Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **5**, September **2018**

Centroid-Based Audio Steganography Scheme in Wavelet Domain

Hanlin Liu, Jingju Liu, Xuehu Yan, Pengfei Xue, Song Wan

National University of Defense Technology
230037 Hefei China
hanlinliu1993@126.com, publictiger@126.com

Li Li

Shenzhen Institute of Information Technology,
518172 Shenzhen, China

Received Auguest, 2017; revised April, 2018

Abstract. *This paper proposes an audio steganography scheme based on centroid in wavelet domain which can be used to hide secret information in digital audio. The major contribution of the proposed scheme is that secret information is represented by the XOR operation and adaptive hiding. First, the original audio is decomposed by 2-level integer wavelet transform which then is divided into many segments to calculate the centroid of each segment. Then secret information is embedded into each low frequency wavelet coefficient by the XOR result of four bits. If the XOR result is not equal to secret information, one bit will be selected and then be flipped according to centroid. The experimental results and comparison with existing techniques show that the proposed scheme has considerable capacity and good imperceptibility. Furthermore, the robustness of the algorithm is well done in resisting common attacks. Please write down the abstract of your paper here....*

Keywords: Audio Steganography, Wavelet domain, Centroid, Robustness

1. **Introduction.** With the rapid development of Internet and multimedia technology, the transmission, storage and processing of multimedia information, such as text, images, audio and video, are also more convenient and fast. At the same time, the security of multimedia information has also suffered an unprecedented challenge. The traditional encryption method encrypts secret information into unreadable ciphertext, and to a certain extent, which ensures the security of secret information. However, the unreadable ciphertext is more likely to cause the attacker's curiosity, and then deciphered by the attacker. Steganography, as an information security technique for hiding information, is a good solution to the above problem. The dark forest rule in the science fiction—*The Three-Body Problem* [1] illustrates the point that any life that exposes itself will soon be wiped out. Compared with encryption techniques, steganography reduces the possibility of exposure by concealing the existence of secret information to ensure the security. So, steganography is a powerful technique which enhances security in data transferring and archiving.

Throughout history, a multitude of methods have been used to hide information. In the last twenty years, digital steganography technique which is used to embed secret information into digital multimedia is gradually rising. So far, various researchers on steganography have been carried out on storage media, such as text, image, audio and video. In these areas, image steganography technology has been mature, but the research of audio steganography is relatively slow. Furthermore, auditory system is the largest source of information in addition to the human visual system(HVS). Therefore, the study of audio steganography is of great significance and has broad application scenarios.

Like other multimedia steganography, audio steganography also has the same three performance evaluation criterias [2]:

1. Capacity means the amount of secret information that can be embedded into the original audio without affecting the perceptual quality of audio.
2. Imperceptibility evaluates how well a secret message is embedded into the cover audio. The difference between audio after hiding and audio before hiding should remain negligible.
3. Robustness indicates the ability of secret messages to resist against attacks.

In audio steganography, Human Auditory System(HAS) is used to hide information in the audio. Because the human auditory system has more precision than HVS, audio steganography has more challenges than image steganography [3].

In this paper, we propose an audio steganography scheme which is based on wavelet packet transform with adaptive embedding. The secret information is represented by the XOR result of four bits and the adaptive hiding is determined by the wavelet domain centroid. The secret information can be recovered without original audio. Experimental results and comparison with existing methods show that the proposed scheme has an advantage of good robustness while maintaining imperceptibility and the same capacity as the Least Significant Bit(LSB) method.

The remainder of this paper is organized as follows: Section II discusses some existing methods. Section III introduces some preliminary knowledge. The proposed scheme is introduced in Section IV. In Section V experimental results of proposed scheme are discussed. Finally, Section VI concludes this paper.

2. **Related Work.** Steganography can be classified into methods in time domain and transform domain. The commonly used methods in time domain include LSB [4], Echo Hiding [5], Spread Spectrum [6] Method and etc. Besides, methods in transform domain contain steganographic method based on Discrete Fourier Transform(DFT), Discrete Cosine Transform(DCT) [7] and Discrete Wavelet Transform(DWT) [8, 9]. The method mentioned above is some of the traditional steganography methods, and if only one of the above methods is used to hide information, its performance will be poor. Thus, most steganography methods combine other techniques to improve the performance.

There are many steganography methods based on LSB method [10–12]. In [10], Cvejic and Seppanen proposed an algorithm which uses perfect reconstruction filter banks and embedded additional information in the wavelet domain of the audio signal by modifying LSB of wavelet coefficients. Bhowal etc. [11] proposed an effective data hiding technique to embed a secret image and extract them in a bit-exact manner by altering the magnitudes of the coefficient of DWT of audio signals. But the robustness should be enhanced. In [12], Krishnan etc. presents enhanced security in audio steganography by using Higher LSB to improve security and robustness by embedding bits of secret message in higher LSB of a cover audio, which is proven to increase the robustness using higher LSB and will be useful as a basis for audio steganography. Steganography based on LSB has large capacity, but its robustness is poor. The current solution is to embed secret information in higher bit.

In audio steganography, some methods use audio transform domain features to hide secret information. In [13], an audio watermark algorithm based on frequency centroid and histogram is proposed by modifying the frequency coefficients. The experimental results show that the algorithm is very robust to resample TSM and a variety of common attacks. Wang etc. [14] proposed a novel centroid-based semi-fragile audio watermarking scheme in hybrid domain. In the proposed scheme, first, the centroid of each audio frame is computed, then Hash function is performed on the obtained centroid to get the watermark, after that, the audio sub-band which carries centroid of audio frame is performed with DWT and DCT, and finally the encrypted watermark bits are embedded into the hybrid domain. The advantage of using audio features for steganography is that it can achieve adaptive steganography, or improve the performance of the method.

Based on LSB and audio feature, in this paper, we propose a scheme based on wavelet domain centroid, in which secret information is embedded into original audio through XOR operation and the altered bit is determined by centroid in wavelet domain. The advantage of the scheme is that the XOR operation can enhance the robustness and XOR operation is simple. Besides, it has good imperceptibility and transparency while keeping the capacity equivalent to that of LSB method.

3. **Preliminaries.**

In this section, we exhibit some preliminaries for our work. Since this scheme is based on the wavelet domain centroid, the wavelet transform and centroid are briefly introduced in this section.

3.1. **Wavelet Transform.**

DWT is a discipline capable of giving time frequency representation of the signal. Assuming that the input signal is an audio signal, after DWT the audio signal is decomposed into two sets of coefficients. One is the approximate coefficient, which represents the low frequency component, and the other is the detail coefficient, which represents the high frequency component. The low frequency part can be further decomposed into low and high frequency. Figure. 1 shows a 3-level DWT decomposition of an audio signal. In Figure. 1, CA_i represents the approximate coefficients and CD_i represents the detail coefficients. The reconstruction of the original signal is the inverse of the decomposition process and the original audio signal can be reconstructed using the inverse DWT process.

The advantage of discrete wavelet transform is that the original signal can be decomposed into different resolution components, and the components with different resolutions represent information of different importance. Most energy of the original signal is concentrated in the low frequency part, which has better robustness. And even if the high frequency part is changed, the impact on the original signal is very small, so the high frequency part has better imperceptibility. In our scheme, we select the 2-level integer wavelet transform to decompose the original audio, which is taken into account the two reasons: (1) The obtained coefficients after integer wavelet transformation are still integers, which is convenient for the embedding process; (2) Embedding secret information into the low frequency coefficients after two level decomposition can give play to the tradeoff between robustness and imperceptibility.

3.2. **Centroid.**

In the wavelet domain, the centroid is defined as the center of the energy distribution. Since the centroid of the audio signal is different in different time periods, the centroid can serve as a typical feature that reflects the nonstationary nature of the audio signal. In addition, because the centroid location is less affected when attacked, the wavelet domain centroid is considered into the category of influencing factors of embedding

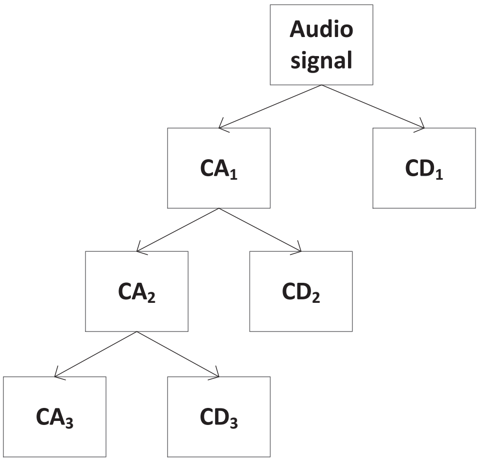


FIGURE 1. 3-level DWT decomposition

location selection. And the centroid is calculated as follows:

$$centroid = \frac{\sum_{i=1}^{N} i|x(i)|^2}{\sum_{i=1}^{N} |x(i)|^2} \tag{1}$$

In Eq. 1, $x(i)$ denotes the *ith* coefficient, and N is the total number of wavelet coefficients. The result of the calculation represents a position. The closer to this position, the more centralized the wavelet energy is .

4. The Proposed Scheme.

In this section, we propose the embedding and extraction processes of our audio steganography scheme based on wavelet domain centroid.

4.1. Embedding Process.

The embedding process of the introduced scheme is shown in Fig. 2 and described in Algorithm 1.

Algorithm 1. The embedding process of the proposed scheme

Input: secret information and original audio
Output: stego audio

Step 1: Binarize the secret information and get binary information S.
Step 2: Decompose the original audio using the 2-level integer wavelet transform and get the approximate coefficients CA_2.
Step 3: Segment the approximate coefficients CA_2, and calculate the centroid of each segment according to Eq. 1.
Step 4: If XOR($CA_2^{(7)}(i)$, $CA_2^{(8)}(i)$, $CA_2^{(9)}(i)$, $CA_2^{(10)}(i)$) is not equal to $S(i)$,$i = 1, 2, \cdots, length(S)$, then select one $CA_2^{(fpos)}(i)(fpos \in \{7, 8, 9, 10\})$ from the above four bits and flip it. The flip position is indicated by $fpos$, and $fpos = \lfloor(centroid/200)\rfloor + 7$
Step 5: Repeat Steps 3 and 4, until all the binary secret information has been embedded into CA_2. Then performing inverse integer wavelet transform and obtain stego audio.

In Algorithm 1, we remark that.

1. Original audio is 44.1 KHz with 16 bits. And the reason why we select seventh, eighth, ninth, and tenth bits to carry secret information is that if the selected bit is too low, the algorithm is poorly robust; otherwise, the perceptibility will decrease.
2. In step 3, we set the length of each segment to 512, and *centroid* $\in$ [0,512].

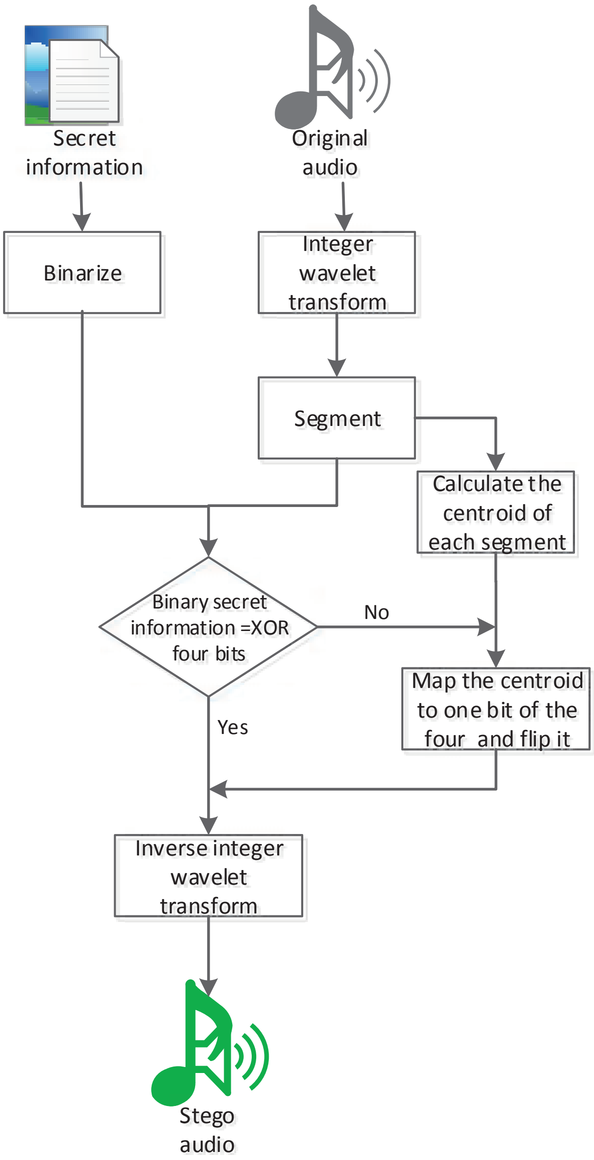

Secret information
Original audio
Binarize
Integer wavelet transform
Segment
Calculate the centroid of each segment
Binary secret information =XOR four bits
No
Yes
Map the centroid to one bit of the four and flip it
Inverse integer wavelet transform
Stego audio

FIGURE 2. The embedding process of the proposed scheme

3. In step 4, as long as the coefficients are within the segment, if the flipping position needs to be selected, it is determined by the centroid of the segment. And $\lfloor(\cdot)\rfloor$ is floor function.
4. In step 4, the flip method is as follows: if XOR$(CA_2^{(7)}(i), CA_2^{(8)}(i), CA_2^{(9)}(i), CA_2^{(10)}(i)) \neq S(i)$, $CA_2^{(fpos)}(i) = mod((CA_2^{(fpos)}(i)) + 1, 2)$.
5. Here, $length(\cdot)$ is used to calculate the length of the binary secret messages.

4.2. **Extraction Process.** The extraction process is somewhat similar to the embedding process, and the extraction process of the introduced scheme is shown in Fig. 3 and described in Algorithm 2.

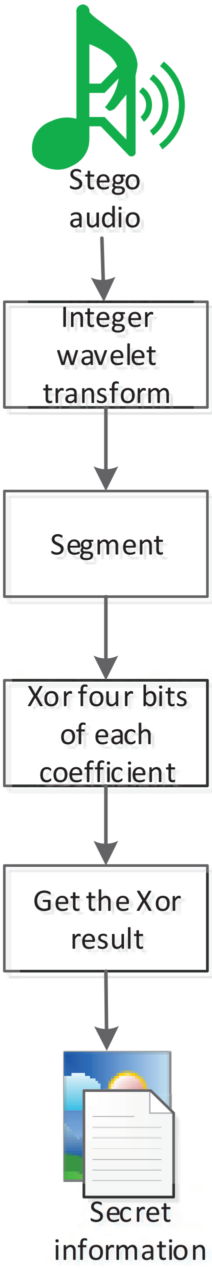


FIGURE 3. The extraction process of the proposed scheme

Algorithm 2. The extraction process of the proposed scheme
Input: stego audio **Output**: secret information
Step 1: Decompose the stego audio using the 2-level integer wavelet transform and get the approximate coefficients CA_2. **Step 2:** Segment the approximate coefficients CA_2. **Step 3:** Perform XOR($CA_2^{(7)}(i)$, $CA_2^{(8)}(i)$, $CA_2^{(9)}(i)$, $CA_2^{(10)}(i)$), $i = 1, 2, \cdots, length(S)$, until all the binary secret information has been extracted. **Step 4:** Convert binary secret information into decimal secret information, and get the original secret information.

In Algorithm 2, it should be noted that:.

1. In step 2, the length of each segment must be equal to 512.

2. Here, $length(\cdot)$ is used to calculate the length of the binary secret messages.

5. **Experimental Results and Analysis.** In order to test the properties of the proposed scheme, we design the following experiments from three aspects: imperceptibility, robustness and capacity.

The origianal audio belongs to category of mono audio file named Pop and secret information is a 256×256 gray image named Lena. Pop is sampled at 44.1 KHz and quantized by 16 bits. It should be noted that other audio files as original audio can also get the similar results. The proposed scheme is implemented using MATLAB (2016b) programming.

In this section, we utilize BER (Bit Error Rate), PSNR (Peak Signal to Noise Ratio), PESQ (Perceptual Evaluation of Speech Quality) and ODG(Objective Difference Grade) to evaluate the performance. Assessment of the quality of stego audio is made through P.862 (PESQ) which is recommended by ITU and in the range [1,4.5]. If PESQ>3.5, the quality of audio accord with the standard of telephone. ODG is a suitable measurement for audio distortions. ODG = 0 means no degradation occurred in embedded audio signal and ODG = -4 means a very annoying distortion occurred in embedded audio. The ODG calculations are done using the advanced ITU-R BS.1387 standard by Thiede et al [15] and implemented by the software tool EAQUAL by A.Lerch [16]. ODG values of the embedded audio are between -0.11 and -0.47 that concludes their good qualities. Two formulas for calculating BER and PSNR are given as follows, respectively.

BER refers to the bit error rate, which is the ratio of the number of error bits in the transmission process to the total number of bits and used to evaluate the robustness of the algorithm. The formula is as Eq. (2).

$$BER = \left(\frac{l}{L}\right) \times 100\% \tag{2}$$

Where l is the number of error bits, and L is the total number of secret information bits. In this paper, L is equal to $length(S)$.

Peak Signal to Noise Ratio (PSNR) is also used to evaluate the robustness, which is calculated using Eq. (3).

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \tag{3}$$

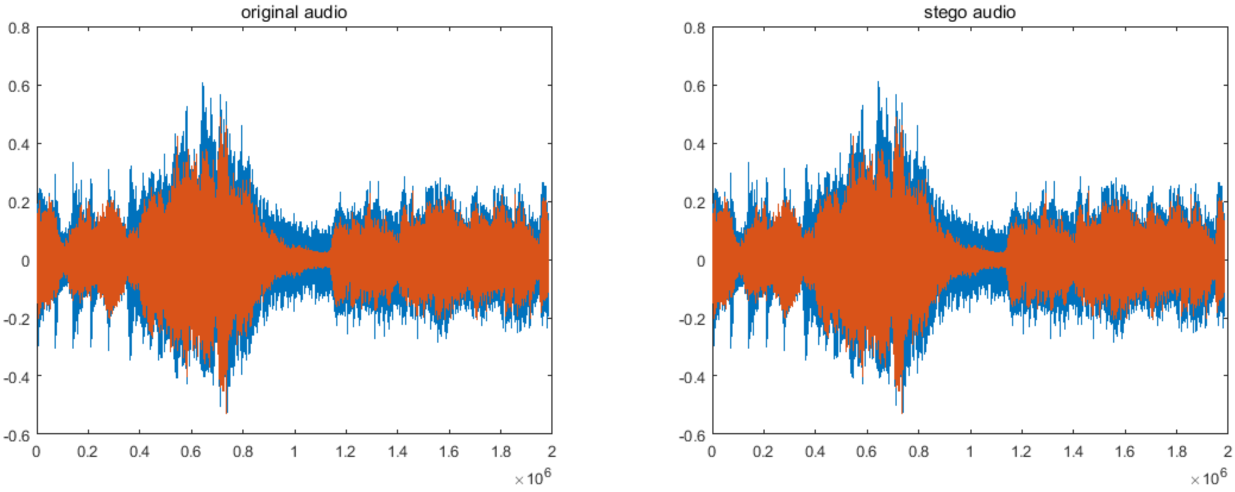
Where MSE is the Mean Square Error, and it referred to the difference value between original secret information and extracted information, which is given in Eq. (4).

$$MSE = \frac{1}{W_1 \times H_1}\sum_{i=1}^{W_1}\sum_{j=1}^{H_1}\left[P\left(i,j\right) - P_a\left(i,j\right)\right]^2 \tag{4}$$

Where $P(i,j)$and P_a(i,j) are pixel values of the original secret image and extracted secret image. W_1 and H_1 are the width and height of the original secret image (the extracted secret image have the same size).

5.1. **Imperceptibility Test.** In the condition of no attack, the waveform of original audio and stego audio are shown in Fig. 4.

From the experimental results, we can see that it is difficult to find the differences between Fig. 4(a) and Fig. 4(b) with the naked eyes. Furthermore, the PESQ value we calculated is 4.30(4.3>3.5) and the ODG Value is -0.2043(-0.2043$\in[-0.11, -0.47]$), which indicates that there is almost no effect on the quality of stego audio.



(a)

(b)

Figure 4. The waveform of original audio and stego audio. (a) Original host audio; (b) Stego audio in the condition of no attack.

5.2. **Robustness Test.** Robustness is an important criterion to evaluate the performance of algorithms. In the experiment, common audio processing including noise addition, re-quantization, and resampling are tested. And the test results are shown in Fig. 5 and Table 1.

1. Noise Addition: white noise with 55 db, 50 db, 40 db and 30 db SNR is added to stego audio.
2. Re-quantization: we tested the process of re-quantization of a 16-bit stego audio to 8-bit and back to 16-bit.Then tested the process of re-quantization of a 16-bit stego audio to 32-bit and back to 16-bit.
3. Resampling: stego audio with original sampling rate 44.1 kHz have been down-sampled to 22.05kHz and up-sampled back to 44.1kHz.Then up-sampled to 88.2kHz and down-sampled back to 44.1kHz.

From Table 1, we can see that the scheme proposed in this paper has good robustness against noise addition, re-quantization of a 16-bit stego audio to 32-bit and back to 16-bit and down-resampling. But it is poor to the resisting of re-quantization of a 16-bit stego audio to 8-bit and back to 16-bit and up-resampling. The above experimental result is because of a 16-bit stego audio to 8-bit and back to 16-bit and down-resampling change sampling points and coefficients a lot. Although adding noise has some influence on stego audio, it mainly affects low bit, and XOR operation also improves robustness.

5.3. **Capacity Analysis.** As secret information is embedded into wavelet coefficients, and each coefficient can be embedded into 1 bit, so the total hiding capacity is closely equal to the number of CA_2. Theoretically, the capacity of this scheme is equal to the capacity of the LSB method. However, the total hiding capacity varies from original audio to original audio, which makes it unfair to measure this performance. Thus, *bps* (bits per second) is presented as a standard to measure hiding capacity. In this experiment, the sampling rate of WAV audio is 44.1 KHz, so the capacity is 44100 *bps*. But in practice, considering other performance, some zero coefficients will not be used to embed secret information. So the capacity is lower than theoretical value.

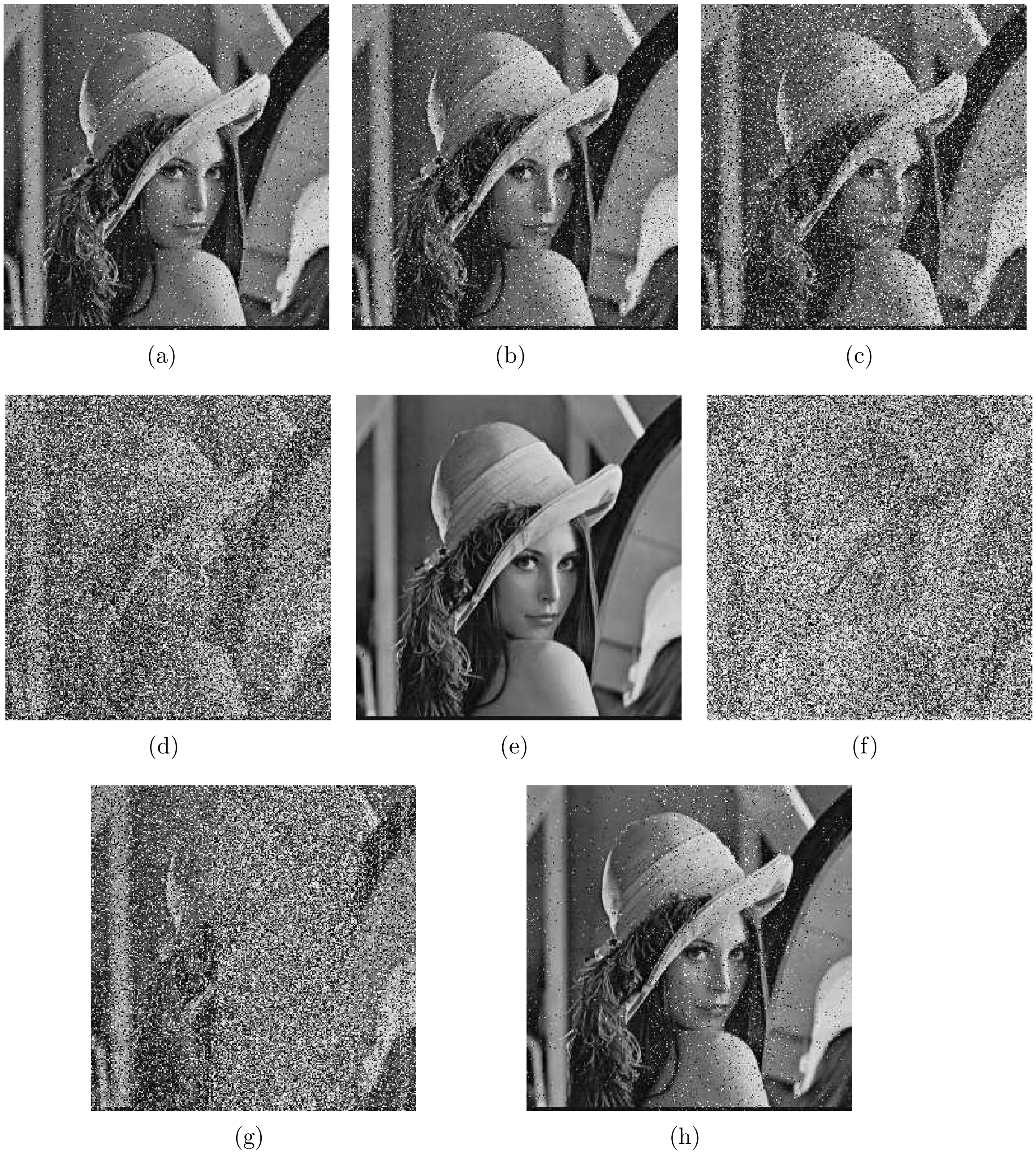


Figure 5. The robustness experimental results. (a)−(d) The secret image extracted from stego audio with 55 db, 50 db, 40 db and 30 db SNR white noise, respectively; (e)−(f) The secret image extracted from stego audio after re-quantization of a 16-bit stego audio to 32-bit and back to 16-bit and re-quantization of a 16-bit stego audio to 8-bit and back to 16-bit; (g)−(h) The secret image extracted from stego audio after up-sampled and down-sampled.

5.4. **Comparison with Related Scheme.** As we all know, every method has different properties and characteristics and it is very difficult to find an impartial comparison of the proposed scheme with some audio steganography schemes. Here, we compare the proposed scheme with existing methods in [11, 17, 18]. And the results of the comparison are shown in Table 2.

Table 1. Test Results of the Proposed Scheme

Attacks	BER(%)	PSNR(db)
55 db SNR white noise	0.0203	20.946
50 db SNR white noise	0.0366	18.500
40 db SNR white noise	0.1149	13.643
30 db SNR white noise	0.3335	9.365
re-quantization (16 bits∼32bits∼16bits)	0	∞
re-quantization (16 bits∼8bits∼16bits)	0.6214	6.9932
up-sampling	0.3253	9.7303
down-sampling	0.0158	22.095

Table 2. Test results comparing with existing methods

Attacks	Scheme in this paper	Scheme in [17]	Scheme in [18]	Scheme in [11]
Capacity	44100	2000−6000	8	44100
SNR	95.46	–	–	93.52
ODG	-0.2043	-0.6—-1.7	-3—-1	-0.11—-0.47

By comparing capacity and imperceptibility of this scheme with existing methods, we can see that the proposed schemes has a greater advantage in imperceptibility and capacity than schemes in [17, 18]. Although our scheme and schemes in [11] have the similar capacity and imperceptibility, but considering that our scheme has better robustness against noise attacks, which is still a worth considering scheme.

6. **Conclusion.** This paper proposed an audio steganography scheme based on wavelet domain centroid. The secret information is embedded through XORing four bits of low frequency wavelet coefficients. And when the XOR result is not equal to binary secret information, one bit will be selected and then be flipped according to the centroid. Experiments show that the proposed scheme has good robustness. Moreover, compared with existing schemes, it is evident that the scheme not only has large capacity, but also has good imperceptibility. Although this method has the disadvantage of poor resistance to re-quantization of a 16-bit stego audio to 8-bit and back to 16-bit and up-resampling, in summary, it is still a practical audio steganography scheme.

Acknowledgement. The authors would like to thank the anonymous reviewers for their valuable comments. This work is supported by the National Natural Science Foundation of China (Grant Number: 61602491), the Key Program of the National University of

Defense Technology (Grant Number: ZK-17-02-07), and Natural Science Foundation of Guangdong Province, China (Grant Number: 2015A030310172).

REFERENCES

[1] C. Liu, T. B. K. Liu, *The three-body problem*, Macmillan.

[2] S. S. Divya, M. R. M. Reddy, *Hiding text in audio using multiple lsb steganography and provide security using cryptography.*

[3] S. Shirali-Shahreza, M.T. Manzuri-Shalmani, Adaptive wavelet domain audio steganography with high capacity and low error rate, *In: International Conference on Information and Emerging Technologies*, pp. 1 – 5, 2007.

[4] N. Cvejic, T. Seppanen, Increasing robustness of lsb audio steganography using a novel embedding method, *In: International Conference on Information Technology: Coding and Computing, 2004. Proceedings. Itcc.*, vol.2, pp.533–537, 2004.

[5] T.C. Chen, W.C. Wu, Highly robust, secure, and perceptual-quality echo hiding scheme, *Audio Speech & Language Processing IEEE Transactions on*, vol.16, no.3, pp. 629–638, 2008.

[6] Rupanshi, Preeti, Vandana, Audio steganography by direct sequence spread spectrum, *International Journal of Computer Trends & Technology*, vol.13, no.2, 2014.

[7] M. Chen, R. Zhang, F.F. Liu, X.X. Niu, Y.X. Yang, Audio steganography by quantization index modulation in the dct domain, *Journal on Communications*, vol.30, no.8, pp.105–111, 2009.

[8] M. Sheikhan, K. Asadollahi, R. Shahnazi, Improvement of embedding capacity and quality of dwt-based audio steganography systems, *World Applied Sciences Journal*, vol.13, no.3, 2011.

[9] M.Zhao, J. S. Pan and S. T. Chen, Optimal SNR of Audio Watermarking by Wavelet and Compact PSO Methods, *Journal of Information Hiding and Multimedia Signal Processing*, vol.6, no.5, pp833–846, 2015.

[10] N. Cvejic, T. Seppanen, Increasing the capacity of lsb-based audio steganography, *In: Multimedia Signal Processing, 2002 IEEE Workshop on*, pp. 336–338, 2002.

[11] K. Bhowal, D.C. Sarkar, S. Biswas, P.P. Sarkar, A steganographic approach to hide information in audio signal using discrete wavelet transforms, 2016.

[12] S. Krishnan, M. S. Abdullah, Enhanced security audio steganography by using higher least significant bit.

[13] X. Zhang, X. Yin, Audio Watermarking Algorithm Based on Centroid and Statistical Features. Springer, Berlin-Heidelberg, Germany, 2007.

[14] H.X. Wang, M.Q. Fan, Centroid-based semi-fragile audio watermarking in hybrid domain, *Science China* , vol.53, no.3, pp. 619–633, 2010.

[15] T. Thiede, Peaq—the itu standard for objective measurement of perceived audio quality, *Journal of the Audio Engineering Society Audio Engineering Society*, vol.48, no.1, pp. 3–29, 2000.

[16] A. Totok, Progressive watermarking techniques using genetic algorithms, *http://cs.nyu.edu/totok/professional/software/tpcw/tpcw.html* , 2002.

[17] M. Fallahpour, D. Megas, High Capacity Method for Real-Time Audio Data Hiding Using the FFT Transform, *Advances in Information Security and Its Application*, 2009.

[18] A. Nishimura, Audio data hiding that is robust with respect to aerial transmission and speech codecs, *In: International Conference on Intelligent Information Hiding and*, pp.1389–1400, 2010.

[19] W. Qiuling, , W.o.E..I.T. Meng, A new method of voice information hiding based on wavelet transform, *Journal of Electronics & Information Technology*, vol.38, no.4, pp.834–840, 2016.

[20] L. Tan, B. Wu, Z. Liu, M.T. Zhou, An audio information hiding algorithm with high-capacity which based on chaotic and wavelet transform, *Acta Electronica Sinica*, vol.38, no.8, pp.1812–1811, 2010.