Steinebach, Zmudzinski, Countermeasure for collusion attacks against digital watermarking; Security, Steganography, and Watermarking of Multimedia Contents VIII, Edward J. Delp III, Ping W. Wong, Editors, January 2006, S. 470-479, ISBN: 0-8194-6112-1

Countermeasure for collusion attacks against digital watermarking

Martin Steinebach, Sascha Zmudzinski
Fraunhofer Integrated Publication and Information Systems Institute (IPSI)
Dolivostraße 15, 64293 Darmstadt, Germany

ABSTRACT

Customer identification watermarking today is one of the most promising application domains of digital watermarking. It enables to identify individual copies of otherwise indistinguishable digital copies. If done without any precautions, those individual watermarking are vulnerable to a number of specialized attacks based on an attacker collecting more than one individual copy. Fingerprinting algorithms are used to create watermarks robust against these attacks, but the resulting watermarks require a high payload of the watermarking algorithm. As soon as a large number of copies need to be distinguished and more than two copies are available to the attacker, the watermarks are too long to be embedded with current algorithms. We present a novel alternative method to fight attacks aimed at individual customer identification watermarks. This is achieved by modifying the watermarked material in a way collusion attacks produce artifacts which significantly reduce the perceived quality while they do not affect the quality of the individual copies.

Keywords: Watermarking, collusion attack, active fingerprinting, countermeasure

1 MOTIVATION

The expansion of digital networks all over the world allows extensive access on, and reuse of, digital content. Problems include unauthorized distribution and duplication, which might lead to loss of sales for the producers, distributors and creators. The Internet has become in many cases a trading place for illegal copies of movies, music and software. Thus, systems are required which help to protect digital media data. Digital watermarking in combination with active fingerprinting algorithms offers a solution to trace illegal copies by embedding individual customer identification codes transparently into the content.

1.1 The nature of digital copies

While digital content has a lot of advantages, with respect to copyright protection also a number of challenges came up:

1. Digital content can be copied without a loss of quality
2. Digital content can be copied and transmitted at high speed via Internet or Darknet [BEP+2002]
3. Digital copies are identical to each other as well as to the original

The issues (1) and (2) together enable an efficient distribution of high-quality illegal copies of e.g. movies or music. Issue (3) makes it very hard to identify the source of an illegal copy. If a web shop sells mp3 copies of musical pieces without additional security measures, each customer gets an identical copy of the song. If one of the customers distributes his copy, there is no way to identify him. Digital watermarking can help to solve this problem by embedding individual customer identification codes into the songs.

1.2 Watermarking vs. Digital Rights Management

Digital watermarking [CoMB02], [Dit00] invisibly embeds information into a cover with the help of a secret key. This information refers to the cover, and provides additional information about it accessible only by those who own the watermarking algorithm and the secret key. The most common application for digital watermarking is copy protection or customer tracing. Both can be seen as an alternative to digital rights management (DRM) (see e.g. [BuKü04]). On of the

most important difference between DRM and watermarking is the fact that content can escape from DRM environments based on cryptography and access control rendering it unprotected. One example for this is when the content is transmitted via an analogue channel. A well-designed digital watermark on the other hand stays in the content even after playback and analogue recording or strong mp3 compression. So when watermarked content is found in an illegal environment, copyright claims can be proven or original customers can be identified. Therefore DRM and watermarking differs in one fundamental aspect: In DRM the challenge is to keep material in a protected environment while in watermarking the protections stays within the content.

1.3 Attacks against digital watermarking for customer identification

While the concept of embedding individual customer identification codes into the media files has many advantages versus DRM, also serious challenges regarding its security exist. All individually marked copies of a media file differ at some small degree. These differences can be used by attackers to identify the position of the watermark and to attack it (see e.g. [Wu2005] or [SDS2002]). Even more simple, creating a copy based on mixing two or more individually marked copies may destroy or change the embedded information.

In this paper we discuss known approaches to counter the attacks on individual customer identification watermarking. First we describe watermarking and fingerprinting in section 2. In section 3 we introduce an alternative to using fingerprinting schemes by creating individual copies which will produce significant quality loss when used for attacks against individual watermarks. In section 4 we describe an example for PCM audio. In section 5 we provide test results. We conclude in section 6 presenting possible applications and identifying needs for further research.

2 STATE OF THE ART

In this section we briefly describe the challenge of collusion attacks against customer identification watermarks.

2.1 Digital watermarking

Digital watermarking can inseparably embed information into digital media such as images, movies or songs. These watermarks are similar to an unperceivable and not removable stamp added to the media. Digital watermarking enables to identify single copies of an original media file by e.g. embedding a transaction code or a simple continuous number into each copy. Whenever a copy is found, the watermark can be retrieved and the source can be identified by the embedded individual information. The watermarks survive format conversion, editing and also analogue copies such as microphone recordings in the case of audio data or printing and scanning for images.

With digital watermarking restrictive protection solutions are not necessary to secure digital media. The customer is free to use and consume the media data he bought in any way and on any device he likes. But if he passes the content into illegal environments and copies are found, he can be identified. This discourages misuse of the media files without reducing the usability for honest customers. Digital watermarking can be applied in various scenarios, like e.g. in online shop systems. Here digital media files can be individually marked at the time the customer decides to buy and download them. Another application is the individual marking of review copies sent to journalists, enabling to trace back possible pre-sales copies to their origin.

2.2 Customer identification by digital watermarking

To achieve customer identification directly connected to the copy of the media to be protected, embedding a robust watermark with a customer identification number - called ID from hereon - is a first solution. A simple example: The content provider wants to sell four copies A-D of an audio file to his customers. To be able to trace the source of an illegal distribution of the file, he embeds a different bit sequence in each copy. For customer A, he embeds “00”, for customer B “01”, for customer C “10” and for customer D “11”.

If he finds a copy of the audio file only sold to those four customers, he could try to retrieve the watermark from the copy. As he uses a robust watermarking algorithm, he is able to find the watermark and to identify the source. For example, he detects the watermark “01” and concludes the source is customer B. In the case of very strong attacks – which would reduce

the quality and make the copies less attractive – he may not be able to detect the watermark, but when the copy is of little value, he does not worry about this.

But due to watermarking characteristics, a much more dangerous situation can occur: Imagine A and D know each other and want to distribute illegal copies of the audio file. They know the file is watermarked and have a certain level of understanding regarding this technology. Therefore they compare both copies to each other, showing differences at certain positions. Knowing most watermarking algorithms can be confused in this way, they now mix both copies, creating a copy consisting of both customers' copies (referred to as *collusion attack* or *coalition attack*). This could render the watermarking algorithm unable to detect the watermarking information embedded. The illegal copy would be of good quality but still not traceable. Even worse, this could lead to accusing at a third customer who is innocent: For example, if copies of customers A (watermark "00") and D (watermark "11") are mixed, depending on the attacking algorithm and the watermarking method, it can happen that "01" or "10" is detected and B or C are accused..

2.2 Active Fingerprinting

Active fingerprinting is a more advanced method of embedding customer identification codes into a media files. While the watermarking algorithm may be same as in the previous section, a more sophisticated customer identification code, often called a fingerprint, is embedded. An active fingerprinting scheme therefore includes a watermarking algorithm and a fingerprinting algorithm. Well known fingerprinting methods are the Boneh-Shaw fingerprint and the Schwenk-Ueberberg fingerprint algorithm [BoSh95], [DBS+99]. Both algorithms offer the possibility to find the customers, which have committed the coalition attack. The drawback here is that the fingerprint's bit length becomes very long (i.e. thousands of bits) if a large number of copies needs to be distinguishable and if the fingerprints should be robust against a collusion attack of more than two pirates.. It is likely that robust audio watermarking does not provide sufficient payload for the embedding of such active fingerprint.

In [Lu2004] an approach to counter collusion attacks is discussed. It is based on embedding watermarking bits with different embedding strength. If a "0" is always embedded stronger than a "1", it is shown that in collusion attacks the result of the attack is not random anymore, but the resulting bit is usually the one embedded with more watermarking strength. While this is a certain improvement to the security of active fingerprinting, the required embedding strength may lead to a reduction of perceived quality of the content. More approaches have been discussed in the literature, like [SKH2005], which all aim at an optimization for resistant watermark embedding.

To summarize, while active fingerprinting is in theory a solution to collusion attacks, in current applications the usage of this technology is often impossible due to high payload demands and the risk of quality loss.

3 GENERAL SOLUTION

In this section we introduce an alternative solution to the challenge of collusion attacks. We improve the security of customer identification watermarks by discouraging from simple collusion attacks.

Our proposed solution to this problem is based on slight modifications of the media data the watermarks are embedded into. The nature of these changes is twofold: On the one hand, the modifications do not lead to perceived loss of quality in the individual copies. On the other hand, *after* running a collusion attack, a significant loss of quality occurs. Figure 1 illustrates the concept.

Slight modifications of the running length of the media data by adding or removing some small elements of the media data like pixels, samples or frames are one possible method to achieve this. The resulting differences between the individual copies caused by the modifications mask those differences caused by the embedding of the watermark. After adding or removing only a few elements, in most passages of the media files differences will be detected because elements other than those originally present at the current position are compared. It also produces significant artifacts when mixing two individually marked copies into each other.

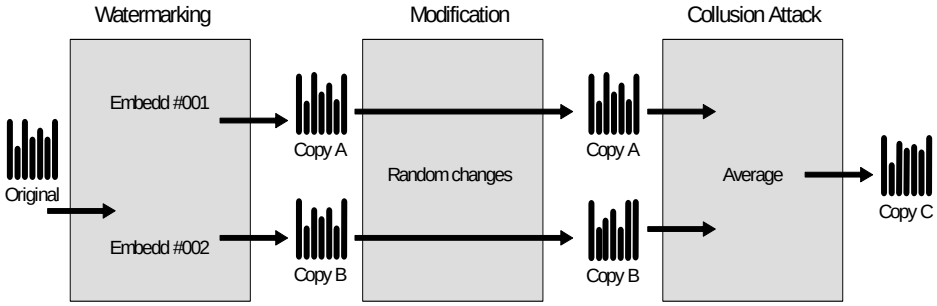

Figure 1: Countermeasure concept: Slight random changes modify copies A and B in such a way that after a collusion attack the resulting copy C is of low quality.

Another suitable modification is changing phase information leading to phase errors after attacks. These are only two examples, any change of the original content leading to minimal changes in a single cover but causing significant artifacts after collusion attacks can be utilized. It is also possible to use more than one method in parallel.

The requirements on a successful modification can be summarized as follows:

1. controls the modifications in such a way that no loss of quality occurs when using this mechanism on a single media copy
2. It creates modified media files which show significant quality losses after collusion attacks by one or more attackers
3. The modification uses the knowledge about the watermark embedding positions to identify suited positions in the file to add these modifications

4 DIGITAL AUDIO EXAMPLE

In this section we describe as an example a mechanism for PCM (Pulse Code Modulation) audio fingerprint protection.

We apply three methods for slight modification in the PCM data in our protection mechanism, which are inaudible in each individual copy but result in significant quality losses after collusion attacks:

Inserting Delays (Sample Addition and Removal): Samples are duplicated or deleted pseudo-randomly causing a short delay. We only select suited samples, which we identify by small sample values. These sample changes are not audible in an individual copy if suitable positions are selected. Additionally, it can easily be realized that the number of deleted and duplicated samples is identical to preserve total file length. When applying a collusion attack, samples from different, i.e. delayed, points of time are processed. This leads to changes in the perceived audio data, producing a flanging effect, similar to a hiss. Modifying an audio file multiple times in this way increases the effect.

Inversion of phase in time-domain (Sample flipping): Audio segments of short length are inverted, i.e the phase of the signal is switched by π=180°, so that collusion attacks lead to clicking effects or annihilation of the audio signals. A first approach is based on inverting/ flipping the signal in the time-domain (i.e. multiplying by -1). To prevent clicking artifacts, the inverted segment is faded with the original data by interpolation and the segment begins and ends at appropriate file positions, i.e. at minimal sound pressure levels (see Figure 2). To avoid flanging effects in stereo files the inversion must be applied on both stereo channels simultaneously. Thus, appropriate file positions must be identified in both channels strictly synchronously, prior to inversion.

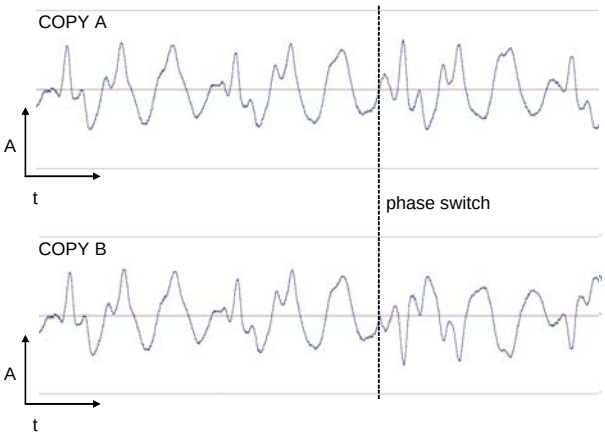

Figure 2: The two copies A and B are identical until a phase switch is applied

Inversion of phase in Fourier-domain (Phase shifting): To avoid the problem of identifying appropriate file positions we introduce a second approach based on FFT phase information. The files are processed in fixed-length audio frames (e.g. of 46ms play time, Hamming windowed) and transformed to the complex Fourier domain represented by magnitude and phase. From frame to frame with increasing play time we either increase or decrease all FFT phase components based on a secret We limit the relative phase shift between adjacent frames to a small amount, e.g. $\pi/100$ as the human ear is insensitive to such small relative variations of the FFT phase. After a sufficient number of frames the phase changes sum up to an amount relevant for discouraging collusion attacks (see Figure 53, example of 90° phase change)

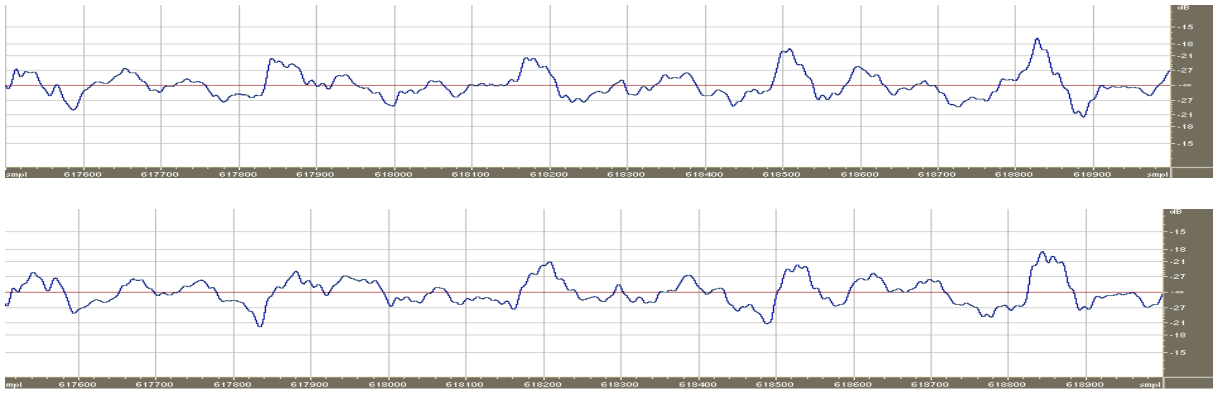

Figure 3: copy A (above) and copy B (below) are FFT phase shifted by $\pi/2$,

Both phase inversion approaches produce copies of an audio file where the phases are inverted to each other at some positions. When averaging the samples, inverse phases annihilate each other partly or completely, creating silent gaps in the attacked audio file (see Figure 4). When randomly selecting samples from two or more audio files as an attack, the inverse phases lead to phase errors and audible clicking artifacts.

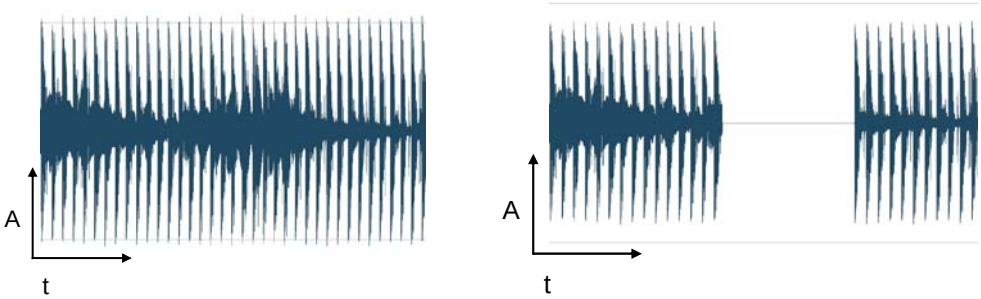

Figure 4: Individual copy before collusion attack (left) and after collusion attack (right)

The methods mentioned before can be combined into a system which individually modifies each copy of a fingerprinted audio file in a way that any collusion attack using one or more of these files will lead to audible artifacts rendering the attacked copy useless. While the fingerprint may not be readable from this attacked copy, its value is also very limited due to reduced sound quality. The modifications can be optimized in a way that they take into account the minimal amount of samples needed to embed a watermarking bit, leading to especially efficient implementations.

5 TEST RESULTS

This section provides experimental results with respect to perceptual sound quality of the protected files before and after collusion attacks. The assessment of the sound quality is expressed in terms of ODGs (*objective difference grades*, see table 1). The ODG values are calculated using the *Opera* system (by Opticom Inc.), an automated software-based analysis system for perceptual evaluation of audio quality.

ODG	perceptual quality
0	No perceived difference
-1	Different but not annoying
-2	Slightly annoying
-3	Annoying
-4	Very annoying

Table 1: ODG scale for sound quality assessment

The evaluation is based on a set of PCM audio files of various type, i.e. pop music, classical music, movie soundtracks, speech, noise, synthetic signals (97 files, 48.5 minutes total, 44.1 kHz, 16 bit, mono). The files were first watermarked using a robust audio watermarking approach presented in [Stei2003]. Two copies with different watermarks were created simulating two different customer fingerprints.

The watermark settings were selected to provide robustness against mp3@40 kBit/s and analog recording. As can be seen from Figure 5 the watermark does not lead to annoying distortions, i.e. an ODG of -1 or greater, except for a few outliers. Further analysis showed that the outliers, i.e. files with an ODG below -1, occur for audio material that contain segments with extreme sound properties. Those segments can easily be identified (e.g. very silent, strong staccato, dominant FFT peaks) and easily be skipped to provide optimal sound quality.

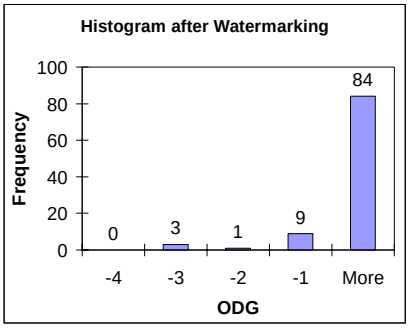

Figure 5: histogram of ODG values after watermarking; ODG mean value: -0.44;

We then applied the following standard collusion attacks on the watermarked audio files:

- Mean: Averaging the samples in both files
- Mix: Exchanging single samples interchangeably between the files
- Mosaic: Exchanging frames of 1024 samples between the files

The distribution of the ODG values for the attacked files show that for most of the files these attacks do not lead to annoying distortions (see Figure 6). That is, those collusion attacks are in fact well suited for an attack on fingerprint watermarks with respect to perceptual quality.

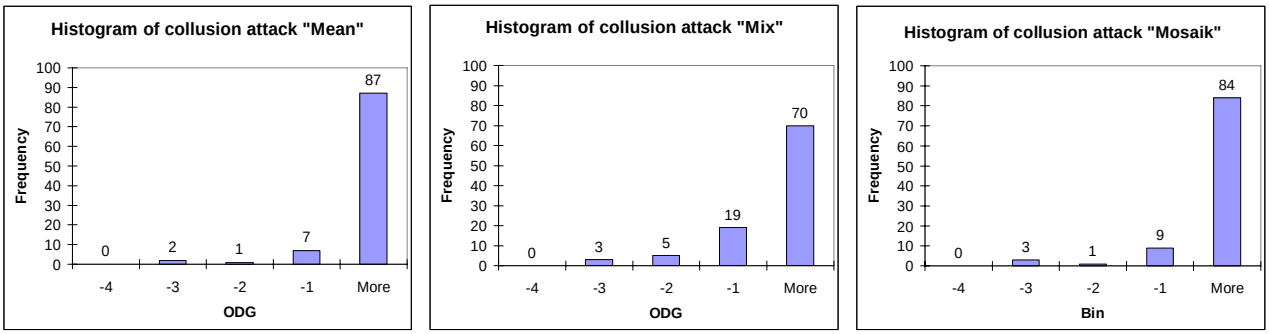

Figure 6: histogram of ODG values after collusion attacks on watermarked files; ODG mean values (f.l.t.r): -0.38, -0.71, -0.46

We then applied the modifications introduced in section 4 on the watermarked files. The modifications we introduced do not lead to annoying distortions for most of the files. Results for sample removing (Figure 7, left) and FFT phase shifting (Figure 7, right) are more promising than those for sample flipping (Figure 7, middle). Nevertheless, annoying distortions are introduced for some of the files which is suspect to further research for improvement.

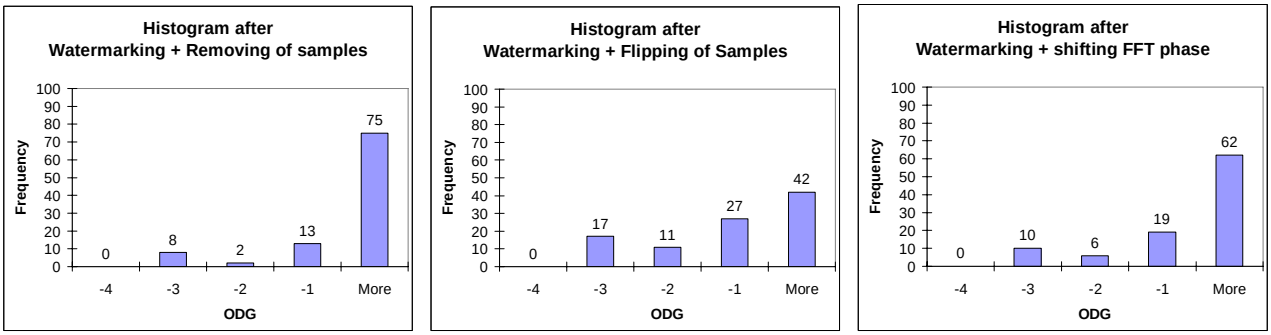

Figure 7: ODG values after applying modifications on watermarked files; ODG mean values (f.l.t.r): -0.71, -1.49, -1.05

In the next stage we applied the collusion attacks on the modified files (see Figure 8 to 10). For all modifications averaging of samples renders the audio files annoyingly distorted by introducing flanging effects or silent gaps (attack "Mean"). The reason for silent gaps is the annihilation of the signals at those positions where the phase difference is near odd multiples of π , When randomly exchanging samples or frames from two or more audio files as an attack, the inverse phases lead to phase errors and audible clicking artifacts as both signal do not "match" (attack "Mix"). The "Mosaic" attack renders a few files in acceptable quality, though. The reason is give by the fact that exchanging complete frames of 1024 samples only affects samples at the borders and leaves most of the samples unmodified. Especially for very silent files these discontinuities cannot be perceived.

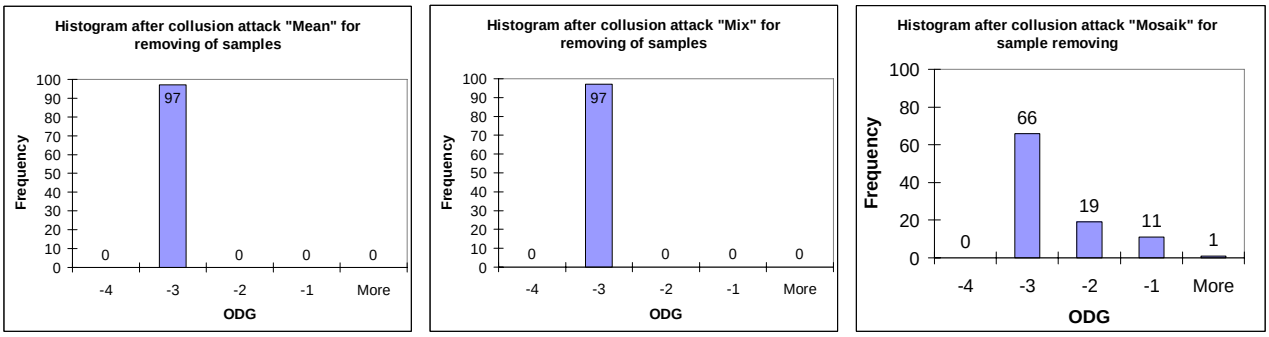

Figure 8: histogram of ODG values after collusion attacks for sample removing;

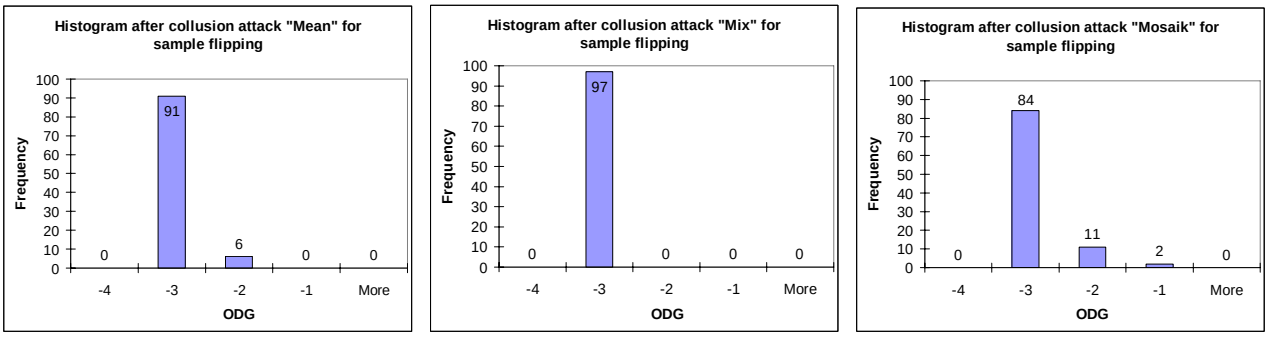

Figure 9: histogram of ODG values after collusion attacks for sample flipping

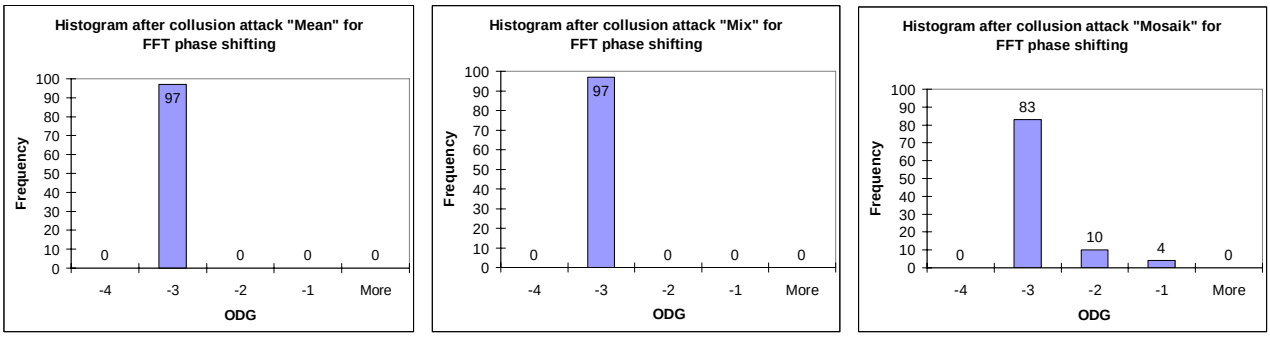

Figure 10: histogram of ODG values after collusion attacks for FFT phase shifting

The modifications we introduced give a few examples for an effective protection against collusion attacks on digital watermarking. As can be seen from the results our approaches meet the requirements we formulated in section 3. This provides a number of promising applications. For example, an increasing number of mp3 music online shops uses digital watermarking for the enforcement of copyrights and to discourage their end customers of passing on illegal copies via file sharing networks. Digital watermarking helps to solve this problem by embedding individual customer identification codes into the songs.

The modifications we introduced could be integrated in the watermarking process to discourage users, from applying collusion attacks. The same holds for individualized Audio CDs containing promotional content sent to music journalists or DJs for reviewing or for sample libraries for professional music production.

Possible counter attacks on our approaches can be developed based on the knowledge on our algorithms. For example, to circumvent/ reverse the phase modifications one could use one of the two copies as a master for the absolute value of the Fourier phase. All other modified copies can be modified with respect to phase by minimizing the energy of the difference between the master and the signal. This leads to an inversion of the original phase shifting modification. Nevertheless, this counter attack requires at least basic knowledge of audio signal processing etc. which is not available for most of the customers of music online shops. Thus, an increased security of the algorithm is highly required. Thus, an improvement of our approaches against attacks should be realized by combining our approaches with other kinds of modification.

6 SUMMARY AND CONCLUSION

Digital watermarking is an efficient mechanism for enforcement of copyrights and tracing distribution channels of illegal copies. It is increasingly integrated in online music shops etc. as a measure to discourage the customers of the online shop to pass on illegal copies via filesharing networks. This is done by embedding individual customer information as a digital watermark into the music files. Such customer watermark is vulnerable to various collusion attacks which are directed in destroying the watermark by mixing several copies. We presented an approach to discourage users from applying such collusion attacks. It is based on applying modifications of the phase information of the audio data. Our experimental results

show on the one hand that such modifications do not lead to audible distortions when applied on single copies. On the other hand, annoying distortions do occur when a number of modified copies are mixed by standard collusion attacks. Thus, our approaches can be used to increase the level of protection of music downloads. As the modifications can be inverted if several copies are available an improvement of the security is still highly required. That is, alternative approaches should be investigated with respect to their ability for preventing collusion attacks.

ACKNOWLEDGEMENTS

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

[BEP+2002] Peter Biddle, Paul England, Marcus Peinado, and Bryan Willman, *The Darknet and the Future of Content Distribution*, Microsoft Corporation, http://msl1.mit.edu/ESD10/docs/darknet5.pdf

[BS1995] Boneh, D., Shaw, J. (1995). *Collusion-Secure Fingerprinting for Digital Data*, Proc. CRYPTO'95, Springer LNCS 963, pp. 452-465

[BuKü04] Marcellus Buchheit and Rüdiger Kügler, *Secure Music Content Standard – Content Protection with CodeMeter*, WIBU-SYSTEMS AG, Germany, Virtual Goods 2004 - International Workshop for Technology, Economy, Social and Legal Aspects of Virtual Goods, May 27 - 29 2004 in Ilmenau, Germany.

[CoMB02] Cox, Miller, Bloom; *Digital Watermarking*, Academic Press, San Diego, USA, ISBN 1-55860-714-5, 2002

[DBS+1999] Dittmann, J., Behr, A., Stabenau, M., Schmitt, P., Schwenk, J., Ueberberg, J. (1999). *Combining digital Watermarks and collusion secure Fingerprints for digital Images*, Proceedings of SPIE Vol. 3657, [3657-51], Electronic Imaging '99, San Jose USA,

[Ditt00] Dittmann; *Digitale Wasserzeichen*, Springer Verlag, Berlin/ Heidelberg, ISBN 3-540-66661-3, 2000

[Lu2004] Chun-Shien Lu (edt.), *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property*, Idea Group Publishing, ISBN: 1-59140-275-1, 2004

[SDS2002] Steinebach, Dittmann, Saar; *Combined Fingerprinting Attacks against Digital Audio Watermarking: Methods, Results and Solutions*, Advanced Communications and Multimedia Security, Edited by Borka Jerman-Blazic and Tomaz Klobucar, IFIP TC6 / TC11 6th Joint Working Conference on Communications and Multimedia Security, September 26 - 27, 2002, Portoroz, Slovenia, Kluwer Academic Publishers, Boston / Dordrecht / London, S. 197 - 212, ISBN 1-4020-7206-6, 2002

[Stei2003] Steinebach, *Digitale Wasserzeichen für Audiodaten*, Shaker Verlag Aachen, ISBN 3832225072

[SKH2005] K. Su, D. Kundur and D. Hatzinakos, *Statistical Invisibility for Collusion-resistant Digital Video Watermarking*, *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 43-51, February 2005

[Wu2005] Yongdong Wu, *Linear Combination Collusion Attack and its Application on an Anti-Collusion Fingerprinting*, 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005