Expert Systems With Applications 91 (2018) 211–222

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Authenticity verification of audio signals based on fragile watermarking for audio forensics

Diego Renza*, Dora M. Ballesteros L., Camilo Lemus

Universidad Militar Nueva Granada, Carrera 11 No. 101-80, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 26 August 2016
Revised 14 August 2017
Accepted 1 September 2017
Available online 6 September 2017

Keywords:
Audio authenticity
Fragile watermarking
Tampering detection
Imperceptibility
Audio forensics

ABSTRACT

This paper presents a new fragile watermarking method for digital audio authenticity for audio forensics purposes. The aim is to verify if an audio proof has been tampered and to locate the segments where the signal was modified. Our proposal is based on an embedding process of a text that is encoded through OVSF (Orthogonal Variable Spreading Factor) codes and spread into the entire signal using automatic adjustment. Several tests were performed in order to quantify the accuracy and the reliability of the tampering detection against four classical attacks (cropping, replacement, additive noise and amplitude reduction) by using kappa index, sensitivity and specificity. It was demonstrated that even if a small number of samples is modified, the system correctly labels the audio proof as manipulated, and locates both the start and end of the manipulation; the kappa index (reliability) is around 0.96, sensitivity is always 1, and specificity is around 0.995. The proposed algorithm could be used as a decision support tool for audio forensics verification purposes, that allows to identify if an audio proof has been modified, and the time segments in which it has been modified.

1. Introduction

In recent years, multimedia data has become a mainstay of modern life. It can support different daily life activities, such as entertainment (photos, music, films), work (slices, audio and video recordings, etc.), education (books, lessons, tutorials, etc.), government (dissemination of policies, political campaigns), or even in forensic audio (proofs, recordings). In some cases, particularly in the latter, it is required to prevent tampering of the evidence, i.e. to guarantee the Chain of Custody (CoC), because digital multimedia data can be easily be manipulated (Zmudzinski & Steinebach, 2009).

Generally, there are two groups of methods that allow evaluating the integrity and authenticity of multimedia data: content-based identification and information hiding (Gomez, Cano, Gomes, Batlle, & Bonnet, 2002). The first group consists in extracting significant characteristics from the data, giving a kind of digital signature of the multimedia data as a result (Gomez et al., 2002). The goal of the second group is to insert useful information into multimedia data in either an imperceptible or a secure way (Lin & Abdulla, 2014).

In content-based identification, a common technique for data authentication is cryptographic hash function, which is designed to ensure that every bit in the data stream is unmodified (Wu & Kuo, 2002; Zmudzinski & Steinebach, 2009). The mathematical hash function takes a variable length message as an input and maps it to an output message of fixed length known as hash value or message digest (Zmudzinski & Steinebach, 2009). They are appropriate when the multimedia data will not be modified at all, since a single bit flip is sufficient to change the digest (Gomez et al., 2002). Message-digest algorithms (MD) and Secure Hash Algorithms (SHA) are examples of traditional cryptographic hash functions. Some popular versions of SHA algorithms are SHA0, SHA1 and SHA2, while in MD algorithms the most popular version is MD5. Nevertheless, algorithms such as MD5, SHA0 and SHA1 have been attacked (Biham et al., 2005; Sotirov et al., 2008), whereby in some cases it is recommended to use the most recent member of these families.

On the other hand, information hiding is a general concept of concealing data into other contents, and it refers to either keeping in secret the existence of embedded information (steganography) or marking the content (watermarking) (Ballesteros L & Moreno A, 2012; Garcia-Hernandez, Parra-Michel, Feregrino-Uribe, & Cumplido, 2013). Watermarking techniques allow embedding a message such as a signature or any other information without noticeable perceptual distortion (Lin & Abdulla, 2014; Renza, Ballesteros, & Ortiz, 2016). Usually, in a blind detection scheme, data

* Corresponding author.
E-mail addresses: diego.renza@unimilitar.edu.co, diegoerre@gmail.com (D. Renza), dora.ballesteros@unimilitar.edu.co (D.M. Ballesteros L.), lemus.camilo@gmail.com (C. Lemus).

http://dx.doi.org/10.1016/j.eswa.2017.09.003
0957-4174/© 2017 Elsevier Ltd. All rights reserved.

authenticity is verified through the secret keys of the embedding and extraction process; if these keys are equal, it means the data have not been modified. Digital watermarking can be classified into robust watermarking, semi-fragile watermarking and fragile watermarking (Qi, Chen, & Xu, 2015; Wang & Fan, 2010). A robust mark is designed to resist attacks that attempt to remove or destroy the mark, and generally they are used for copyright protection (Lei, Soon, & Li, 2011; Yalcin & Vandewalle, 2002). Fragile watermarking has very limited robustness and it is designed to detect slight changes of the watermarked data, therefore it is sensitive to all kind of malicious and non-malicious manipulations (Acevedo, 2007; Chen & Wang, 2009; Yalcin & Vandewalle, 2002). In between, it is semi-fragile watermarking, which presents high robustness against malicious manipulations and limited robustness against non-malicious manipulations (Li, 2005).

In audio forensics applications, there are some aspects of cryptographic hash functions to take into account. First, additional meta-data is required (the original digest) in the integrity-check phase, which implies inserting the fingerprint in a database or in a header (not appropriate in some cases) (Gomez et al., 2002). Second, they allow identifying if the multimedia data has changed or not (binary answer), but they cannot identify the places where the data has changed. A solution to the latter would be to divide the multimedia data into smaller data frames to obtain different hash values. The point here has to do with the resolution at which the original data are divided. On the other hand, a same signature (key) can be used in watermarking to mark different multimedia files, which is not possible with cryptographic hash functions (a hash value for each multimedia file); besides, by using fragile watermarking, locations where data was modified can be detected (Serra-Ruiz & Megías, 2011). This property is of special interest in audio forensics because data is susceptible to be modified partially, and it is desired to guarantee the CoC of the evidence; in cases of tampering, it can be important to detect the time ranges of the manipulations before the recording is used as evidence.

In accordance with the previous explanation, fragile marks have been used to detect slight modifications of multimedia data (very useful in audio forensics), whereby if the mark persists on the data, there is a high probability that the data has not been altered (Acevedo, 2007). Some previous works related to fragile watermarking on audio have been proposed in the literature. In Wu and Kuo (2002), two schemes that use a simplified masking model to embed the mark in the DFT (Discrete Fourier Transform) magnitude domain are proposed: exponential-scale odd/even modulation and linear additive speech watermarking; these two approaches have a trade-off between tamper localisation and bit rates, and their security includes a pseudo-random sequence of real numbers generated from the secret key. Spread spectrum and replica watermarks are proposed in Li (2005) to detect and locate manipulations of audio recordings; when spread spectrum is used, a pseudo-random sequence (shaped in time and frequency to match the original audio) is multiplied by a weighted data signal with amplitude limited to ± 1; when a replica watermark is used, the pseudo-random sequence is replaced by a frequency and time shift of the original signal. In any case, the resulting signal is added to the original one. The disadvantage of these schemes is the perceptibility of the watermark given principally by the addition operation. The proposal in Gomez et al. (2002) consists in extracting the fingerprint of a signal and embedding it in the signal using watermarking. In Zmudzinski and Steinebach (2009), a robust audio hash is embedded using a blind spread-spectrum watermarking approach in the frequency domain (magnitude coefficients); however, only a human-perception-based was used for authentication, therefore a semantic assessment was not done. Recently, a fragile watermarking scheme for speech content authentication was proposed; the method consists in generating two marks from a hash function and speech sample points; then, these marks are embedded into the wavelet coefficients (Qian, Wang, Hu, Zhou, & Li, 2015). The main disadvantage of these proposals is related to guaranteeing that the fingerprint of the watermarked signal remains the same as the original signal.

In terms of expert systems, there are many data hiding proposals applied to images and audio signals. For instance, the mark can be inserted by a pre-processing step based on sparse coding (Tareef & Al-Ani, 2015), or it can be embedded in specific zones of the audio signal by an adjusted process (Wu, Lin, Hu, & Chen, 2011), or by considering the energy relationship between sub-frames to insert the mark (Peng, Li, Luo, Wang, & Zhang, 2013). In the first case, the mark is very robust and resists attacks; however, it does not provide tampering detection. In the second case, the inserted mark is imperceptible, but it does not protect the integrity in all zones of the audio proof. In the latter case, high imperceptibility and robustness are achieved, but again, it does not provide information about tampering detection. In any case, a decision support tool for audio forensics verification purposes should identify not only whether the signal has been modified, but also the time segments in which it has been modified.

According to the above, this paper presents a new fragile watermarking method for digital audio authenticity which can be used for audio forensics purposes. The main characteristic of our proposal is that the embedding process is adjusted according to the length/value of the mark and the length/amplitude of the audio signal. The embedding process is done in wavelet domain through quantisation index modulation (QIM) that uses a low value of quantisation step to increase the fragility. The secret key is defined as a variable length text string. Besides, in order to enhance the tampering detection, the text input signal is expanded through OVSF (Orthogonal Variable Spreading Factor) codes and adaptive redundancy is applied to the expanded signal.

2. Basic concepts

This section explains briefly some concepts used in the proposed scheme: Discrete Wavelet Transform and Orthogonal Variable Spreading factor Codes.

2.1. Discrete Wavelet Transform (DWT)

DWT provides a simple and fast method to analyse a signal at different scales; it uses functions (such as wavelets) that automatically adapt to the different components of the signal. Fig. 1(a) shows the analysis filter bank (FB) for input signal decomposition $H(n)$, obtaining two sub-bands, $a(n)$ and $d(l, n)$. These signals represent the low frequency/high resolution part (approximation, $a(n)$) and the high frequency/low resolution part (details, $d(l, n)$). The process is repeated using $a(n)$ according to the desired number of decomposition levels (L). At each decomposition level, the time resolution is halved and the frequency resolution is doubled (Lin & Abdulla, 2014). This FB uses a low-pass filter ($g_0(n)$), and a high-pass filter ($g_1(n)$) where the output of each filter is down-sampled by two to obtain a half sample rate in each component. The corresponding synthesis FB is showed in Fig. 1(b).

2.2. Orthogonal Variable Spreading Factor (OVSF) codes

Originally, OVSF codes are used to enable multiple access in UMTS (Universal Mobile Telecommunications System). They are formally specified as $C_{SF,\,n}$ where SF stands for spreading factor, and n is the code number. To construct OVSF codes, an iterative tree can be used. The codes start with the $C_{1,0} = [1]$ code. The second level of the tree has two branches, the upper branch $C_{2,0} = [C_{1,0} \quad C_{1,0}]$ and the lower branch $C_{2,1} = [C_{1,0} \quad -C_{1,0}]$. Namely, the