Research Article

Robust and Blind Audio Watermarking Scheme Based on Genetic Algorithm in Dual Transform Domain

Qiuling Wu , Aiyan Qu, Dandan Huang, and Lejun Ma

School of Cyber Security, Jinling Institute of Technology, Nanjing 211169, China

Correspondence should be addressed to Qiuling Wu; wuqiuling@jit.edu.cn

Received 21 July 2021; Revised 23 September 2021; Accepted 21 October 2021; Published 3 November 2021

Academic Editor: Paolo Crippa

Copyright © 2021 Qiuling Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to protect the copyright of audio media in cyberspace, a robust and blind audio watermarking scheme based on the genetic algorithm (GA) is proposed in a dual transform domain. A formula for calculating the embedding depth is developed, and two embedding depths with different values are used to represent the “1” and “0” states of the binary watermark, respectively. In the extracting process, the embedding depth in each audio fragment will be calculated and compared with the average embedding depth to determine the watermark bit by bit, so this scheme can blindly extract the watermark without the original audio. GA will be applied to optimize the algorithm parameters for meeting the performance requirements in different applications. Besides, the embedding rule is further optimized to enhance the transparency based on the principle of minimal modification to the audio. Experimental results prove that the payload capacity reaches 172.27 bps, the bit error rate (BER) is 0.1% under the premise that its transparency is higher than 25 dB, and its robustness is strong against many attacks. Significantly, this scheme can adaptively select the algorithm parameters to satisfy the specific performance requirements.

1. Introduction

In modern society, the development of network technology has greatly promoted the rapid dissemination of network resources. People can efficiently and conveniently obtain network resources. At the same time, they are also worried about the copyright infringement of these network resources [1]. Therefore, how to protect the copyright of these network resources has aroused many scholars' interest. Encryption is a traditional technology to protect the information, but ciphertext cannot be spread more widely compared with plaintext [2]. Sometimes, if the hacker fails to get the correct password, he may corrupt it in a violent manner. Therefore, encryption may put information at risk. Information hiding technology is a novel way to secretly embed information into some digital media that can be made public for achieving the purpose of protecting the copyright of the digital media, transmitting confidential information, labelling information, and so on [3]. It not only conceals the information content but also hides the transmission behaviour, which reduces the possibility of being attacked by paralyzing the human perception system. It can be seen from the above

analysis that encryption prevents unauthorized hackers from obtaining the information, while information hiding technology conceals the behaviour of transmission [4]. Therefore, information hiding technology provides a novel mode for protecting information in cyberspace, and it has been widely used in copyright protection, secret communication, content anticounterfeiting, military intelligence, identity authentication, and so on [5]. There are four main indicators, including transparency, robustness, security, and payload capacity, to evaluate the performance of information hiding technology. In fact, they are contradictory, so researchers usually focus on some indicators according to the actual application requirements when they develop the hiding scheme. Steganography [6, 7] and digital watermarking are the two main branches of information hiding technology. They both must have high transparency; otherwise, the quality of the carriers that may be images [8], videos [9], audios [10, 11] and so on will be degraded because of the embedded information. Steganography usually has large payload capacity to carry large confidential information. However, digital watermarking technology pays more attention to the robustness against external attacks.

Digital watermarking technology protects the copyright of the carrier by embedding the watermark that is difficult to be sensed by the human perception system [12]. In the process of using those carriers, watermarks hidden in the carriers may be destroyed or lost due to some signal processing operations, so watermarking scheme must have strong robustness to withstand those operations. In addition, to make the developed watermarking scheme convenient for practical application, the scheme with blind extraction is more popular. In recent years, digital watermarking technology has been widely used in the field of information security, so many scholars have invested in the research of watermarking scheme, and the core issue is how to effectively improve the overall performance of the watermarking scheme in practical application.

Audio media is one of the most common multimedia resources. There are a lot of online songs, conversations, and other audio types in cyberspace every day. How to protect the copyright of these audio media has aroused the interest of many researchers. However, there are few research results on audio watermarking schemes, mainly due to the following difficulties. The human ear is very sensitive to the drop in audio quality caused by the presence of the watermark, which will affect the normal use of audio media [13, 14]. In addition, many audios editing software can be used to modify audio signals conveniently, which will cause the watermark hidden in the audio to be destroyed and even lost partially or totally. Many existing audio watermarking schemes have not overcome these difficulties well, so there are still many technical issues to be solved in improving their performance.

1.1. Related Works. According to the different domains, audio watermarking schemes are mainly divided into two categories: time domain methods [15, 16] and transform domain methods [17–21]. In order to prevent the loss of the information hidden in the audio carriers which may suffer from various attacks, watermarking schemes developed in the transform domain usually have strong robustness. There are many transform domain schemes, such as discrete cosine transform (DCT) [17], discrete wavelet transform (DWT) [18, 19], discrete Fourier transform (DFT) [20], and singular value decomposition (SVD) [21]. DCT focuses most of the energy of the audio fragment on its low-frequency part, so many scholars use DCT to design an audio watermarking scheme. Hu [17] proposed a high-capacity audio watermarking scheme that embedded watermark into the low-frequency coefficients according to the masking characteristics of the human auditory system in the DCT domain, but its robustness was not strong against some conventional signal processing operations. DWT has multiresolution characteristics, so it is often used to analyse the main component of the signal in both time domain and frequency domain [22]. Huang [18] presented an adaptive watermarking scheme which modified the DWT coefficients using signal-to-noise ratio (SNR) of the audio, and this scheme had good transparency because its embedding formula was optimized by minimizing the difference between the original coefficient and the modified coefficient. Hsu [19] designed

an audio watermarking scheme based on the two-stage Lagrange principle and minimum-energy scaling optimisation in the wavelet domain. This scheme also had good transparency, but their BER was high indicating weak robustness against low-pass filtering, time-scaling, and resampling attacks. In [20], a stereo audio watermarking scheme was proposed in discrete Fourier transform (DFT) domain. This scheme calculated the similarity of the audio signals in two stereo channels to develop the embedding and extracting rules. Although the payload capacity of the scheme was low, it had strong robustness when withstanding attacks. In [21], a blind audio watermarking scheme was proposed based on singular value decomposition (SVD) by mixing the watermark with the diagonal matrix of singular value. In recent years, there are many watermarking schemes in multitransform domains, which are usually more robust than the schemes developed in a single transform domain. Lei [23] proposed an audio watermarking scheme in the dual transform domain consisting of SVD and DCT. This scheme processed the audio signal into two-dimensional data blocks which would be applied by SVD and DCT in turn, and finally the watermark was embedded into these SVD-DCT coefficients with larger values. Some audio watermarking schemes based on DWT and SVD are proposed in [24, 25]. In addition, some scholars have proposed other watermarking schemes. Wang [26] proposed an audio watermarking scheme with blind extraction by using exponential modulation, and this scheme extracted exponential distance feature parameters by mapping the audio into two-dimensional dates and then embedded watermark on these parameters. These above watermarking schemes have promoted the development of watermarking technology, but most audio watermarking schemes still have many shortcomings, such as poor transparency which may cause a decline in audio carriers, weak robustness which may lead to loss of the watermark, low security, nonblind extraction, and insufficient capacity. In addition, the parameters of the most existing schemes are set by the designers according to their own experience, which makes those schemes unable to determine their parameters adaptively in different application, thus unable to stimulate their best performance.

In our work, a robust audio watermarking scheme is proposed by combining the energy concentration characteristics of DCT with the multiresolution characteristics of DWT. Firstly, apply DWT on the audio carrier for choosing the data in a specific frequency band, which can improve transparency. Secondly, Apply DCT on the data to focus energy on the low-frequency component to carry the watermark, which can improve robustness. In order to solve the problem of parameter setting, a genetic algorithm is used to optimize the algorithm parameters in different applications.

1.2. Contributions. It can be seen from the above introduction that there are still many problems in the audio watermarking scheme that need further research. In this paper, a robust and blind audio watermarking scheme based on GA is proposed in the DCT-DWT dual transform domain. Our contributions are as follows:

- (1) Our proposed scheme is developed in the DWT-DCT dual transform domain, and it has strong robustness to prevent from losing the watermark hidden in the carried audio which may suffer from various attacks. Besides, the embedding rules are further optimized based on the principle of minimal modification to the audio to improve transparency.
- (2) Our proposed scheme is blind when extracting the watermark. This scheme employs two different embedding depths to represent two states of the binary watermark. In the extracting process, the embedding depth in each audio fragment will be calculated and compared with the average embedding depth to determine the watermark information bit by bit, so the extracting process does not need the original audio.
- (3) Our proposed scheme uses GA to optimize the important parameters adaptively for meeting the performance requirements in different applications, which can stimulate the optimal performance of the scheme. In many existing watermarking schemes, the algorithm parameters are often set by the designer according to their own experience, which cannot fully stimulate the performance of the algorithm and also cannot adjust the algorithm parameters adaptively in different applications.

The remainder of this paper is organized as follows: the principle of the proposed scheme is described in Section 2, including chaotic encryption of the watermark, principle of the embedding algorithm, principle of the extracting algorithm, and optimisation of the parameters based on GA. In Section 3, the performance of this proposed scheme is evaluated in terms of transparency, payload capacity, and robustness. Furthermore, this proposed scheme is compared with some related schemes in recent years according to their experimental results. Finally, we summarize our work and introduce the future research focus in Section 4.

2. Principle of the Proposed Scheme

Due to the auditory masking effects, the human auditory system cannot effectively capture the extremely small changes from the frequency components in audio media, so watermarks can be embedded in audio media to protect the copyright of audio media. Figure 1 is the principle diagram of the watermarking scheme.

In embedding processing, watermarks will be encrypted firstly and then embedded into the audio media using the proposed embedding algorithm. Finally, the carried audio with the watermark will be uploaded to the Internet. When it is necessary to prove the copyright of the audio media, this carried audio will be implemented with the corresponding extracting algorithm to extract the encrypted watermark which will be decrypted using the correct key. In this scheme, the embedding and extracting algorithms are the core particularly. It can be seen that the extracting process and decryption are symmetric with the embedding process and encryption, respectively, and only those who have the

corresponding extracting algorithm and the correct key can obtain the watermark.

2.1. Chaotic Encryption of Watermark. In order to enhance the security of the watermark, it is necessary to encrypt the watermark before embedding it into the audio carrier. Because of its nonperiodical, continuous broadband, noise-like and long-term unpredictability, chaotic encryption is an information security protection technology that has developed rapidly in recent years and is especially suitable for security communications and other related fields [27].

Assume that the watermark can be converted into a binary stream W_1 , as follows:

$$W_1 = \{w_1(q), 1 \leq q \leq L_w\}, \quad (1)$$

where $w_1(q) \in \{0, 1\}$, L_w is the length of W_1 , and q is the serial number of the element in W_1 . Apply logistic mapping equation to generate a chaotic sequence $s(q)$ with the same size as W_1 as follows:

$$x_{q+1} = \alpha x_q (1 - x_q), \quad (2)$$

$$s(q) = \begin{cases} 1, & x_q \geq \delta, \\ 0, & x_q < \delta, \end{cases} \quad (3)$$

where $0 < x_q < 1$, $x_1 \in (0, 1)$ is the initial value when $q = 1$ and $\delta \in (0, 1)$ is a threshold to get $s(q)$. The logistic system is in chaos when $3.5699456 \leq \alpha \leq 4$. Exclusive OR operation is performed on W_1 and $s(q)$ to obtain the encrypted information W_2 shown in equation (4), where $\oplus$ represents the exclusive OR operator. The triple key $\text{Ch}(x_1, \alpha, \delta)$ will be the unique key that can be used to decrypt W_2 :

$$W_2 = \{w_2(q) = w_1(q) \oplus s(q), 1 \leq q \leq L_w\}. \quad (4)$$

In order to mark the start and end positions of W_2 , a synchronization code should be added to W_2 . For instance, add "1111 1111 0000 0000" in front of W_2 as the start flag, and add "1111 1111 0000 0000" after W_2 as the end flag.

2.2. Principle of the Embedding Algorithm. Suppose that A is the original audio with L_a sample points as follows:

$$A = \{a(k), 1 \leq k \leq L_a\}, \quad (5)$$

where k is the serial number of the element in A and $a(k)$ represents the k^{th} sample point. A is evenly divided into M audio fragments which can be expressed as A_l ($1 \leq l \leq M$, $M \geq L_w$); then, the length of each audio fragment is $N = \text{floor}(L_a/M)$, where $\text{floor}(\cdot)$ indicates that the data in brackets is rounded down. Apply r -level DWT on A_l to obtain a set of the approximation coefficient AC_r and r sets of the detail coefficient DC_i ($i = 1, 2, \dots, r$). AC_r contains the main frequency components in A_l , which may cause a serious decline in audio quality when there are minor changes in AC_r , so the watermark usually is not concealed into AC_r . DC_i contains the higher frequency components in A_l , so the watermark often can be concealed into these frequency band because of the less influence on audio

FIGURE 1: Principle diagram of the watermarking scheme.

quality. DC_1 represents the highest frequency band of the audio signal, and the degradation of audio quality caused by embedding the watermark in this frequency band is usually hard for human ear to perceive. Since the high-frequency components are vulnerable to attack, the detail coefficient close to AC_r is usually used to carry the watermark. In the following description, the principle of embedding algorithm will be illustrated by taking how to embed 1-bit binary information into the r -level detail coefficient DC_r of A_l as an example.

Divide DC_r into two data blocks, one is the former block $DC_r(j)$ ($j = 1, 2, \dots, N/2^{r+1}$) and the other is the latter block $DC_r(j + N/2^{r+1})$, where j is the serial number of the element in DC_r . Perform DCT on those two data blocks to obtain two sets of transform domain coefficients (TDC) which can be expressed as $TDC_1(r, j)$ and $TDC_2(r, j)$, respectively. The low-frequency components of $TDC_1(r, j)$ and $TDC_2(r, j)$ can be used to carry the watermark because they contain most of the energy of TDC, which is beneficial in improving the scheme's robustness. M_{l1} and M_{l2} are the average amplitudes of the low-frequency components of $TDC_1(r, j)$ and $TDC_2(r, j)$ as follows:

$$M_{l1} = \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} |TDC_1(r, p)|, \quad (6)$$

$$M_{l2} = \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} |TDC_2(r, p)|, \quad (7)$$

where $p = 1, 2, \dots, N/2^{r+2}$ and M_l is the average value of M_{l1} and M_{l2} , as shown in equation (8). In order to embed 1-bit binary information into the audio, $TDC_1(r, p)$ and $TDC_2(r, p)$ should be modified by the rules in equations (9) and (10), where λ is the embedding depth in the range of (0, 1). $TDC'_1(r, p)$ and $TDC'_2(r, p)$ are the modified coefficients.

$$M_l = \frac{1}{2} (M_{l1} + M_{l2}), \quad (8)$$

$$TDC'_1(r, p) = TDC_1(r, p) \times \frac{(1 + \lambda)M_l}{M_{l1}}, \quad (9)$$

$$TDC'_2(r, p) = TDC_2(r, p) \times \frac{(1 - \lambda)M_l}{M_{l2}}, \quad (10)$$

where M'_{l1} and M'_{l2} are the average amplitude of the modified coefficients $TDC'_1(r, p)$ and $TDC'_2(r, p)$ and can be calculated in the following equations:

$$\begin{aligned} M'_{l1} &= \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} |TDC'_1(r, p)| \\ &= \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} \left| TDC_1(r, p) \times \frac{(1 + \lambda)M_l}{M_{l1}} \right| \\ &= (1 + \lambda)M_l, \end{aligned} \quad (11)$$

$$\begin{aligned} M'_{l2} &= \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} |TDC'_2(r, p)| \\ &= \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} \left| TDC_2(r, p) \times \frac{(1 - \lambda)M_l}{M_{l2}} \right| \\ &= (1 - \lambda)M_l. \end{aligned} \quad (12)$$

The average value M'_l of M'_{l1} and M'_{l2} can be calculated in equation (13). Then, the variation Δ of M'_{l1} and M'_{l2} can be expressed in equation (14). Thus, the embedding depth λ of A_l can be calculated according to equation (15):

$$M'_l = \frac{M'_{l1} + M'_{l2}}{2} = \frac{(1 + \lambda)M_l + (1 - \lambda)M_l}{2} = M_l, \quad (13)$$

$$\begin{aligned} \Delta &= |M'_{l1} - M'_{l2}| = |(1 + \lambda)M_l - (1 - \lambda)M_l| \\ &= 2\lambda M_l = 2\lambda M'_l, \end{aligned} \quad (14)$$

$$\lambda = \frac{|M'_{l1} - M'_{l2}|}{2M'_l} = \frac{|M_{l1} - M_{l2}|}{2M_l}. \quad (15)$$

Two embedding depths with different values can be used to represent the status of 1-bit binary information "1" and "0," so the embedding rules can be designed in the following equation when embedding the q^{th} bit of the watermark into A_l , where $0 < \lambda_1 < \lambda_2 < 1$, $l = l_0 + q$, and A_{l_0+1} is the first audio fragment to carry the watermark:

$$\lambda(q) = \begin{cases} \lambda_1, & w_2(q) = 1, \\ \lambda_2, & w_2(q) = 0. \end{cases} \quad (16)$$

The rules shown in equations (9) and (10) can be named as the first rule, and the second rule can be shown as follows:

$$TDC'_1(r, p) = TDC_1(r, p) \times \frac{(1 - \lambda)M_l}{M_{l1}}, \quad (17)$$

$$TDC'_2(r, p) = TDC_2(r, p) \times \frac{(1 + \lambda)M_l}{M_{l2}}. \quad (18)$$

Then, the average amplitudes M'_{l1} and M'_{l2} of $C'_1(r, p)$ and $C'_2(r, p)$ will be calculated in equations (19) and (20) according to the second rule. When $M_{l1} \geq M_{l2}$, $\Delta_1 = |M_{l1} - M'_{l1}|$, and $\Delta_2 = |M_{l2} - M'_{l2}|$ are the variations, as shown in Figure 2(a) according to the first rule and also can be shown in Figure 2(b) according to the second rule:

$$M'_{l1} = \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} |C'_1(r, p)| = (1 - \lambda)M_l, \quad (19)$$

$$M'_{l2} = \frac{2^{r+2}}{N} \sum_{p=1}^{N/2^{r+2}} |C'_2(r, p)| = (1 + \lambda)M_l. \quad (20)$$

Both Δ_1 and Δ_2 in Figure 2(a) are smaller than those in Figure 2(b), which indicates that the first rule is better than the second rule in improving the scheme transparency when $M_{l1} \geq M_{l2}$. Similarly, the comparison of the two graphs in Figure 3 about Δ_1 and Δ_2 also implies that the second rule should be chosen to embed the information bit when $M_{l1} < M_{l2}$.

Finally, the inverse discrete cosine transform (IDCT) and the inverse discrete wavelet transform (IDWT) are performed on the modified coefficients in turn to obtain the carried audio fragment A'_l . The flowchart of the embedding algorithm is shown in Figure 4. The embedding process can be described as follows:

- Step 1: convert the watermark into a binary stream W_1 and encrypt it to obtain W_2 .
- Step 2: add synchronization code at the beginning and end of W_2 .
- Step 3: divide the original audio A into M audio fragments A_l ($1 \leq l \leq M$).
- Step 4: apply DWT on A_l to obtain the r -level detail coefficient DC_r .
- Step 5: divide DC_r to obtain the former block $DC_r(j)$ and the latter block $DC_r(j + N/2^{r+1})$.
- Step 6: apply DCT on two data blocks to obtain $TDC_1(r, j)$ and $TDC_2(r, j)$, respectively.
- Step 7: calculate M_{l1} , M_{l2} , and M_l according to equations (6)–(8).
- Step 8: if $M_{l1} \geq M_{l2}$, modify $TDC_1(r, j)$ and $TDC_2(r, j)$ by the first rule; otherwise, use the second rule. If $W_2(q)$ is equal to 1, then $\lambda = \lambda_1$; otherwise, $\lambda = \lambda_2$.
- Step 9: apply IDCT and IDWT in turn to recover the carried audio fragment A'_l .
- Step 10: repeat Step 4 to Step 7 until all watermark bits are concealed.
- Step 11: recombine all audio fragments to recover the carried audio A' .

2.3. Principle of the Extracting Algorithm. The extracting process is symmetric with the embedding process. When extracting the watermark from the carried audio A' , the embedding depth in each audio fragment will be calculated

according to equation (15), and then, the 1-bit binary information will be determined by comparing it with the overall average embedding depth λ_0 which can be calculated according to equation (21). The extracting rule is expressed in equation (22):

$$\lambda_0 = \frac{1}{L_w} \sum_{q=1}^{L_w} \lambda(q), \quad (21)$$

$$w'_2(q) = \begin{cases} 1, & \lambda(q) \leq \lambda_0, \\ 0, & \lambda(q) > \lambda_0. \end{cases} \quad (22)$$

When all binary bits are extracted from the audio fragments A'_l ($1 \leq l \leq M$), the data between the start flag and the end flag is the encrypted W'_2 . Remove the synchronization code, and then decrypt W'_2 using the triple key $Ch(x_1, \alpha, \delta)$ to obtain the watermark. It can be seen from the principle of the extracting algorithm that this scheme has high security because only those who have the corresponding extracting algorithm and the correct key can access the watermark. The flowchart of the extracting algorithm can be shown in Figure 5. The process of extracting the watermark can be described as follows:

- Step 1: divide the carried audio A' into M audio fragments A'_l .
- Step 2: apply DWT on A'_l to obtain the r -level detail coefficient DC'_r .
- Step 3: divide DC'_r to obtain the former block $DC'_r(j)$ and the latter block $DC'_r(j + N/2^{r+1})$.
- Step 4: apply DCT on two data blocks to obtain $TDC'_1(r, j)$ and $TDC'_2(r, j)$, respectively.
- Step 5: calculate M'_{l1} , M'_{l2} , and M'_l .
- Step 6: calculate λ according to equation (15).
- Step 7: repeat Step 2 to Step 6 until all λ are calculated.
- Step 8: calculate λ_0 according to equation (21).
- Step 9: extract all binary bits according to equation (22).
- Step 10: remove the synchronization code and decrypt W'_2 to obtain the watermark.

2.4. Optimization of the Parameters Based on GA. Three important parameters (r, λ_1, λ_2) of this scheme have an important effect on the overall performance of the scheme. To stimulate the best performance in different applications, GA is used to search for the optimal parameters according to the specific performance indicators. The fitness function Fitness can be constructed with transparency, payload capacity, and robustness as follows:

$$\begin{cases} \text{SNR} > \text{SNR}_0, \\ \text{Cap} > \text{Cap}_0, \\ \text{fitness} = \frac{1}{\text{BER}}. \end{cases} \quad (23)$$

SNR and BER can be expressed in equations (24) and (25), respectively, and Cap represents the payload capacity. SNR_0 and Cap_0 are the thresholds of transparency and payload capacity:

FIGURE 2: Coefficient variations under different rules when $M_{I1} \geq M_{I2}$: (a) the first rule; (b) the second rule.

FIGURE 3: Coefficient variations under rules when $M_{I1} < M_{I2}$: (a) the first rule; (b) the second rule.

$$\text{SNR}(A, A') = 101g\left(\frac{\sum_{j=1}^K A^2}{\sum_{j=1}^K (A' - A)^2}\right), \quad (24)$$

$$\text{BER} = \left(\frac{\sum_{q=1}^{L_w} w(q) \oplus w'(q)}{L_w} \times 100\%\right), \quad (25)$$

where A and A' represent the original audio and the carried audio, respectively, and $w(q)$ and $w'(q)$ represent the original watermark and the extracted watermark. The population POP consists of C_1 chromosomes, and the length of each chromosome is C_2 . Chromosomes will be encoded by using a binary encoding approach, as shown in equation (26), where Chrom_1 represents the first chromosome; $\text{Bin}_1(r)$, $\text{Bin}_1(\lambda_1)$, and $\text{Bin}_1(\lambda_2)$ represent the binary of r , λ_1 , and λ_2 ; and their lengths are L_r , L_{λ_1} , and L_{λ_2} , respectively:

$$\text{POP} = \begin{bmatrix} \text{Chrom}_1 \\ \vdots \\ \text{Chrom}_{C_1} \end{bmatrix}_{L_1 \times K_1} = \begin{bmatrix} \text{Bin}_1(r) & \text{Bin}_1(\lambda_1) & \text{Bin}_2(\lambda_2) \\ \vdots & \vdots & \vdots \\ \text{Bin}_{C_1}(r) & \text{Bin}_{C_2}(\lambda_1) & \text{Bin}_{C_1}(\lambda_2) \end{bmatrix}_{C_1 \times C_2}, \quad (26)$$

where r can be obtained by Converting $\text{Bin}_1(r)$ from binary to decimal. The transformation relationship between the chromosome and the two parameters (λ_1, λ_2) can be shown as follows:

$$\lambda_1 = \lambda_{01} + (\lambda_{02} - \lambda_{01}) \frac{B2D[\text{Bin}(\lambda_1)]}{B2D(L_{\lambda_1})}, \quad (27)$$

$$\lambda_2 = \lambda_{03} + (\lambda_{04} - \lambda_{03}) \frac{B2D[\text{Bin}_1(\lambda_2)]}{B2D(L_{\lambda_2})}, \quad (28)$$

where $B2D[\]$ means converting the data in brackets from binary to decimal and $\lambda_{01} < \lambda_{02} < \lambda_{03} < \lambda_{04} \in (0, 1)$. The detailed process can be described as follows:

Step 1: parameter initialization. Set the crossover probability p_c , the mutation probability p_m , λ_{01} , λ_{02} ,

λ_{03} , λ_{04} , Cap_0 and SNR_0 and then generate an initial population POP_0 .

Step 2: calculate the four parameters $(r, \lambda_1, \lambda_2)$ and then execute the embedding algorithm in Section 2.2 to obtain the carried audio.

Step 3: attack test. Apply some attacks on the carried audio, and calculate SNR according to equation (24).

Step 4: choose all qualified chromosomes that meet $\text{SNR} > \text{SNR}_0$ and $\text{Cap} > \text{Cap}_0$. Then, execute the extracting algorithm to calculate BER according to equation (25).

Step 5: calculate the fitness value according to equation (23) to obtain the best chromosome with the largest fitness value.

Step 6: apply selection operation by roulette on the best chromosome to generate the transition population POP'_0 .

Step 7: apply crossover operation on two adjacent chromosomes except for the best chromosome to obtain a new transition population POP''_0 .

Step 8: apply mutation operation on each chromosome except for the best chromosome to obtain the next generation population POP_1 .

Step 9: repeat Step 2 to Step 8 until the global optimal chromosome appears.

3. Performance Evaluation

In this section, the performance of this scheme is evaluated, including transparency, security, capacity, robustness, and complexity. The detailed experimental parameters can be described as follows: (1) $\text{SNR}_0 = 25$, $\text{Cap}_0 = 172.27$, $p_c = 0.8$, $p_m = 0.1$, $C_1 = 20$, $L_r = 4$, $L_{\lambda_1} = 8$, $L_{\lambda_2} = 8$; (2) $\lambda_{01} = 0.1$, $\lambda_{02} = 0.3$, $\lambda_{03} = 0.5$, and $\lambda_{04} = 0.9$; (3) the triple key is $\text{Ch}(0.2, 3.9, 0.5)$; (4) use a binary image as the watermark, as shown in Figure 6; and (5) the original audio carriers are 10 male voices and 10 female voices, all of which come from TIMIT standard database.

FIGURE 4: Flowchart of the embedding algorithm.

FIGURE 5: Flowchart of the extracting algorithm.

FIGURE 6: A binary image as the watermark. (a) The original image with sizes of 100×100 . (b) The encrypted image by chaotic encryption.

SNR, the subjective difference grades (SDG), and the object difference grade (ODG) will be used to evaluate the transparency of this proposed scheme. SDG refers to the original audio and the carried audio being provided to the same group of listeners to distinguish the difference and give a subjective score. The closer the average score is to 0, the better the audio quality will be. ODG is one of the output values obtained from the perceptual evaluation of audio quality (PEAQ), so it can be used to give an objective score from -5 to 0 for audio. BER refers to the ratio of the number of erroneous bits to the total number of bits in the extracted information, which can be used to evaluate the robustness of the scheme. Correlation coefficient (NC) refers to the similarity between the original information and the extracted information, as defined in equation (29). It is located in the range of $[0, 1]$. The larger the NC, the more similar the original information and the extracted information, and the stronger the robustness of the scheme.

$$NC(w, w') = \frac{\sum_{q=1}^{L_w} w(q) \times w'(q)}{\sqrt{\sum_{q=1}^{L_w} w(q)^2 \sum_{q=1}^{L_w} w'(q)^2}}. \quad (29)$$

3.1. Transparency and Capacity. To illustrate the transparency of this scheme, an audio clip lasting for about 3 seconds was picked up randomly from the tested audio signals to compare the waveform and spectrum before and after embedding information, as shown in Figure 7. It can be seen from the figures that the waveforms of the original audio are very similar to those of the carried audio, and there is no obvious difference between them, so this scheme has high transparency. In the case of no attack, the average experimental results are listed in Table 1.

The experimental results in Table 1 confirm that this scheme has higher transparency and robustness compared with [22, 26], which can be seen from the results about SNR,

SDG, ODG, BER, and NC. In the case of providing the capacity of 172.27 bps, although the transparency of our scheme is slightly lower than that of [19] and BER of our scheme is significantly better than that of [19], which shows that the robustness of our scheme is stronger.

3.2. Robustness. In this section, robustness evaluation can be carried out through the following three steps. Firstly, implement various types of attacks on the carried audio, then use the developed extracting algorithm to extract the watermark in the carried audio, and finally calculate BER and NC of the watermark. The attack types considered in our test are shown in Table 2.

Figure 8 shows the waveform comparison of the original audio and the carried audio that they are both compressed by MP3 with 64 kbps (only show an audio clip lasting about 3 seconds randomly). It can be seen from Figure 8 that there is no obvious abnormality in their waveforms. The waveform comparison diagrams after they suffered other attacks are similar to this, which will not be shown here one by one. Table 3 shows the result about BER of the watermark extracted from the carried audio under the above attacks. The extracted watermarks and their NC values are listed in Figure 9. According to the experimental results in Tables 1 and 3 and Figure 9, the payload capacity and SNR of this scheme reach 172.27 bps and 25 dB, respectively, which indicates that this scheme has large capacity and high transparency, so it can be used to protect the copyright of audio media without affecting the audio quality itself. The extracted watermarks are very similar to the original watermark under most attacks except for additive noise with 20 dB, the values of BER are below 1.60%, and NC values are above 0.9237, which implies that this scheme has strong robustness, so the watermark will not be destroyed or lost due to some conventional signal processing operations in the process of using the audio.

FIGURE 7: Waveform comparison. (a) The original audio. (b) The carried audio.

TABLE 1: Experimental results (no attack).

Performance		Our scheme	Reference [19]	Reference [22]	Reference [26]
SNR	Male	25.06	28.9	24.35	22.45
	Female	25.02	26.5	24.31	22.42
SDG	Male	-0.10	-0.08	-0.25	-1.01
	Female	-0.13	-0.16	-0.27	-1.21
ODG	Male	-1.10	-0.87	-1.35	-2.84
	Female	-1.24	-0.98	-1.46	-2.92
BER	Male	0.12	1.32	0.13	0.00
	Female	0.10	1.13	0.11	0.00
NC	Male	0.998	0.948	0.996	1
	Female	0.999	0.962	0.998	1
Cap	Male	172.27	172.27	172.27	125
	Female	172.27	172.27	172.27	125

TABLE 2: Attack types.

Item	Attack type	Description
A	Additive noise	Add Gaussian noise with 35 dB on the carried audio
B	Additive noise	Add Gaussian noise with 30 dB on the carried audio
C	Additive noise	Add Gaussian noise with 20 dB on the carried audio
D	Amplitude scaling	Scale the amplitude of the carried audio to 0.8 times
E	Amplitude scaling	Scale the amplitude of the carried audio to 1.2 times
F	Low-pass filtering	Apply the low-pass filter with a cutoff frequency of 4 kHz
G	Low-pass filtering	Apply the low-pass filter with a cutoff frequency of 8 kHz
H	MP3 compression	Apply MP3 compression with 64 kbps
I	MP3 compression	Apply MP3 compression with 128 kbps
J	Echo jamming	Add an echo with 10 ms to the carried audio
K	Echo jamming	Add an echo with 50 ms to the carried audio
L	Resampling	Modify the sampling rate with 44.1 kHz-22.05 kHz-44.1 kHz
M	Requantizing	Modify the quantizing bit with 16 bits-8 bits-16 bits per sample

Compared with the schemes in the other references, this scheme has better overall performance. Although the transparency of our scheme is slightly lower than that in [19], our scheme is more robust when resisting all the attacks listed in Table 2. Under the same capacity, this scheme has stronger robustness than that in [22] when against most signal processing operations. This proposed scheme also has a larger capacity and better robustness than that in [26]. The average values of BER in [26] reach 2.87% and 17.92% when resisting

amplitude scaling, while those of this proposed scheme are only 0.19% and 0.34%. Amplitude scaling is the most common way that audio media may suffer from in the process of being used, so this proposed scheme is more practical.

3.3. Security and Complexity. According to the Kerckhoffs criterion, the security of the scheme should not only depend on the scheme itself but also be further strengthened with

FIGURE 8: Waveform comparison (64 kbps MP3 compression): (a) The original audio. (b) The carried audio.

TABLE 3: Experimental results of BER (%) under attacks.

Item	Our scheme	Reference [19]	Reference [22]	Reference [26]
A	0.84	2.46	0.01	0.78
B	1.60	3.16	0.08	1.02
C	4.70	5.14	0.43	3.56
D	0.19	1.12	0.13	2.87
E	0.20	1.12	0.13	17.92
F	0.19	23.31	6.44	0.39
G	0.18	8.54	5.32	0.00
H	1.05	3.68	4.12	1.95
I	0.21	0.74	0.11	0.34
J	0.22	5.14	0.41	0.01
K	0.25	7.56	0.83	0.52
L	0.18	4.83	0.68	0.00
M	0.34	5.68	0.42	0.78

FIGURE 9: Continued.

FIGURE 9: The extracted watermarks under different attacks: (a) additive noise (35 dB); (b) additive noise (30 dB); (c) additive noise (20 dB); (d) amplitude scaling (0.8); (e) amplitude scaling (1.2); (f) low-pass filtering (4 kHz); (g) low-pass filtering (8 kHz); (h) MP3 compression (64 kbps); (i) MP3 compression (128 kbps); (j) echo jamming (10 s); (k) echo jamming (50 s); (l) resampling; (m) requantizing; (n) no attack.

FIGURE 10: Spectrogram comparison: (a) the original audio; (b) the carried audio.

encryption technology. Therefore, the size of the key space determines the security degree of the scheme. This scheme uses a triple key to encrypt the watermark in chaos. Three parameters of this triple key are all taken in the real field, so this scheme has infinite key space in theory, but in fact, they are affected by the word length of the computer system; thus, their key space is limited. In addition, the comparison of the spectrogram in Figure 10 also indicates that the characteristics of the carried audio have some changes because of carrying watermarks.

The running time of the algorithm can be used to test its complexity. In our scheme, GA is used to optimize the main parameters of this algorithm, so the overall running time is also related to the search efficiency of GA. The

average embedding time of this algorithm is 35.4561 seconds, and the average extraction time is 11.6896 seconds.

From the above analysis, it can be seen that our proposed scheme has better performance, mainly because it combines the advantages of DWT and DCT to enhance the robustness, optimizes the important parameters of the scheme by using GA, and adjusts the embedding rules to improve the transparency based on the principle of least modification of TDC. However, the complexity of this scheme is high because it takes more time to search for the best algorithm parameters using GA. In addition, slight changes in the spectrogram may also expose the watermark hidden in the audio media.

4. Conclusions

In this paper, a robust and blind audio watermarking scheme based on GA is proposed in the dual transform domain, which can be used to protect the copyright of the audio media in cyberspace. This scheme is developed in the DWT-DCT dual transform domain, so it has strong robustness to prevent from losing the watermark hidden in the carried audio which may suffer from various attacks. Furthermore, this scheme utilizes GA to optimize the important parameters adaptively for meeting the performance requirements in different applications. Besides, this scheme adjusts the embedding rules based on the principle of minimal modification to the carried audio to improve transparency. When embedding watermarks, firstly, the carried audio will be divided into many audio fragments, and DWT will be performed on each audio fragment to obtain one set of appropriate wavelet coefficients which will be used to carry the watermark. Secondly, those appropriate wavelet coefficients will be divided into two groups of data blocks which will be implemented by DCT to obtain two groups of TDC, respectively. Finally, two different embedding depths are used to modify these two groups of TDC for embedding the binary watermark according to the designed embedding equation. When extracting the watermark, the embedding depth of each audio fragment will be calculated firstly and then compared with the overall average embedding depth to extract the binary watermark according to the designed extracting equation, so this scheme is blind because it can extract the watermark without the carried audio.

Experimental results confirm that this proposed scheme can be used to embed watermarks into audio media without affecting the audio to be used normally and blindly detect it. Compared with other schemes in the relevant references, it has achieved excellent performance, such as the strong robustness when withstanding MP3 compression, additive noise, low-pass filtering, requantizing, resampling, amplitude scaling, and echo jamming. The SNR of the carried audio reaches more than 25 dB in the case of the payload capacity of 172.27 bps, which indicates that this proposed scheme has good transparency and a large capacity. However, due to the intervention of the genetic algorithm, this algorithm has high complexity. In the next research work, we will strive to reduce complexity and improve the security of the scheme.

Data Availability

All audio signals and images tested in our experiment can be used under the public platform.

Conflicts of Interest

All authors declare no conflicts of interest.

Acknowledgments

This research work was funded by the High-Level Talent Scientific Research Foundation of Jinling Institute of Technology, China (Grant no. jit-b-201918), the Natural

Science Foundation of the Jiangsu Higher Education Institutions (Grant no. 20KJB110004), the National Natural Science Foundation of China (Grant no. 11601202), the “333 Project” Scientific Research Support Project of Jiangsu in 2020 (Research on Privacy Protection Technology for Big-Data), and the Scientific Research Fund Incubation Project of Jinling Institute of Technology in 2020 (Research on Multipath TCP Attack and Defense Mechanism).

References

- [1] F. Khelifi, T. Brahimi, J. Han, and X. Li, “Secure and privacy-preserving data sharing in the cloud based on lossless image coding,” *Signal Processing*, vol. 148, pp. 91–101, 2018.
- [2] Z. Liu, J. Huang, X. Sun, and C. Qi, “A security watermark scheme used for digital speech forensics,” *Multimedia Tools and Applications*, vol. 76, no. 7, pp. 9297–9317, 2017.
- [3] W. Jiang, X. Huang, and Y. Quan, “Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division,” *Signal Processing*, vol. 162, pp. 153–160, 2019.
- [4] Q. L. Wu and M. Wu, “A novel robust audio watermarking algorithm by modifying the average amplitude in transform domain,” *Applied Sciences*, vol. 8, pp. 1–19, 2018.
- [5] Q. Qian, H. Wang, X. Sun, Y. Cui, H. Wang, and C. Shi, “Speech authentication and content recovery scheme for security communication and storage,” *Telecommunication Systems*, vol. 67, no. 4, pp. 635–649, 2018.
- [6] A. Sukumar, V. Subramaniaswamy, V. Vijayakumar, and L. Ravi, “A secure multimedia steganography scheme using hybrid transform and support vector machine for cloud-based storage,” *Multimedia Tools and Applications*, vol. 79, pp. 10825–10849, 2020.
- [7] M. Y. Valandar, M. J. Barani, P. Ayubi, and M. Aghazadeh, “An integer wavelet transform image steganography method based on 3D sine chaotic map,” *Multimedia Tools and Applications*, vol. 78, pp. 9971–9989, 2019.
- [8] Y. Y. Wang, “An adaptive watermark detection algorithm for vector geographic data,” *KSI Transactions on Internet and Information Systems*, vol. 14, pp. 323–343, 2020.
- [9] Z. Li, L. Meng, X. Jiang, and Z. Li, “High capacity HEVC video hiding algorithm based on EMD coded PU partition modes,” *Symmetry*, vol. 11, pp. 1–18, 2019.
- [10] I. Natgunanathan and Y. Xiang, “Patchwork-based multilayer audio watermarking,” *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 25, pp. 2176–2187, 2017.
- [11] B. Lei, I. Song, and S. A. Rahman, “Robust and secure watermarking scheme for breath sound,” *Journal of Systems and Software*, vol. 86, pp. 1638–1649, 2013.
- [12] Z. P. Su, G. F. Zhang, and F. Yue, “SNR-constrained heuristics for optimizing the scaling parameter of robust audio watermarking,” *IEEE Transactions on Multimedia*, vol. 20, pp. 2631–2644, 2018.
- [13] M. A. Nematollahi, S. Al-Haddad, and F. Zarafshan, “Blind digital speech watermarking based on Eigen-value quantization in DWT,” *Journal of King Saud University- Computer and Information Sciences*, vol. 27, pp. 58–67, 2015.
- [14] H. T. Hu and L. Y. Hsu, “Incorporating spectral shaping filtering into DWT based vector modulation to improve blind audio watermarking,” *Wireless Personal Communications*, vol. 94, no. 2, pp. 221–240, 2017.

- [15] R. Ronaldo, G. Pedro, and C. Mylène, "Detecting tampering in audio-visual content using QIM watermarking," *Information Sciences*, vol. 328, pp. 127–143, 2016.
- [16] Y. Erfani and S. Siahpoush, "Robust audio watermarking using improved TS echo hiding," *Digital Signal Processing*, vol. 19, pp. 809–814, 2009.
- [17] H. T. Hu and L. Y. Hsu, "Robust transparent and high capacity audio watermarking in DCT domain," *Signal Processing*, vol. 109, pp. 226–235, 2015.
- [18] H. N. Huang and S. T. Chen, "Optimization-based embedding for wavelet-domain audio watermarking," *Journal of Signal Processing Systems*, vol. 80, pp. 197–208, 2015.
- [19] C. Y. Hsu, S. Y. Tu, and C. T. Yang, "Digital audio signal watermarking using minimum-energy scaling optimisation in the wavelet domain," *IET Signal Processing*, vol. 14, pp. 791–802, 2021.
- [20] I. Natgunanathan, Y. Xiang, and Y. Rong, "Robust patchwork-based watermarking method for stereo audio signals," *Multimedia Tools and Applications*, vol. 72, pp. 1387–1410, 2014.
- [21] E. Fathi, "An efficient singular value decomposition algorithm for digital audio watermarking," *International Journal of Speech Technology*, vol. 12, pp. 27–45, 2009.
- [22] J. F. Li, H. X. Wang, T. Wu, X. M. Sun, and Q. Qian, "Norm ratio-based audio watermarking scheme in DWT domain," *Multimedia Tools and Applications*, vol. 77, pp. 14481–14497, 2018.
- [23] B. Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT," *Signal Processing*, vol. 91, pp. 1973–1984, 2011.
- [24] B. K. Vivekananda, I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Digital Signal Process*, vol. 20, pp. 1547–1558, 2010.
- [25] P. K. Dhar and T. Shimamure, "Blind audio watermarking in transform domain based on singular value decomposition and exponential-log operations," *Radioengineering*, vol. 26, pp. 552–561, 2017.
- [26] X. Y. Wang, Q. L. Shi, S. M. Wang, and H. Y. Yang, "A blind robust digital watermarking using invariant exponent moments," *AEU-International Journal of Electronics and Communications*, vol. 70, pp. 416–426, 2016.
- [27] Y. V. Milad, A. Peyman, and J. B. Milad, "A new transform domain steganography based on modified logistic chaotic map for color images," *Journal of Information Security and Applications*, vol. 34, pp. 142–151, 2017.