# High Capacity and Resistance to Additive Noise Audio Steganography Algorithm

Haider Shahadi<sup>1</sup>, Razali Jidin<sup>2</sup>

<sup>1</sup> Electronic and Communication Department, University Tenaga National, 43000 kajang, Selangor, Malaysia

<sup>2</sup> Electronic and Communication Department, University Tenaga National, 43000 kajang, Selangor, Malaysia

#### Abstract

Steganography is the art of message hiding in a cover signal without attracting attention. The requirements of the good steganography algorithm are security, capacity, robustness and imperceptibility, all them are contradictory, therefore, satisfying all together is not easy especially in audio cover signal because human auditory system (HAS) has high sensitivity to audio modification. In this paper, we proposed a high capacity audio steganography algorithm with good resistance to additive noise. The proposed algorithm is based on wavelet packet transform and blocks matching. It has capacity above 35% of the input audio file size with acceptable signal to noise ratio. Also, it is resistance to additive Gaussian noise to about 25 db. Furthermore, the reconstruction of actual secret messages does not require the original cover audio signal.

**Keywords:** Audio Steganography, Wavelet Packet Transform, Block Matching, Robust Stego-algorithm, High capacity Steganography.

# **1. Introduction**

The rapid development in information technology, especially in the field of computer networks and wireless communication has enabled hackers to eavesdrop on communication. Eavesdropping can be prevented using techniques of cryptography or/and steganography [1]. Steganography means embedding information in a cover signal to prevent illegal detection [2]. Steganography hides secret messages, while in cryptography, secret message is visible, therefore unauthorized users can access to the secret encrypted data [3].

Steganography is often used in covert communication such as military and government communication and it requires relatively high payload, while a watermarking which is often used in copyright protection however does not require high capacity [4]. The major requirements that must be satisfied for good steganography algorithms are perceptual transparency, payload or capacity and robustness [5]. Steganography, unlike watermarking, needs high capacity and is considered as an important factor, but robustness should be dominant for watermarking. Higher targets for one of the mentioned requirements will negatively affect the other performances for the same stego-algorithm as they are contradictory according to magic triangle [6].

In recently years many techniques appeared to develop information hiding [4, 7, 8], and most techniques used image and video media but lacking of using audio signal as a cover signal especially in high rate data embedding, most likely due to Human Auditory System (HAS) is more sensitive than Human Visual System (HVS) [9]. Although adopting audio signals as cover signals may yield inferior inaudible performance, there are still suitable features such as transitory and unpredictability that makes sound signal as suitable secure cover signal.

In this paper we proposed a new steganography algorithm that has high capacity and resistance to additive noise characteristics, and it is based on both discrete wavelet packet transform (DWPT) and blocks matching. The proposed algorithm is based on matching between the cover and message blocks. Then it uses blocks scaling and replacement in embedded process to keep power of embedded message high in order to resist the additive noise. The output stego-signal has an acceptable quality with high capacity that can reach more than 35% of the input audio size (about 250 kb/sec). The message recovery process does not require original audio cover signal, and also the algorithm provides high security due to the arbitrary distribution of the secret message blocks in embedding process instead a known embedded message distribution in most stego-algorithms.

# 2. Related Work

Generally audio steganography and watermarking can be classified according to embedding domain either time domain or transform domain. The common and simplest hiding technique in time domain with acceptable capacity is the Least Significant Bits (LSB), but it is vulnerable due to any changes in LSB that can destroy the embedded message [6]. In transform domain, there are many transform methods can be employed in information hiding such as Fourier domain [6, 10], discrete cosine domain [6, 11], and wavelet domain [6, 9, 12, 13]. Each domain has its features in signal processing and in information hiding [14], however, the wavelet domain has a major advantage over the others because it divides a signal into different frequency components with different resolution and each component can be used in embedded process according to its power. DWT behaves similar to the time-frequency characteristics of human ears that has high time and low frequency resolution for high frequencies, and high frequency and low time resolutions for the low frequencies components [15].

In general, Discrete Wavelet Transform (DWT) decomposes a signal at a level of decomposition into two components, high and low frequency components. Most power of the input signal is concentrated in low frequency component or called approximation signal, while little power spread over coefficients of high frequency component or called details signal. The decomposition process starts by decompose a signal into two components, high and low frequency components, then will further decompose each the low frequency component into another two components, and the process repeats for further levels of decomposition [15]. The reconstruction of original signal is performed by the Inverse Discrete Wavelet Transform (IDWT), and it bases on approximation and details signals.

The modification in the details signals has little effect on the reconstructed signal, rather depends on the number of levels of decomposition and which selected details level was modified. However modification in the approximation signal or low frequency component may affect significantly on the reconstructed signal. Therefore using details signals as a cover for information embedded process enable high payload and acceptable quality, when it is used in steganography [9]. However, information embedding in details signal has weakness in robustness for watermarking applications such as visible watermarking of TV channels and copyright protection because it is possible to remove a secret message by signal processing attackers for example by resetting details coefficients. Therefore in watermarking applications, it is better to use low frequency component in embedding process because watermarking does not need high capacity. Watermarking however, requires robust algorithm to embed information as a part of the cover signal, render it almost impossible to be removed by attackers [12, 16].

The LSB is the most common method employed in embedding process in DWT domain [9, 13, and 17]. This method has superior stego-signal quality and capacity, but it is sensitive to additive noise especially when data is embedded in high frequencies components which have low power.

In this paper a high capacity and noise resistance audio steganography algorithm will be described. The algorithm uses Discrete Wavelet Packet Transform (DWPT) to decompose audio cover signal. The DWPT is similar to DWT except decomposition is performed for both high frequency and low frequency components. The decomposed signal by DWPT for selected *L* levels, yields  $(2^L)$  components with equal lengths, but only one represents the approximation signal that has the highest power, while the others are details signals with decreasing power, starting from lowest frequency details component to the highest frequencies details component. Subsequently after several steps, the Inverse of Discrete Wavelet Packet Transform (IDWPT) is used to reconstruct the original cover signal.

DWPT decomposition provides two benefits, first, separation the signal components according to their frequency and power, and second, making all the components in the same length and that will facilitate the task of block matching which to be used in our proposed scheme. The proposed scheme however does not use the approximation signal in embedding process to maintain the quality of output of stego-signal. The lowest frequency component of the details signal which has highest power with respect to other details signal is used to embed main key that is generated in matching process. The remaining  $(2^L - 2)$  details signals are arranged as blocks and they are used in message embedding process after block matching.

The proposed block matching is based on minimum distance between blocks of secret message after primary scaling. The primary scaling will make the amplitudes of the message signal within the range of the amplitudes of the cover signal. The block matching output is used to generate embedding keys that will be used later in both embedding and extraction processes. The embedding key is encrypted using stream cipher and then it is embedded in lowest frequency details signal as shown in section (3.5).

The proposed algorithm uses block replacement after matching and scaling in the embedding process instead of LSB to satisfy three contradictory things simultaneously, firstly, to maintain message power high to resist the additive noise, second, to keep quality of modified cover signal closes to original input signal, and thirdly, to provide high rate of payload instead of less bits in LSB method. Another advantage for this algorithm over most algorithms is that it cannot estimate the positions of embedded secret message because the blocks are distributed arbitrary according to minimum distance between embedding and replacing blocks and this feature make secret message more protected against detection and tampering. Furthermore the secret message recovery algorithm is does not need the original audio cover signal.

#### 3. Proposed Audio Steganography Scheme

Fig. 1 shows the general block diagram of the proposed hiding algorithm. The main stages of the algorithm are as follow:

#### 3.1 Cover signal decomposition

The input cover audio signal  $C = \{c \ (i) \ 0 \le i \le Z, Z\}$ number of samples is decomposed by using DWPT to L level. Haar filter was chosen in decomposition process because of it is a quadrature mirror filter type and it has finite impulse response. The output from DWPT decomposition stage are  $2^{L}$  signals, each signal has  $Z/2^{L}$ sample. These signals have different power and different frequencies. One of them which has the greatest power and lowest frequency represents the approximation signal A={a (i),  $0 < i \leq Z/2^{L}$ }, this signal is kept without any modification to maintain the quality of output stego-signal. The other signals have decreasing powers starting from lowest to highest frequencies details components D = $\{d_i(i), j=1,2,..2^{L}-1, 0 < i \le Z/2^{L}\}$ . Except for the lowest frequency component, other details signal are arranged in the 3D matrix named Det.

 $Det = \{ \det_{i,k}(i) = 1, 2, \dots, V, k = 1, 2, \dots, W, i = 1, 2, \dots, N \}$ (1)

Where  $V = 2^{L} \cdot 2$  represents the number of details signals that are being used in the embedding process,  $W = Z'(2^{L}N)$ represents number of blocks in each details signal, and Nrepresents number of samples in each block. This matrix will be fed to block matching process for information embedding process while keeping the lowest frequency details component  $(D_2^{L}_{-1})$  to be used for embedding the key.

#### 3. 2 Secret message pre-processing

The secret message M pre-processing depends on the type of the message. If it is an audio signal has number of bit per sample similar to cover signal, directly it is segmented to Q blocks each block has N samples as shown in Eq. (2).  $M = \{m(i, j), i = 1, 2, ..., Q, j = 1, 2, ..., N\}$  (2) If the secret message is an image or text, it will have 8 bits per pixel or character. While the audio cover signal for example has 16 bits per sample, therefore it will be combined each two neighbor pixels or character into one sample of 16 bits to reduce number of samples to half. After that arranging the produced message in a one dimension vector, and then segmenting the resultant vector into Q blocks, each block has N samples such as given in Eq. (2).

The produced matrix in Eq. (2) is scaled depends on *Det* matrix which is produced in Eq. (1) as follows:

$$df = mean(M) / mean(Det)$$
(3)

$$SM = M / df \tag{4}$$

Where df is a division factor of the primary scale to make message blocks samples within range of details signal matrix. The function *mean* is used to compute average of *M* and *Det* matrices, and *SM* is a scaled message matrix.

#### 3.3 Block matching Process

Block matching process construct the matching matrix by computing minimum distance between MS and Det blocks. The matching matrix (*Match*) is a 3D matrix that includes Q layers, where each layer contains  $V \times W$  elements. Each element represents a minimum distance between block i from the MS matrix and block k from details signal j according to Eq. (5).

$$match(i, j, k) = \frac{1}{N} \sum_{r=1}^{N} \sqrt{(MS_i(r) - Det_{j,k}(r))^2}$$
(5)

$$Match = \{match(i, j, k), i = 1, 2, \dots, Q, j = 1, 2, \dots, V, k = 1, 2, \dots, W\}$$
(6)

## 3.4 Key generating and Information Embedding

There are two functions in this stage, first, to generate the key that contains positions of embedding blocks and main scale factors between the message blocks and replacing cover blocks. Second function is embedding of the secret message blocks. The steps of this stage are as follows:

• For each layer *i* within the *Match* matrix, compute the positions  $m_{1,i}$ , and  $m_{2,i}$  for minimum distance which represent minimum element in a layer. The indices  $m_{1,i}$  and  $m_{2,i}$  give the position of the block in *Det* matrix

that has maximum matching with the block i from the scaled message matrix MS.

• Perform scaling between message block ( $MS_i$ ) and replacement block ( $Det_{ml,i}, M2,i$ ). This is done as follows:

$$sf_i = mean(MS_i - Det_{m_{i,i}, m_{i,j}})$$
<sup>(7)</sup>

$$mss_{i,j} = ms_{i,j} - sf_i \tag{8}$$

$$MSS_i = \{mss_i(j), j = 1, 2, \dots, N\}$$
 (9)

Where  $sf_i$  is the subtraction factor that is obtained by computing the average error between  $MS_i$  and  $Det_{m_{i,i},m_{2,i}}$ 

blocks, where  $ms_{i,j}$  represents sample *j* from block  $MS_i$ ,  $mss_{i,j}$  represents sample *j* from scaled block  $MSS_i$ .

• Construct main key vector for each message block *i* as follows:

 $Key_i = [m_{1,i}, m_{2,i}, sf_i]$ (10)

- Replace elements of block  $Det_{m_{1,i},m_{2,i}}$  by elements of block  $MSS_i$ .
- Modify each element that has position  $m_{1,i}$ ,  $m_{2,i}$  in *Match* matrix layers i+1 to Q equal to large number (for example 10<sup>6</sup>) to avoid embedding overwrite if there is any similarity in matching.
- Repeat all the above steps Q times until embedding process of all message blocks completed.

Subsequently, the main key matrix (Key = { $Key_i(j), j = 1, 2, ..., Q$ }) and modified details matrix (*Det*) will be generated.

#### 3.5 Main key encryption and embedding

The generated key in previous stage is encrypted and embedded in the lowest frequency details component  $(D_2^{L}I)$ . We chose this component for embedding the main key matrix because it has power higher than other details signals and therefore the embedded key will be more resistance to distortion or lost. In the proposed scheme any type of encryption can be used, either secret or public encryption.

In this work we use a simple stream cipher based pseudo random number generator (*PRN*) to produce a random number vector that has length of  $3 \times Q$  samples, same length of the main key vector that was produced by converting *Key* matrix to 1D vector as shown in Fig 2. Next, convert both vectors to binary vectors with resolution of 16 bit per sample to obtain random and key binary vectors. The encrypted key (*Enkey*) vector with length of  $3 \times 16 \times Q$  bits is produced by making XOR between two binary vectors. The encrypted key vector is embedded by using the same embedding technique which was used in [9] with some modification to be resistance to the additive noise. All the lowest frequency details component samples will be converted to integers with scale based on the maximum value. In other word the maximum value will be the highest 16-bit integer. Then for each of these integers  $d_{i}$ , we need to find the biggest  $P_i$  which satisfies the following inequality

$$2^{P_i} < d_i < 2^{P_i+1} \tag{11}$$

For each sample *i* of the above integers, we have  $nb_i$  bits that to be used for embedding encrypted key. We can compute  $nb_i$  according to Eq. (12).

$$nb_i = P_i - sb_1 - sb_2 \tag{12}$$

Where  $sb_1$  is safety bits with direction from the most significant bit to maintain good output quality and  $sb_2$  is the safety bits considered from the least significant bit in order to make embedded bits resistance to the additive noise. Any sample  $d_i$  that has zero or negative  $nb_i$  will not be used in the key embedding process.

#### 3.6 Stego-Signal Reconstruction

The final stage in the proposed hiding algorithm is the stego-signal reconstruction using Haar IDWPT. The reconstructed process uses modified Det matrix and modified  $D_2^{L}$ -1 vector as well as non-modified approximation signal as inputs to IDWPT to obtain on the output stego-signal.

## 4. Recovery Algorithm of Secret Message

Fig. 3 shows the general block diagram of the secret message recovery algorithm. The algorithm starts by decomposing stego-signal using Haar DWPT, then followed by rearrangement of all the details signal except lowest frequency details signal in 3D matrix *Det* similar to section (3.1) Eq. (1).

The lowest frequency details signal  $(D_2^{L}-I)$  enters the key recovery stage which starts by scaling the signal, then search for the largest power of 2 for each scaled sample with inequality (11) to be satisfied. Depend on the positive values of the  $nb_is$  which is computed using Eq. (12), the main key vector *Keyvec* is reconstructed using Eq. (13).

$$Keyvec = \{d_i(j), j = 1, ..., nb_i, i = 1, 2, ..., \frac{Z}{2^L}\}$$
(13)

Where  $d_i$  is a binary vector for *i* sample in vector  $D_2^{L-1}$  that has positive  $nb_i$ .

The key vector is deciphered in decryption stage using same key and procedure that was used in key encryption process in section (3.5) except main key vector *Keyvec* will not be converted to a 1D and binary because it is already a binary vector. The output from this stage is converted to 2D matrix to obtain the main key matrix (*Key*).

The main key and *Det* matrices are fed to message block recovery stage to reconstruct the secret message blocks as Eq. (14):

$$[m_{1,i}, m_{2,i}, sf_i] = Key_i$$

$$MSS_i = Det_{m_1, i, m_{2,i}}$$

$$MS_i = \{mss_i(j) + sf_i, j = 1, 2, \dots, N\}$$
(14)

Where  $mss_i$  (*j*) is the sample *j* from block *i* in *MSS* matrix. Finally, the secret message is scaled and arranged by inversing procedure in section (3.2) to obtain the required secret message.



Fig 1 General Block Diagram of Proposed Hiding Scheme



Fig. 2 Main Key Encryption Process



Fig. 3 Recovery Algorithm of Secret Message

# 5. Results and Discussion

The proposed algorithm was tested using three audio cover signals: male speaker, female speaker and music. Each signal has 16 bit per sample and 44100 samples/sec. Also three data type: audio, image (as given in Fig. 4), and text are used as secret messages in the tests. The quality of output signal in each test was computed using signal to noise ratio (*SNR*) according to Eq. (15).

$$SNR = 10\log_{10} \frac{\sum_{k=1}^{Z} C^{2}(k)}{\sum_{k=1}^{Z} [C(k) - C'(k)]^{2}}$$
(15)

Where C and C are input cover signal and output stegosignal respectively, Z represents the number of samples in each one of them.

The similarity between extracted secret message and original secret message was computed using normalized correlation NC between them according to Eq. (16).

$$NC(M,M') = \frac{\sum_{k=1}^{QN} M(k)M'(k)}{\sqrt{\sum_{k=1}^{QN} M(k)^2} \sqrt{\sum_{k=1}^{QN} M'(k)^2}}$$
(16)

Where M and M' are original and extracted secret message respectively, QN represents number of samples in each one of them.

The proposed scheme was tested for different hiding capacity and the results show that it has an acceptable SNR (35 db and above according to ref. [18]) with capacity about of 250 kb/sec. This capacity represents more than 35% from input audio file size. Fig. 5 show the relationship between SNR and embedding capacity for three different message type and three different cover signals. In all experiments of the testing, the following attributes were used: length of block N=100 sample/sec, Haar type DWPT, L=2, and  $sf_1=sf_2=2$  bits. From Fig. 5 we can recognize small differences in SNR with respect to capacity for different cover signal, and this is due to difference in frequencies. Lower frequency signal has better SNR for same capacity than higher frequency signal because it has more percentage power in its non-modified approximation signal.

The extracted secret message is recovered using message recovery algorithm in section (4). The similarity between extracted and original messages was computed using Eq. (16). Without adding noise to the stego-signal, the recovered message is exactly similar to the original message (NC (M, M') = 1) for different message types and different cover signals. The results show that the proposed algorithm has significant resistance to Additive White Gaussian Noise (AWGN). Fig. 6-8 compares AWGN resistivity of our proposed scheme to the LSB-based DWT scheme. In LSB-based DWT scheme, the secret message will lost nearly about 65 db of AWGN, while in the proposed scheme secret message can be recovered about 25 db of AWGN. This superior performance of the proposed scheme is due to the power of embedding message is much higher compared to that of the LSB-based DWT scheme.

#### **6-** Conclusions

We present in this paper a high capacity and AWGN resistance audio steganography algorithm based on DWPT and block matching. Because of the arbitrary block matching that depends on minimum distance between blocks, an arbitrary key is produced that made the proposed algorithm having strong security. The results show that ability of proposed algorithm to embed high capacity of data of different type that can reach to more

than 35% from the input audio file size with an acceptable *SNR*. Also it has significant resistance to AWGN. The proposed algorithm can recover a recognized secret message until up to 25 db of AWGN.









Fig. 5 The Relationship betweersigner and Embedding Capacity



Fig. 6 Relationship between AWGN and NC for Male Speaker Cover Signal







Fig. 7 Relationship between AWGN and *NC* for Female Speaker Cover Signal





Fig. 8 Relationship between AWGN and NC for Music Cover Signal

# References

- K. Kumar, K. Raja, R. Chhotaray, S. Pattnaik, "Steganography Based on Payload Transformation", IJCSI International Journal of Computer Science Issues, Vol. 8, No. 2, March, 2011, pp. 241-248.
- [2] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography", IEEE Security and Privacy Magazine, Vol. 1, No. 3, June 2003, pp. 32–44.
- [3] H. Wang, and S. Wang, "Cyber warfare: Steganography vs. Steganalysis", Communications of the ACM magazine, Vol. 47, No.10, October 2004, pp. 76-82.
- [4] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "Digital image steganography: Survey and analysis of current methods", Journal of Signal Processing, Vol. 90, No. 3, Mar. 2010, pp.727–752.
- [5] Y. Wang; P.Moulin, "Perfectly Secure Steganography: Capacity,Error Exponents, and Code Constructions", Information Theory IEEE Transactions, Vol. 54, No. 6, Jun 2008, pp. 2706 – 2722.
- [6] N. Cvejic, "Algorithms for Audio Watermarking and Steganography", Msc. thesis, Department of Electrical and Information Engineering, Information Processing Laboratory, University of Oulu, Finland Oulu, Finland, 2004.
- [7] K. Bailey, and K. Curran, "An Evaluation of image based Steganography methods" Journal of multimedia Tools and Applications, Vol 30, No. 1, July, 2006, pp. 55-88.
- [8] N. Meghanathan, and L. Nayak. "Steganalysis algorithms for detecting the Hidden information in image, audio and Video cover media", International Journal of Network Security & Its Application (IJNSA), Vol.2, No.1, January 2010, pp. 43-55.
- [9] S. Shahreza, and M. Shalmani, "High capacity error free wavelet Domain Speech Steganography", IEEE International conference on acoustics, speech, and signal processing, March 31 - April 4, 2008, pp. 1729 - 1732.
- [10] A. Khashandarag, A. Oskuei, H. Mohammadi and M. Mirnia, "A Hybrid Method for Color Image Steganography in Spatial and Frequency Domain",

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pp. 113-120.

- [11] Z. Zhou, and L. Zhou "A Novel Algorithm for Robust Audio Watermarking Based on Quantification DCT Domain", Third International Conference on International Information Hiding and Multimedia Signal Processing, vol. 1, 26-28 November, 2007, pp. 441-444.
- [12] S. Wu, J. Huang, D. Huang and Y. Q. Shi, "A Self-Synchronized Audio Watermarking in DWT Domain", Circuits and Systems, ISCAS 4th. Proceedings of the 2004 International Symposium, Vol. 5, 23-26 May 2004, pp. 712-715.
- [13] N. Cvejic, and T. Seppanen, "A wavelet domain LSB insertion algorithm for high capacity audio steganography", Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. Proceedings of 2002 IEEE 10th, 13-16 October 2002., pp. 53-55.
- [14] M. Hasija1, A. Jindal2," Contrast of Watermarking Techniques in different domains", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pp. 559-563.
- [15] G. Tzanetakis, G. Essl, and P. Cook "Audio Analysis using the Discrete Wavelet Transform " In Proceeding WSES International Conference, Acoustics and Music: Theory and Applications (AMTA 2001), kiathos, Greece, 2001. pp. 318-323.
- [16] X. yang, W. P. Niu, M. Lu, , "A Robust Digital Audio Watermarking Scheme using Wavelet Moment Invariance", The Journal of Systems and Software, Vol 84, No. 8, 2011, pp. 1408-1421
- [17] R. Santosa, and P. Bao, "Audio to image wavelet transform based audio steganography", Proceeding of 47th International Symposium, ELMAR, June 2005, pp. 209-212.
- [18] T. Wu, H. Lin, W. Hu, and Y. Chen, "Audio Watermarking Scheme with Dynamic Adjustment in Mute Period", Expert System with Applications, Vol. 38, No. 3, pp.6787-6792.

183





**M.Sc. Haider I smael Shahadi** received B.Sc. in Information Engineering in 2001, and M.Sc. in Electronic and Communication Engineering in 2004 from Baghdad university-Iraq. Work as a lecturer in Babylon University since 2005 until 2010. Currently, he is a Ph.D. student in University Tenaga National (UNITEN)-Malaysia. His special fields of interest include Signal and Image Processing, Communication Security, and Artificial Intelligent system.He

has 12 publishing in local journal and conference in Iraq.



Assoc. Prof. Dr. Razali bin Jidin received both BSc and MSc in Electrical Engineering in 1986 from Univ. of Bridgeport, Conn, USA. After a career as Computer/Instrument engineer and short teaching, he continued his studies and earned his PhD (in Computer Science and Electrical Engineering) in 2005 from University of Kansas, Lawrence, Kansas, USA. Currently Dr. Razali is a faculty member at Univ. Tenaga in Malaysia.

His interest is in Embedded System, Control & Instrumentation in Power Generation, Hardware-Software Co-design. He has published more than twenty five papers at local and international conferences.