ELectronics EXpress

IEICE Electronics Express, Vol.7, No.14, 995–1001

Improving the security of Fallahpour's audio watermarking scheme

Jose Juan Garcia-Hernandez[1a)], **Claudia Feregrino-Uribe**[1b)], **Rene Cumplido**[1c)], **and Ramon Parra-Michel**[2d)]

[1] *National Institute for Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico*

[2] *Department of Electrical Engineering, Communications Section, CINVESTAV-IPN, Guadalajara, Mexico*

a) *jjuan@ccc.inaoep.mx*

b) *cferegrino@ccc.inaoep.mx*

c) *rcumplido@ccc.inaoep.mx*

d) *rparra@gdl.cinvestav.mx*

Abstract: The audio watermarking scheme recently proposed by Fallahpour in [1] is one of the schemes with highest payload published to date. In this letter a key-based security improvement is proposed for that scheme. It is achieved by adding a Pseudo-Random Number Sequence (PRNS) in the frequency domain to the data samples, before applying the insertion algorithm. Experimental results show that the proposed enhancement keeps the perceptual transparency and the robustness to all attacks originally reported by Fallahpour's scheme and the payload is not significantly affected.

Keywords: watermarking, spline interpolation, audio signals, key-based security

Classification: Science and engineering for electronics

References

[1] M. Fallahpour and D. Megias, "High capacity audio watermarking using fft amplitude interpolation," *IEICE Electron. Express*, vol. 6, no. 14, pp. 1057–1063, 2009.

[2] D. Kirovski and H. Malvar, "Spread spectrum watermarking of audio signals," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1020–1033, April 2003.

[3] B. S. Ko, R. Nishimura, and Y. Suzuki, "Time spread echo method for digital audio watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 212–221, 2005.

[4] H. Kang, K. Yamaguchi, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Full index embedding patchwork algorihm for audio watermarking," *IEICE Trans. Inf. & Syst.*, vol. E91-D, no. 11, pp. 2731–2734, 2008.

[5] J. J. Garcia-Hernandez, M. Nakano, and H. Perez, "Data hiding in audio signals using rational dither modulation," *IEICE Electron. Express*, vol. 5, no. 7, pp. 217–222, 2008.

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010

[6] A. Deshpande and K. M. Prabhu, "A substitution by interpolation algorithm for watermarking audio," *Signal Processing*, Elsevier, vol. 89, no. 2, pp. 218–225, 2009.
[7] H. T. Sencar, M. Ramkumar, and A. N. Akansu, *Data Hiding Fundamentals and Applications*, Elsevier Academic Press, 2004.
[8] N. Cvelic and T. Seppanen, "Improving audio watermarking scheme using psychoacoustic watermark filtering," *Proc. IEEE Int. Symp. Signal Process. Inf. Technol.*, pp. 169–172, 2001.
[9] F. P. Gonzalez, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: a high rate data-hiding method invariant to gain attacks," *IEEE Trans. Signal Process.*, vol. 53, pp. 3960–3975, Oct. 2005.
[10] A. Lang, "Stirmark benchmark for audio." [Online] http://wwwiti.cs.uni-magdeburg.de/~alang/smba.php

1 Introduction

The quick growth of Internet has increased the easy reproduction and retransmission of multimedia contents which has facilitated the proliferation of unauthorized data-manipulation. One possible solution to this piracy problem is the use of watermarking techniques, in which an imperceptible and statistically undetectable signature is added to the multimedia content to be protected. A watermark must completely characterize the person who embedded it in order to be useful for proving the intellectual property of an audio material. The watermark has to be imperceptible to preserve the original signal quality and robust to signal processing operations such as: filtering, compression, resampling, noise corruption, etc., which are classical no-intentional attacks against watermarking schemes. Several schemes for watermarking embedding and detection of watermarks in audio sequences have been proposed, where the main used techniques are spread spectrum [2], echo [3], patchwork [4], rational dither modulation [5] and interpolation [6]. In order to keep the perceptual transparency, these techniques take into consideration the properties of the Human Auditory System (HAS). From the proposed schemes, rational dither modulation and interpolation systems are those with the highest payload. Recently, an audio watermarking scheme based on spline interpolation was proposed in [1], reporting one of the highest throughputs among the audio watermarking schemes published to date. However, this scheme lacks of a basic security aspect required in the data hiding schemes: key-based security [7], because it is not sufficient to keep the parameters secret to guarantee security against hidden data estimation. In this letter it is proposed a security improvement to Fallahpour's audio watermarking scheme that is performed by adding and substracting a PRNS to the frequency domain samples (those obtained through the use of a Discrete Fourier Transform (DFT)) just before and after applying, respectively, the original insertion algorithm. This modification adds key-based security to Fallahpour' scheme. Experimental results show that our solutions do not affect the perceptual transparency of the original Fallahpour' scheme and

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010

keeps the same robustness to all others attacks originally reported.

2 Fallahpour's audio watermarking scheme

The scheme proposed by Fallahpour in [1] is based on the difference between the original and the interpolated amplitudes of the DFT samples as obtained by spline interpolation. A sample is selected for embedding secret information if the difference is lower than a given fraction of the interpolated value. To obtain the marked DFT samples, the interpolated value is changed according to the secret bit. The embedding steps are as follows:

$k = 1;$
for $i = low_{band}$ to $high_{band}$
  if $\mod(i, 2) == 0$
    $e_i = f_i - I_i;$
    if $|e_i| > 2\alpha I_i$
      $f_i^{'} = f_i;$
    elsif $b_k == 0$
      $f_i^{'} = I_i;\ k = k + 1;$
    elsif $\{(b_k == 1) \wedge (e_i \geq 0)\}$
      $f_i^{'} = I_i(1 + \alpha);\ k = k + 1;$
    else
      $f_i^{'} = I_i(1 - \alpha);\ k = k + 1;$
    end;
  end;
end;

where f_i is the magnitude of the i^{th} sample of the DFT spectrum, low_{band} and $high_{band}$ are the lower and higher limits of a selected band for embedding secret information, I_i is the interpolated value of f_i, α is a threshold, b_k is the k^{th} bit of the secret bit stream, and $f_i^{'}$ is the watermarked value of f_i.

The extraction process is as follows:

$k = 1;$
for $i = low_{band}$ to $high_{band}$
  if $\mod(i, 2) == 0$
    $e_i^{'} = f_i^{'} - I_i;$
    if $|e_i^{'}| < 0.5\alpha I_i$
      $b_k^{'} = 0;\ k = k + 1;$
    elsif $(|e_i^{'}| \geq 0.5\alpha I_i) \wedge (|e_i^{'}| \leq 1.5\alpha I_i)$
      $b_k^{'} = 1;\ k = k + 1;$
    end;
  end;
end;

where $b_k^{'}$ is the k^{th} bit of secret bit sequence.

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010

3 Estimation of secret bit sequence and the proposed solution

In secret communications applications such as military communications, ownership proof, etc., it is fundamental to insert and extract the information in the multimedia data using a secret-key [7]. The Fallahpour' scheme lacks in that requirement. In the next subsection we introduce an attack that extracts the secret bit sequence without knowledge of the parameters α, low_{band} and $high_{band}$. After that we propose a solution for this drawback.

3.1 Estimating the secret bit sequence

If an attacker only knows the number of embedded bits (as it is the case for practical watermarking applications) and the watermarked audio signal has not suffered other attacks, it is possible to undercover the secret bit sequence as follows:

- For each frame of the audio clip find the lower and the higher index values i when the watermarked sample $f_i^{'}$ is the same that the interpolated sample I_i (it happens when a 0 bit was inserted in that DFT sample), save them in the vectors $mins$ and $maxs$ respectively. The lower value of $mins$ corresponds to low_{band} and the higher value of max corresponds to $high_{band}$.

- In the extraction algorithm, carry out a brute-force attack for the α parameter from 0.1 to 0.5 in intervals of 0.01. Considering that $e_i^{'}$ will always be zero, the values of α for the brute-force attack do not affect the number of recovered 0 bits. Therefore, it is necessary to find an α value such that the number of recovered 1 bits matches the number of embedded bits minus the number of 0 bits.

In our experiments this attack was successful in about 93% of the trials which is a high score for any estimation attack.

3.2 The proposed solution

In order to improve the security of Fallahpour's scheme, we propose to add and subtract a PRNS to the DFT magnitudes just before and after, respectively, applying the original insertion algorithm. It has been shown in other watermarking schemes such as spread spectrum, patchwork and rational dither modulation, that, despite its simplicity, adding a PRNS improves significantly the security of the scheme. In order to keep the transparency, the sequence is filtered by a system with an impulse response modeled for the HAS in quiet conditions [8]. In our experiments this solution does not affect significantly the payload while maintaining transparency and robustness as presented by the original Fallahpour's scheme.

Nevertheless, this proposed solution is weak against gain attack (also known as volume attack). This attack consists on applying a fixed gain to the watermarked clip before the extraction process. Due to the PRNS in the extractor system is not scaling at the same fashion than the attacked audio clip, the extraction process fails. A solution to this issue is to attach a pilot

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010

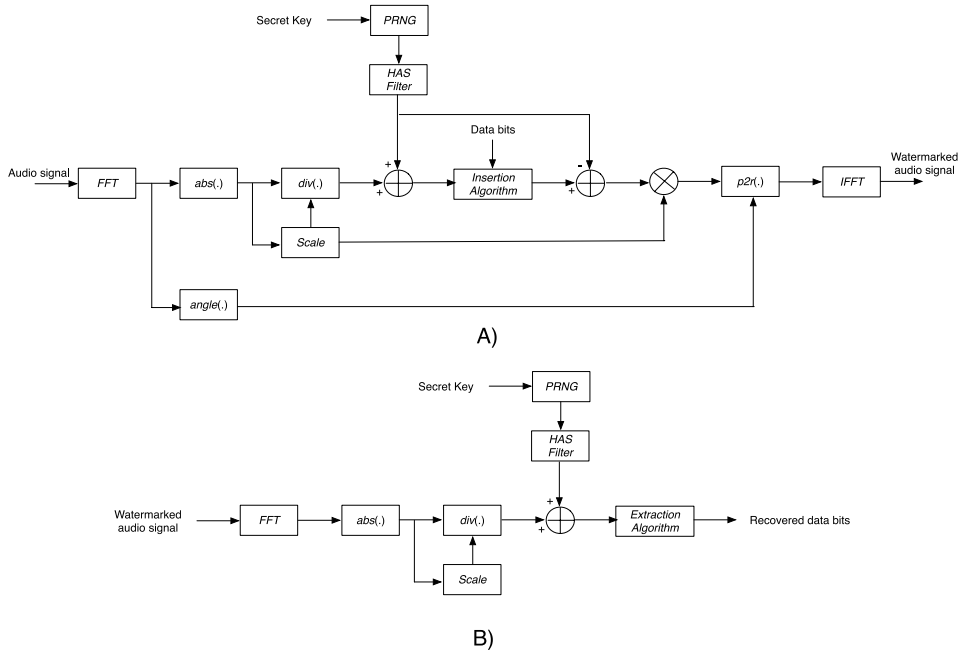

Fig. 1. The key-based improved Fallahpour's audio watermarking system. A) insertion, B) extraction

signal that is known to the extractor system and which is able to compensate for the gain attack. This solution could be ineffective in channels with impulse noise, to compensate for this, another modification is introduced. This second modification is based on the idea behind Rational Dither Modulation (RDM) [9], where the Dither Modulation algorithm considers the ratio of the current host sample and the previously generated watermarked sample for embedding the information. By using this algorithm the ratio will always be constant independently of the gain factor of the attack, therefore, the watermarked information will remain unaltered. This results in a robust scheme that is both self-cointained and no-sensitive to impulse noise. In this letter we propose to apply the Fallahpour's insertion algorithm to the values resulting from adding the PRNS to the ratio of the current DFT sample and the previous odd sample, instead of applying it to the DFT samples as Fallahpour proposed.

Figure 1 shows the insertion and extraction processes, where *Scale* is a vector formed by the DFT magnitudes (which are calculated using a Fast Fourier Transform (FFT) in the block FFT) as follows: $Scale = [f_1, f_1, f_1, f_3, f_3, f_5, f_5, ..., f_{2n-1}, f_{2n-1}]$, $abs(.)$ is the absolute function, $angle(.)$ is the angle function (the last two functions are used in order to transform the DFT samples from rectangular to polar coordinates), $div(.)$ is the division operation and $p2r(.)$ is the polar to rectangular function. The two blocks, *insertion algorithm* and *extraction algorithm*, are the original algorithms proposed by Fallahpour.

The solution using the ratio of the current DFT sample and the previous odd sample reduces about 20% of the original scheme payload, however this solution is effective against gain and impulse noise attacks.

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010

By using a secret key, both solutions are secure because although the attacker may know the parameters α, low_{band} and $high_{band}$ it is not possible to extract the secret bit sequence without knowing the key.

4 Experimental results

For the experiments we use a fragment of the song *The game of love* by Carlos Santana. As in [1] we obtain the SNR, ODG and payload measurements for comparison purposes. In order to carry out our experiments we considered the same reported parameters in [1] such as: $\alpha = 0.3$ and the frequency band $0.5 - 5\,\text{kHz}$. Since the size of the DFT window was not reported, we utilized 4096 samples due to different size values do not affect significantly the results. For each window, different PRNS were applied. The PRNS values were generated with uniform distribution on the interval [0,0.005], employing the *rand(.)* function of Matlab 7.9 running in a Mac Os X workstation, the sequence is double-precision valued. Table I shows the transparency and payload results for the solution with and without the rational stage. As it can be seen, the SNR value varies slightly for the two solutions, the ODG value is kept in the acceptable region of transparency (0 to -1.0) and the payload varies 3% for solution without rational stage, which is negligible, and 20% for the solution with rational stage, which is the cost of achieving robustness

Table I. Transparency and payload results

Proposed solutions	Duration (seg)	SNR (dB)	ODG	Payload (bps)
Original scheme	60	31	-0.35	2812
Solution without rational stage	60	31	-0.42	2755
Solution with rational stage	60	30	-0.51	2248

Table II. Robustness against Stirmark attacks

	Original scheme		No rational stage		With rational stage	
Attack	ODG	BER	ODG	BER %	ODG	BER %
AddBrumm	-3.8	0.0	-3.4	0.0	-3.6	0.1
AddDynNoise	-2.7	0.1	-2.7	0.2	-2.5	2.0
ADDFFTNoise	-0.6	0.8	-0.6	1.0	-0.6	1.8
Addnoise	-0.2	0.0	-0.4	0.1	-0.8	0.3
AddSinus	-0.2	0.0	-0.2	0.0	-0.2	0.0
Amplify	-0.1	0.0	-0.1	0.0	-0.1	0.0
BassBoost	-3.0	0.0	-3.2	0.0	-3.4	0.0
Echo	-3.0	0.0	-3.0	0.0	-3.1	0.5
FFT_HLPassQuick	-2.9	2.0	-3.0	2.0	-3.1	3.2
FFT_Invert	-3.4	1.6	-3.4	2.0	-3.6	3.0
Invert	-2.8	0.0	-2.9	0.0	-3.0	0.0
Resampling	-1.6	4.8	-1.6	5.0	-1.8	5.2
LSBZero	0.0	0.0	-0.2	0.0	-0.2	0.0
MP3	-0.1	0.0	-0.2	0.0	-0.3	0.7
Noise_Max	-3.0	0.2	-3.0	0.2	-3.2	1.0
Pitchscale	-0.1	0.0	-0.1	0.0	-0.2	0.9
RC_HighPass	-2.7	0.0	-2.7	0.0	-3.0	1.0
RC_LowPass	-2.8	0.0	-3.0	0.0	-3.2	0.0

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010

to gain attack and impulsive noise at the same time.

Table II shows the effect of various attacks from the Stirmark Benchmark [10], those attacks are the same reported in the Fallahpour's paper. From this table it is possible to claim that our solutions keep the robustness and the transparency of the original scheme. The results for ODG and BER tests present slight changes compared to the Fallahpour' scheme, which are negligible. These experimental results show that in spite of improving the security of Fallahpour' scheme, the transparency, payload and robustness of the scheme are not significantly affected.

5 Conclusions

In this letter it was corroborated that the recently proposed Fallahpour audio watermarking scheme is able to insert high payloads, however, a weakness in terms of security was demonstrated with a simple attack. This work presents a security improvement that allows to use Fallahpour's algorithm for practical applications. The proposed modification consist on incorporating a PRNS in the frequency domain to the data samples before applying the insertion algorithm. Simulation results show that transparency and robustness of the watermarked data remains practically unchanged while the payload has been secured.

Acknowledgments

The authors would like to thank CONACyT for financial support under grant CB-2007-1-84668.

DOI: 10.1587/elex.7.995
Received April 13, 2010
Accepted June 22, 2010
Published July 25, 2010