Hindawi
Security and Communication Networks
Volume 2021, Article ID 4471117, 10 pages
https://doi.org/10.1155/2021/4471117

Research Article

A Robust MP3 Steganographic Method against Multiple Compressions Based on Modified Discrete Cosine Transform

Yunzhao Yang,[1,2] **Xiaowei Yi**,[1,2] **Xianfeng Zhao**,[1,2] **and Jinghong Zhang**[1,2]

[1]*State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*
[2]*School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100093, China*

Correspondence should be addressed to Xianfeng Zhao; zhaoxianfeng@iie.ac.cn

Received 25 August 2021; Accepted 23 November 2021; Published 21 December 2021

Academic Editor: Jinwei Wang

MP3 appears in various social networking sites wildly, and it is very suitable to be applied for covert communication indeed. However, almost all social networking sites recompress the uploaded MP3 files, which leads to the ineffectiveness of the existing MP3 steganographic methods. In this paper, a robust MP3 steganographic algorithm is proposed with the ability of multiple compressions resistance. First, we discover a new embedding domain with strong robustness. The scalefactor bands of higher energy are applied as the embedding bands. The message bits are embedded by adjusting the position of the MDCT coefficients with the largest magnitude in the embedding bands. Besides, the embedding and extraction operations are realized in the process of MP3 decoding at the same time. Experimental results illustrate that our proposed method is of strong robustness against multiple MP3 compressions. The bit error rate is less than 1% at the MP3 bitrate of 320 kbps. It is worth mentioning that the proposed method is proved to be applicable to social networking sites, such as SoundCloud, for covert communication. Our method achieves a satisfactory level of embedding capacity, imperceptibility, and undetectability.

1. Introduction

Steganography is the art of embedding secret messages into innocent-looking digital files for cover communication. To avoid the suspicion of the other party, in addition to ensuring the undetectability of the carrier files, the covert communication behavior also needs to be hidden. Using social networking sites (SNS) as steganographic channels can better cover up the trace of the communication. However, many social networking sites transcode the uploaded files for higher transmission efficiency, which will bring an unpredictable impact on the embedding elements of the carrier files. Thus, the steganographic methods must be robust enough against the transcoding while achieving a satisfactory level of undetectability performance.

Hitherto, there is no MP3 steganographic method that can resist transcoding on social networking sites. Existing MP3 steganographic methods can be divided into two major categories: encoding parameters-based methods and encoding data-based ones. For the encoding parameters-based methods, the messages are embedded by adjusting window type [1], quantization step [2, 3], Huffman table index [4], and so forth. The modification of these parameters will not cause sharp disturbances to the audio signal. For the encoding data-based methods, the embedding operation is limited by the principle of the MP3 encoding. Without disturbing the frame-offset effect, the sign bits [5], linbits [6, 7], and the Huffman code [8–12] in the encoding data can be applied to embed messages. It is noteworthy that most audio sharing platforms, such as SoundCloud and Himalayan, will compress the uploaded audio files with their fixed bitrate, even if the uploaded MP3 file meets the platform's specific rules. However, the MP3 encoding is a kind of lossy compression, which means that the encoding parameters and data will be changed greatly after the transcoding. Therefore, these methods cannot be used for covert communication on SNS directly.

MP3 watermarking methods are also robust to social networking sites to some extent. According to the embedding position, they can be divided into three categories: time domain watermark [13–15], transform domain watermark [16–18], and hybrid domain watermark [19–21]. Unlike watermarking, the design of robust MP3 steganography does not need to consider clipping, resampling, and filtering attacks. After all, steganography algorithms are applied to covert communication, so they require a lower bit error rate (BER) and higher embedding capacity on the premise of compression attacks resistance. Thus, the MP3 watermarking methods cannot be seamlessly applied to robust MP3 steganography.

To fill this gap, a robust MP3 steganographic method (RMSM) is proposed in this paper, which is implemented during the MP3 decoding process. Firstly, considering that the energy relations are often robust to various digital signal processings, among the 21 scalefactor bands in each frame, the first k scalefactor bands with the largest energy are selected as the embedding bands. In this way, even after multiple compressions, the corresponding embedding bands can be located to extract the secret messages. Secondly, MDCT coefficients with larger energy are of robustness due to their concentration on most of the signal information. So the MDCT coefficients with the largest magnitude in each embedding band are selected as the embedding elements. Thirdly, by adjusting the position of these embedding elements in the scalefactor band, the parity of the coefficients position index is matched with the messages. Finally, the generated stego audio files are uploaded to the SNS for covert communication test. Experimental results demonstrate that the embedding capacity of our proposed method reaches 479 bits/s, which meets the need of covert communication. Moreover, the proposed method provides strong robustness against multiple compressions attacks, so it can be applied to covert communication based on social networking sites. Meanwhile, the proposed method has a good performance against the attacks of the steganalysis based on the statistical characteristics of MDCT coefficients.

The contribution of this paper is twofold: (1) We propose a robust embedding domain to resist MP3 multiple compressions. In addition, this domain is able to make a good balance between embedding capacity and impeccability. (2) We are the first to implement the robust MP3 steganographic method on SNS. The proposed RMSM is verified with the properties of good robustness and security for covert communication on SNS such as SoundCloud.

The rest of this paper is organized as follows. Section 2 presents some preliminaries about MP3 encoding and decoding. The RMSM is presented in Section 3. Section 4 proposes a specific method suitable for SoundCloud. Experiments are illustrated in Section 5. Finally, the paper is concluded in Section 6.

2. Overview of MP3 Encoding and Decoding

The procedure of the MP3 encoding is shown in Figure 1, which is mainly divided into five parts [22]: analysis filter bank, psychoacoustic model, scaler, and quantizer, Huffman coding, and bitstream formatting. The original audio is encoded by pulse code modulation (PCM), which is WAV format. The WAV audio is split into frames of 1152 samples. The frames are further segmented into two granules of 576 samples each. The frames are sent to both the analysis filter bank and the Fast Fourier Transform (FFT). The analysis filter bank divides the audio signal into 32 uniform frequency subbands. The psychoacoustic model (PAM) process analyzes the audio signal via FFT, which determines the type of windows used in the MDCT process. In addition, the masking thresholds for each scalefactor band are calculated in the process of PAM, which is used to control the quantization noise. The subband samples are transformed to the frequency domain by the MDCT process with four types of windows. Then the nonuniform quantization iteratively adjusts the quantization parameters and scalefactors until the quantization noise level of each scalefactor band falls below the masking thresholds. The quantized MDCT coefficients are encoded into Huffman code losslessly. Finally, the codestream and the side information are formatted into the MP3 file.

Scalefactor bands are groups of frequency lines that are scaled by a single factor and they can approximate the auditory critical bands. Hence, the masking thresholds and quantization noise are calculated in the scalefactor bands. After the MDCT process, each granule obtains 576 frequency lines (i.e., MDCT coefficients), which are ordered in terms of increasing frequency. There are predetermined tables which indicate the mapping between these coefficients and the scalefactor bands. For each sampling rate, there are 21 bands for long windows and 12 bands for short bands. Taking 44.1 kHz sampling rate as an example, the scalefactor band division of long window is shown in Table 1.

MP3 decoding is the reverse process of its encoding, and the basic diagram of the decoder is illustrated in Figure 2. The first unit is reading the MP3 data frame by frame according to the synchronization word in the header part of the frame. The next CRC check word is used for error detection. According to the stored side information, the Huffman codes are decoded into QMDCT coefficients. Then the MDCT coefficients are obtained by inverse quantization. This is done through a descaling operation which is based on the QMDCT coefficients obtained by Huffman decoding and the scalefactors extracted from the scalefactor decoder. After that, the MDCT coefficients are transformed using the IMDCT block and the frequency inversion to the time domain. Finally, the PCM samples that can be used for playback are obtained by synthesis polyphase filter bank.

3. The Proposed Method

Compared with the MP3 encoder, the MP3 decoder is implemented without the quantization loop and psychoacoustic model, which brings about lower computational complexity and shorter execution time. In consideration of practicality and facility, the MP3 robust steganographic method is implemented in the decoding process. More importantly, a new embedding domain is applied to enhance the robustness of the steganographic method.

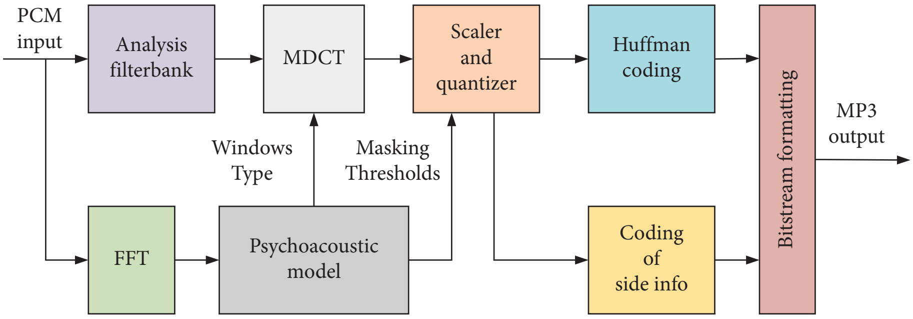


Figure 1: Sketch of the basic structure of the MP3 encoder.

Table 1: Frequency lines division of scalefactor band at 44.1 kHz.

No.	Width of band	Index of start	Index of end	No.	Width of band	Index of start	Index of end
0	4	0	3	11	12	62	73
1	4	4	7	12	16	74	89
2	4	8	11	15	20	90	109
3	4	12	15	14	24	110	133
4	4	16	19	15	28	134	161
5	4	20	23	16	34	162	195
6	6	24	29	17	42	196	237
7	6	30	35	18	50	238	287
8	8	36	43	19	54	288	341
9	8	44	51	20	76	342	417
10	10	52	61				

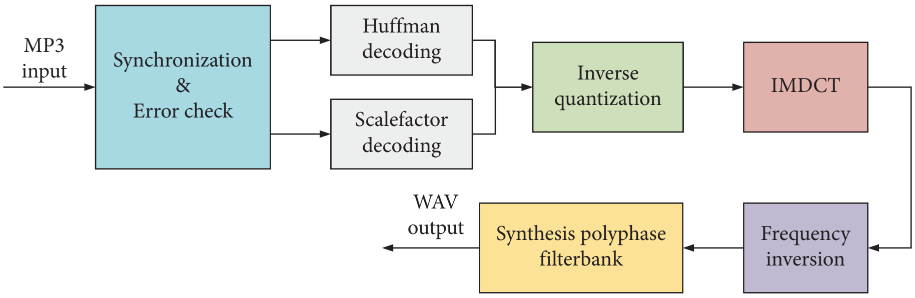


Figure 2: Sketch of the basic structure of the MP3 decoder.

3.1. Embedding Domain Construction. The key to the robust steganographic method is the embedding domain construction. A qualified embedded domain means that the corresponding embedding elements can be found during extraction even going through various attacks. According to the characteristics of signal processing, the energy relation of each region will remain stable after multiple compressions. In the psychoacoustic model of MP3 encoding, the scalefactor band is the basic unit used for the masking threshold calculation. Thus, we conclude that the energy relation among the scalefactor bands has strong robustness.

To verify whether the conjecture is reasonable, we perform a simulation as follows. Firstly, an MP3 clip with a sampling rate of 44.1 kHz and a duration of 10 s is randomly selected as a cover file. Then, the cover file is repeatedly decompressed and compressed 5 times by LAME [23] at the two common bitrates of 128 kbps and 320 kbps. In the decompressing process, the MDCT coefficients are extracted and the energy of each scalefactor band is calculated. Finally, the energy of 21 scalefactor bands within the first frame after each decompressing/recompressing process is shown in Figure 3. Ci denotes the energy curve after the i-th compression. It can be observed from Figure 3 that the energy peak in the energy curve can remain stable. Especially at 320 kbps bitrate, $C1 \sim C5$ are basically coincided. This is because the high bitrate can allocate more bits to store signal information and energy. To sum up, the first k bands with the largest energy are still the ones with the largest energy after multiple compressions. Hence, by selecting the first k bands with the largest energy as embedding bands to hide messages, we can still locate these scalefactor bands to extract messages even after multiple compressions.

The next difficulty is how to embed the message binary stream into the embedding bands. For image and video robust steganographic methods, the messages are embedded into the maximum perceptual components, which

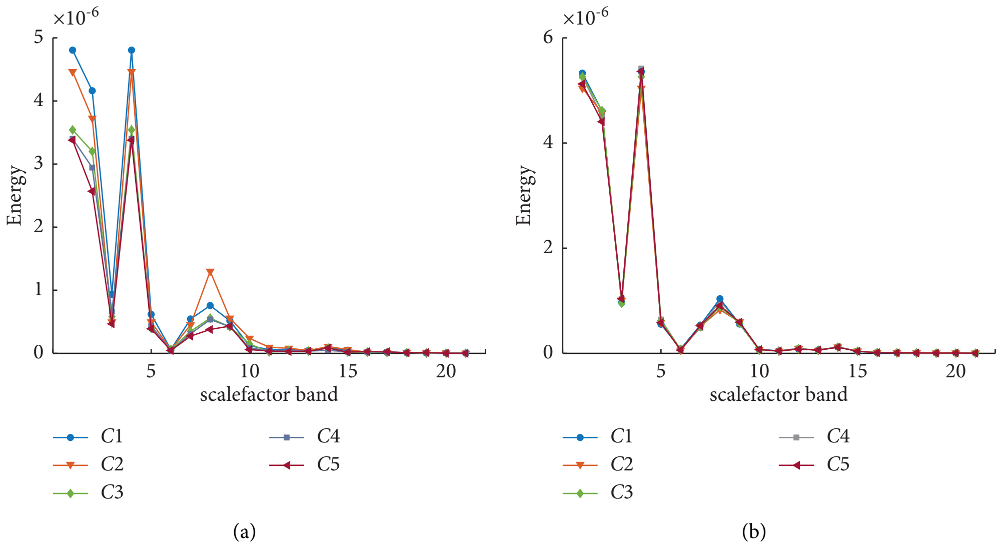


Figure 3: Energy of scalefactor band after 5 decompressing/recompressing attacks. (a) 128 kbps. (b) 320 kbps.

concentrate most of the signal energy. Likewise, DCT has the characteristic of energy concentration; that is to say, a small number of MDCT coefficients concentrate most of the audio signal energy. Thus, we further select the MDCT coefficients with the largest magnitude in each scalefactor band as the embedding elements. Generally speaking, there are two ways to embed messages into MDCT coefficients: sign bits flipping and the ±1 modification. However, flipping the sign bits of the maximum MDCT coefficient will cause a larger modification magnitude, which leads to obvious perceptual noise. Utilizing the ±1 modification of the MDCT coefficients to embed messages has insufficient robustness under the attack of multiple compressions.

Given these conditions, the positions of embedding elements are investigated in this paper. We rank the 21 scalefactor bands of each frame in order of energy and observe the position change of embedding elements in each band after 5 decompressing/recompressing attacks. As illustrated in Figure 4, the position changes occur mainly in the bands with less energy. When the bitrate is 320 kbps, there is almost no position change of the embedding elements in the high energy bands. At the bitrate of 128 kbps, there is a phenomenon of change aggregation. In the high energy bands, the number of the embedding elements with position changes is not large, and the resulting message bit error can be corrected by using BCH codes. Thus, the positions of the maximum MDCT coefficients in the embedding bands have sufficient robustness as the embedding domain. The embedding operation is implemented by adjusting the position of the maximum MDCT coefficients in the embedding bands to make their index parity consistent with the message.

3.2. The Procedure of Embedding and Extraction

3.2.1. Embedding Procedure. As the embedding process is implemented during the MP3 decoding, the input file is the MP3 format and the output file is WAV format. The framework of the proposed robust steganographic method is shown in Figure 5, which is mainly divided into five steps:

Step 1: BCH encoding of message bitstream. In order to control the errors caused by multiple compressions, the message bitstream is encoded with a BCH code. Next, the encrypted messages are scrambled to generate random errors instead of continuous errors, which is convenient for error correction.

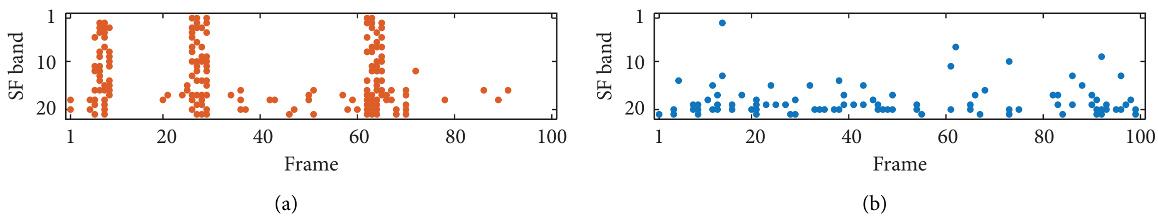
Step 2: MDCT coefficients are obtained via semi-decoding. According to the type of the current frame, it is judged whether the frame is a long block or a short block. If it is a long block, the MDCT coefficients are divided into 21 scalefactor bands. Otherwise, the coefficients are divided into 36 bands. The number of coefficients contained in each band is determined by the sampling rate, which can be found by referring to the MP3 standard.

Step 3: the energy of each scalefactor band is calculated as in (1). The first k bands with the largest energy among all scalefactor bands are used as embedding bands. Then, the MDCT coefficients with the largest magnitude in each embedding band are selected as the embedding elements.

$$en(sb) = \frac{1}{bw(sb)} \sum_{i=lbl(sb)}^{lbl(sb)+bw(sb)-1} \left|xr_i^2\right|, \qquad (1)$$

where $bw(sb)$ denotes the number of coefficients in the sb-th scalefactor band and $lbl(sb)$ indicates the first MDCT coefficients belonging to the scalefactor band. xr_i is the value of the MDCT coefficients.

Step 4: Judge whether the parity of the location index of the embedding element is consistent with the message bit. If so, skip the current embedding element and process the next. Otherwise, the coefficients in the sb-th scalefactor band are divided into two subsets, $\Pi_o^{(sb)}$ and



(a)

(b)

FIGURE 4: Visualization of the change on the position of the maximum MDCT coefficients. The dots represent the coefficients of the position change. (a) 128 kbps. (b) 320 kbps.

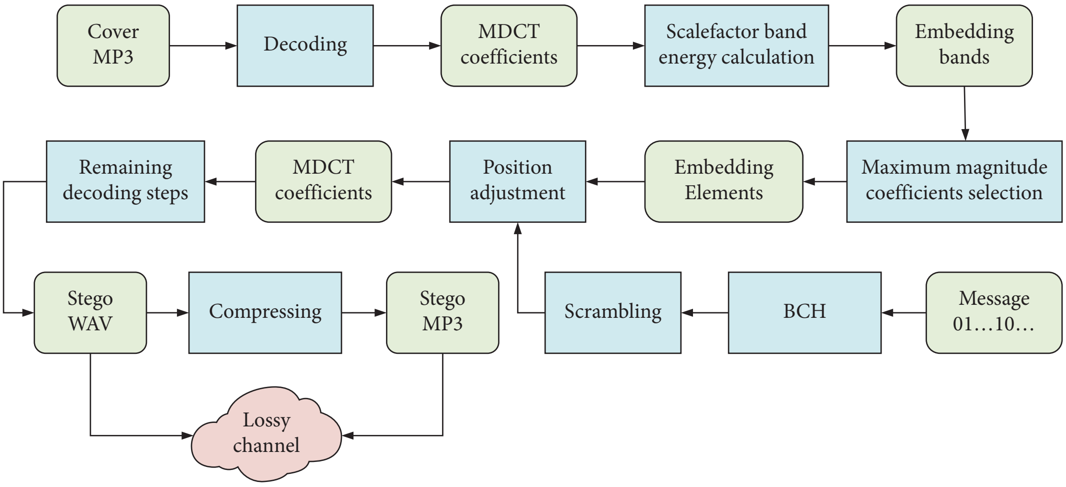


FIGURE 5: Framework of the proposed method.

$\Pi_e^{(sb)}$. $\Pi_o^{(sb)}$ and $\Pi_e^{(sb)}$ denote the set of coefficients with an odd and even position index, respectively. According to the parity of the message bit, the coefficient $xr_{\max}$ with the largest magnitude in $\Pi_o^{(sb)}$ or $\Pi_e^{(sb)}$ is selected to replace the embedding element xr_{new}. $xr_{\max}$ becomes the coefficient with the largest magnitude in the sb-th scalefactor band by adjusting the values of xr_{new} and $xr_{\max}$, which are formulated as follows:

$$xr'_{\max} = \begin{cases} xr_{\max} - \alpha\left(\left|xr_{\max}\right| - \left|xr_{\text{new}}\right|\right), & \text{if } xr_{\max} > 0, \\ xr_{\max} + \beta\left(\left|xr_{\max}\right| - \left|xr_{\text{new}}\right|\right), & \text{if } xr_{\max} < 0, \end{cases}$$
$$xr'_{\text{new}} = \begin{cases} xr_{\text{new}} + \alpha\left(\left|xr_{\max}\right| - \left|xr_{\text{new}}\right|\right), & \text{if } xr_{\max} > 0, \\ xr_{\text{new}} - \beta\left(\left|xr_{\max}\right| - \left|xr_{\text{new}}\right|\right), & \text{if } xr_{\max} < 0, \end{cases} \quad (2)$$

where α and β are the adjustment parameters in order to control the modified magnitude. When $\alpha, \beta = 1$, the values of xr_{new} and $xr_{\max}$ are exchanged actually. For stronger robustness, we can increase the values of α and β so as to enlarge the magnitude difference between xr'_{new} and $xr'_{\max}$.

Step 5: The modified MDCT coefficients are then processed by the remaining decoding steps to generate PCM signals in WAV format. Determine whether the stego WAV file is compressed into MP3 format according to the audio format supported by the lossy channel.

3.2.2. Extraction Procedure.

3.2.2. Extraction Procedure. The extraction procedure is also completed in the MP3 decoding process. The first 4 steps are the same as in the embedding procedure. After obtaining the embedding element $xr_{\max}$ in each embedding band, the message bits are extracted as in (4). Finally, antiscramble the extracted bitstream and decode it by BCH code to generate the secret message.

$$m_i = \text{find}\left(xr_{\max}\right) \bmod 2, \quad (3)$$

where find $(xr_{\max})$ is a function used to calculate the position index of $xr_{\max}$ in the embedding bands.

3.3. Application on SoundCloud. In order to apply our robust method in practice, SoundCloud is used as the lossy transcoding channel for investigation. SoundCloud is one of the largest music streaming services, reaching over 175 million monthly users worldwide. In addition, SoundCloud supports both WAV and MP3 formats, which is more representative.

However, the source code of SoundCloud is nonpublic, which means that it is a black box to us. The only way to study its channel characteristics is by comparing the uploaded and downloaded audios. First, the WAV files and MP3 files with various bitrates are uploaded to SoundCloud.

Then, the transcoded audios are downloaded for comparison. The channel characteristics of SoundCloud are shown in Table 2. It is obvious that the audios are transcoded into MP3 format with 128 kbps no matter what formats are uploaded. Moreover, whether the audio is transcoded or not is determined according to changes of MDCT coefficients. It can be found that the uploaded MP3 file is transcoded even under 128 kbps, and the rate of change reaches 20.39%. Compared with 192, 256, and 320 kbps, the uploaded MP3 with 128 kbps has the lowest rate of change. Therefore, when the upload format is MP3, its bitrate is best consistent with the download bitrate (i.e., 128 kbps) specified by SoundCloud to reduce the change of the MDCT coefficients.

Considering that some channels can only support MP3 or WAV format, Figure 6 illustrates the simulation of the covert communication based on SoundCloud as an example. If the upload format is WAV, it only goes through 1 decompressing/recompressing attack. Meanwhile if the upload format is MP3, it has to go through 2 decompressing/recompressing attacks. This is due to the transcoding mechanism of SoundCloud; the uploaded MP3 file is decoded to WAV format first and then encoded to MP3 format under 128 kbps. Although various encoders and decoders are used in the different lossy channels, the principles of the encoder and decoder are almost the same, just the difference of some parameters. So this phenomenon is common in other lossy transcoding channels. As mentioned above, the proposed embedding domain has been proved to be robust enough against multiple decompressing/recompressing attacks. Thus, our method supports both WAV and MP3 formats audio upload to the lossy transcoding channel.

4. Experiments

To evaluate the performance of the proposed method, several experiments including robustness, embedding capacity, undetectability, and imperceptibility are performed in this section. A nonrobust adaptive steganographic method (AHCM [12]) and an audio watermarking method based on hybrid domain (DWT-SVD-QIM [19]) are compared with the proposed method. In [12], a generalized adaptive Huffman code mapping framework is proposed to obtain a higher secure payload. In [19], an audio watermarking method based on SVD in the DWT domain using synchronization code is presented. QIM is used to embed the watermarking bits into the SVs of the wavelet blocks. A dataset that consists of 2000 stereo MP3 clips with a sampling rate of 44.1 kHz and a duration of 10 s is constructed. In addition, various music styles are included in this dataset. The local encoder and decoder are based on LAME. For stronger robustness, parameters α and β are set as 1.1.

4.1. Robustness. In this experiment, we first evaluate the robustness of our method under the local simulation of one and two decompressing/recompressing attacks without BCH code. The embedding rate (R_e) is set as 450 bits/s. As illustrated in Figure 7(a), even without the help of error-correcting codes, the BER of our method is lower than 2.5% within various bitrates after 1 decompressing/recompressing attack. Moreover, the BER decreases with the increase of the bitrate of MP3 files, which is consistent with experimental results in Section 3.1. Figure 7(b) shows that the BER increases slightly under 2 decompressing/recompressing attacks, but it is still less than 3%. As the principle of MP3 codecs is basically the same, we can confirm that the proposed method is effective to the lossy channels supporting different upload formats (WAV and MP3) and bitrates. Owing to the low BER in the local simulation, BCH code (7, 4, 1) is used for error-correcting with the consideration of embedding capacity.

To further evaluate the effectiveness of the proposed method, the stego audios with different formats are uploaded to SoundCloud. Next, the transcoded MP3 files are downloaded to calculate the BER. As shown in Figure 8, the BER of our method is the lowest under the same embedding rate (200 bits/s), which is close to 0. Although the BER of DWT-SVD-QIM is less than 5%, as a watermarking method, the acoustic fidelity is destroyed seriously when the embedding rate is 200 bits/s, which can be perceived by the human ear clearly. AHCM is a nonrobust method, so its BER is up to 20% in some cases. In general, the error-correcting code is invalid when the BER is more than 10%. It means that AHCM cannot resist social networking transcoding. The experimental results show that the proposed method can provide strong robustness against social networking transcoding.

4.2. Embedding Capacity. As mentioned above, the first k scalefactor bands with the largest energy are selected as embedding bands. Each embedding band can hide 1 bit of message. For stereo MP3 files, each frame contains two granules and each granule includes two channels. The embedding capacity is calculated as follows:

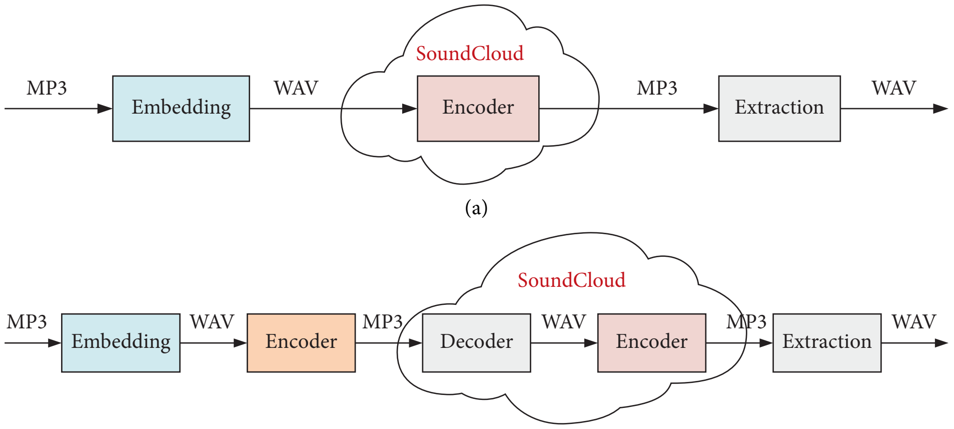
$$R_c = \frac{4 * k}{D_{\text{frame}}} = \frac{4 * k}{(1152/R_s)}, \tag{4}$$

where D_{frame} denotes the duration of each frame. As the samples per frame are fixed at 1152, D_{frame} is determined by the sampling rate (R_s). As illustrated in Figure 3, the first 7 scalefactor bands with the largest energy are still the energy peak after multiple decompressing/recompressing attacks. Taking the balance between the robustness and embedding capacity into account, k is set as 3 and 5 in this paper.

The experimental results are shown in Figure 9. The embedding capacity of our method is calculated after BCH encoding. Due to the fact that AHCM is a nonrobust method and the Huffman codes are used as embedding domain, it has the highest embedding capacity. The embedding capacity of DWT-SVD-QIM is 45.9 bit/s [19], which is superior to other audio watermarking algorithms. The embedding capacity of our proposed method is 5 to 9 times that of DWT-SVD-QIM and the maximum is 439 bits/s, which is able to satisfy the needs of covert communication based on lossy channels.

Table 2: The investigation of channel characteristics on SoundCloud.

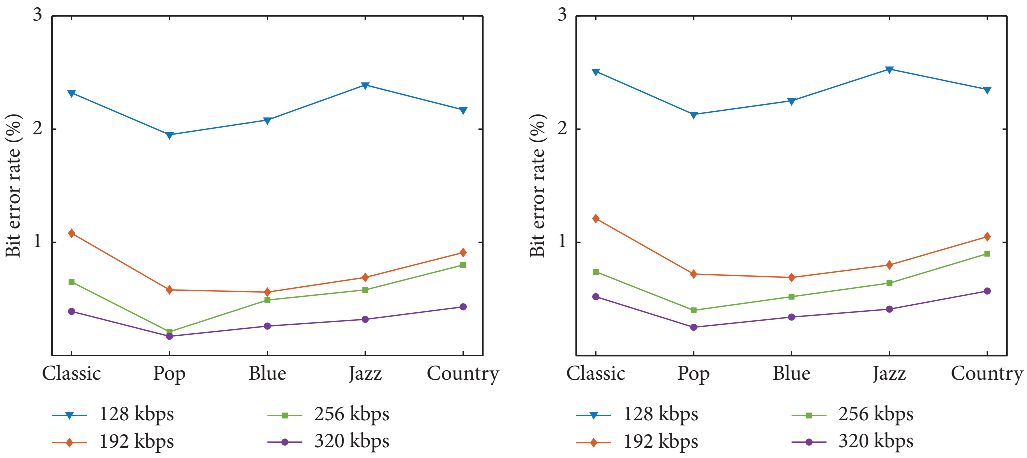
Upload	WAV	MP3 (kbps)			
		128	192	256	320
Download	128	128	128	128	128
Transcode	Y	Y	Y	Y	Y
Rate of change	—	20.39%	26.71%	27.79%	28.65%



(a)

(b)

Figure 6: Simulation of the covert communication based on SoundCloud. (a) Upload WAV. (b) Upload MP3.



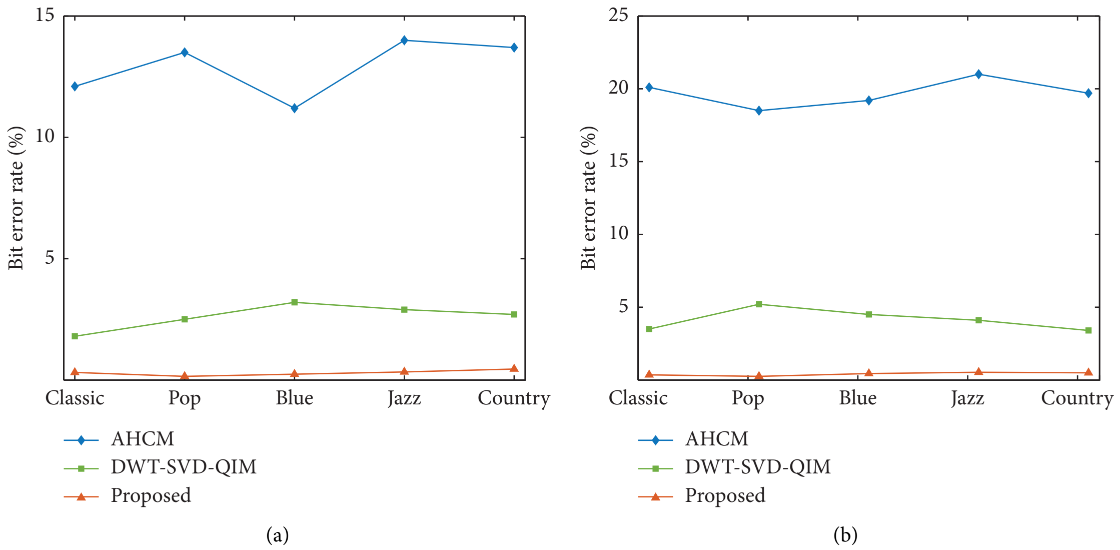
(a)

(b)

Figure 7: The BER of the proposed method under the local simulation of one and two decompressing/recompressing attacks without BCH code. (a) 1 decompressing/recompressing. (b) 2 decompressing/recompressing.

4.3. Imperceptibility. In this experiment, the perceptual evaluation of audio quality (PEAQ [24]) of the ITU standard is adopted for the objective measurement of imperceptibility. The objective difference grade (ODG) is the output of the PEAQ algorithm, and it belongs to $[-4, 0]$. The higher the ODG value, the closer the acoustic quality of the cover audio to its corresponding stego audio. The stego audios are transcoded by SoundCloud. When the cover audio is of enough acoustic similarity, the value of ODG may be greater than 0.

The ODG values of AHCM, DWT-SVD-QIM, and the proposed method are calculated at the embedding rate of 200 bit/s. As shown in Figure 10, the ODG values of AHCM are all higher than those of the other two methods. In order to ensure robustness, DWT-SVD-QIM and the proposed method embed the messages into the high-energy domains, which has a greater impact on the acoustic quality. However, the ODG values of the proposed method are higher than 0 within any musical style, which illustrates that our method is of good performance on the imperceptibility.



(a)

(b)

Figure 8: The BER of WAV and MP3 formats after transcoding by SoundCloud. (a) WAV. (b) MP3, 128 kbps.

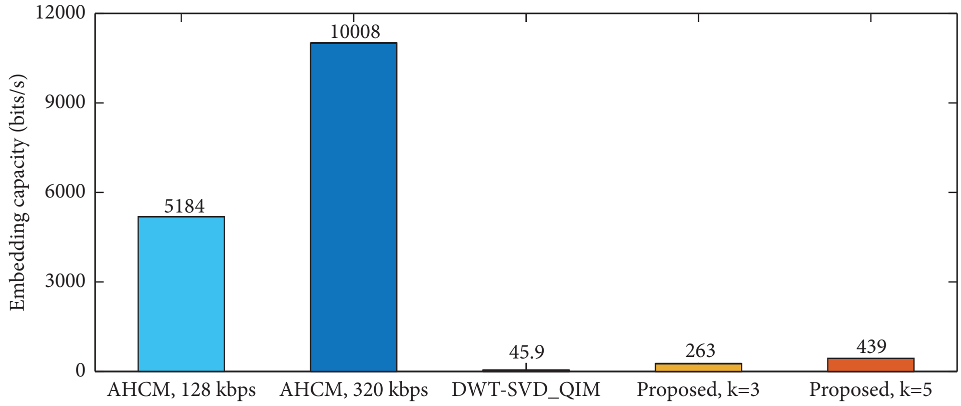


Figure 9: Comparison of the embedding capacity.

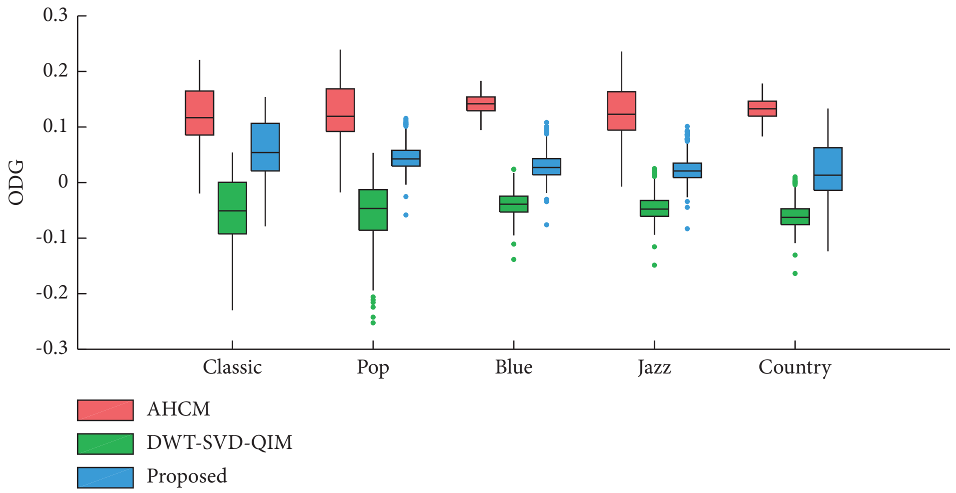


Figure 10: Comparison of ODG values.

Table 3: Detection accuracy (%) for all steganographic methods.

	AHCM			DWT-SVD-QIM			Proposed		
	128	320	SC	128	320	SC	128	320	SC
200 bits/s	52.39	55.67	54.44	85.62	88.12	86.25	51.35	53.36	52.32
300 bits/s	53.85	57.15	56.73	87.50	89.75	87.88	52.45	53.98	53.03
400 bits/s	54.75	58.77	57.04	90.31	94.23	91.16	53.64	55.25	54.85

4.4. Undetectability. To evaluate the security of the proposed method, a state-of-the-art steganographic feature set (MDI2 [25]) is adopted. In MDI2, Markov transition probability and accumulative neighboring joint density are extracted from the multiorder differential coefficients of intra- and interframe for steganalysis. Meanwhile other steganalysis features are based on the statistical characteristics of QMDCT coefficients, such as D2MA [26], ADOTP [27], and JPBC [28], which are not suitable for detecting the embedding methods based on MDCT coefficients. Three embedding methods, AHCM, DWT-SVD-QIM, and the proposed method, are implemented to generate the stego audio files under the different embedding rates. Then stego files are recompressed by the local codec with bitrates of 128 kbps and 320 kbps. In addition, the stego files are transcoded by SoundCloud (SC) to reflect the security in practice. 60% of the cover MP3 files and their corresponding stego files are used for training and the rest 40% for testing.

As shown in Table 3, the detection accuracy of the proposed method is the lowest in all cases. When the embedding rate reaches 400 bits/s, the detection accuracy is still lower than 55% even when the stego audios are transcoded by SoundCloud. This indicates that our method can achieve satisfactory performance on undetectability.

5. Conclusion and Future Work

A robust MP3 steganographic method against social networking transcoding is proposed in this paper. The algorithm is implemented in the process of MP3 decoding, which is convenient and fast. To enhance the robustness, a new embedding domain is proposed. The scalefactor bands with large energy in each frame are selected as embedding domain. The message bits are embedded by adjusting the position of the MDCT coefficient with the largest magnitude in the embedding bands. Experiments show that this embedding domain can provide strong robustness even after multiple decompressing/recompressing attacks. Moreover, the proposed method achieves a satisfactory level of embedding capacity, imperceptibility, and undetectability.

As illustrated in Figure 4, the error aggregation phenomenon occurs at the bitrate of 128 kbps. The frame selection is adopted to avoid error aggregation for better robustness in the feature extraction. The main difficulty is the construction of the side channel, which is used to transmit the information of selected frames. This means we have to find a new embedding domain with stronger robustness. Therefore, it is an important work for us to construct a side channel in the future work.

Data Availability

Data are available at https://github.com/yangyunzhao2sy/database_scn.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this work.

Acknowledgments

This work was supported by the NSFC under Grants nos. 61 902 391, U1736214, 61 972 390, and 61 872 356 and National Key Technology Research and Development Program under Grant no. 2019QY0701.

References

[1] D. Yan, R. Wang, X. Yu, and J. Zhu, "Steganography for MP3 audio by exploiting the rule of window switching," *Computers & Security*, vol. 31, no. 5, pp. 704–716, 2012.

[2] F. Petitcolas, "MP3stego," 1998, http://www.petitcolas.net/steganography/mp3stego/.

[3] D. Yan, R. Wang, and L. Zhang, "Quantization step parity-based steganography for mp3 audio," *Fundamenta Informaticae*, vol. 97, no. 1-2, pp. 1–14, 2009.

[4] D. Yan and R. Wang, "Huffman table swapping-based steganograpy for mp3 audio," *Multimedia Tools and Applications*, vol. 52, no. 2-3, pp. 291–305, 2011.

[5] Y. Yang, Y. Wang, X. Yi, X. Zhao, and Y. Ma, "Defining joint embedding distortion for adaptive MP3 steganography," in *Proceedings of the 7th ACM Workshop on Information Hiding and Multimedia Security*, pp. 14–24, ACM, Paris, France, July 2019.

[6] Y. Dong, *Research on information hiding based on Mp3*, Ph.D. Thesis, Beijing University of Posts and Telecommunications (BUPT), Beijing, China, 2015.

[7] Y. Yang, H. Yu, X. Zhao, and X. Yi, "An adaptive double-layered embedding scheme for MP3 steganography," *IEEE Signal Processing Letters*, vol. 27, pp. 1984–1988, 2020.

[8] H. Gao, "The MP3 steganography algorithm based on huffman coding," *Acta Scientiarum Naturalium Universitatis Sunyatseni*, vol. 46, no. 4, pp. 32–35, 2007.

[9] X. Liu and L. Guo, "High capacity audio steganography in MP3 bitstreams," *Computer Simulation*, vol. 24, no. 5, pp. 110–113, 2007.

[10] D. Yan, R. Wang, and L. Zhang, "A large capacity MP3 steganography algorithm based on huffman coding," *Journal of Sichuan University (Medical Science Edition)*, vol. 48, no. 6, pp. 1281–1286, 2011.

[11] K. Yang, X. Yi, X. Zhao, and L. Zhou, "Adaptive MP3 steganography using equal length entropy codes substitution," in *Proceedings of the 16th International Workshop on Digital*

Forensics and Watermarking, pp. 202–216, Springer, Magdeburg, Germany, August 2017.

[12] X. Yi, K. Yang, X. Zhao, Y. Wang, and H. Yu, "Ahcm: Adaptive huffman code mapping for audio steganography based on psychoacoustic model," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 2217–2231, 2019.

[13] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[14] W. N. Lie and L. C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Transactions on Multimedia*, vol. 8, no. 1, pp. 46–59, 2006.

[15] E. Yousof and S. Shadi, "Robust audio watermarking using improved ts echo hiding," *Digital Signal Processing*, vol. 19, pp. 809–814, 2009.

[16] E.-S. Fea, "An efficient singular value decomposition algorithm for digital audio watermarking," *International Journal of Speech Technology*, vol. 12, pp. 27–45, 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] N. Iynkaran, Y. Xiang, Y. Rong, and D. Peng, "Robust patchwork-based watermarking method for stereo audio signals," *Multimedia Tools and Applications*, vol. 72, pp. 1–24, 2014.

[19] V. Bhat K, I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Digital Signal Processing*, vol. 20, no. 6, pp. 1547–1558, 2010.

[20] H. T. Hu, H. H. Chou, C. Yu, and L. Y. Hsu, "Incorporation of perceptually adaptive qim with singular value decomposition for blind audio watermarking," *Journal on Advances in Signal Processing*, vol. 2014, pp. 1–12, 2014.

[21] B. 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.

[22] J. O. Smith and J. S. Abel, "ISO/IEC 11172-3: information technology-coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s-part 3," 1993.

[23] R. Hegemann, A. Leidinger, and R. Brito, "LAME (Lame Aint an MP3 Encoder)," 1998, https://sourceforge.net/projects/lame/files/lame/.

[24] T. Thiede, W. Treurniet, R. Bitto et al., "Peaq-the itu standard for objective measurementof perceived audio quality," *Audio Engineering Society*, vol. 48, pp. 3–29, 2000.

[25] Y. Ren, Q. Xiong, and L. Wang, "A steganalysis scheme for aac audio based on mdct difference between intra and inter frame," in *Proceedings of the 16th International Workshop on Digital Forensics and Watermarking*, pp. 217–231, Springer, Magdeburg, Germany, August 2017.

[26] M. Qiao, A. H. Sung, and Q. Liu, "MP3 audio steganalysis," *Information Sciences*, vol. 231, pp. 123–134, 2013.

[27] C. Jin, R. Wang, and D. Yan, "Steganalysis of MP3 stego with low embedding-rate using Markov feature," *Multimedia Tools and Applications*, vol. 76, no. 5, pp. 6143–6158, 2017.

[28] Y. Wang, X. Yi, and X. Zhao, "MP3 steganalysis based on joint point-wise and block-wise correlations," *Information Sciences*, vol. 512, pp. 1118–1133, 2020.