

A New High-Capacity Audio Watermarking Based on Wavelet Transform using the Golden Ratio and TLBO Algorithm

Ali Zeidi Joudaki¹, Marjan Abdeyazdan^{2*}, Mohammad Mosleh¹, Mohammad Khayrandish¹

¹. Department of Computer Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran

². Department of Computer Engineering, Mahshahr Branch, Islamic Azad University, Mahshahr, Iran

Received: 26 Apr 2021/ Revised: 16 Dec 2021/ Accepted: 14 Jan 2022

DOI:

Abstract

Digital watermarking is one of the best solutions for copyright infringement, copying, data verification, and illegal distribution of digital media. Recently, the protection of digital audio signals has received much attention as one of the fascinating topics for researchers and scholars. In this paper, we presented a new high-capacity, clear, and robust audio signaling scheme based on the DWT conversion synergy and golden ratio advantages using the TLBO algorithm. We used the TLBO algorithm to determine the effective frame length and embedded range, and the golden ratio to determine the appropriate embedded locations for each frame. First, the main audio signal was broken down into several sub-bands using a DWT in a specific frequency range. Since the human auditory system is not sensitive to changes in high-frequency bands, to increase the clarity and capacity of these sub-bands to embed bits we used the watermark signal. Moreover, to increase the resistance to common attacks, we framed the high-frequency bandwidth and then used the average of the frames as a key value. Our main idea was to embed an 8-bit signal simultaneously in the host signal. Experimental results showed that the proposed method is free from significant noticeable distortion (SNR about 29.68dB) and increases the resistance to common signal processing attacks such as high pass filter, echo, resampling, MPEG (MP3), etc.

Keywords: Audio Watermarking; Discrete Wavelet Transforms (DWT); High Capacity; TLBO Optimization Algorithm.

1- Introduction

With the rapid and increasing growth of the Internet and the spread of digital multimedia technologies, it has become very easy to copy and exchange digital multimedia data. For this reason, data protection against copyright and digital multimedia data has become an important issue. One of the ways to prevent these problems is to use the watermarking technique. In watermarking, a hidden signal called a watermark is embedded directly inside the host signal and always stays inside it unrecognizably. This technology can be used in various fields such as audio, video, multimedia, etc. [1-2] and since the human auditory system is more sensitive than its visual system, it is more difficult to embed watermark bits clearly in the host signal (audio signal) than video and multimedia [3-4]. This hiding of the watermark signal should not reduce the quality of the host signal. The difference between watermarking and cryptography is that in watermarking, the information was

used in the same way after the watermarking operation and it is not necessary to extract the watermark signal bits to use them. While in encryption, the encrypted signal is not usable and its use requires the extraction of the password. In general, watermarking operations do not restrict access to information. However, the purpose of cryptography is to restrict the access of unauthorized persons to information [5]. The important features of a reliable and efficient watermarking process are transparency, flexibility, and capacity. In a transparent watermark system, hiding the watermark signal does not affect the quality of the original signal. In the case of audio watermarking, the audio quality of the original audio signal and the watermarked audio signal should not be significantly different for the average listener. According to the International Federation of Sound Industry (IFPI), a signal should operate with a signal-to-noise ratio (SNR) greater than 20 dB [6-7]. Resistance: The watermarking system is resistant to common types of attacks and common processes. An audio watermarking system must be resistant to attacks such as MP3 compression, adding echoes, changing the sampling rate, and converting digital to analog and analog to digital.

Capacity: Refers to the number of bits that are embedded in a medium or in a unit of time in that medium. The storage volume of a photo is equal to the number of bits in which the photo is stored. But the capacity of an audio signal refers to the number of bits that audio is stored per second. In marking methods, a compromise must always be made between the above factors [8]. So far, no specific method has been proposed that can simultaneously meet all three requirements detailed above. Audio watermarking methods are always optimized in terms of the mentioned requirements. Moreover, after performing the watermarking operation, we have to extract the watermark signal [9]. Watermark signal extraction methods can be divided into the following three categories:

- Blind extraction: In this type of extraction, the main signal is not needed and the watermark mark is extracted directly from the watermarked signal. The algorithm proposed in this paper was falls under this extraction category.
- Semi-blind extraction: where some of the main signal properties are required for extraction.
- Non-blind extraction: This method requires the main signal for extraction [10-12].

Methods that use blind extraction are more applicable than non-blind methods. Fallahpour et al. [13] proposed an audio signal watermarking algorithm by embedding and extracting data by the FFT spectrum resizing method. The main point was to choose a frequency band for embedding based on the comparison between the compressed signal and the original and MP3 and a scaling factor. Their results showed the embedding of 5000 bps. Furthermore, the average signal to noise (SNR) was higher than 20 db. The ODG parameter is was also calculated in this paper. The value obtained for this parameter was -0.25. Prayoth Kumsawat [14] presented a blind, robust, transparent audio signal watermarking method with acceptable load capacity. They took advantage of artificial intelligence and wavelet transform. Watermark bits were embedded in the low-frequency coefficients of the signal determined by discrete wavelet transform. The embedding method was developed through the quantization process. A genetic algorithm was also used to optimize the quality of the watermarked signal. Their results confirmed the embedding of 34.14 bps. The average signal to noise (SNR) in this article was higher than 27.50 db. Meanwhile, the average BER value extracted in this study was about 0 to 4.0039%. The most important limitation of this system was its very low capacity. Martinez et al. [15] proposed a semi-blind, resistant, transparent audio signal watermarking method. They used the benefits of the LDPC technique. The watermark signal bits were embedded in the fifth level of the host audio signal specified by the wavelet transform. Their results demonstrated the embedding of 229 bps. The average signal to noise (SNR) in this article was higher than 40 db. Meanwhile, the average BER value was about $1.233 \times 10^{-4}\%$. The most important

limitation of this system was its very low capacity. Bath et al. [16] proposed a blind, transparent, and strong audio signal watermarking method. They used the advantages of SVD and QIM (quantization). The watermark insertion and extraction methods were designed based on quantization of the norm of singular values of the blocks. Their results showed the embedding of 196 bps. Besides, the average signal to noise (SNR) was higher than 20 db. Meanwhile, the average BER value was about 0 to 0.56%. The most important limitation of this system was its very low capacity. Fallahpour et al. [17] presented a blind extraction audio watermarking method using human auditory amplitude logarithmic domain (HAS). Their results showed the embedding of 7000 to 8000 bps. The average signal to noise (SNR) varied from 21 to 36 db. Meanwhile, the average BER value was about 0 to 13%. The ODG parameter was also calculated in this paper with the corresponding value of -1 to -0.1. Mohsenfar et al. [18] presented a semi-blind, robust, transparent sound signal watermarking scheme with adequate load capacity. They took advantage of QR decomposition (QR factorization) and genetic algorithms. They split the main audio signal into several frames. Each frame was then decomposed using the QR decomposition method, and then the most suitable location for embedding watermarks signal bits with high resistance to potential attacks was determined using a genetic algorithm. Their results confirmed the embedding of 159 bps. The average signal to noise (SNR) in this article was higher than 24 db. Meanwhile, the average BER value was about 0 to 24.18%. The ODG parameter was another calculated in this paper, with a value of -0.36 to -0.81. The most important limitation of this system was its very low capacity. Hu et al. [19] presented a blind, robust, transparent sound signal watermarking scheme with acceptable load capacity. They used variable dimensional vector modulation (VDVM) and the benefits of DWT to balance the two main parameters of capacity and transparency. Their results demonstrated the embedding of 301.46 bps. The average signal to noise (SNR) in this article was higher 20.400 db. Meanwhile, the average BER value extracted in this study was about 1.05%. ODG was also calculated in this paper. The value obtained for this parameter was -0.151[+-0.133]. The most important limitation of this system was its very low capacity. Chen et al. [20] proposed a semi-blind digital audio signal watermarking method. They embedded the watermark signal in the low-frequency coefficients of the discrete wavelet transform. They also used the Karush-Kuhn-Tucker (KKT) advantage to minimize the difference between the main signal coefficients and the watermarked signal. Their results showed the embedding of 1000 to 2000 bps. The average signal to noise (SNR) in this article was 20.21 db. Meanwhile, the average BER value was about 0.11%. Uma R. Nair et al. [21] presented a blind, robust, transparent sound signal watermarking scheme with a high capacity. They used the benefits of discrete wavelet transform (DWT).

First, they processed the original audio signal using wavelet transform. They then divided them into frames of appropriate size. Fibonacci numbers were used to embed the watermark signal bits. Their results confirmed the embedding of 2100 to 3125 bps. The average signal to noise (SNR) was higher than 58 to 69 db. Moreover, the average BER value extracted in this was about 0.0059%. Erfani et al. [22] proposed an audio signal watermarking method. The benefits of the Fourier domain were used to embed the watermark signal by changing the phase of the signal coefficients. Moreover, watermark bits are embedded in the upper domain. Their results showed the embedding of 56.5 bps. The average signal to noise (SNR) in this article was 20.1 db. Meanwhile, the average BER value was about -0.23 to -1%. Mustapha Hemis et al. [23] presented a sound signal watermarking method, semi-blind, robust, transparent, and with appropriate capacity. Using discrete wavelet converter (DWPT) and (DC-DM) converters, they were able to use DWPT to split audio frames into multiple frequency bands. Use the psychoacoustic model to determine the appropriate sub-bands for embedding watermark signal bits. Moreover, the DC-DM method was used to embed watermark bits in DWPT coefficients. For our compromise between strength and capacity, the sync code technique was used to resist synchronization attacks in the proposed method. Their results showed the embedding of 2500 bps. The average signal to noise (SNR) in this article was higher 35.95 db. Meanwhile, the average BER value was about 0 to 0.35%. The most important limitation of this system was its very low capacity. Arashdeep et al. [24] proposed a sound signal watermarking method with blind extraction, using the advantages of the wavelet transform. Their results showed the embedding of 4884 bps. The average signal to noise (SNR) in this article was 19.88 to 37.92 DB. Meanwhile, the average BER value was smaller than 0 to 5.117%. Sadie et al. [25] Proposed an audio signal marking scheme with blind extraction. Arnold's conversion was used to maintain detection security. They first segmented the host signal and then used the advantages of discrete wavelet transform (DWT) and discrete cosine conversion (DCT) per frame. Their results showed the embedding of 41.19 to 53.87 bps. The average signal to noise (SNR) in this article was 31.0786 db. In addition, the average BER value varied from 00. to 5.0781%. The most important limitation of this system was its very low capacity. Pourhashemi et al. [26] presented a new blind, resistant, transparent sound marking design with high capacity. Using the advantages of Lucas' regular mathematical sequence, they also proposed a 2-bit embedding method. An intelligent recursive adjustment process was also employed to determine the frame size and frequency band values. Their results approved the embedding of 1 to 8000 bps. The average signal to noise (SNR) in this paper varied from 33 to 58 db. Additionally, the average BER value was about 4.908%. The ODG parameter was estimated and the resulting value varied from

-0.35 to -1.57. The main weakness of is the proposed design was its low resistance to some attacks. Mehdi Mosleh et al. [27] presented a new blind audio watermarking design that could use the synergy between fuzzy inference system, one-value analysis (SVD), and the Fibonacci sequence in discrete cosine conversion (DCT) between power transparency and compromise and create capacity. Their results demonstrated an embedding of 593.34 bps. The average signal-to-noise ratio (SNR) in this paper was 49.8093 db. Moreover, the average BER value was 1.3644. ODG was another parameter calculated in this paper with the resulting value varying from -0.18 to -0.46. The main disadvantage of the design developed in this article was its low capacity. Ahmad Hussein Ali et al. [28] proposed a high-capacity audio steganography model based on fractal coding and uniform coefficient modulation. Their model was able to use the HASFC approach based on latent coating mapping, uniform coefficient modulation, and hybrid turbulence mapping techniques in the field of wavelet transform. The average signal-to-noise ratio (SNR) in this paper was 50 db. Moreover, the average BER value was 0.0369. The ODG parameter calculated in this paper varied from 4.6 to 4.8. The main disadvantage of the developed model was its very low capacity as well as its low resistance to some attacks.

1-1- Discrete Wavelet Transformation (DWT)

The authors used the advantages of the TLBO algorithm to determine the frame length and the optimal location of the signal bits, and the optimal target performance in the embedding and extraction steps. In a marked signal, SNR was increased in the embedding process and the bit error was reduced in the extraction process. The main idea was to embed an 8-bit signal simultaneously in the host signal. As mentioned above, one of the most important challenges in the auditory signal system is to provide an effective way to increase the correlation between the three criteria of transparency, strength, and load capacity. To this end, this study presented a new audio signal scheme based on the benefits of DWT and the TLBO optimization algorithm, and the golden ratio. The other sections of the article are structured as follows: In the first section, the concept of watermarking and its applications are introduced. In the second section, the DWT conversion, the golden ratio, and then the TLBO algorithm are discussed. In the third section, the proposed method is introduced in detail. Simulation results and experimental data are presented in the fourth section of the paper. Finally, the article ends with a concluding section.

2- Preliminaries

The general outline of this paper is as follows: Initially, the host program reads the audio signal and the signal (image) separately. In the second step, wavelet transform is applied to the main audio signal in three levels and CD3 coefficients are generated. Then, using the TLBO algorithm, the CD3 coefficients are framed and the amount of embedding distance is also determined. Then, using the golden ratio sequence, the signal bits are embedded in appropriate suitable place in the host signal. Finally, a watermark signal is generated. In the extraction step, the specified signal (Lena image) is extracted. An inverted wavelet transform is applied to the specified signal in three steps. The coefficients below the high-frequency band are also extracted and framed. Finally, the binary equivalent of the extracted data is calculated and the pattern signal is extracted. signal at different scales.

Wavelet transform, like Fourier transform and cosine transform, models the data [29]. compute coefficients to indicate how similar a particular function is to applied data. Most wavelet analyses add a parameter that specifies how often this function is applied to the wavelet function, and the remaining intervals of the function are zeroed by this wavelet. What does not make sense in the sine and cosine waves is a powerful and flexible tool for solving many complex problems. Here are some of the features listed below: • Wavelet is a tool that can display a signal with different degrees of resolution and can be used to examine the signal at different scales. • Wavelet transform can display the signal in only a few non-zero sentences. The principles of discrete wavelet transforming go back to a method called sub-band bandwidth [30-31]. In the discrete mode, filters with different cut-off frequencies are used to analyze the as the signal passes through the upper and lower filters, its different frequencies are analyzed. In discrete wavelet transform, the signal resolution is controlled by the operation of the filters, and the scale changes with low sampling or high sampling. Processing begins with a discrete wavelet transform. Initially, the signal passes through a low-voltage digital half-band filter with a shock response of $h[n]$, so the filter output is equal to the input convolution and the filter shock response. As a result of this filtering operation, all frequency components greater than half of the highest frequency in the signal are removed. Since the maximum output signal frequency of the filter is equal to $\pi/2$ radians, half of the samples can be omitted. Therefore, by deleting one of the samples, the signal length is halved without losing any piece of information. A similar process is performed using a semi-band digital high pass filter with a shock response of $g[n]$. As a result, at the output of the first stage of the wavelet conversion operation, two versions, one high-pass and one low-pass, with a reduced length (half) of the initial signal are obtained as follows:

$$y_{\text{high}}[k] = \sum_n x[n] \cdot g[2k - n] \quad (1)$$

$$y_{\text{low}}[k] = \sum_n x[n] \cdot h[2k - n] \quad (2)$$

This halves the temporal resolution and doubles the frequency resolution. This process can be re-applied to the low-pass version, doubling the frequency resolution at each stage by reducing the time resolution to half of the previous stage. This idea is known as the filter bank method for calculating discrete wavelet transform. The output coefficients of the low-pass filter follow the original shape of the signal. Hence, these coefficients are called approximations. High pass filter output coefficients also include high-frequency details. Hence, these coefficients are called details. As the number of conversion steps increases, so does the amount of details. It should be noted that the number of steps required to convert a discrete wavelet depends on the frequency characteristics of the analyzed signal. Finally, discrete wavelet wave conversion is obtained by juxtaposing the filter outputs from the first stage of the filter. Therefore, the number of wavelet conversion coefficients will be equal to the number of input discrete signal samples. Three levels of discrete wavelet transform decomposition are shown in Figure 3. As can be seen, A1 and D1 are the coefficients of the first level of signal decomposition $X[n]$. In the second level, A1 is decomposed into A2 and D2 and, in the third level, A2 is decomposed into A3 and D3:

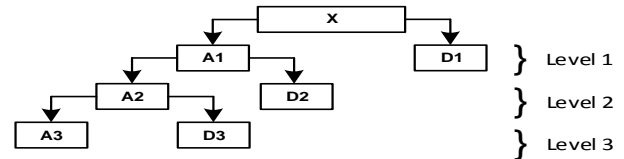


Fig.1 The three levels of the discrete wavelet transform decomposition

2-1- Golden Ratio

In mathematics and art, the golden ratio occurs when the ratio of the larger part to the smaller part is equal to the ratio of the whole to the larger part. It geometric definition was used to obtain the golden ratio, $\frac{a+b}{a} = \frac{a}{b} = \phi$. Following this equation, which is the definition of the number (ϕ), we can conclude from the equation on the right: $a = b \times \phi$. Thus, we will have: $\frac{b \times \phi + b}{b \times \phi} = \frac{b \times \phi}{b}$. By removing b from both sides., we have: $\frac{\phi + 1}{\phi} = \phi$. The positive answer is as follows:

$$\phi = \frac{1 + \sqrt{5}}{2} \approx 1.6180339887 \quad (3)$$

2-2- TLBO (Tteacher Learning Based Optimization) Algorithm

In recent years, meta-heuristic algorithms have been used to optimize engineering problems. These algorithms are either modeled based on natural phenomena (such as ant colony and birds algorithms) or sample human social exchanges (such as Imperial competition algorithms and teacher learning algorithms). The most important advantage of these algorithms is that they are simple and do not require complex mathematical problems such as derivatives and integrals. Teacher learning based optimization algorithm is an interesting algorithm for optimizing engineering issues which are modeled based on the teacher training in the classroom. This algorithm has two training steps. The first step is based on the teacher training and the second one is based on the student debate after the end of the class. In the first phase, the person who provides the best answer in the population is selected as the teacher (X_{teacher}) and other members of the population are known as students (X_i). In the following section, the average position of the students (X_{mean}) is calculated. The reason for calculating the student knowledge average is that the teacher gives the training according to the average level of the class. By considering “r” as a random number and T_f as a constant coefficient, it is possible to model the movement of students in the first step by the following equation [32-34].

$$X_{\text{new}} = X_i + r \times (X_{\text{teacher}} - T_f \cdot X_{\text{mean}}) \quad (4)$$

Where X_i and X_{new} are the current and the new situation of the students, respectively, and T_f is a training factor that is considered as 2.

In the second stage, the teaching process is assigned to the students, so that each student selects another student randomly and shares knowledge with each other and also updates his/her position; thus trying to use the other students' information to raise his/her level of awareness and knowledge. This phase can be modeled using the following formulations [13-16]:

$$X_{\text{new}} = X_i + r(X_i - X_j) \quad \text{if } f(X_i) < f(X_j) \quad (5)$$

$$X_{\text{new}} = X_i + r(X_j - X_i) \quad \text{if } f(X_j) < f(X_i) \quad (6)$$

In this stage, the move is made if the new position is better than the previous position. Moreover, the condition for the termination of this algorithm is to reach the end of the iteration. The algorithm will continue until the termination condition is met. The method proposed in this study is a blind audio watermarking technique that is developed by applying the discrete wavelet transform (DWT) on the digital audio signal. The algorithm consists of two procedures; The watermarking embedding procedure and watermarking extraction procedure.

3- The Proposed Watermarking Scheme

The method presented in this article was a new approach to increasing the encryption capacity of audio signals for covert telecommunication purposes. In a wavelet transform using two filters, the signal is decomposed into several sub-bands in a specific frequency range. Because the human auditory system is not sensitive to changes in high-frequency bands, we use these sub-bands to embed blue signals. This increases the embedding capacity. The method discussed in this paper uses high-frequency coefficients above the surface of the third wavelet, to which the human ear is less sensitive than embedded blue markers. As mentioned, we need to find a way that fits the three criteria of transparency, flexibility, and capacity. Determining the appropriate sub-band plays a significant role in enhancing capacity and transparency. In this paper, we have determined the appropriate sub-band by wavelet transform. Moreover, the size of the frame is directly related to the amount of resistance. Therefore, the longer the frame, the more resistant it will be and the less capacity it will have, and vice versa. The appropriate frame length was determined using the TLBO algorithm. To further increase the resistance, the coefficients of this sub-band were framed and the average length of each frame was calculated and placed in the first cell of each frame. The signal was converted to a bit sequence, and each of the eight signal bits was placed in the appropriate range determined by the TLBO algorithm using the golden ratio sequence.

3-1- Watermark Embedding Procedure:

In this section, a sound-blind marking scheme with high-capacity and clear and robust features is developed based on the TLBO algorithm and wavelet transform. The embedding process is performed through 6 main operations:

Inputs: Watermark signal, suitable sub-band coefficients by DWT The frame length size is determined by the TLBO algorithm.

Output: Watermark signal the proposed method is detailed as follows:

Begin

Step 1: Before inserting the pattern into audio, the pattern is converted into a one-dimensional bit sequence. To convert bit patterns into a sequence of bits, we need to convert a two-dimensional array into a one-dimensional array. To do this, we scroll the bit pattern line by line and convert the bit values of the bit pattern into a bit sequence.

Step 2: We used the Dobchi 8 wavelet transform to analyze the host sound. Host sound analysis is performed at 3 levels. The lower the resolution level, the lower the resistance of the algorithm and the higher the watermark capacity, and vice versa, because the human ear is less

sensitive to the coefficients of detail (high frequency). In this method, we model on the CD3 sub-band.

Step 3: The CD3 coefficients are divided by the TLBO optimization algorithm into the number of frames of a given length. Then, the average (m_i) for each frame is calculated based on the average absolute values of the samples per frame. It should be noted that the greater the number of frames or the smaller the frame size, the greater the transparency of the algorithm and the lower the resistance, and the smaller the number of frames, the greater the power of the algorithm. Then, the average absolute value of the samples of each frame is calculated using the formula (7) and store it in the first cell of each frame.

$$m_i = \frac{1}{s} \sum_{j=(i-1)s+1}^{i \times s} |C_j| \quad (7)$$

Where (C_j) is the sub-band detail coefficient of the third wavelet surface, s is the size of each frame, and (m_i) is the mean of the (i) frame. Using Equation (8) below, the number of watermark signal bits is calculated and placed in the third level coefficients. The parameter k is the embedding distance, which is less than one. If the value of this parameter increases, the amount of embedded capacity increases leading to a severe decrease in the power of the algorithm. In this paper, the advantages of the TLBO algorithm are used to select the optimal value of K factors that have reached the signal in proportion to the noise. The signal to noise ratio (SNR) is a measure of statistical difference used to determine the similarity between the original distorted audio signal and the distorted audio signal.

Step 4: Using the golden ratio, the appropriate locations of each frame for embedding watermark signal data are determined.

Step 5: The number of bits to be embedded is determined via Eq. (10). The power of the proposed algorithm is that at each appropriate factor (n) bits of the bit string of the watermark signal are embedded. This idea increases the cryptographic capacity. For example, in this paper, we consider the value of $n = 8$. The embedding capacity is equal to 2^n . The distance between $(K \times m_i)$ and $-(K \times m_i)$ is divided by 256 levels, and then the decimal equivalent of each of the 8 watermark signals is stored in the variable eq. In the reconstruction process, the binary equivalent of the variable eq is used to reconstruct the desired printing mark.

$$C(i) = -k \times m(i) + \frac{2^{\times k \times m(i) \times \text{double}(xn)}}{2^n} \quad (8)$$

Step 6: Finally, after applying the embedding process using the reverse wavelet converter, a watermark signal is generated.

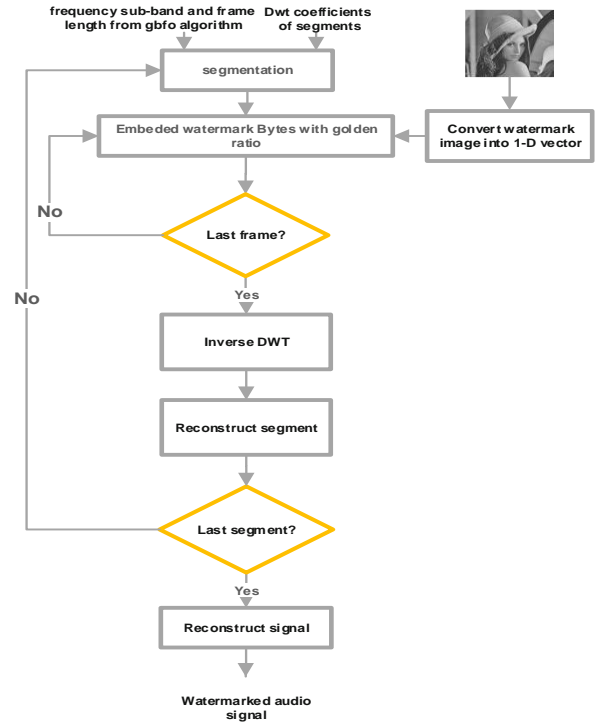


Fig.2 The watermark embedding procedure

3-2- Watermark extraction procedure

The detection process is the inverse signal of the embedding operation, this process is blind because it does not require a host audio signal. Two main operations are performed at this stage: Frequency analysis based on discrete wavelet transform and cryptographic detection of the third level details of the audio signal. The detection process can be summarized in the following steps:

Inputs: The received audio signals[n]

Outputs: The extracted watermark

Begin

Step 1: The watermark audio signal received $s [n]$

Step 2: Three levels of wavelet transform are applied to the watermarked signal.

Step 3: After extracting the sub-band coefficients from the details of the third level, they are framed according to the fact that in the pasting stage, after reformatting and converting decimal data, the calculated average was stored at the location of the first cell of each frame. Decimal data stored in its binary equivalent can retrieve our data.

Step 4: Using the following equation (9), the string of cryptographic bits embedded in the signal is detected.

$$Ex = (C'(i, j) + K \times C'(i, 1)) \times \frac{(2^8)}{2 \times K \times C'(i, 1)} + K \quad (9)$$

C' High-frequency sub-band samples of the third level of watermark signal wavelet decomposition.

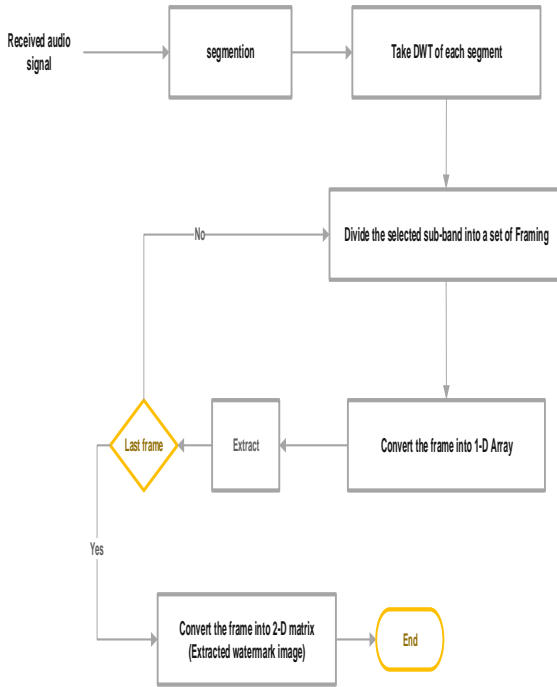


Fig.3 Block diagram of watermark extraction algorithm

In this paper, we define the objective function as comparatively equal to Formula 10.

The teacher training optimization algorithm performs the optimization by minimizing the objective function. Thus, the objective function is defined as follows:

$$Fitness = A_1 \times SNR - A_2 \times BER + A_3 \times NC \quad (10)$$

$$A_1 = 1 + \frac{\beta_1}{1 + e^{\frac{SNR}{\alpha_1}}} \quad (11)$$

$$A_2 = 1 + \frac{\beta_2}{1 + e^{\frac{-BER}{\alpha_2}}} \quad (12)$$

$$A_3 = 1 + \frac{\beta_3}{1 + e^{\frac{NC}{\alpha_3}}} \quad (13)$$

In the above equation, the coefficients A_1 , A_2 , and A_3 which are related to the parameters SNR, BER, and NC are variable and they adapt their program function to the objective function. According to coefficient A_1 , if the SNR value decreases for any reason, the denominator of the fraction decreases in the same way. As a result, A_1 increases, and the algorithm increases the SNR. In coefficient A_3 , if the NC value decreases for any reason, the denominator of the fraction decreases in the same way. As a result, A_3 increases, and the algorithm increases the NC. But the coefficient A_2 works the opposite of the coefficients A_1 and A_3 . Thus, if the value of BER increases,

the denominator of the fraction also decreases and thus the coefficient A_2 increases by a negative sign, indicating that the algorithm decreases the value of BER. The values of coefficients $\beta_1, \beta_2, \beta_3$ are selected experimentally equal to 1, 50, and 20, and coefficients $\alpha_1, \alpha_2, \alpha_3$ are set as 25, 0.05, and 1, respectively.

4- Experimental Results

The method described in the previous section was implemented with Matlab 2019 software. This section presents the results of various tests performed on the proposed method. First, the sounds on which the cryptography was performed are identified [29]. These sounds are selected as common sounds on which other methods of cryptography are used in the experiments. We used the Lena image as the watermark image shown in Figure 4:



Fig.4 The watermark image

Loopy Music is an audio signal in which the carrier signal is chosen. To implement the watermarking procedure, an audio signal with 44100 HZ frequency is sampled and quantized with 16 bits. Then, by applying a three-level wavelet transform, the audio signal is divided into four sub-bands called are CA3, CD3, CD2, and CD1, respectively. Here, the CD3 sub-band was used to hide watermarks. We used the 8-coefficient Daubechies wavelet (db8) to decompose the audio signal. The simulation results also showed that by the best robustness against attacks would be achieved using Daubechies filter (db8). Using Equation 14, the signal-to-noise (SNR) value can be calculated. As it is clear from the formula, this parameter shows the similarity of the watermarked signal to the main signal. According to the International Federation Phonographic Industry (IFPI) standard, the SNR value must be at least 20dB and is calculated using Equation 14:

$$SNR = 10 \log_{10} \frac{\sum_{n=0}^{N-1} X_o^2(n)}{\sum_{n=0}^{N-1} [X_o(n) - X_w(n)]^2} \quad (14)$$

In Equation (14) $X_o(n)$, $X_w(n)$ signals are the main signal and the watermarked signal, respectively. The minimum acceptable value for this parameter is 20 db.

4-1- Subjective Testing

We must test the final quality of the results of the original signal and the watermarked signal by the human auditory system. To do this, we select five volunteers and ask them to tell us the difference between the original signal and the watermarked signal. Degree of mental difference Degree (MDG). In Table 1, MDG = 5.0 has the concept of excellent quality and degree MDG = 1.0 indicates poor quality.

Table 1 ODG and MDG values with their descriptions

explanation	Quality	MDG values	ODG values
Inconspicuous	Excellent	5.0	0.0
Tangible	Good	4.0	-1
Some Pesky	Exhibition	3.0	-2
Pesky	Impaired	2.0	-3
Vary Pesky	Bad	1.0	-4

4-2- Objective Testing

The degree of objective difference (ODG) and the signal-to-noise ratio (SNR) are two important parameters for the objective evaluation of watermarked signals. In this article, we intend to compare and evaluate the proposed method with a focus on the previously presented methods and an assessment of these parameters. Using Equation 14, the signal-to-noise (SNR) ratio can be calculated. As is shown, this parameter shows the similarity of the watermarked signal to the main signal. According to the International Federation Phonographic Industry (IFPI) standard, the SNR value must be at least 20 dB according to Equation 14. The SNR values of the different signals that have been watermarked by the method presented as can be seen in Table 3. The ODG parameter is measured with EQUAL software [30]. The value of this parameter can vary from -4 (indicating that the applied changes are very annoying) to 0 (indicating that the applied changes are imperceptible). The SNR values of the different signals that have been watermarked by the method presented as can be seen in Table 2:

Table 2 The results of objective and subjective tests performed on the proposed method

Audio files	Objective tests	ODG	Subjective test
	SNR (dB)		MDG
Blues	29.42	- 0.56	4.19
Electronics	30.31	- 0.16	4.25
Average	29.32	-0.321	4.36

4-3- Robustness Testing

The BER (bit error rate) and NC (normalized correlation) are two important parameters used for evaluating the resistance of watermarked signals. Equations 15 and 16 show how to calculate these two parameters:

$$NC = \frac{\sum_{i=1}^{Mlength} \sum_{j=1}^{Mwidth} [M(i,j)M'(i,j)]}{\sum_{i=1}^{Mlength} \sum_{j=1}^{Mwidth} [M(i,j)]^2} \tag{15}$$

$$BER(W, \bar{W}) = \frac{\sum_{i=1}^M \sum_{j=1}^M W(i,j) \oplus \bar{W}(i,j)}{M \times M} \tag{16}$$

Table 3 The assessment of the resistance of the proposed design against Stir-mark attacks for BER and NC

Audio file	Attack Type	NC	BER (%)
Electronic	Nothing	0.9989	0.1879
	Lsbzero	0.9831	0.9412
	fft_real_reverse	0.9642	0.6721
	fft_invert	0.9721	5.3446
	Addsinus	0.9916	0.9412
	addbrumm_2100	0.9933	0.7514
	rc_highpass	0.9845	0.9654
Average		0.9839	1.4005
Blues	Nothing	0.9899	0.7325
	Lsbzero	0.9731	1.8123
	fft_real_reverse	0.9754	2.9478
	fft_invert	0.9768	1.0841
	Addsinus	0.9842	1.7367
	addbrumm_2100	0.9657	0.8579
	rc_highpass	0.9731	1.1124
Average		0.9768	1.4689

4-4- Data Ccapacity Result

Capacity refers to the number of bits stored per second of the host sound and is expressed as bits per second. Capacity is calculated via the following equation:

$$payload = \frac{\text{Total bits of watermark}}{\text{time of host audio}} \tag{17}$$

Table 4 Comparison the proposed design with other methods in terms of capacity and intangibility

Watermarking schemes	capacity (bps)	Average of ODG	Average of SNR (dB)
[13]	8005	N/A	21.03
[14]	34.14	N/A	34.14
[15]	229	N/A	> 40
[16]	196	N/A	> 20
[17]	7000 to 8000	N/A	21 to 36
[18]	159	-0.36 to -0.81	24
[19]	301.46	-0.151[±0.133]	20.14
[20]	1000 to 2000	N/A	20.21
[21]	2100 to 3125	N/A	58 to 69
[22]	56.5	-0.21 to -1	> 20
[23]	2500	-0.35	35.95
[24]	4884	N/A	37.92
[25]	41.19 to 53.87	N/A	31.0786
[26]	1 to 8000	-0.35 to -1.57	33 to 58
[27]	598.34	-0.18 to -0.46	49.8093
[28]	562.4	4.6 to 4.8	50.4
Proposed	13000	- 0.56 to - 0.16	29.68

N/A means no report is found in the watermarking scheme

Table 5 Comparing the proposed watermarking scheme with the previous works

Scheme	BER (%)	Payload capacity (bps)	Average of SNR (dB)	Blind	Average of ODG	Opti-method
[13]	0.0 to 30	8005	21.03	Blind	N/A	N/A
[14]	0 to 4.0039	34.14	34.14	Blind	N/A	GA
[15]	10323×10^{-4}	229	> 40	Samir - Blind	N/A	N/A
[16]	0.0 to 0.56	196	> 20	Blind	N/A	SVD-Quantization
[17]	0 to 13	7000 to 8000	21 to 36	Blind	N/A	N/A
[18]	0 to 24.18	159	24	Samir - Blind	-0.36 to -0.81	GA
[19]	1005	301.46	20.14	Blind	0.151[± 0.133]	VDVM
[20]	0.11	1000 to 2000	20.21		N/A	N/A
[21]	0.0059	2100 to 3125	58 to 69	Blind	N/A	N/A
[22]	0.0 to 5	56.5	> 20	Blind	-0.21 to -1	N/A
[23]	0.0 to 0.35	2500	35.95	Samir - Blind	-0.35	DWPT
[24]	0.0 to 5.17	4884	37.92	Blind	N/A	N/A
[25]	0.0 to 5.0781	41.19 to 53.87	31.0786	Blind	N/A	N/A
[26]	-0.32 to -1.57	1 to 8000	33 to 58	Blind	-0.35 to -1.57	N/A
[27]	1.3644	598.34	49.8093	Blind	-0.18 to -0.46	Fuzzy
[28]	0.0369	562.4	50.4	None - Blind	N/A	N/A
Proposed	1.4698	13000	29.68	Blind	-0.56 to -0.16	TLBO
N/A means no report is found in the watermarking scheme						

5- Conclusion

This paper presented a new, high-capacity, DWT-based watermark audio signal-blind approach. In the proposed algorithm, to further increase power and transparency, the third-level sub-band coefficients were framed and then the average of each frame was calculated. The TLBO algorithm was used to determine the appropriate frame length and appropriate embedding range. In addition, the

proposed method is blind method, which means that when the print signal is retrieved, the original signal is not required. The experimental results showed that the proposed method has an excellent SNR signal quality of about 29.68 dB and a very good capacity (about 13 kbps). It also has good resistance against sound processing attacks such as MP3 compression attacks, filtering, sampling reduction, etc.

References

- [1] Swanson MD, Zhu B, Tewfik AH, Boney L. Robust audio watermarking using perceptual masking. Signal processing. 1998, pp.337-355.
- [2] Liu W, Hu AQ. A sub-band excitation substitute based scheme for narrowband speech watermarking. Frontiers of Information Technology & Electronic Engineering. 2017, pp.627-643.
- [3] Xia Z, Wang X, Zhang L, Qin Z, Sun X, Ren K. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. IEEE transactions on information forensics and security. 2016, pp.2594-2608.
- [4] Nguyen TS, Chang CC, Yang XQ. A reversible image authentication scheme based on fragile watermarking in the discrete wavelet transform domain. AEU-International Journal of Electronics and Communications. 2016, pp.1055-1061.
- [5] Sun L, Xu J, Liu S, Zhang S, Li Y, Shen CA. A robust image watermarking scheme using Arnold transform and BP neural network. Neural Computing and Applications. 2018, pp.2425-2440.
- [6] slam M, Roy A, Laskar RH. SVM-based robust image watermarking technique in LWT domain using different sub-bands. Neural Computing and Applications. 2020, pp.1379-1403.
- [7] Katzenbeisser S, Petitcolas FA. Digital watermarking. Artech House, London. 2000; 2.
- [8] Nejad MY, Mosleh M, Heikalabad SR. LSB-based quantum audio watermarking using MSB as arbiter. International Journal of Theoretical Physics. 2019, pp.3828-3851.
- [9] Rasti P, Samiei S, Agoyi M, Escalera S, Anbarjafari G. Robust non-blind color video watermarking using QR decomposition and entropy analysis. Journal of Visual Communication and Image Representation. 2016, pp.838-847.
- [10] Bhardwaj A, Verma VS, Jha RK. Robust video watermarking using significant frame selection based on the coefficient difference of lifting wavelet transform. Multimedia Tools and Applications. 2018, pp.19659-19678.
- [11] Ali M, Ahn CW. An optimal image watermarking approach through cuckoo search algorithm in the wavelet domain. International Journal of System Assurance Engineering and Management. 2018, pp.602-611.
- [12] Mehta R, Rajpal N, Vishwakarma VP. Robust image watermarking scheme in lifting wavelet domain using GA-LSVR hybridization. International Journal of Machine Learning and Cybernetics. 2018, pp.145-161.
- [13] Fallahpour M, Megias D. Robust high-capacity audio watermarking based on FFT amplitude modification. IEICE TRANSACTIONS on Information and Systems. 2010, pp.87-93.

- [14] Kumsawat P. An efficient digital audio watermarking scheme based on genetic algorithm. In 2010 10th International Symposium on Communications and Information Technologies 2010, pp. 481-485.
- [15] Martínez-Noriega R, Nakano M, Kurkoski B, Yamaguchi K. High payload audio watermarking: toward channel characterization of MP3 compression. *Journal of Information Hiding and Multimedia Signal Processing*. 2011, pp. 91-107.
- [16] Bath V, Sengupta I, Das A. A new audio watermarking scheme based on singular value decomposition and quantization. *Circuits, Systems, and Signal Processing*. 2011, pp.915-927.
- [17] Fallahpour M, Megías D. High capacity logarithmic audio watermarking based on the human auditory system. In 2012 IEEE International Symposium on Multimedia 2012, pp. 28-31.
- [18] Mohsenfar SM, Mosleh M, Barati A. Audio watermarking method using QR decomposition and genetic algorithm. *Multimedia Tools and Applications*. 2015, pp.759-779.
- [19] Hu HT, Hsu LY, Chou HH. Perceptual-based DWPT-DCT framework for selective blind audio watermarking. *Signal processing*. 2014, pp.316-327.
- [20] Chen ST, Hsu CY, Huang HN. Wavelet-domain audio watermarking using optimal modification on low-frequency amplitude. *IET signal processing*. 2015, pp.166-176.
- [21] Nair UR, Birajdar GK. Audio watermarking in wavelet domain using Fibonacci numbers. In 2016 International Conference on Signal and Information Processing (IconSIP), 2016, pp. 1-5.
- [22] Erfani Y, Pichevar R, Rouat J. Audio watermarking using spikegram and a two-dictionary approach. *IEEE transactions on information forensics and security*. 2016, pp.840-852.
- [23] Hemis M, Boudraa B, Megías D, Merazi-Meksen T. Adjustable audio watermarking algorithm based on DWPT and psychoacoustic modeling. *Multimedia Tools and Applications*. 2018, pp.11693-11725.
- [24] Kaur A, Dutta MK. An optimized high payload audio watermarking algorithm based on LU-factorization. *Multimedia Systems*. 2018, pp.341-353.
- [25] Saadi S, Merrad A, Benziane A. Novel secured scheme for blind audio/speech norm-space watermarking by Arnold algorithm. *Signal Processing*. 2019, pp.74-86.
- [26] Pourhashemi SM, Mosleh M, Erfani Y. Audio watermarking based on the synergy between Lucas regular sequence and Fast Fourier Transform. *Multimedia Tools and Applications*. 2019, pp.22883-22908.
- [27] Mahdi Mosleh, Saeed Setayeshi, Behrang Barekatin, Mohammad Mosleh. A novel audio watermarking scheme based on fuzzy inference system in DCT domain. *Multimedia Tools and Applications*. 2021, pp. 20423–20447
- [28] Ahmed Hussain Ali, Loay Edwar George, Mohd Rosmadi Mokhtar. An Adaptive High Capacity Model for Secure Audio Communication Based on Fractal Coding and Uniform Coefficient Modulation. *Circuits, Systems, and Signal Processing*. 2020. pp. pages5198–5225.
- [29] Cai, Y.M., Guo, W.Q. and Ding, H.Y., 2013. Dwt-svd. *Journal of Software*, 8(7), p.1801.
- [30] Brannock, E., Weeks, M. and Harrison, R., 2008, April. Watermarking with wavelets: simplicity leads to robustness. In *IEEE SoutheastCon 2008*, pp. 587-592.
- [31] Pandey, V., 2014. Analysis of image compression using wavelets. *International Journal of Computer Applications*, 103(17).
- [32] Lerch, A. (2002). Zplane development, EAQUAL-Evaluate Audio QUALity, version: 0.1. 3alpha. Retrieved July 2018, from <http://www.mp3-tech.org/programmer/misc.html>.
- [33] Akhlaghi M, Emami F, Nozhat N. Binary TLBO algorithm assisted for designing plasmonic nano bi-pyramids-based absorption coefficient. *Journal of Modern Optics*. 2014, pp.1092-1096.
- [34] Akhlaghi M, Emami F, Nozhat N, "Binary TLBO algorithm assisted for designing plasmonic nano bi-pyramids-based absorption coefficient", *Journal of Modern Optics*, 2014, pp.1092-1096.
- [35] Akhlaghi M. Optimization of the plasmonic nano-rods-based absorption coefficient using TLBO algorithm. *Optik*. 2015, pp.5033-5037.
- [36] Balvasi M, Akhlaghi M, Shahmirzaee H, "Binary TLBO algorithm assisted to investigate the supper scattering plasmonic nanotubes", *Superlattices and Microstructures*. 2016, pp.26-33.
- [37] Kaboli M, Akhlaghi M, "Binary teaching-learning-based optimization algorithm was used to investigate the superscattering plasmonic nanodisk", *Optics and Spectroscopy*. 2016, pp.958-63.