signal $\underline{r}$, allows one to decide whether its PDS is flat or resembles the shape of $\Phi_{ss}(\Omega)$. Obviously, permutation of the host signal is necessary, since the hypothesis test is designed on the assumption of a flat PDS of $\underline{x}$.

Note that the watermark sequence $\underline{w}$ need not to be known for watermark detection. The shape of $\Phi_{ss}(\Omega)$ is the public key that allows watermark detection. In general, it is impossible to find $\underline{w}$ from $\Phi_{ss}(\Omega)$. Thus, the described scheme was thought to be secure against malicious attacks.

However, there is a way of making public watermark detection impossible without knowing $\underline{w}$. One simply has to filter the permuted public signal $\underline{s}$ so that its PDS is whitened. In personal communications, Teddy Furon pointed out that such filtering has to be implemented carefully, since phase modifications of the public signal can have a perceivable effect on the signal quality. One way of implementing the attack is to compute the Fourier spectrum of the public signal, modify only its absolute values and use the inversely transformed data as the attacked public signal $\underline{r}$. This attack is successful without decreasing the signal quality since mainly watermark components are filtered out.

4.3 Legendre Watermarking

Van Schyndel et al. [12] proposed an asymmetric watermarking scheme based on a length-N Legendre sequence $\underline{a}$. Legendre sequences have a simple relationship to their DFT, namely $\mathbf{G}_{\mathcal{DFT}}\underline{a} = \underline{A} = A_1\ \underline{a}^\star$, where $\mathbf{G}_{\mathcal{DFT}}$ is the DFT matrix, the scalar A_1 can be complex, and $\underline{a}^\star$ denotes the conjugate Legendre sequence. Large letters, e.g. $\underline{A}$, denote frequency-domain values. The Fourier invariance of the Legendre sequence is a property that does not hold for general sequences. Therefore, van Schyndel et al. proposed to use the Legendre sequence as a watermark $\underline{w} = \underline{a}$, so that the public signal is $\underline{s} = \underline{x} + \underline{a}$. The watermark is detected in the received signal $\underline{r} = \underline{x} + \underline{v} + \underline{a}$ by correlating $\underline{r}$ with its conjugate Fourier transform $(\mathbf{G}_{\mathcal{DFT}}\underline{r})^\star = \underline{R}^\star = (\underline{X} + \underline{V})^\star + A_1^\star\underline{a}$. A large correlation value $c = \underline{r}^H\mathbf{G}_{\mathcal{DFT}}\underline{r}/N$ indicates the existence of the embedded Legendre watermark. Here, $\underline{r}^H$ denotes the conjugate transpose of $\underline{r}$. The detection works reliably if $(\underline{x} + \underline{v})^T(\underline{X} + \underline{V})^\star \approx 0$ is fulfilled.

The embedded Legendre sequence $\underline{a}$ need not to be known explicitly for the described detection process. Thus, the watermarking scheme can be used as a public-key scheme, where the embedded Legendre sequence $\underline{a}$ serves as the private key and the sequence length N is the public key. 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. There are also malicious attacks against the Legendre watermarking scheme [5]. Those will be discussed in the next subsection.

4.4 Eigenvector Watermarking

The key idea of the Legendre watermaking scheme is that the DFT maps a Legendre sequence back to itself, except for conjugation and a scale factor. We looked at modifications of this approach and proposed an asymmetric watermarking scheme using eigenvectors of linear transforms [6].

The eigenvector watermarking scheme is based on a $N \times N$ transform matrix $\mathbf{G}$ and a watermark vector $\underline{w}$ with the property $\mathbf{G}\underline{w} = \lambda_0\underline{w}$, thus $\underline{w}$ is an eigenvector of

$\mathbf{G}$, and λ_0 the corresponding eigenvalue. The watermark sequence $\underline{w}$ must be so small that $\underline{x}$ and $\underline{s}$ are perceptually equal.

For watermark detection, the correlation $c = \underline{r}^H \mathbf{G}\underline{r}/N$ between the received signal $\underline{r}$ and its transformed signal $\mathbf{G}\underline{r}$ is measured. A large correlation value c indicates that the received signal $\underline{r}$ contains an eigenvector of $\mathbf{G}$. The described watermarking scheme is asymmetric, since the embedded watermark signal $\underline{w}$ is not needed in the detection process. The matrix $\mathbf{G}$ serves as the public key.

An analysis of the properties of Legendre watermarking and eigenvector watermarking can be found in [6]. Due to space constraints, we can only summarize the major results:

- Legendre watermarking and eigenvector watermarking suffer significantly from host-signal interference. Compared to symmetric spread-spectrum watermarking, the watermark signal length N has to be increased dramatically, in particular when small watermark-to-document ratios (WDRs) are desired.
- Legendre watermarks are not secure against exhaustive search for the embedded sequence. Eigenvector watermarks can be much more secure, if the eigenvector belongs to an eigenvalue of $\mathbf{G}$ with a large geometric multiplicity. However, attacks like the sensitivity attack described by Cox and Linnartz in [4] might be successful if good objective quality measurements without reference to the original signal are known.
- Instead of removing an embedded watermark, an attacker can try to confuse the public watermark detector by adding another signal $\underline{z}$ with the property $\mathbf{G}\underline{z} = -\beta\lambda_0\underline{z}$, where $\beta > 0$. The additional distortion of a successful confusion attack depends on the eigenvalues of $\mathbf{G}$. The largest distortion penalty of $D_{A,\mathtt{min}}/D_E = 3 \equiv 4.771$ dB was found for $\mathbf{G}$ being a certain permutation matrix. The confusion attack can also be used with minor modification against the Legendre watermarking scheme. In this case the distortion penalty is $D_{A,\mathtt{min}}/D_E = 2 \equiv 3$ dB.
- Anybody can embed a watermark $\underline{w}$ that is publicly detectable by $\mathbf{G}$. Thus, eigenvector watermarking is only usefull for certain applications. One application might be copy control. A signal is not copied if it contains an eigenvector watermark. No pirate would intentionally embed such a watermark.

Teddy Furon has found an effective attack against the eigenvector watermarking scheme. The public signal $\underline{s}$ is projected onto the subspace defined by all eigenvectors of $\mathbf{G}$ belonging to the eigenvalue λ_0. This projection $\underline{p}$ is scaled by a factor ν and subtracted from $\underline{s}$ to obtain the attacked signal $\underline{r} = \underline{s} - \nu\underline{p}$. The factor ν can be found analytically or simply by trial-and-error until the public detector no longer works. The distortion introduced by this attack is acceptable. Depending on the WDR, it is even possible to obtain a signal $\underline{r}$ having a higher quality than $\underline{s}$.

5 Conclusions

We presented attacks for two symmetric watermarking schemes combined with public detectors. Spread-spectrum watermarks can be easily removed by simultaneously improving the signal quality. QIM watermarks can be destroyed only by decreasing the

average signal quality; however, the distortion penalty for a successful attack is small, about 2-2.5 dB.

We reviewed some proposals for asymmetric watermarking schemes and discussed their pros and cons. It was found that none of the schemes is sufficiently robust against malicious attacks.

For future research, it seems appropriate to develop a stronger theoretical foundation of asymmetric watermarking so that fundamental limits can be found. It is still not clear whether asymmetric watermarking might ever lead to secure public watermark detection. However, the topic is highly relevant since multimedia content providers already complain about huge financial losses due to illegal copying of the digital data.

6 Acknowledgement

We thank Teddy Furon for the frequent email discussion concerning the topic of asymmetric watermarking schemes.

References

1. B. Chen and G. W. Wornell. Digital watermarking and information embedding using dither modulation. In *Proc. of IEEE Workshop on Multimedia Signal Processing* , Redondo Beach, CA, USA, December 1998.
2. B. Chen and G. W. Wornell. Dither modulation: a new approach to digital watermarking and information embedding. In *Proc. of SPIE Vol. 3657: Security and Watermarking of Multimedia Contents*, San Jose, January 1999.
3. I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 1997.
4. I.J. Cox and J.-P. Linnartz. Some general methods for tampering with watermarks. *IEEE Journal on Selected Areas in Communications*, 16:587–593, May 1998.
5. J. J. Eggers and B. Girod. Robustness of Public Key Watermarking Schemes. In V^3D^2 *Watermarking Workshop*, Erlangen, Germany, October 1999.
6. J. J. Eggers, J. K. Su, and B. Girod. Public Key Watermarking by Eigenvectors of Linear Transforms. In *Proc. of European Signal Processing Conf.*, Tampere, Finland, April 2000. To appear.
7. T. Furon and P. Duhamel. An Asymmetric Public Detection Watermarking Technique. In *Workshop on Information Hiding*, Dresden, Germany, October 1999.
8. T. Furon and P. Duhamel. Copy Protection of Distributed Contents: An Application of Watermarking Technique. In *Workshop COST 254: Friendly Exchange through the net*, Bordeaux, France, March 2000.
9. T. Furon, N. Moreau, and P. Duhamel. Audio Public Key Watermarking Technique. In *Proc. of the IEEE Intl. Conf. on Speech and Signal Processing 2000* , Istanbul, Turkey, June 2000. To appear.
10. F. Hartung and B. Girod. Fast Public-Key Watermarking of Compressed Video. In *Proc. of the IEEE Intl. Conf. on Image Processing 1997* , Santa Barbara, CA, USA, October 1997.
11. F. Hartung and B. Girod. Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3):283–301, May 1998.
12. R. G. van Schyndel, A. Z. Tirkel, and I. D. Svalbe. Key independent watermark detection. In *Proc. of the IEEE Intl. Conf. on Multimedia Computing and Systems*, volume 1, Florence, Italy, June 1999.