World Applied Sciences Journal 10(12): 1501-1507, 2010
ISSN 1818-4952

High Quality Audio Steganography by Floating Substitution of LSBS in Wavelet Domain

Mansour Sheikhan, Kazem Asadollahi and Ehsan Hemmati

Electrical Engineering Department, Islamic Azad University, South Tehran Branch, Tehran, Iran

Abstract: Steganography is the art of hiding information in a host like audio, image, text and video to secure data transmission over Internet. In other words, in steganography the secret message is embedded in the cover object and transmitted in such a way that the existence of information is undetectable. Various methods have been proposed for audio steganography using different domains like time, frequency and wavelet. In this paper, an audio steganography method is proposed which uses packet wavelet subband coefficients to hide messages. In this way, message signal is embedded in least significant bits (LSBs) of wavelet coefficients that are selected based on a new method suggested in this work. Inverse of packet wavelet decomposition is applied to modified coefficients to result the stego audio signal in time domain. This method has acceptable capacity and is also efficient in terms of signal to noise ratio (SNR) and mean opinion score (MOS).

Key words: Steganography · Audio signal · Packet wavelet decomposition · LSB

INTRODUCTION

The technology and networking development has driven the interest among computer security researchers to overcome the threats for secured data transmission. Information hiding in digital audio can be used for such diverse applications as proof of ownership, authentication, integrity, secret communication, broadcast monitoring and event annotation. There are two well-known special cases of information hiding: digital watermarking and steganography [1].

The steganography is derived from the Greek words and means "covered writing". In steganography, the secret message is embedded in the covert object or host and transmitted in such a way that the existence of information is undetectable. Various hosts can be used for steganography, such as digital images, videos, sound files and other computer files. In other words, the main purpose of steganography is to hide information in a media such as audio or image signal in such a way that only the intended receiver can recover the secret message. The setanography makes the presence of secret data appear invisible to eaves droppers. This is the main difference between steganography and watermarking [2]. Also, in watermarking the embedded message is a short one, but robust against signal processing techniques such as filtering and compression [3]. But in steganography, there is no such a restriction about the message size and generally high embedding capacity is needed. However, steganography is not robust against signal processing techniques, as compared to watermarking.

In audio steganography, the behavior of human auditory system (HAS) is considered, when information is hidden in the audio [4]. The main aspect of steganography is to achieve high capacity, security and robustness.

The steganography implementation methods can be classified as follow:

- Time-domain substitution, in which only the least significant bits (LSBs) of the host, is replaced by message bits without significant affect on the host.
- Transform-domain substitution, in which various transform domains such as fast Fourier transform (FFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT) are used to hide information in transform coefficients of the covert object [5].
- Spread spectrum, in which the message is spread over a wide frequency bandwidth than the minimum required bandwidth to send the information.
- Statistical, in which the cover is divided into blocks and the message bits are hidden in each block.

Corresponding Author: Dr. Mansour Sheikhan, P.O. Box 11365/4435, Post-Graduate Center, South Tehran Branch, Islamic Azad University, Iran.
Tel: +98-21-88605565, Fax: +98 21-33717140, E-mail: msheikhn@azad.ac.ir.

- Distortion, in which the information is stored by signal distortion. The encoder adds sequence of changes to the cover and the decoder checks for the various differences between the original cover and the distorted cover to recover the secret message.

Time-domain substitution techniques are not robust against simple processing techniques, like compression and transform. So, we focus on transform-domain substitution method in this study. However, this method has its advantages and disadvantages. For example, in frequency or wavelet transform domains, the capacity is higher. But, it has problem in extracting the message from the host, due to the generated errors that result in low quality. To overcome this problem, we propose a modified version of substitution technique (floating LSB) in wavelet domain, in which the error is minimized and high quality in terms of signal to noise ratio (SNR) and mean opinion score (MOS) is achieved.

Related Works: Different methods have been proposed for hiding information in digital hosts, such as audio signals. Most of these methods make use of HAS weaknesses.

Many researches are based on LSB method [4-6]. When this method is used along with wavelets [5] or with minimum error replacement (MER) technique [6], higher quality and capacity of steganography are achieved. Comparing LSB technique in time domain with wavelet and Fourier domains shows that more bits can be hidden in frequency domain for similar SNR conditions.

In this way, a high bit rate LSB audio data hiding method, that reduced embedding distortion of the host audio, has been proposed in [6]. This paper proposed a two-step algorithm; in which the hidden bits has been embedded into the higher LSB layers, resulting in increased robustness against noise addition. The key idea of the algorithm is watermark bit embedding that causes minimal embedding distortion of the host audio.

In [5], a method for digital audio steganography, which encrypted covert data is embedded into the wavelet coefficients of host audio signal, has been proposed. To avoid the extraction error, lifting wavelet transform (LWT) has been used. For using the maximum capacity of audio signals, hearing threshold has been calculated in wavelet domain. Then according to this threshold, data bits have been embedded in the least significant bits of lifting wavelet coefficients. Inverse lifting wavelet transform has been applied to modified coefficients to construct stego signal in time domain.

Also, high capacity and security steganography using discrete wavelet transform (HCSSD) has been proposed in [7]. The wavelet coefficients of both the covert and secret message have been fused into a single image. The covert and secret message has been preprocessed to reduce the pixel range to ensure the accurate recovery of message at the destination. It has been observed that the capacity and security increased with acceptable peak SNR (PSNR) in the proposed algorithm as compared to the existing algorithms.

As noted earlier, the capacity of the classic LSB insertion method can be increased by performing the embedding process in the wavelet domain. In this way, in [8] an algorithm has been proposed which uses perfect reconstruction filter banks and embedded additional information inside wavelet domain of audio signal by modifying LSB of wavelet coefficients.

As a hybrid approach, simultaneous low bit rate encoding and information hiding for highly compressed audio signals has been reported in [9]. The tests with an extended MPEG4 advanced audio coding (AAC) encoder confirm the robustness of the method.

On the other hand, the most important type of threat is the potential for concealing steganographic writing within computerized images and audios [10, 11]. The best way to guard against all the mentioned hazards of modern steganography is "steganalysis". In this way, a practical forensic steganalysis tool has been proposed in [10], which can properly analyze the statistics disturbed by stego embedding and classify them to selected current steganographic methods.

Also, detecting the presence of LSB steganographic messages in a voice secure communication system has been reported in [12]. In this way, hiding the text messages in wave files in frequency domain format using FFT technique has been reported in [13]. Substitution of secret speech data bits for line prediction coefficients (LPCs) has been reported in [14], as well.

Wavelet Transform: In this section, we briefly introduce the discrete and packet wavelet transforms. The discrete wavelet transform (DWT) is a discipline capable of giving time-frequency representation of signal. Starting from the original audio signal *S*, DWT produces two sets of coefficients (Fig. 1): the approximation coefficients *A* (low frequencies) and the detail coefficients *D* (high frequencies).

Depending on the application and also the duration of signal, the low frequencies part might be further decomposed into two parts of high and low frequencies.

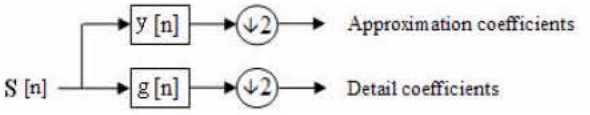

Fig. 1: One-level DWT decomposition

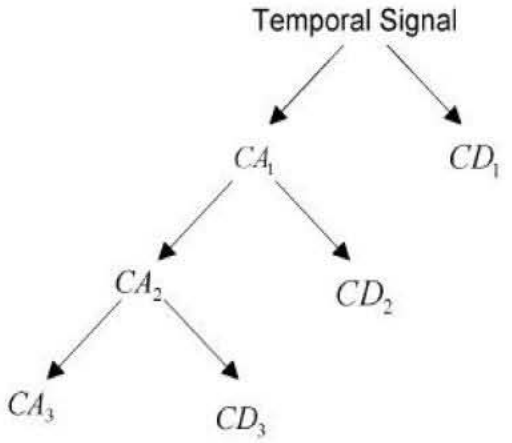

Fig. 2: Three-level DWT decomposition

Fig. 2 shows a 3-level DWT decomposition of signal S. As can be seen, the output will be a string with 4 subbands. Each of these belongs to a particular resolution level of wavelet transform. The original signal S can be reconstructed using the inverse DWT process.

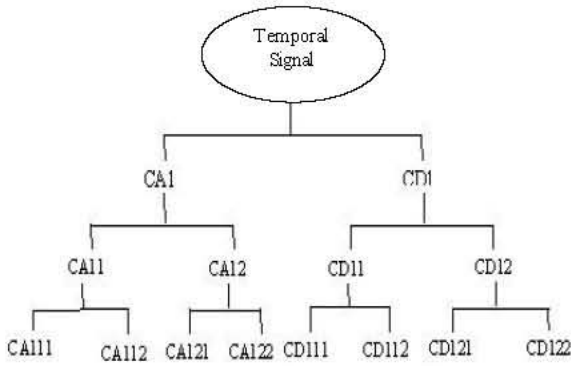

Fig. 3: Packet wavelet implementation in 3 resolution levels

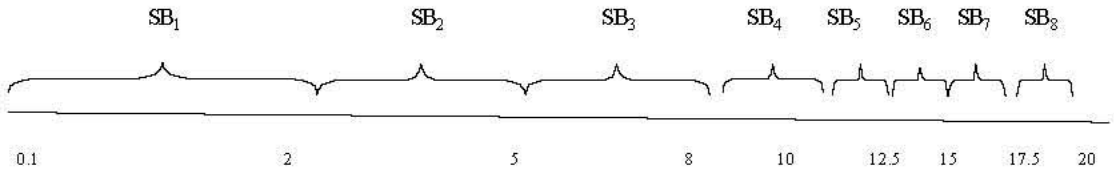

Fig. 4: Frequency segments for wavelet subbands (SBs) (f/kHz)

On the other hand, packet wavelet decomposition (PWD) (sometimes known as just wavelet packets) is a wavelet transform where the signal is passed through more filters than the DWT. In the DWT, each level is calculated by passing only the previous approximation coefficients (CA_j) through low and high pass filters. However, in the PWD, both the detail and approximation coefficients are decomposed (Fig. 3).

Proposed Algorithm: To do steganography, the subbands which have lower energy, as compared to audio threshold, are extracted from wavelet analysis. For this purpose, we use packet wavelet transform in such a way that CA_1 and CD_1 coefficients are extracted and again this analysis is applied to CA_1 and CD_1 to achieve CA_{11}, CA_{12}, CD_{11} and CD_{12}. This process is repeated up to one more level. The frequencies that form an audio signal are ranged between 20 Hz to 20 kHz. When wavelet transform is applied to audio signal, it is divided to separate frequency segments by using high pass and low pass filters. The frequency segments for each wavelet subband in this work are shown in Fig. 4. As can be seen, the first subband has the largest frequency range and most of the signal energy.

On the other hand, energy level calculation of each subband shows that the maximum level of audio signal energy is in the first subband, which is about 70% of total energy [2]. As shown in Fig. 5, the frequencies which are perceivable by human ear with low sound pressure level (SPL) are in the range of 20 Hz to 2 kHz [15]. In the present work, this is called subband 1. The other seven subbands are placed in 2 kHz to 20 kHz frequency range.

Due to the HAS behavior, small changes of high subbands energy level has no significant effect on listener. In this work, we assume that up to 2% modification in spectrum magnitude of the mentioned subbands is permissible. The block diagram of the proposed method is shown in Fig. 6.

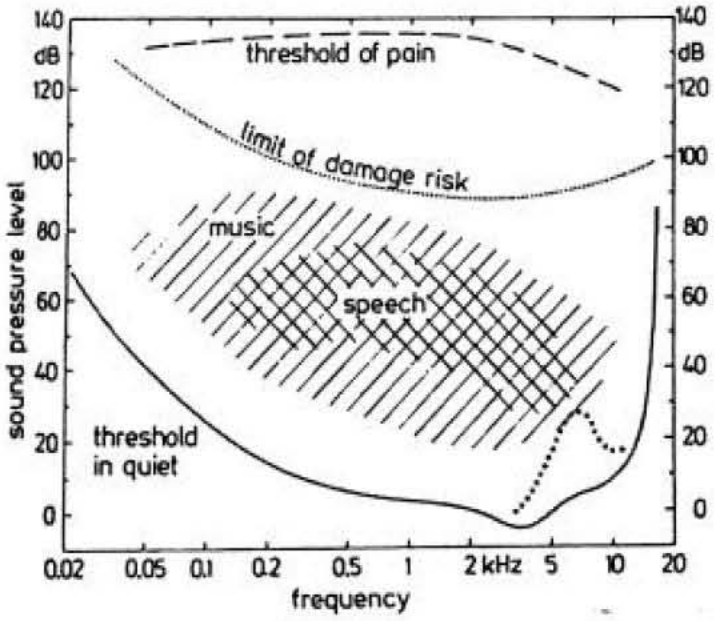

Fig. 5: Human hearing thresholds [15]

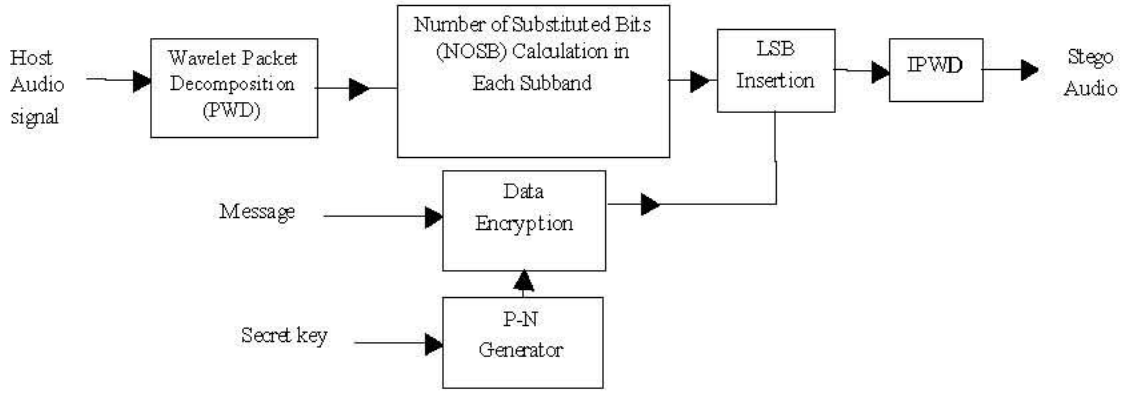

Fig. 6: Block diagram of proposed steganography method

At the first block, PWD is applied to the host signal. This transform is performed up to the third level. Among different investigated wavelet functions, Haar wavelet function achieves higher performance, in terms of SNR. It is noted that, Haar wavelet is a piecewise wavelet that provides orthogonal decomposition.

In this work, the number of bits used for substitution in each subband is floating; rather than fixed and is calculated using the following equation:

$$NOSB = \text{round}\left(\log_2\left(\text{subband energy level} \times 0.02\right)\right) \quad (1)$$

in which (subband energy level×0.02) is the permissible fluctuation in the energy of each subband, which is unperceivable by human ear. So, the number of bits for each subband is different depending on its energy. Furthermore, in order to increase the security, data encryption is also proposed by using a pseudo-noise (P-N) string.

So, the steps of data hiding algorithm are as follows:

- Selecting host and message audio signals;
- Wavelet transforming up to the third level to have 2^L subbands (in this paper, L=3) and extracting subbands CA_{112} to CD_{122} (Fig. 3);
- Calculating the number of bits for substitution in each of upper 7 subbands;
- Data encryption using P-N string;

- Embedding message signal in the host signal with the specified number of bits;
- Applying inverse PWD to achieve stego audio signal.

Experimental Results: To measure the performance of proposed method, 6 music and 2 speech signals are used. The sampling rate is set to 44100 samples per second each sample is encoded in 16 bits. The duration of these audio signals are 1 to 10 sec.

Voice quality measurement can be carried out using either subjective or objective methods. MOS is the most widely used subjective measure of voice quality and is recommended by the International Telecommunication Union (ITU) [16]. A MOS value is normally obtained as an average opinion of quality based on asking people to grade the quality of speech signals on a five-point scale (Excellent, Good, Fair, Poor and Bad) under controlled conditions.

On the other hand, SNR is an important factor in determining the quality of audio data, as an objective measure. This is particularly important in speech recognition [17, 18] and coding technology [19-21], since it is well known that recognition performance is strongly influenced by the SNR [22]:

$$SNR = 10\log\left(\frac{\sum_{n} x^2(n)}{\sum_{n} (x(n) - y(n))^2}\right) \quad (2)$$

where x(n) is the input signal to system and y(n) is the output signal.

The steganography signal and main host signal are played for 15 listeners repeatedly and scored by them. The average value of SNR, MOS and embedding capacity are reported in Table 1.

The performance of the proposed method is compared with some other similar works in Table 2. As can be seen, the proposed method has the highest average SNR. But, its capacity is low.

There are two differences between the proposed method and its competitive method, reported in [5]:

- Steganography has been performed for 32 subbands with different number of substituted bits for all of the subbands in [5], but it is performed for 7 selected subbands in the proposed method;
- Steganography is not performed for silence parts of audio signal in the proposed method, but it has been performed for all of the audio frames in [5].

If the duration of silence parts is long, then the proposed method shows its preference, especially in speech signals.

The substitution specifications in different subbands of the proposed method and 32 subbands reported in [5] are listed in Table 3.

As can be seen, the average number of substituted bits in the two methods is roughly the same for some of the subbands. However, the quality of proposed method is higher.

Table 1: Performance of proposed audio steganography method

Host signal	Average SNR (dB)	MOS	Embedding capacity (%)
Song of a woman-music No. 1	51.83	4.9	18.39
Song of a woman-music No. 2	46.70	4.7	21.51
Song of a man-music No. 1	74.83	4.9	19.02
Song of a man-music No. 2	45.46	4.5	17.37
Pop music No. 1	54.08	4.7	9.90
Pop music No. 2	52.25	4.9	11.46
Speech No. 1	49.50	4.8	9.50
Speech No. 2	49.98	4.7	9.60

Table 2: Performance comparison of proposed method with some other similar methods

Researchers/Method	Average SNR (dB)	Embedding capacity (%)
Pooyan and Delforouzi [5]	39	20.0
Cvejic and Seppanen [6]	42	17.5
Cvejic and Seppanen [23]	39	32.0
Bao and Ma [24]	36	2.0
Proposed in this paper	49	14.3

Table 3: Substitution specifications in different subbands of the proposed method and a similar work [5]

Subbands [5]	Average No. of substituted bits in subbands [5]	Subband in the proposed method	Average No. of substituted bits in subband of the proposed method
5 to 8	4.25	2	4.6
9 to 12	5.25	3	4.5
13 to 16	5.00	4	6.1
17 to 20	5.25	5	4.0
21 to 24	5.75	6	5.2
25 to 28	6.00	7	5.8
29 to 32	7.25	8	7.2

CONCLUSION

In this paper, an audio steganography method has been proposed which uses packet wavelet subband coefficients to hide messages. For this purpose, message signal has been embedded in LSBs of packet wavelet coefficients. The performance of proposed method is compared, in terms of embedding capacity, SNR and MOS, with some other similar works. Experimental results have shown that the proposed method has high quality in terms of average SNR and MOS, when the tests are performed on music and speech signals. This method offeres acceptable capacity, as well. In addition, by using a modified version of substitution technique (floating LSB) we can overcome the weakness of substitution methods in resisting even simple attacks, such as compression.

REFERENCES

1. Özer, H., B. Sankur, N. Memon and I. Avcıbas, 2006, Detection of Audio Covert Channels Using Statistical Footprints of Hidden Messages. Digital Signal Processing, 16: 389-401.
2. Taanuja, T.C. and R.N. Agaraj, 2008. Schemes for Evaluating Signal Processing Properties of Audio Watermarking. Intl. J. Computer Science and Network Security, 8: 242-246.
3. Akhaee, M.A., N. Khademi Kalantari and F. Marvasti, 2010. Robust Audio and Speech Watermarking Using Gaussian and Laplacian Modeling. Signal Processing, 90: 2487-2497.
4. Gopalan, K., 2003. Audio Steganography Using Bit Modification. In the Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2: 421-424.
5. Pooyan, M. and A. Delforouzi, 2007. LSB-Based Audio Steganography Method Based on Lifting Wavelet Transform. In the Proceedings of IEEE Intl. Symposium on Signal Processing and Information Technol., pp: 600-603.
6. Cvejic, N. and T. Seppanen, 2004. Increasing Robustness of LSB Audio Steganography Using a Novel Embedding Method. In the Proceedings of IEEE International Conference on Information Technology: Coding and Computing, 2: 533-537.
7. Majunatha Reddy, H.S. and K.B. Raja, 2009. High Capacity and Security Steganography Using Discrete Wavelet Transform. Intl. J. Computer Sci. Security, 3: 462-472.
8. Cvejic, N. and T. Seppanen, 2002. Increasing Capacity of LSB-Based Audio Steganography. In the Proceedings of IEEE Workshop on Multimedia Signal Processing, pp: 336-338.
9. Xu, S., P. Zhang, P. Wang and H. Yang, 2009. Performance Analysis of Data Hiding in MPEG-4 AAC Audio. Tsinghua Sci. Technol., 14: 55-61.
10. Geetha, S., N. Ishwarya and N. Kamaraj, 2010. Audio Steganalysis with Hausdorff Distance Higher Order Statistics Using a Rule Based Decision Tree Paradigm. Expert Sys. with Appl., 37: 7469-7482.
11. Geetha, S., N. Ishwarya and N. Kamaraj, 2010. Evolving Decision Tree Rule Based System for Audio Stego Anomalies Detection Based on Hausdorff Distance Statistics. Information Sci., 180: 2540-2559.
12. Deng, Z.Y., X. Shao and Z. Yang, 2007. LSB Steganalysis of Speech Data Based on Distance Measure and Ml Decision. J. China Universities of Posts and Telecommun., 14: 103-107.
13. Viswanathan, V., 2008. Information Hiding in Wave Files through Frequency Domain. Appl. Mathematics and Computation, 201: 121-127.
14. Wu, Z.J., W. Gao and W. Yang, 2009. LPC Parameters Substitution for Speech Information Hiding. J. China Universities of Posts and Telecommun., 16: 103-112.
15. Zwicker, E. and H. Fastl, 1999. Psychoacoustics: Facts and Model. Springer-Verlag, Berlin.

16. ITU-T, P.800 Recommendation, 1996. Methods for Subjective Determination of Transmission Quality.
17. Sheikhan, M., M. Tebyani and M. Lotfizad, 1997. Continuous Speech Recognition and Syntactic Processing in Iranian Farsi Language. Intl. J. Speech Technol., 1: 135-141.
18. Sheikhan, M., 2003. Suboptimum Extracted Features and Classifier for Speaker-Independent Farsi Digit Recognizer. In the Proceedings of the Intl. Symposium on Telecommun., pp: 246-249.
19. Sheikhan, M., V. Tabataba Vakili and S. Garoucy, 2009. Complexity Reduction of LD-CELP Speech Coding in Prediction of Gain Using Neural Networks. World Appl. Sci. J., 7(Special Issue of Computer and IT): 38-44.
20. Sheikhan, M., V. Tabataba Vakili and S. Garoucy, 2009. Codebook Search in LD-CELP Speech Coding Algorithmin Based on Multi-SOM Structure. World Appl. Sci. J., 7(Special Issue of Computer and IT): 59-68.
21. Sheikhan, M. and S. Garoucy, 2010. Reducing the Codebook Search Time in G.728 Speech Coder Using Fuzzy ARTMAP Neural Networks. World Appl. Sci. J., 8: 1260-1266.
22. Deller, J.R., J.H.L. Hansen and J.G. Proakis, 2000. Discrete-Time Processing of Speech Signals, IEEE Press.
23. Cvejic, N. and T. Seppanen, 2002. A Wavelet Domain LSB Insertion Algorithm for High Capacity Audio Steganograghy. In the Proceedings of IEEE Workshop on Digital Signal Processing, pp: 53-55.
24. Bao, P. and X. Ma, 2004. MP3-Resistant Music Steganography Based on Dynamic Range Transform. In the Proceedings of IEEE International Symposium on Intelligent Signal Processing and Commun. Sys., pp: 266-271.