Total Disclosure of the Embedding and Detection Algorithms for a Secure Digital Watermarking Scheme for Audio

David Megías, Jordi Herrera-Joancomartí, and Julià Minguillón

Estudis d'Informàtica i Multimèdia,
Universitat Oberta de Catalunya,
Av. Tibidabo 39–43, 08035 Barcelona
Tel. (+34) 93 253 7523, Fax (+34) 93 417 6495
{dmegias, jordiherrera, jminguillona}@uoc.edu

Abstract. This paper discusses the modification of a robust digital audio watermarking scheme to allow the disclosure of the embedding and detection algorithms. The chosen scheme uses MPEG 1 Layer 3 compression to determine the position of the mark bits in the frequency domain. The marking positions would be exposed if the original embedding algorithm was disclosed. In fact, it is shown that even if an attacker did not know the exact tuning parameters used for embedding, he or she could still produce an approximate superset of the marking frequencies from only a marked copy and successfully attack the file. To avoid this problem, a secret key is introduced in the embedding and detection processes. The secret key includes the seed of a pseudo-random number generator which is used to compute the exact marking positions. The modification is then analysed in terms of capacity, imperceptibility, robustness and security. The experiments show that the modified scheme preserves most of the properties of the original one, such as robustness against MP3 compression for the most frequently used bit rates, and does introduce additional security as the mark is more difficult to erase when the embedding and detection algorithms are disclosed.

Keywords: Audio Watermarking, Information Hiding, Intellectual Property Protection.

1 Introduction

Digital watermarking deals with the problem of embedding information (a mark) into a digital object (the cover object). Depending on the application, digital watermarking has different goals. For instance, for copyright protection, the embedded mark should not be removed when modifying the cover object unless the cover object itself becomes unusable. For authentication purposes, the watermark should be fragile in the sense that a minor change in the cover object should produce the loss of the mark.

The first few digital watermarking applications were focused on the copyright protection problem. Within this scenario, the major concern about watermarking was robustness, since the embedded mark should not be able to be removed. To measure the robustness of watermarking schemes, different benchmark tools were developed

S. Qing et al. (Eds.): ICICS 2005, LNCS 3783, pp. 427–440, 2005.

(for example [6]) and the schemes were exhaustively tested against the attacks included in those benchmarks. At this time, no existing watermarking scheme supports the vast range of attacks included in all those benchmarks. However, a deep study on the survived attacks shows that some of the suggested watermarking schemes are robust against a moderate number of attacks. In fact, the survived attacks are enough for many specific applications. In the light of these results, the next benchmark generation is focused on defining application-oriented benchmarks [15].

At this first stage, the problem of removing the mark from a marked object was only dealt from the robustness point of view. This means that only attacks produced in an unintentional manner were considered. However, it was then pointed out that other attacks could be envisaged, such as the *sensitive attack* [4], in which the knowledge of the watermarking system could be exploited to erase the mark. In fact, in [4], attacks were classified between signal transformations and intentional attacks. With this classification, there are different approaches to define watermarking security [13,10,1,3], but there is no consensus at this time about this issue.

The main problem is whether the set of robustness attacks and the set of security attacks are disjoint or not and, then, some robustness attacks can be considered also as security attacks. Such a situation arises when we consider the definition proposed in [10], where the difference between robustness and security is defined in terms of the intentionality of the attack. Clearly, with this definition, the intersection between robustness and security attacks is not empty, since their classification only depends on the intention but not on the attack itself. However, some type of attacks can be uniquely classified using the distinction based on intentionality. For instance, any attack which exploits the knowledge of the watermarking embedding or detecting algorithm is intentional and can be labeled as a security attack. From a security point of view, the best strategy to protect any secure system against attacks which exploit the knowledge of the scheme's construction is to ensure the Kerckhoffs' principle [14]. Such principle establishes that the security of any system (cryptosystems in particular) must only depend on a secret key whereas all the other information concerning the system is public.

This paper is organised as follows. Section 2 describes the audio watermarking scheme. In Section 3, the ad-hoc attack for the watermarking scheme is presented and the modification to overcome this attack is suggested. Section 4 presents the performance of the modified scheme in terms of imperceptibility, capacity, robustness and security. Finally, Section 5 summarises the conclusions and the future research.

2 Audio Watermarking Scheme

The watermarking scheme (referred to as **W**atermarking of **Au**dio **C**ontent, WAUC) presented in [18] (and improved in [19]) is described in the following sections.

2.1 Mark Embedding

Let the signal S to be marked be a collection of PCM samples. If the signal to be marked is stereo: $S_{\text{stereo}} = [S_{\text{left}}, S_{\text{right}}]$, both channels (left and right) must be added into a new "working" signal $S = S_{\text{left}} + S_{\text{right}}$. In the case of a mono signal, this step is not required. The spectrum of S, denoted as S_F, is computed with a Fast Fourier Transform (FFT) algorithm. Then, the signal S is compressed using an MP3 algorithm with a rate of R kbps and decompressed again to PCM format. The result of this compression/decompression operation is a new signal S', and its spectrum S'_F is obtained[1]. In the stereo case, the modified signal S' is obtained by adding the S'_{left} and S'_{right} which result after the compression and decompression operation.

Now, the set of marking frequencies F_{mark} is chosen as follows. Firstly, all $f_{\mathrm{mark}} \in F_{\mathrm{mark}}$ must belong to the relevant frequencies F_{rel} of the original signal S_F:

$$F_{\mathrm{rel}} = \{f \in [0, f_{\max}] : |S_F(f)| \geq (p/100)\, |S_F|_{\max}\}\,, \tag{1}$$

where $f_{\max}$ denotes the maximum frequency of the spectrum, which depends on the sampling rate and the sampling theorem[2], $p \in [0, 100]$ is a percentage and $|S_F|_{\max}$ is the maximum magnitude of the spectrum S_F. Note that the spectrum values in the interval $[-f_{\max}, 0]$ are the complex-conjugate of those in $[0, f_{\max}]$.

Secondly, the frequencies to be marked are those for which the magnitude remains "unchanged" after lossy compression and decompression, where "unchanged" means a relative error below a given threshold ε :

$$F_{\mathrm{mark}} = \{f_1, f_2, \ldots, f_n\} = \{f \in F_{\mathrm{rel}} : |(S_F(f) - S'_F(f))\,/S_F(f)| < \varepsilon\}\,. \tag{2}$$

Similarly as done in the image watermarking scheme of [8], a 70-bit stream mark, W ($|W| = 70$), is firstly extended to a 434-bit stream W_{ECC} ($|W_{\mathrm{ECC}}| = 434$) using a dual Hamming Error Correcting Code (ECC). This coding makes it possible to apply the watermarking scheme as a fingerprinting scheme robust against collusion of two buyers [7]. Finally, a pseudo-random binary stream (PRBS), generated with a cryptographic key k, is added to the extended mark as it is embedded into the original signal.

Once the frequencies in F_{mark} have been chosen, the spectrum of the marked signal is computed as:

$$\hat{S}_F(f) = \begin{cases} S_F(f), & f \notin F_{\mathrm{mark}}, \\ S_F(f) \cdot 10^{\pm d/20} & f \in F_{\mathrm{mark}}, \text{ to embed `1' } (+d/20) \text{ or `0' } (-d/20). \end{cases}$$

Since spectrum components in S_F are paired (pairs of complex-conjugate values), the same transformation (increase or decrease d dB) must be performed to $S_F(f_{\mathrm{mark}})$ and to its conjugate. In this process, the mark W_{ECC} is replicated as many times as required. In the stereo case, the magnitude modification step is applied to both S_{left} and S_{right} independently **at the same frequencies**. Finally, the marked audio signal is converted to the time domain $\hat{S}$ applying an inverse FFT (IFFT) algorithm. As discussed in [18], this scheme has been designed to provide with "natural" robustness against lossy compression attacks.

2.2 Mark Detection

The objective of the mark detection algorithm is to determine whether an audio test signal T is a (possibly attacked) version of the marked signal $\hat{S}$. It is assumed that T is

[1] Throughout this paper, the Blade codec [12] (**co**der/**dec**oder) for the MP3 algorithm has been chosen and, thus, the psychoacoustic model of this codec is implicitly used.

[2] $f_{\max} = \frac{1}{2T_s}$, where T_s is the sampling time.

in PCM format or can be converted to it. Note that working signals adding the left and the right channels must be used in the stereo case.

First of all, the spectrum T_F is obtained applying the FFT algorithm and, then, $|T_F(f_{\text{mark}})|$, the magnitude at the marking frequencies, is computed for all $f_{\text{mark}} \in F_{\text{mark}}$. Note that this method is strictly positional and, because of this, it is required that the number of samples in $\hat{S}$ and T is the same. If there is only a little difference in the number of samples, it is possible to complete the sequences with zeroes.

When the magnitudes $|T_F(f_{\text{mark}})|$ are available, a scaling (Least Squares) step can be undertaken in order to minimize the distance between the sequences $\lambda\,|T_F(f_{\text{mark}})|$ and $\left|\hat{S}_F(f_{\text{mark}})\right|$ (see [18] for details). This LS step implicitly uses the embedded mark (since $S_F(f_{\text{mark}})$ is needed) but it can be omitted ($\lambda = 1$) or performed with the original signal $S_F(f_{\text{mark}})$ instead of the marked one $\hat{S}(f_{\text{mark}})$.

Now, the ratios $\boldsymbol{r}_i = \lambda\,|T_F(f_i)|\,/\,|S_F(f_i)|$, are computed to decide whether a '0', a '1' or a '*' (not identified) might be embedded at the i-th position. Given the interval

$$I = \left[10^{\frac{d}{20}}(100-q)/100, 10^{\frac{d}{20}}(100+q)/100\right],$$

if $\boldsymbol{r}_i \in I \Rightarrow \hat{\boldsymbol{b}}_i :=$ '1', if $1/\boldsymbol{r}_i \in I \Rightarrow \hat{\boldsymbol{b}}_i :=$ '0' and, otherwise, $\hat{\boldsymbol{b}}_i :=$ '*'. Here, $q \in [0, 100]$ is a percentage and $\hat{\boldsymbol{b}}_i$ is the i-th component of the vector $\hat{\boldsymbol{b}}$ which contains a sequence of "detected bits". Finally, the PRBS signal is removed from the bits $\hat{\boldsymbol{b}}$ to recover the true embedded bits $\boldsymbol{b}$. This operation must preserve unaltered the '*' marks.

Once $\boldsymbol{b}$ has been obtained, its length n will be greater than the length of the extended mark. Hence, each bit of the mark appears at different positions in $\boldsymbol{b}$. A *voting* scheme (see [18] for details) is applied to choose whether the i-th bit of the mark is '1', '0' or unidentified ('*'). As a result of this voting scheme, an identified extended mark W'_{ECC} is obtained and the error correcting algorithm is used to recover an identified 70-bit stream mark, W', which will be compared with the true mark W.

The suggested scheme is informed (not blind) since the original signal is needed by the mark detection process. However, the bit sequence which forms the embedded mark is not required for detection (if the LS step is omitted or performed using S_F), which makes this method suitable also for fingerprinting [2].

3 Security Issues

In this section, we focus on attacks which exploit the knowledge of the embedding and detection algorithms. More specifically, we consider the case referred to as *Watermark Only Attack*, in which the attacker has only access to marked contents [5,1]. An ad-hoc strategy can be specifically defined for the WAUC watermarking scheme once the mark embedding and detection algorithms are disclosed.

Concerning the WAUC watermarking scheme presented in the previous section, the disclosure of the mark embedding and detection algorithms has an obvious drawback from a security point of view: given the embedding parameters R, ε and p, and the MP3 encoder/decoder, the position of the embedded bits (F_{mark}) is absolutely determined. Therefore, a malicious attacker (Mallory) with knowledge about the embedding process, could design an ad-hoc attack to disturb the spectrum of $\hat{S}$ at those frequencies and try to erase the mark. The following section presents such attack.

3.1 Ad-Hoc Security Attack

Assuming that Mallory knows the embedding and detection algorithms, an ad-hoc security attack for this watermarking scheme can be described in the following way:

1. Mallory obtains the marked signal $\hat{S}$.
2. Mallory computes the spectrum $\hat{S}_F$ applying the FFT.
3. Mallory encodes/decodes the marked signal $\hat{S}$ with an MP3 encoder/decoder and gets a modified signal $\hat{S}'$. Here, he uses the bit rate R' for the MP3 encoder/decoder. Now, he applies the FFT to the signal $\hat{S}'$ and gets the spectrum $\hat{S}'_F$.
4. Mallory computes the set $\hat{F}_{\text{mark}}$ applying the criteria of Equations 1 and 2 using the spectra $\hat{S}_F$ instead of S_F, using $\hat{S}'_F$ instead of S'_F, and the parameters p' and ε'. Note that Mallory does not have the original signal S neither the modified signal S'.
5. Finally, Mallory disturbs the magnitude of the spectrum at the frequencies. $\hat{F}_{\text{mark}}$. For example, he could decide to disturb $\pm d'$ dB at those frequencies randomly.

Note that, even if Mallory knew the mark W and the extended mark W_{ECC}, the use of the PRBS generated with a **secret** key k in the embedding process prevents Mallory from knowing which exact bit is embedded at each position. So, even if Mallory got the exact set F_{mark} (which is impossible unless he had the original signal), he would not know whether if a binary ‘0’ or a binary ‘1’ is embedded at each position. Therefore, the best strategy to disturb the marked signal is to add or subtract d' dB randomly.

Of course, Mallory should gain knowledge of the embedding parameters in order to have all the information required to construct his approximation to F_{mark}. If the parameters R, p and ε were public, the attack would be easier, since Mallory could set $R' = R$, $p' = p$ and $\varepsilon' = \varepsilon$. In fact, the sets F_{mark} and $\hat{F}_{\text{mark}}$ cannot be exactly the same, since Mallory should have access to the original signal S to obtain F_{mark}. However, as $\hat{S}$ is expected to be quite similar to S (since the WAUC scheme has a good imperceptibility level), the constructed set $\hat{F}_{\text{mark}}$ will contain many of the frequencies in the original F_{mark}, possibly enough to be able to delete the mark.

In addition, it must be noticed that although Mallory does not know the exact values of the embedding parameters, he can construct an approximate superset $\hat{F}_{\text{mark}}$ following these guidelines:

1. Choose a large enough parameter R'. The larger R' is, the more similar $\hat{S}$ and $\hat{S}'$ become and, thus, the more frequencies will be included in $\hat{F}_{\text{mark}}$.
2. Choose a small enough percentage p'. This way, more frequencies satisfy the criterion of Equation 1.
3. Choose a large enough relative error ε'. This way, more frequencies satisfy the criterion of Equation 2.

Here, the parameter d' should be chosen in such a way that the binary ‘1’s and ‘0’s are erased. Thus, an advisable choice for d' is $d' > d$. Since the imperceptibility of the mark requires that d is not very large [20], usually $d \leq 1$ dB will be chosen for mark embedding. Hence, Mallory decides to use $d' = 2$ dB. As a consequence of Mallory's attack, the perceptual quality of T will be reduced with respect to that of $\hat{S}$, but this is the price an attacker has to pay in order to delete the mark.

3.2 Security Enhancement

The main idea to solve the security problem described above is to "hide" the marking positions as much as possible by computing them using a secret key, as detailed below. Then, the watermarking scheme can be considered secure under the Kerckhoffs' assumption. On the other hand, such modification should preserve as many properties as possible compared to the original "non-secure" scheme. Special attention should be devoted to robustness, imperceptibility and capacity.

If these two conditions are met, it would be possible to make the watermarking scheme publicly known except for the secret key. The unavailability of the key should make it very difficult to proceed with the attack presented in section 3.1. In order to meet these conditions, the embedding process can be modified in such a way that the marking frequencies depend on a secret key as follows:

- Proceed with the mark embedding process described in Section 2.1 until the set F_{mark} is obtained. Now, define $f_M = \max F_{\text{rel}}$ (the maximum frequency that satisfies the criterion of Equation 1) and $F_{\text{perc}} = F_{\text{mark}}$. The set F_{mark} obtained in Section 2.1 is a temporary variable (F_{perc}) which stores the most perceptually relevant part of the spectrum, hence the subindex *perc*.
- Define the set of candidate marking frequencies as $F_{\text{cand}} = [0, f_M]$. This prevents very high frequencies, which are not usually good for embedding the mark as robustness is concerned, to be chosen. Let m be the cardinality of the set F_{cand}, *i.e.* $m = |F_{\text{cand}}|$. Note that $F_{\text{perc}} \subseteq F_{\text{cand}}$[3].
- Choose a **pseudo-random** number generator in the range $[0, 1]$ and a **secret key** k_{sec} as the (initial) seed[4].
- Choose two probabilities $p_1, p_2 \in [0, 1]$ such that a given frequency which belongs to the set F_{perc} will be chosen for marking with a probability p_1, and p_2 is the probability of choosing a frequency in the set $F_{\text{cand}} - F_{\text{perc}}$, where "$-$" stands for the set subtraction operation.
- Reset the random number generator with the seed k_{sec}. Let F_{mark} be the empty set, and proceed as follows. For all the frequencies f in the set F_{cand} do:
  1. Generate a random number $r \in [0, 1]$.
  2. If ($f \in F_{\text{perc}}$ **and** $r < p_1$) **or** ($f \notin F_{\text{perc}}$ **and** $r < p_2$) then $F_{\text{mark}} := F_{\text{mark}} \cup \{f\}$ (the frequency f is included into the set of marking frequencies).
  3. Otherwise, discard the frequency f.

Once the set F_{mark} has been generated, it is possible to apply the mark embedding process presented in Section 2.1 with this new set. The mark reconstruction process

[3] Usually, the set F_{cand} has many more elements than F_{perc}, *i.e.* $|F_{\text{perc}}| \ll |F_{\text{cand}}|$.

[4] The subindex *sec* is used to distinguish from the secret key k which was already used in the embedding process described in Section 2.1.

should be also modified accordingly, since the mark embedding detection must repeat the first few steps of the mark embedding process in order to obtain F_{mark} (and $\hat{S}$).

It is worth pointing out some remarks about the modification suggested above. Firstly, note that the frequencies in F_{perc} are included into the set F_{mark} with a probability p_1 and those in $F_{\text{cand}} - F_{\text{perc}}$ are included with a probability p_2. Secondly, any pseudo-random number generator can be used and, thus, the length of the secret key k_{sec} will depend on it. In this paper, the Mersenne Twist method presented in [17] is used. In addition, in the original watermarking scheme presented in Section 2, the number of marked bits is $n = |F_{\text{perc}}|$ (remember that F_{mark} is now referred to as F_{perc}). It is possible to define p_1 and p_2 such that the expected value of the final number of elements in F_{mark} is the same as in the scheme presented in Section 2. The expected value of the number of elements in F_{mark} is the following: $\mathrm{E}\left(|F_{\text{mark}}|\right) = np_1 + (m-n)p_2$. In order to get $\mathrm{E}\left(|F_{\text{mark}}|\right) = n$, p_2 should be chosen as:

$$np_1 + (m-n)p_2 = n \Leftrightarrow p_2 = (1-p_1)\frac{n}{m-n},$$

which is considered as the **default value** for p_2 hereafter. With this default value for p_2, the number of marking frequencies $n' = |F_{\text{mark}}|$ will be (in average) equal to n (since $\mathrm{E}(n') = n$). In addition, the ratios of frequencies in F_{mark} belonging to F_{perc} and $F_{\text{cand}} - F_{\text{perc}}$ will be (in average) p_1 and $1 - p_1$, respectively. Thus, p_1 determines the balance between the marking frequencies inside and outside F_{perc}. For example, with $p_1 = 0.2$ (and p_2 equal to the default value), a 20% (in average) of the frequencies in F_{mark} belong to the set F_{perc} and the other 80% belong to $F_{\text{cand}} - F_{\text{perc}}$.

Note, also, that if $p_1 = 1$ and the default value is chosen for p_2, then $p_2 = 0$. In this case, the set F_{mark} becomes identical to F_{perc} and the modified watermarking scheme becomes identical to the one presented in [19].

The robustness of the modified scheme will depend on the value of p_1. *A priori*, the frequencies in the set F_{perc} are better for mark embedding, since they have been chosen in such a way that MP3 compression attacks can be overcome [19]. It is expected that $p_1 = 1$ is the best value for robustness, but such a value is not advisable for security. Thus, a trade-off solution between robustness and security must be attained. As imperceptibility is concerned, the frequencies F_{perc} refer to the most perceptible part of the spectrum and, thus, the lower p_1 is, the better imperceptibility is expected.

One may think that Mallory could distort not only the frequencies in the set F_{perc}, but all the frequencies in the set F_{cand}. Note that, if it were the case, the mark would very probably be erased, but the distortion introduced to the signal would be so large (the spectrum would be distorted at all the low frequencies) that the attacked signal would be unusable in most typical situations.

Last, but not least, note that p_1 should not be chosen too small or, at least, Mallory should not have knowledge about p_1. If p_1 is so small that most of the marking frequencies lie outside the set F_{perc}, then Mallory would do better attacking at the frequencies in $F_{\text{cand}} - \hat{F}_{\text{perc}}$. If he attacked at those frequencies, he would be able to delete more bits and, in addition, he would disturb the least perceptible part of the spectrum resulting in an attacked signal with very good audio quality.

The secret key K of the modified watermarking scheme in order to increase security should be formed, at least, by the pseudo-random seeds k and k_{sec}, and the probabilities p_1 and p_2. Here, we consider that the secret key is the following: $K = \{R, p, \varepsilon, d, k, k_{\text{sec}}, p_1, p_2\}$, which is required for both mark embedding and detection. The decision whether R, p, ε and d are part of the secret key or public values can depend on the application. In any case, these parameters should not be used to enhance security but rather for tuning reasons [20]. Note, also, that the parameter q does not affect the mark embedding process. Thus, q in not a part of the secret key. Hereafter, the modified scheme is referred to as WAUC-sec.

4 Performance Evaluation

In this section, the WAUC and the WAUC-sec schemes are evaluated in terms of capacity, imperceptibility, robustness and security. Both the WAUC and the WAUC-sec schemes have been implemented using a dual binary Hamming code $DH(31, 5)$ as ECC and the PRBS has been generated with the a DES cryptosystem in an Output Feedback (OFB) mode. A 70-bit mark W ($|W_{\text{ECC}}| = 434$) was embedded. In addition, the following values have been chosen for the embedding and detection parameters: $R = 128$ kbps, $p = 2$, $\varepsilon = 0.05$, $d = 1$ dB and $q = 10$.

To test the performance of the audio watermarking schemes described in the previous sections, different audio files provided in the Sound Quality Assessment Material (SQAM) page [9] have been used. In order to summarise the results as much as possible, only the experiments performed for the (stereo) violoncello (melodious phase) file[5] are shown. Completely analogous results have been obtained for the other files in the SQAM corpus set. The properties of the WAUC-sec scheme have been tested for nine values of the probability $p_1 \in \{0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1\}$ and the default value for p_2. In addition, eight different values of the secret key k_{sec} have been chosen randomly. Different values of k_{sec} are used in order to avoid biased results which could arise with some sequences of pseudo-random numbers. Note that the original WAUC scheme is also considered, since it is identical to WAUC-sec with $p_1 = 1$ and $p_2 = 0$.

4.1 Capacity

Capacity (C) is the amount of information that may be embedded and recovered in the audio stream and it is measured in bits per second (bps). It must be taken into account that the true capacity C is not the number of marked bits n' (the size of the set F_{mark}). Hence, $C = 70 \cdot n'/(434 \cdot l)$, where l is the length of the marked signal.

Table 1 shows the capacity results obtained for the WAUC (the column with $p_1 = 1$) and the WAUC-sec ($0 \leq p_1 \leq 0.875$) schemes. Since eight different values of the key k_{sec} have been chosen, three different measures are shown in the table: the maximum, the minimum and the average. It can be observed that all the values are very similar, as the modification has been designed in such a way that the capacity of the original scheme is preserved when the default value is chosen for p_2. If more than eight values of the secret key k_{sec} had been used, the average value of C in each column would be

[5] In fact, only the first ten seconds (441000×2 samples) of the file have been taken into account.

Table 1. Capacity results for WAUC and WAUC-sec

Capacity	**WAUC-sec**								**WAUC**
	p_1								$p_1 = 1$
	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	
Maximum C (bps)	62.56	62.42	61.68	61.39	61.76	61.69	62.05	61.76	61.08
Minimum C (bps)	59.06	58.60	59.32	59.31	59.47	59.55	59.73	60.34	61.08
Average C (bps)	60.93	60.86	60.70	60.46	60.64	60.68	60.85	61.07	61.08

closer to that of the original scheme (61.08 bps). It must be taken into account that capacity also depends on the original signal, as discussed in [19].

4.2 Imperceptibility

Imperceptibility is concerned with the audio quality of the marked signal $\hat{S}$ with respect to S. Here, to measure such property, the Objective Difference Grade (ODG) based on the ITU-R Recommendation standard BS.1387 [11] and the signal-to-noise ratio (SNR) are used. The BS.1287 standard is used for Perceptual Evaluation of Audio Quality (PEAQ) [21]. In particular, the implementation provided in the tool EAQUAL [16] has been used in this paper. The computed ODG values are in the range $[-4, 0]$, where 0 means imperceptible, -1 means perceptible but not annoying, -2 means slightly annoying, -3 means annoying and -4 means very annoying. The SNR values make it possible to compare these results with those presented in previous papers [19,20].

Table 2. Imperceptibility results for WAUC and WAUC-sec

Imperceptibility	**WAUC-sec**								**WAUC**
	p_1								$p_1 = 1$
	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	
Maximum ODG	-0.46	-0.88	-1.17	-1.31	-1.53	-1.72	-1.81	-1.85	-1.96
Minimum ODG	-0.53	-0.96	-1.30	-1.58	-1.70	-1.80	-1.92	-2.01	-1.96
Average ODG	-0.50	-0.92	-1.24	-1.45	-1.64	-1.77	-1.87	-1.95	-1.96
Maximum SNR (dB)	39.76	28.58	25.77	23.95	22.23	21.15	20.42	19.71	18.95
Minimum SNR (dB)	39.18	26.69	24.50	23.13	21.86	20.75	19.94	19.47	18.95
Average SNR (dB)	39.47	27.85	25.10	23.49	22.00	20.96	20.19	19.55	18.95

In Table 2, the SNR and ODG measures obtained with both WAUC and WAUC-sec are shown. It is observed that the ODG values increase from imperceptible for $p_1 = 0$ to slightly annoying for $p_1 = 1$, and analogous results are obtained in terms of SNR. These results are quite satisfying in general. For values of $p_1 \leq 0.5$, a listener is not expected to notice any remarkable difference between the original and the marked files, since the ODG values range between imperceptible and perceptible but not annoying.

Imperceptibility has been improved with respect to the original scheme. Thus, the modifications do not only enhance security, but also imperceptibility. These results were expectable, since the changes introduced to the scheme decrease the number of relevant frequencies at which the magnitude is disturbed. Finally, note that the imperceptibility results might be further improved by tuning the embedding parameters carefully [20].

4.3 Robustness

Robustness is the resistance to accidental removal of the embedded bits. The robustness of the resulting scheme has been tested against the version 0.2 of the StirMark Benchmark for Audio (SMBA) [6] and also against MP3 compression. In particular, 43 different attacks of the SMBA have been tested, since the attacks which modify the number of samples in a significant way cannot be tested with the current version of the scheme. Robustness has been assessed using a correlation measure between the embedded and the identified marks (W and W'):

$$\text{Correlation} = \frac{1}{|W|}\sum_{i=1}^{|W|}\beta_i.$$

where $\beta_i = 1$ if $W_i = W'_i$ and -1 otherwise. In this paper, we consider that the watermarking scheme survives an attack if the Correlation ≥ 0.8.

Table 3. Robustness results against the SMBA for WAUC and WAUC-sec

	WAUC-sec								**WAUC**
Robustness	p_1								$p_1 = 1$
	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	
SMBA test	$34/43$	$35/43$	$36/43$	$36/43$	$36/43$	$37/43$	$37/43$	$35/43$	$35/43$

The robustness of WAUC and WAUC-sec are not very different with respect to the SMBA. Table 3 shows the robustness results obtaiend for the WAUC and WAUC-sec schemes with a value for the key k_{sec} (analogous results have been obtained with other keys). The values in the table are given in a $x/43$ ratio meaning how many SMBA attacks out of the 43 attacks performed have been survived. It can be noticed that the survival ratio varies from $34/43$ to $37/43$ but there is not a monotonic pattern for this variation. This result is not very surprising since the SMBA attacks are not specifically designed to disturb the most perceptually significant frequencies of the spectrum.

However, the robustness of the WAUC and WAUC-sec schemes against MP3 compression attacks are quite different. Table 4 summarises the robustness results against MP3 compression using all eight random values for the key k_{sec}. The attacks have been performed with a Blade codec and all the allowed bit rates: 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 and 320 kbps. The scores given in the table are in the form: Score $=$ # Survived attacks/# Performed attacks. Therefore, scores of 1 mean that the scheme survives the compression attack for all the values of the key k_{sec}, whereas 0 means that the compression attack has successfully erased the mark for all values of k_{sec}. In the WAUC scheme (which does not depend on the key k_{sec}), the only possible scores are 0 and 1, whereas intermediate results are possible for the WAUC-sec scheme. As it can be observed, both schemes are able to overcome MP3 compression attacks for all bit rates of 128 kbps and higher. The robustness against the other bit rates increases as p_1 is larger, as expected, since more marking frequencies belong to the most perceptually significant part of spectrum.

Table 4. Robustness results against MP3 compression attacks for WAUC and WAUC-sec

	WAUC-sec								**WAUC**
MP3 bit rate	p_1								$p_1 = 1$
	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	
32 kbps	0	0	0	0	0	0	0	0	0
40 kbps	0	0	0	0	0	0	0	0.25	0
48 kbps	0	0	0	0	0	0	0.375	0.625	1
56 kbps	0	0	0	0	0.125	0.5	0.625	1	1
64 kbps	0	0	0	0.125	0.75	1	1	1	1
80 kbps	0	0	0.5	1	1	1	1	1	1
96 kbps	0.125	1	1	1	1	1	1	1	1
112 kbps	0.375	1	1	1	1	1	1	1	1
$[128, 320]$ kbps	1	1	1	1	1	1	1	1	1

Table 5. Security results for WAUC and WAUC-sec

		WAUC-sec		**WAUC**
		p_1		$p_1 = 1$
		0.25	0.5	
Attack #1	Correlation	$1 > 0.8$ (*)	$0.9077 > 0.8$ (*)	$-0.0462 < 0.8$
	ODG	-2.1845	-2.0159	-1.6894
Attack #2	Correlation	$1 > 0.8$ (*)	$0.6308 < 0.8$	$0.2308 < 0.8$
	ODG	-2.4946	-2.3492	-2.1868

As MP3 compression is concerned, the WAUC scheme performs better than WAUC-sec, since it survives attacks for all bit rates greater to or equal to 48 kbps. The WAUC-sec scheme is not that robust, but it still produces good enough results for many values of p_1. For example, with $p_1 = 0.5$ the WAUC-sec scheme is able to overcome all the attacks with bit rates greater to or equal to 80 kbps, and a 75% of the attacks performed with 64 kbps. If high quality is required (for example for music audio files), it is not expected that an attacker would use MP3 bit rates lower than 80 kbps.

4.4 Security

In this section, two different kinds of experiments are presented. Firstly, false positive results are shown and, secondly, the resistance of both WAUC and WAUC-sec against the ad-hoc attack presented in section 3.1 is examined.

As false positives are concerned, the experiments consist of using different values for the key k_{sec} in the embedding and the detection processes. For these experiments, the value $p_1 = 0.5$ has been chosen. The correlation measure described in the previous section has been computed to assess the similarity between the embedded mark W and the recovered one W'. Since eight different values have been used for k_{sec}, seven correlation values are obtained for each k_{sec}. Thus, $8 \times 7 = 56$ experiments have been performed. If any of these 56 correlation values were too close to 1, then the false positive rate would be relatively large. In these 56 experiments, the *maximum* correlation value obtained is 0.3231, quite far from the required survival threshold (0.8). The average of the absolute values of all these 56 correlation measures is lower than 0.1.

The results for two settings of the ad-hoc attack depicted in Section 3.1 are given below. **Attack #1**: performed assuming that Mallory correctly guesses the parameters used by the embedder: $R' = R = 128$ kbps, $p' = p = 2$ and $\varepsilon' = \varepsilon = 0.05$. The magnitude modification parameter $d' = 2$ dB has been chosen. **Attack #2**: it is assumed that Mallory does not know the parameters used by the embedder and he tries to produce an approximate superset of F_{perc} by choosing: $R' = R = 128$ kbps, $p' = 1$ ($p = 2$) and $\varepsilon' = 0.1$ ($\varepsilon = 0.05$). Again, $d' = 2$ dB is used. This attack will affect more frequencies than the previous one.

For these attacks, the LS step mentioned in Section 2.2 has not been used ($\lambda = 1$). The experiments have been carried out for two values of p_1 (0.25 and 0.5) and the same secret key k_{sec}. The results obtained for this ad-hoc attack are summarised in Table 5, where the ODG results do not refer to the original (unmarked) file but to the marked signal. The ODG values with respect to the original file are obviously worse. The star sign "(*)" shows which attacks are not successful, *i.e.* when the correlation is larger than or equal to the threshold 0.8. The WAUC-sec scheme is able to survive Attack #1 for both values of p_1, whereas the original WAUC scheme fails to recover the mark. Note, also that the ODG values are worse than those obtained for mark embedding (see Table 2). Thus the WAUC-sec scheme survives attacks which introduce more distortion into the attacked signal than what the embedder does in the marking process. The situation is a bit more difficult with the Attack #2, since the spectrum is disturbed at more frequencies. Because of this, the ODG values are worse than those of the Attack #1. Now the WAUC-sec scheme does not survive the attack with $p_1 = 0.5$, though the correlation value is not too far from the threshold. These results show that the modification works exactly as predicted in Section 3, since the ad-hoc attack fails to erase the mark as the probability p_1 is decreased.

Two final experiments have been performed to test a worse attack scenario. Firstly, it is assumed that Mallory wants to disturb not only the frequencies in the set F_{perc}, but all the frequencies in F_{cand}. Such attack successfully erases the mark even for $p_1 = 0.25$, since the obtained correlation is -0.1692. However, in such a situation, the ODG value between the marked file and the attacked one is -2.6933, *i.e.* the noise introduced by the attack is quite annoying according to the ODG scale defined above. Secondly, if Mallory guesses (or discovers) that $p_1 = 0.25$ has been chosen, he would do better attacking $F_{\text{cand}} - F_{\text{perc}}$, as already remarked in Section 3.2. Such attack has been performed with $R' = R = 128$ kbps, $p' = p = 2$, $\varepsilon' = \varepsilon = 0.05$ and $d' = 2$ dB, disturbing the spectrum at the frequencies $F_{\text{cand}} - \hat{F}_{\text{perc}}$. In this case, the correlation is 0.3538, but audio quality is very good according to the ODG measure: -1.2195.

The main conclusion after all these experiments is that the value of p_1 should be chosen in some interval centred at 0.5 and too small values should be avoided. For example, p_1 should be in the interval $[0.3, 0.7]$. This way, a good trade-off between robustness and security would be obtained.

5 Conclusions

In this paper, several security issues related to a watermarking scheme for audio (WAUC) are discussed. On the first hand, it has been shown that the original WAUC scheme is not suitable for the disclosure of the embedding and detection algorithms, since the marking positions could be exposed to a malicious user, making it possible to design a successful ad-hoc attack. A modification, WAUC-sec, has been described in such a way that the embedding and the detection processes depend on a secret key without which the marking positions cannot be exactly determined.

The experiments show that it is possible to tune the modification so that the capacity of the original scheme can be preserved. In addition, the WAUC-sec scheme obtains better imperceptibility results (both in ODG and SNR measures) than the original counterpart for the same capacity. As robustness is concerned, both schemes produce similar results against the SMBA, but the original WAUC scheme is more robust against MP3 compression. Finally, concerning security, both false positive experiments and ad-hoc attacks have been performed. On the one hand, it has been shown that false positives are quite improbable if different secret keys are used for embedding and detection. On the other hand, the ad-hoc attacks can be survived by the WAUC-sec scheme whereas they successfully erase the mark when the original WAUC scheme is used. In short, the WAUC-sec scheme provides with a trade-off solution between security and robustness against MP3 compression.

There are several directions to further the research presented in this paper. The first one is to take into account some attacks which are not included in the version 0.2 of the SMBA, such as play speed variance attacks. Secondly, the development of a real application would require working with frames (blocks of samples) instead of the whole file, which would imply some reformulation. Finally, the possibility of obtaining a blind detector should be investigated.

Acknowledgements and Disclaimer

This work is partially supported by the Spanish MCYT and the FEDER funds under grants TIC2003-08604-C04-04 MULTIMARK and SEG2004-04352-C04-04 PROPRIETAS-WIRELESS. The work described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document reflects only the author's views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

1. M. Barni, F. Bartolini, and T. Furon. A general framework for robust watermarking security. *Signal Process.*, 83(10):2069–2084, 2003.
2. D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. In *Advances in Cryptology-CRYPTO'95*, LNCS 963, pages 452–465. Springer-Verlag, 1995.
3. F. Cayre, C. Fontaine, and T. Furon. Watermarking attack: Security of wss techniques. In *Digital Watermarking: Third International Workshop*, volume LNCS 3304, pages 197–208, Seoul, Korea, Oct 2004.
4. I. 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. W. Diffie and M. Hellman. New directions in cryptography. *IEEE Trans. on Information Theory*, 22(6):644–654, November 1976.

6. J. Dittman, M. Steinebach, A. Lang, and S. Zmudzinski. Advanced audio watermarking benchmarking. In *Proceedings of the IS&T/SPIE's 16th Annual Symposium on Electronic Imaging*, volume 5306 - Security, Steganography, and Watermarking of Multimedia Contents VI, Sant Jose, CA, US, January 2004.
7. J. Domingo-Ferrer and J. Herrera-Joancomartí. Short collusion-secure fingerprinting based on dual binary hamming codes. *Electronics Letters*, 36(20):1697–1699, September 2000.
8. J. Domingo-Ferrer and J. Herrera-Joancomartí. Simple collusion-secure fingerprinting schemes for images. In *Proceedings of the Information Technology: Coding and Computing ITCC'2000*, pages 128–132. IEEE Computer Society, 2000.
9. EBU. SQAM - Sound Quality Assessment Material, 2001. http://sound.media.mit.edu/mpeg4/audio/sqam/.
10. T. Furon et al. Security analysis. Deliverable 5. 5, 2002. IST project CERTIMARK (IST-1999-10987).
11. ITU-R. Recommendation BS.1387. Method for objective measurements of perceived audio quality, December 1998.
12. T. Jansson. Homepage for BladeEnc, 2001. http://bladeenc.mp3.no/.
13. T. Kalker. Considerations on watermarking security. In *Proceedings of the IEEE Fourth Workshop on Multimedia Signal Processing*, pages 201–206, Cannes, France, Oct. 2001.
14. A. Kerckhoffs. La cryptographie militaire. *Journal des Sciences Militaires*, 9:5–38, January 1883.
15. A. Lang, J. Dittmann, and E. J. Delp. Application-oriented audio watermark benchmark service. In *Proceedings of the IS&T/SPIE's 17th annual symposium on Electronic Imaging*, volume 5681, San Jose, CA, Jan. 2005.
16. A. Lerch. EAQUAL - Evaluate Audio QUALity. http://www.mp3-tech.org/programmer/sources/eaqual.tgz.
17. M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.
18. D. Megías, J. Herrera-Joancomartí, and J. Minguillón. A robust audio watermarking scheme based on MPEG 1 layer 3 compression. In *Communications and Multimedia Security - CMS 2003*, Lecture Notes in Computer Science 2828, pages 226–238, Turin (Italy), October 2003. Springer-Verlag.
19. D. Megías, J. Herrera-Joancomartí, and J. Minguillón. An audio watermarking scheme robust against stereo attacks. In *Proceedings of the Multimedia and Security Workshop*, pages 206–213, Magdeburg (Germany), September 2004. ACM.
20. D. Megías, J. Herrera-Joancomartí, and J. Minguillón. Robust frequency domain audio watermarking: a tuning analysis. In *International Workshop on Digital Watermarking - IWDW 2004*, Lecture Notes in Computer Science 3304, pages 244–258, Seoul (Korea), November 2004. Springer-Verlag.
21. T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandeburg, and B. Feiten. PEAQ – The ITU standard for objective measurement of perceived audio quality. *J. Audio Eng. Soc.*, 48(1-2):3–29, Jan.-Feb. 2000.