Content-based Message Authentication Coding for Audio Data

Sascha Zmudzinski, Martin Steinebach
Fraunhofer Institute for Secure Information Technology (SIT)
Darmstadt, Germany

sascha.zmudzinski@sit.fraunhofer.de, martin.steinebach@sit.fraunhofer.de

Abstract: Current systems and protocols based on cryptographic methods for integrity and authenticity verification of media data do not distinguish between legitimate signal transformation and malicious tampering that manipulates the content. Furthermore, they usually provide no localization or assessment of the relevance of such manipulations with respect to human perception or semantics. For the verification of digital audio data we present an algorithm for a robust message authentication code in the context of content fragile authentication watermarking. Therefore we introduce an extension of an existing audio fingerprinting approach with respect to security and synchronization with audio watermarking. The experimental results show that the proposed algorithm provides both a high level of distinction between perceptually different audio data and a high robustness against signal transformations that do not change the perceived information.

1 Motivation

Modern computer hardware, software and the Internet provide many ways of production, recording, post processing, editing, archiving and distribution of multimedia data. In many scenarios, digital audio data contains important information, for example, telephone calls to call center agencies (phone banking, emergency calls etc.), air traffic communication or historic documents. An example: With current audio editing software speech recordings of a statement *"I am not guilty"* or of the question *"Am I guilty?"* can easily be changed into the statement *"I am guilty!"* by moving, cropping or changing base frequency or pitch.

Those kinds of audio data are more and more recorded immediately on hard discs and CDs, transmitted in digital form (e.g. Voice over IP), and archived in various media formats, respectively. As the audio data can easily be modified or even manipulated, the *integrity* of the audio content and the *authenticity* of the data its origin are of special interest.

There exist various mechanisms for integrity and/or authenticity verification of digital data. The most important are checksums, error correction codes, digital signatures based on cryptographic hash functions or secret key based message authentication codes. Different approaches are given by speaker detection or digital forensics techniques that provide the detection of manipulations or to recognize the original source of a recording, namely the speaking person or the particular recording device, e.g. the digital camera, scanner or microphone etc. (see for example [AES05, LFG06, KODL07, GKWB07, WF07]).

Most of these mechanisms include a number of drawbacks regarding the properties of multimedia data: Many of the mechanisms mentioned above do not distinguish between inaudible signal transformations that can be tolerated in many scenarios (e.g. file format conversion) from malicious manipulations that actually change what a listener *hears* or even *understands* from the meaning of the recording. Neither multimedia specific properties of the file formats nor perceptive properties of the human auditory system are considered. Furthermore, the verification information is stored separated from the protected digital media itself: the verification is dependent on a security protocol and an existing infrastructure that provides the verification codes.

A promising approach to accept these challenges is given by authentication watermarking approaches, especially *content-fragile watermarking*. This approach, originally introduced by *Dittmann et al* [DSS99] is based on *content-based multimedia retrieval* methods [YI99] for audio data in combination with *digital audio watermarking* [ASW03, CS07].

In the remainder, we focus on the audio feature retrieval process involved in content-fragile watermarking, especially with regard to current audio fingerprinting methods and robust hashing. Therefore, we will introduce a number of necessary security extensions and adaptations of existing audio retrieval methods [HOK01b, HOK01a]. The challenge is to develop a suitable retrieval system that is able to detect audible modifications *and* that can be combined with audio watermarking. This includes crucial adaptations of the existing approaches such that the audio fingerprinting and the audio watermarking do not interfere with each other in a content-fragile watermarking system.

Experimental results show that our fingerprinting approach features a high distinction between tolerable changes and malicious attacks on the audio data. By design, it provides a high degree of non-interference with the audio watermarking process involved.

2 Introduction

This chapter explains how content-based audio feature retrieval methods can serve for the detection of audible modifications of audio data in an authentication watermarking system. Therefore, we explain current authentication watermarking systems in general and the involved audio retrieval approaches, namely audio fingerprinting and robust hashing.

2.1 Content-based Authentication Watermarking

Digital watermarking is a technique for embedding additional information directly into an audio stream using a secret key. The embedding is done by applying inaudible signal transformations to the audio signal. These signal transformations represent the embedded *watermark message* and the watermark can be detected and retrieved at any later date.

Most watermarking approaches are based on adding a pseudo-noise signal in the time domain or in a transformed domain (e.g. representation in Fourier, DCT or Wavelet coefficients) into the audio stream which later can be detected by filtering, comparison to the original, correlation or several statistical methods [CMB02, Dit00].

First audio watermarking approaches for data authentication were given by *fragile watermark* and *semi-fragile watermark* embedding schemes which are, by design, very sensitive to *any* kind of modification of the watermarked data. Here, the watermark can be seen a *digital seal* that is broken at modified parts of the protected audio file [WK02, DSP03, LC04, YLSP05, Zha03, PTW07].

In the remainder of this paper we will focus on an approach for content-based authentication watermarking. It was originally introduced by *Dittmann et al* [DSS99]. The design relies on the extraction of perceptually relevant audio features. Detection of malicious manipulations is done on the basis of changes of these features. The verification is done in two stages:

1. **Protection Stage** The audio file is divided in segments and from every audio segment perceptively relevant audio features are extracted. This is done by means of content-based retrieval methods such as audio fingerprinting and robust hashing. The extracted audio features are then embedded as a robust digital watermark into the audio data, see figure 1.

2. **Verification Stage** The embedded *original* audio features are retrieved at any later date and can be compared with the *current* audio features, see figure 2. As the original features are provided by a robust watermark they can be retrieved even after the protected file was subject to file format changes or DA/AD-conversion etc.

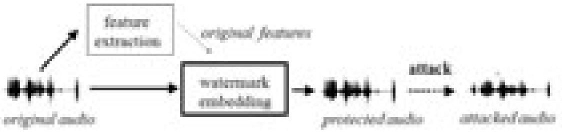

Figure 1: Protection Stage

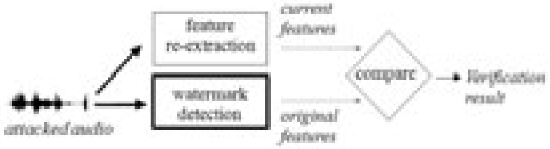

Figure 2: Verification Stage

The challenge of a fingerprint based verification framework is to extract a highly compact representation of the audio data that provides a high distinction between tolerable transformations and malicious manipulations [CGB^{+}02]. Especially, the watermark embedding process in the protection stage must not interfere with the audio feature extraction in the verification stage.

This *content-fragile watermarking* concept is also referred to as *mixed watermarking-fingerprinting* [GCdG$^+$02], *semi-fragile signature watermarking* [FKK04] or *self-embedding* [YH04]. A number of approaches can be found for audio data in the literature, for example feature extraction based on short-time Fourier and time domain statistics [SD03], based on signal statistics in the spectrum after a modified complex lapped transform [MV01], based on psychoacoustic modeling [RM02], extraction of an "alphabet" of acoustic events by means of Mel-frequency cepstral coefficients (MFCCs) and hidden Markov modeling [GCdG$^+$02] or based on multiple embedding [GMS06]. Especially for speech data there are approaches based on the extraction of CELP speech coding coefficients, for example by means of G.729 speech codecs, using the pitch of the speech signal or parameters modeling the shape of the vocal tract [WK01b, WK01a] or based on GSM 610 speech codecs for the protection of voice-over-IP transmission [YH04].

2.2 Robust Hash Functions and Audio Fingerprinting

2.2.1 Definition

A common technique for the purpose of data authentication are cryptographic hash functions. A hash function is a mathematical function that maps a variable length input message to an output message digest of fixed length [Sch96].

Cryptographic hash algorithms by design are extremely sensitive to changes in the input data, that is, even a one bit change leads to a totally different hash value. For audio data this high sensitivity is inappropriate in many scenarios as for example changes of least significant bits in PCM raw audio data on an audio CD will be inaudible for an average human listener.

Thus, for multimedia applications a number of specially designed audio hash algorithms have been introduced that are tolerant against inaudible or moderate transformations of audio signals. Originally, these approaches were designed to search audio data stored in an audio database e.g. to recognize pieces of music that sound *similar* to a given one. For this, these so called *audio fingerprinting* algorithms extract acoustic features from the data that are relevant with respect to the human perception.

As audio fingerprinting approaches allow to recognize acoustic events that sound similar, it is very likely that they can also detect that pieces of audio have sound *dissimilar* from each other. Thus, we will investigate, how fingerprinting can be used to detect audible modifications of audio data for integrity verification.

2.2.2 Requirements

Under more strict conditions, fingerprinting algorithms are referred to as *robust hash* or *perceptual hash* in the literature. In order for a fingerprint algorithm to serve as a *robust* *audio hash* and to integrate it in a content-fragile watermarking system it must meet a number of requirements. The most important are [MV01]:

- Distinction: For the detection of perceptually relevant transformations of the signal we require that the hash values should be different for perceptually different audio signal. This requirement allows the detection of malicious tampering of the signal.
- Robustness: The robust hash must provide the invariance for perceptually similar audio data. The hash should be robust against signal transformations that do not affect the perceptual quality of the data. In addition, the hash must be robust against the distortion introduced by watermarking to prevent false alarms.
- Security: The features must survive attacks that are directly aimed at the feature extraction. Therefore, the robust hash values should be equally distributed among all possible pieces of audio data. Furthermore, the hash values of two perceptually different audio signals should be statistically independent. These requirements should discourage the attacker from generating a hash collision by keeping the effort for an attacker as high as possible.

It should be noted that the security and distinction requirements are similar for cryptographic hash functions. The robustness requirement is specific for robust hash functions for multimedia data.

3 Proposed Authentication Watermarking System

In this section we introduce an authentication watermarking system following the content-fragile approach. It is based on a combination of robust watermarking and secure robust audio hashing, which we will both explain in the following sections.

3.1 Watermark Embedding

For embedding the robust audio hash we use a blind spread spectrum/ patchwork watermarking approach previously presented by us [Ste03]. In the original algorithm, the embedding is basically done in the Fourier domain by modifying the FFT magnitude coefficients. Here, the audio signal is first divided into non overlapping frames and the Fourier spectrum is calculated for each frame. Into each frame one bit of watermarking information is embedded and subsequent bits of the watermark message are embedded into the following frames. Here, dependent on the secret watermark key K_1, a pseudo-randomly selected subset of FFT coefficients is split in two groups A and B. Dependent on the watermark information bit ”one” or ”zero”, the coefficients in one of the two groups are decreased while those in the other group are increased, and vice versa. These modifications enforce a deviation between the mean magnitude in groups A and B.

The degree of modification is controlled by a psychoacoustic model to provide maximum robustness and transparency by considering the perceptual properties of the human auditory system.

3.2 Fingerprinting Algorithm by *Haitsma et al.*

Our approach for message authentication coding is an extension to an existing audio fingerprinting scheme introduced by *Haitsma et al* [HOK01b, HOK01a]. This robust feature extraction uses a time-frequency analysis of the Fourier magnitude coefficients of the audio signal. A similar approach for video data based on average block luminance values has been introduced by the same authors [OKH01].

In the original algorithm, the audio signal is digitally represented by PCM samples and it is divided into overlapping frames $\vec{x}_t$ containing L PCM samples where t denotes the time-step of the frame.

Then, the energy differences of adjacent energy bands k and $k+1$ at a given time t are compared to those with the same band indices in the following time-step $t+1$ as follows:

$$d(k,t) := E(k,t) - E(k,t+1) - [E(k+1,t) - E(k+1,t+1)] \tag{1}$$

where $k = 1, 2, ..., K$. The decision about the fingerprint bits $H(k,t)$ is then given by the sign of those comparisons:

$$H(k,t) = \begin{cases} 1 & \text{if } d(k,t) \geq 0 \\ 0 & \text{if } d(k,t) < 0 \end{cases}$$

This extracted audio feature provides a high level of robustness against encoding to lossy compression,re-sampling, filtering, dynamics compression, noise addition and analog tape recording.

3.3 Proposed Robust Message Authentication Code Algorithm

An analysis of the security of the algorithm, i.e. attacks aimed on the robust hash under knowledge of the algorithm was not given by the authors [HOK01b, HOK01a] as the security in many scenarios is not a relevant requirement for audio retrieval purposes. For the application in an integrity verification scenario, we introduce a number of extensions with respect to security.

3.3.1 Key Dependent Feature Selection

For integrity protection, security must be provided with respect to the requirements given in section 2.2.2 in a way that an attacker can not generate a hash collision.

As pointed out by *Fridrich et al.* the security must be provided by a key dependent feature extraction [FG00, ASW05]. In the original approach (see equation 1) for calculation of every fingerprint bit only energy coefficients from consecutive bands n and $n+1$ and consecutive time-steps t and $t+1$ were used.

Here, we introduce that each code will now be extracted from an audio segment consisting of L consecutive frames, typically a few hundred (representing a number of seconds playing time). Then, we introduce to derive the hash bits from four coefficients at different time-steps t_1, t_2, t_3, t_4 and band indices k_1, k_2, k_3, k_4 in the time-frequency domain. The selection shall be *pseudo-randomly* dependent on a secret key K_2. The security is introduced as it is obscure to an attacker which bands are selected for the hash bit calculation (and which are not).

Because the fingerprint we introduce is dependent on a shared secret K_2 it can be regarded as a message authentication code (MAC). As we will show in the following chapter, it withstands a number of signal transformations. Because of this *robustness* we will denote it as *rMAC* in the remainder of this article.

3.3.2 Normalization of the FFT spectrum

For real-world audio data, for example music or speech recordings, the FFT magnitude coefficients are neither equally distributed nor statistically independent. For example, in most speech recordings, the coefficients related to lower frequencies (e.g. below 1000 Hz) usually have a higher mean energy than coefficients related to higher frequencies. Thus, the key dependent selection of bands for fingerprint extraction raises problems with respect to the security requirements listed in section 2.2.2. For example, if such low frequency coefficient is selected by the rMAC key as the first summand in equation (1) to calculate a particular hash bit, that coefficient will dominate the sum and causing the single hash bit to be more likely a ”one” than a ”zero” *for any kind of data*. Thus, the complete rMAC would not be equally distributed and rMACs from different music segments would not be independent, allowing systematic security attacks on the rMAC.

Therefore, we introduce a *normalization* of the FFT spectrum as follows: We regard all FFT coefficients $e(n,t)$ of a given band index n of all time-steps $t = \{1, ..., L\}$ in an audio segment as a random variable with expectation μ_n and variance σ_n^2.

We introduce a normalization given as follows:

$$e'(n,t) := \frac{e(n,t) - m(n)}{s(n)} \tag{2}$$

where $m(n)$ and $s^2(n)$ denote the empirical mean and variance. As can be seen from the linearity of the variance, the transformed quantities $e'(n,t)$ of a given band index n have mean 0 and variance 1. We will show in the experimental evaluation that this causes the single hash bits to be equally distributed.

3.3.3 Synchronization With Watermark Embedding

As pointed out in section 3.1, in a content-fragile watermarking system it is essential that the original audio data distortions caused by embedding the rMAC as a content-fragile watermark do not significantly change the values of the rMAC. Therefore, we will *synchronize* the process of rMAC feature extraction and the robust watermarking scheme involved.

As described in section 3.1 the embedding is done by modifying a pseudo-randomly selected set of Fourier magnitude coefficients. In order to ensure that the rMAC features are not affected by the watermark embedding, only those magnitude coefficients will be used for rMAC extraction if they had *not* been selected by the watermark key K_1 for watermark embedding. Thus the extraction algorithm is not only dependent on the secret rMAC key K_2 but also on the given secret watermarking key K_1: In other words, both watermark embedding and rMAC extraction are based on a pseudo-random selection of Fourier coefficients in the time-frequency representation that show minimal overlap.

Then, the sign comparison is done out on the normalized FFT coefficients selected as described above:

$$d'_k := e'(n_1, t_1) - e'(n_2, t_2) - (e'(n_3, t_3) - e'(n_4, t_4)) \tag{3}$$

where $n_{1,2,3,4} \in \{1, 2, ..., N\}, t_{1,2,3,4} \in \{1, 2, ...L\}$. In order to provide a sufficient degree of security against brute force attacks, a minimum of 128 rMAC bits will be extracted, i.e. $k = 1...128$.

As we will show later in the experimental results the related modified fingerprint

$$H'_k = \begin{cases} 1 & \text{if } d'_k \geq 0 \\ 0 & \text{if } d'_k < 0 \end{cases}$$

meets the requirements for a robust hash function listed in section 2.2.2 with the additional property of key-dependence.

4 Experimental Evaluation

In this chapter we will present empirical test results to evaluate the proposed extension and adaptation.

4.1 Test Data and Simulated Audio Attacks

The detection success was tested on a set of PCM audio files of different genre and sound quality, i.e. pop music, classical music, audio books, talk radio, movie sound tracks, synthetic sounds etc. (44.1 kHz, 16 bit, mono, total length 3.9 hours). The audio files were divided into segments of 20 seconds and an 128 bit rMAC was extracted from each segment. The frequency range we selected was from 300 Hz to 7000 Hz as this is the part of the spectrum where the human ear is most sensitive with respect to the hearing threshold in silence [ZF90, Moo95]. The FFT frame size is 2048 samples which provides a sufficient frequency resolution.

4.2 Results for Distinction Performance

In a training stage different kinds of content-preserving and content-changing attacks on the hashed audio data were applied:

4.2.1 Robustness to Content-preserving Transformations

We investigated the behavior of the rMAC extraction with respect to increasing distortion of the protected audio signal caused by lossy compression. Therefore, we compared the rMAC of the original audio segments to those of mp3 re-encoded versions at different mp3 bitrates (*Lame 3.97*, average bitrate ABR). As can be seen from the histograms of the Hamming distance between original and attacked audio files, the rMAC shows a good robustness at 160 kbit/s (see figure 3) as the average Hamming distance is 2.9, representing an bit error rate (BER) of $2.9/128 = 0.023$. That is, for perceptually similar files the respective rMACs are actually very similar, as well. Closer analysis of the dependence on the bitrate shows, that the BER significantly increases for bitrates below 32 kBit/s mono (see figure 4). This is mainly given by noticeable artifacts and the resampling done by the mp3 encoder: for bitrates below 32 kBit/s, the audio data is temporarily low-pass filtered with a cut-off frequency at 5.5 kHz and then resampled to 11kHz. As the rMAC is extracted from a range of 0.3 kHz to 7 kHz, many rMAC bits d'_k are affected. Compared

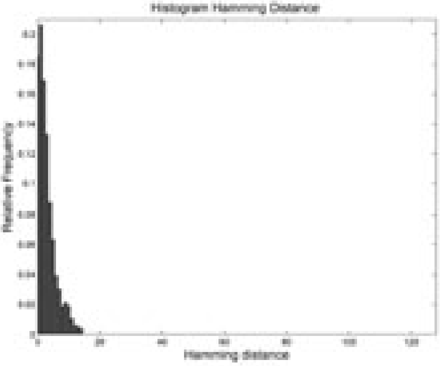

Figure 3: Histogram of rMAC hamming distance for mp3@160 kBit/s mono; mean=2.9

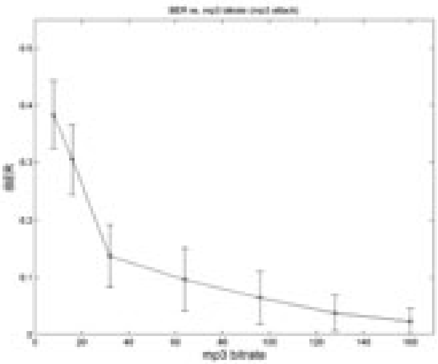

Figure 4: mean BER vs. mp3 compression (error bars: 1σ interval)

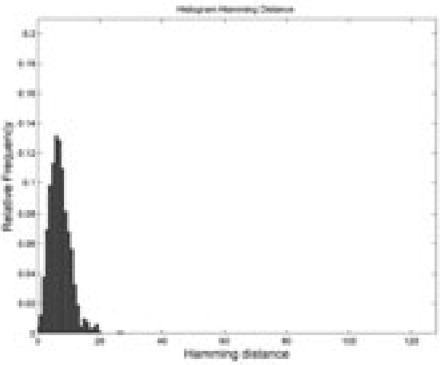

Figure 5: Histogram of rMAC hamming distance after watermark embedding; mean=7.1

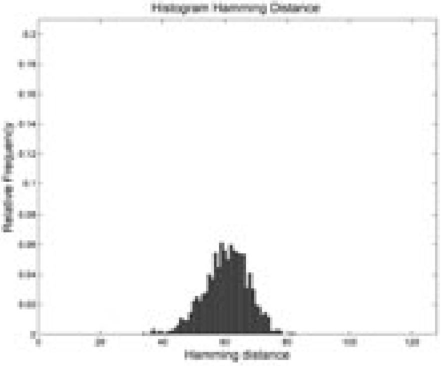

Figure 6: Histogram of rMAC hamming distance for shifting attack; mean=60.5, stddev=7.1

to the results in [HOK01b, HOK01a] our algorithm provides a robustness at the same order of magnitude with respect to bit error rate. This implies also that the key-dependent extraction does not necessarily lead to a decreased robustness.

The files were also marked with a synchronized robust audio watermark algorithm as explained in the previous chapter. The embedding strength was chosen in such a way that the distortions introduced do not exceed the masking threshold by more than 3 dB which is hardly noticeable for an average listener (see [Ste03]). As can be seen from figure 5 the watermark embedding shows very low rMAC bit error rates. This is provided by the rMAC extraction compliant with the embedding strategy as described in section 3.3.3. This demonstrates that our rMAC algorithm is well suited in an *content-fragile watermarking* system.

4.2.2 Sensitivity to Content-changing Attacks

Here, we evaluated the behavior when the audio data is re-encoded at mp3 coding at 16 kbit/s, which imposes perceivable quality degradation. As we can see from figure 4 the rMAC successfully indicates such distortions. Furthermore, we made a comparison of the original audio with a time-shifted copy of the same file (time shift 10 seconds): For speech data this simulates replacing parts of the audio by parts from the same speaker and with similar background noise but different semantic content. This provides an analysis of the extracted hash with respect to distinction between perceptually different audio frames after ”malicious” attacks. For the ”shifting” attack the distribution of the Hamming distance between original rMAC and attacked rMAC is centered around the average of 60.4 ± 7.1 bit errors, thus an bit error rate (BER) of 0.45 ± 0.06, see figure 6. Thus, perceptually different audio frames have identical hash bits only by coincidence. It should be noted that the shape of the histogram and its center near an BER of 0.5 is as it could be expected from the Hamming distance of uncorrelated binary random vectors.

4.2.3 Overall Distinction Performance

To evaluate the overall distinction performance we combined the results listed above (see sections 4.2.1 and 4.2.2). If we define the inaudible watermark embedding and lossy mp3 compression at 160 kBit/s to be ”tolerable” transformations, and define lossy mp3 compression at 16 kBit/s and shifting the sequence as ”malicious” attacks. Now, we obtain the *false positive rates (FPR)* and *false negative rates (FNR)* as given in figure 7 and table 1

threshold	0	1	2	...	18	19	20	21	...	128
FPR	1	0.9051	0.7959	...	0.0060	0.0042	0.0011	0.0008	...	0
FNR	0	0.000	0.000	...	0.0015	0.0026	0.0038	0.0045	...	1

Table 1: Intersection of false positives (FPR) and false negatives (FNR)

The intersection of the FPR and FNR curve shows a equal error rate (EER) of 0.3% if we set the threshold for the hamming distance between original rMAC and atacked rMAC between 19 and 20 bit see table 1).

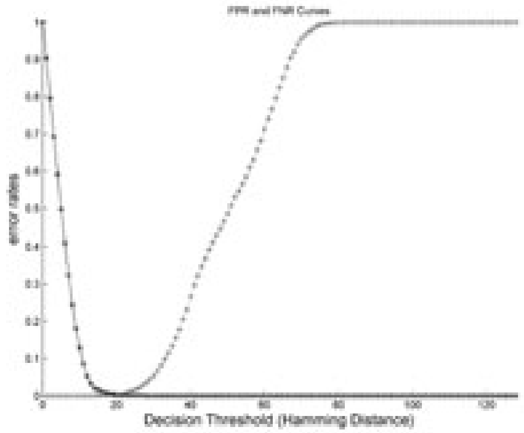

Figure 7: FPR: false positives (asterisk); FNR: false negatives (crosses)

4.3 Results for Key Dependence

To demonstrate the key-dependence of the algorithm we compared the rMAC using two different rMAC keys. The results in figure 8 show that the compared single rMAC bits are identical only by coincidence and most likely 50% of the rMAC bits are flipped. The shape of histogram is as it can be expected from the Hamming distance of uncorrelated binary random vectors. That is, without knowledge of the rMAC key the protected audio can not be verified as authentic.

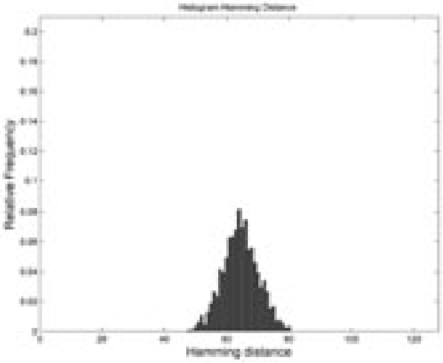

Figure 8: Histogram of rMAC hamming distance for different rMAC keys; mean=64.4

5 Conclusion and Future Work

This paper introduces an approach for a content-based message authentication code for audio data in the context of digital watermarking. Our aim is to detect such tampering that changes the cover audio data perceivably.

Our approach is based on an audio fingerprinting algorithm by *Haitsma et al* [HOK01b] [HOK01a]. We presented an extension with respect to security of the algorithm. Especially, we included the usage of a secret key, thus, extending the original fingerprinting algorithm to a true robust hash and/or robust message authentication coding (rMAC), respectively. Furthermore, we introduce a synchronization strategy to provide a watermarking compliant rMAC extraction and embedding.

Unlike cryptographic hash functions our approach shows a high robustness against signal transformations that do not perceivably change the audio data. At the same time, experimental results show a good distinction between tampering and non-tampering signal changes.

The robustness will be improved by an adaptive quantization of the FFT coefficients and subsequent lossless compression [ASW06]. The perception-based distinction performance will be furthermore improved by integrating psychoacoustic properties of the human auditory system (e.g. frequency and temporal masking) by means of filtering the audio signal prior to the rMAC extraction or by using psycho acoustic properties on selection of relevant/ irrelevant energy coefficients.

Because of its robustness, security and distinction performance our proposed algorithm can provide perception-based audio audio authentication in many scenarios, for example in protecting archives preserving the cultural heritage or other sensitive audio data. It can furthermore improve the security of fingerprint based filter methods for peer-to-peer networks or broadcast monitoring systems as discussed e.g. in [HOK01b, HOK01a].

References

[AES05] AES. *Proceedings of the 26th International AES Conference: Audio Forensics in the Digital Age, Denver, USA, 2005 July 7-9*, Proceedings of AES. AES, 2005.

[ASW03] Michael Arnold, Martin Schmucker, and Stephen D. Wolthusen. *Techniques and Applications of Digital Watermarking and Content Protection*. Artech House, Inc., Norwood, MA, USA, 2003.

[ASW05] Y. Mao A. Swaminathan and M. Wu. Security of Feature Extraction in Image Hashing. In *IEEE Conference on Acoustic, Speech and Signal Processing (ICASSP), Philadelphia, PA, March 2005.*, 2005.

[ASW06] Y. Mao A. Swaminathan and M. Wu. Robust and Secure Image Hashing. *IEEE Transactions on Image Forensics and Security*, 1, June 2006.

[CGB+02] P. Cano, E. Gómez, E. Batlle, L. de Gomes, and M Bonnet. Audio Fingerprinting: Concepts and Applications. In *2002 International Conference on Fuzzy Systems Knowledge Discovery (FSKD'02), Singapore, November 2002*, 2002.

[CMB02] Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom. *Digital Watermarking*. The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufmann Publishers, 2002. ISBN 1-55860-714-5.

[CS07] Nedeljko Cvejic and Tapio Seppnen, editors. *Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks*. Information Science Reference, Hershey, NY, USA, July 2007.

[Dit00] Jana Dittmann. *Digitale Wasserzeichen*. Springer Verlag Berlin Heidelberg, April 2000.

[DSP03] Dittmann, Steinebach, and Pharow. Neue Perspektiven zur Manipulationserkennung in digitalen Medien Elektronische Signaturen kombiniert mit invertierbaren Wasserzeichen. In *DACH Security IT Security & IT Management, Patrick Horster (Eds.), Proceeding DACH Security, DACH Security Bestandsaufnahme und Perspektiven*, 2003.

[DSS99] Jana Dittmann, Arnd Steinmetz, and Ralf Steinmetz. Content-Based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS '99) vol. 2*, page 209, Washington, DC, USA, 1999. IEEE Computer Society.

[FG00] J. Fridrich and M. Goljan. Robust Hash Functions for Digital Watermarking. In *Proc. ITCC 2000, Las Vegas, Nevada, March 27-29, 2000, pp. 173178*, 2000.

[FKK04] Chuhong Fei, Deepa Kundur, and Raymond H. Kwong. Analysis and design of authentication watermarking. In Edward J. Delp and Ping Wah Wong, editors, *Security, Steganography, and Watermarking of Multimedia Contents*, volume 5306 of *Proceedings of SPIE*, pages 760–771. SPIE, 2004.

[GCdG+02] E. Gomez, P. Cano, L. de Gomes, E. Batlle, and M. Bonnet. Mixed Watermarking-Fingerprinting Approach for Integrity Verification of Audio Recordings. In *International Telecommunications Symposium ITS2002, Natal, Brazil*, 2002.

[GKWB07] Thomas Gloe, Matthias Kirchner, Antje Winkler, and Rainer Böhme. Can we trust digital image forensics? In *ACM MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 78–86, New York, NY, USA, 2007. ACM.

[GMS06] Michael Gulbis, Erika Muller, and Martin Steinebach. Audio Integrity Protection and Falsification Estimation by Embedding Multiple Watermarks. In *2006 International Conference on Intelligent Information Hiding and Multimedia*, volume 0, pages 469–472, Los Alamitos, CA, USA, 2006. IEEE Computer Society.

[HOK01a] J.A. Haitsma, J.C. Oostveen, and A.A.C. Kalker. A Highly Robust Audio Fingerprinting System. In *2nd International Symposium of Music Information Retrieval (ISMIR 2001), Indiana University, Bloomington, Indiana, USA October 15-17, 2001*, Online proceeding only: http://ismir2001.ismir.net/proceedings.html (link verified: 2004-05-10), 2001.

[HOK01b] J.A. Haitsma, J.C. Oostveen, and A.A.C. Kalker. Robust Audio hashing for content identification. In *Content based multimedia Indexing (CBMI) 2001, Brescia Italy*, 2001.

[KODL07] Christian Kraetzer, Andrea Oermann, Jana Dittmann, and Andreas Lang. Digital audio forensics: a first practical evaluation on microphone and environment classification. In *MM&Sec '07: Proceedings of the 9th workshop on Multimedia & security*, pages 63–74, New York, NY, USA, 2007. ACM.

[LC04] Chia-Hsiung Liu and O.T.-C Chen. Fragile speech watermarking scheme with recovering speech contents. *The 2004 47th Midwest Symposium on Circuits and Systems, 2004. MWSCAS '04*, 2:165–168, July 2004.

[LFG06] J. Lukáš, J. Fridrich, and M. Goljan. Detecting digital image forgeries using sensor pattern noise. In E. J. Delp, III and P. W. Wong, editors, *Security, Steganography, and Watermarking of Multimedia Contents VIII. Edited by Delp, Edward J., III; Wong, Ping Wah. Proceedings of the SPIE, Volume 6072, pp. 362-372 (2006).*, volume 6072 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 362–372, February 2006.

[Moo95] B. C. J. Moore, editor. *Hearing – Handbook of Perception and Cognition*, volume 1. Academic Press, New York, 1995.

[MV01] M. Kıvanç Mıçak and Ramarathnam Venkatesan. A Perceptual Audio Hashing Algorithm: A Tool For Robust Audio Identification and Information Hiding. In I.S. Moskowitz, editor, *Lecture Notes in Computer Science, 4th International Workshop Information Hiding, IH 2001, Pittsburgh, PA, USA, April 25-27, 2001, ISBN 3540427333*, volume 2137, August 2001.

[OKH01] Job Oostveen, Ton Kalker, and Jaap Haitsma. Visual hashing of video: application and techniques. In Ping Wah Wong and Edward P. Delp III, editors, *IS&T/SPIE 13th Int. Symposium on Electronic Imaging San Jose, Security and Watermakring of Multimedia Contents, CA, USA, Jan. 2001*, volume 4314. SPIE–The International Society for Optical Engeneering, 2001.

[PTW07] Chang-Mok Park, Devinder Thapaa, and Gi-Nam Wang. Speech authentication system using digital watermarking and pattern recovery. *Pattern Recognition Letters*, 28:931–938, June 2007.

[RM02] R. Radhakrishnan and N. D. Memon. Audio content authentication based on psychoacoustic model. In *Proc. SPIE Vol. 4675, p. 110-117, Security and Watermarking of Multimedia Contents IV, Edward J. Delp; Ping W. Wong; Eds.*, pages 110–117, April 2002.

[Sch96] Bruce W. Schneier. *Applied Cryptography*, chapter 18. Wiley, 2 edition, 1996.

[SD03] Martin Steinebach and Jana Dittmann. Watermarking-based Digital Audio Data Authentication. *EURASIP Journal on Applied Signal Processing*, 10:1001–1015, 2003.

[Ste03] Martin Steinebach. *Digitale Wasserzeichen fuer Audiodaten*. PhD thesis, TU Darmstadt, Germany, 2003. ISBN 3832225072.

[WF07] W. Wang and H. Farid. Exposing Digital Forgeries in Video by Detecting Duplication. In *ACM Multimedia and Security Workshop*, Dallas, TX, 2007.

[WK01a] Chung-Ping Wu and C. C. Jay Kuo. Speech Content Authentication Integrated With Celp Speech Coders. In *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo, ICME 2001, August 22-25, 2001, Tokyo, Japan*, 2001.

[WK01b] Chung-Ping Wu and C.-C. Jay Kuo. Speech Content Integrity Verification Integrated with ITU G.723.1 Speech Coding. *itcc*, 00:0680, 2001.

[WK02] Chung-Ping Wu and C.-C. Jay Kuo. Fragile speech watermarking based on exponential scale quantization for tamper detection. In *Proceedings of the International Conference Acoustics Speech and Signal Processing (ICASSP 2001), Orlando, Florida, May 13–17, 2002*, volume 4, pages 2205–3308. IEEE, 2002.

[YH04] Song Yuan and Sorin A. Huss. Audio watermarking algorithm for real-time speech integrity and authentication. In *MM&Sec '04: Proceedings of the 2004 multimedia and security workshop on Multimedia and security*, pages 220–226, New York, NY, USA, 2004. ACM Press.

[YI99] A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, 1999.

[YLSP05] B. Yan, Z.-M. Lu, S.-H Sun, and J.-S. Pan. Speech Authentication by Semi-fragile Watermarking. *Lecture Notes in Computer Science*, 3683:497–504, 2005.

[ZF90] E. Zwicker and H. Fastl. *Psychoacustics – Facts and Models*. Springer, Berlin, 1990.

[Zha03] Ronghui Tu; Jiying Zhao. A novel semi-fragile audio watermarking scheme. *Proceedings of The 2nd IEEE Internatioal Workshop on Haptic, Audio and Visual Environments and Their Applications, 2003. HAVE 2003*, pages 89–94, 20-21 Sept. 2003.