A new kind of MP3 audio steganography algorithm

PAN Fen[1], ZHANG Yao[1,a], SHEN Jun-wei[1], WEN Ren-yi[2]

[1]Engineering University of Armed Police Force Department of Electronic Technology, Xian 710086

[2]Officer College of Armed Police Force Department of Electronic Technology, Sichuan 610000

[a]540921912@qq.com

Keywords: MP3 audio steganography; pre-type algorithm; inlaid-type algorithm; big _value area; count_1 value; Matrix coding;

Abstract. In view of the current international security situation, this paper proposed a new steganography algorithm based on MP3 audio format aimed at solving solve the problem of network security communication. This algorithm combined pre-type algorithm and inlaid-type algorithm of MP3 audio steganography .Pre-type algorithm principle was used to extract sample points in count_1 value area then took matrix code method to embed secret information; The edge information in the process of big_value area programming was modified according to the inlaid-type algorithm principle, in order to embed secret information. Experimental results show that this algorithm solves the problems including the small capacity, poor robustness and low accuracy extraction of MP3 audio steganography algorithm.

Introduction

The convenience of the Internet has brought people with quick information transmission, but also comes about many problems of information security. Since the end of last century, information steganography technology has become a new information security technology and has rasied attention from the international academic circles. It's widely used by the military, intelligence, national security and e-commerce security.

The steganography technique of MP3 audio format is a new field of information hiding technology. Its advantage lies in the commonness and popularity in the Internet. Using this kind of carrier to hide the secret information, has very strong confusion. In addition, the MP3 audio has a larger capacity than the text and digital image file to hide secret information.

Mp3 steganography algorithm can be divided into, pre-type algorithm, inlaid-type algorithm and post-type algorithm (as shown in figure 1) according to the secret information embedding location.

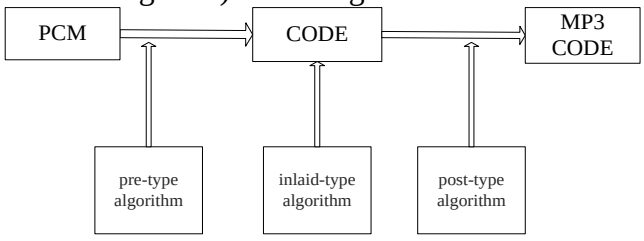

Fig.1 The position of MP3 algorithm

Pre-type algorithms could embed secret information before audio signal encoding, its greastest advantage is the abaility to draw lessons from the existing mature uncompressed audio steganography algorithm. Some of the typical steganography algorithm types are Shanghai university of WANG Shuo-zhong[1] realizes embedding secret information according to the size of the adjacent frames MDCT coefficients energy relations; HOU Li-min[2] et al realizes embedding secret information by modifying the symbol of MDCT coefficients. The key point of the inlaid-type algorithms is to embed secret information in the link of encoding by adjusting some parameters. Such as classic MP3stego[3] algorithm realizes embedding information through adjustment of parameters of inner loop; on the basis of this algorithm , ZHONG Shang-ping[4] improves this classic approach by introducing Logistic chaotic models to select steganography units randomly, thus making improvement of MP3Stego algorithm necessarily. Post-type algorithms realize

embedding information by operating code words after encoding. There are some Post-type algorithms. GAO hai-ying [5]puts forward steganography algorithm that maps Huffman code words of big_value area; LIU Xiu-juan[6] puts forward steganography algorithm that maps Huffman code words of count_1 value area.

The steganography algorithm discussed in this paper combines the pre-type algorithm and inlaid-type algorithm ,which used matrix coding to implement embedding of secret information before count_1 value area encoding in order to improve the efficiency of embedding capacity、efficiency and imperceptibility. While modifying Modify the frame edge information to implement embedding secret information (usually store the secret key, etc.) during the encoding process of big_value area, in order to further improve safety.

Related work

MP3 compression technology and the Huffman coding

MP3(MPEG Audio Layer3) [7] is a way of lossy compression, which was proposed by the ISO (International Standard Organization) in the early ninety's , aimed at video compression、audio compression and synchronization data flow compression . It has been widely applied in the digital audio and radio system. Its data compression ratio could reach 1:12 in fully using of psychoacoustics[8] model and human auditory system (HAS),Figure 2 shows the MP3 encoding procession. First , the voice signal was grouped , each 1152 sampling points as a group, each group has two size granules of 576 sampling points; After filter analysis, output 32 aequilate subbands; Then make MDCT transformation to subdivide each subband 18 second-frequency bands; Next is quantization coding based on the Huffman coding tables and SMR that calculates by psychoacoustic model ; The last step is format encapsulation.

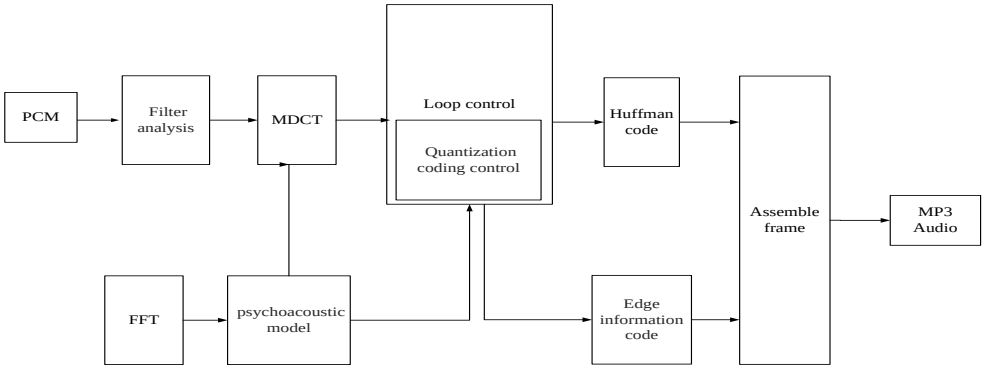

Fig.2 MP3 encoding procession

Huffman coding structures the shortest average length code words based on the probability of occurrence of the coefficients, reaching the purpose of compression[9]. After MDCT transformation for audio, Huffman coder divides each particle's 576 quantitative coefficients into three areas according to the frequency. They are big value area、count_1 area and zeros , divided area shown in figure 3.

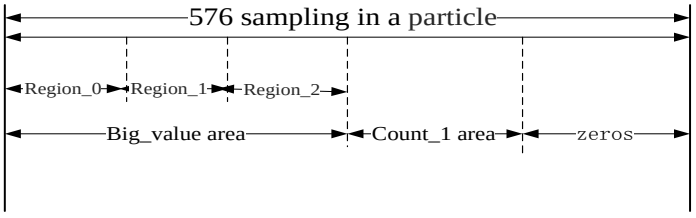

Fig.3 Divided area

Zero corresponds to the high frequency part and value are zeros, no coding. count_1 area corresponds to the intermediate frequency part and values are -1, 0, 1. Four of them as a team use two specified Huffman tables (Ha, Hb) to encode them. Big_ value area corresponds to the low frequency part and has the values from – 8207 to 8207. Big_ value area can be further divided into three sub regions, region 0, region 1 and region 2, each region has the different table (0 ~ 31

number tables list, excluding number14 and 4 tables because they are reserved not participating in coding) to encode them. Big_value area's Huffman tables are shown in table 1.

Table 1 Huffman tables of Big value area

Table index	Maximum code	Table index	Maximum code	Table index	Maximum code	linbits	Table index	Maximum code	linbits
0	0	8	5	16	16	1	24	31	4
1	1	9	5	17	19	2	25	41	5
2	2	10	7	18	23	3	26	79	6
3	2	11	7	19	31	4	27	143	7
4	reserved	12	7	20	79	6	28	271	8
5	3	13	15	21	271	8	29	527	9
6	3	14	reserved	22	1039	10	30	2016	11
7	5	15	15	23	8207	13	31	8207	13

The algorithm design of count_1 area

According to the characteristics of count_1 area, method of matrix coding was applied to embed information,which was put forward by Andreas Westfeld [10]. The advantages including high embedding efficiency, fast speed and low bits' change. Only 1 bit was amended when embedding k bits of information The general form of the matrix is that k bits information was embedded into the n bits carriers($n = 2^k - 1$), only 1 bit of data carrier need to be modified, with a triple group(1, n, k) [11] to describe it.

Corresponding values in Count_1 area are -1, 0,1, and each 4 bits ware coded together, no need of binary conversion and easy to group. Although after Huffman encoding, the code are still composed of -1, 0, 1, embedding of information must be done before Huffman encoding.

Although count_1 area after Huffman encoding, the codes words still consist of -1, 0,1, but we must embed information in front of the Huffman encoding because the code after embedding (amend some bits of the code word according to matrix algorithm rule) may no longer belong to the Ha or Hb tables, thus will cause the decoding error. For example, 0010 code word after the Ha table encoding would be 0100. If first bit need to be modified, the 0100 code will become 1100, but 1100 in Ha has no value to decode, so cause decoding failure.

Considering embedding steganography capacity、efficiency and robustness, this paper chose a triple group(1,3,7) as the parameter to embed information.

The algorithm design of big_value area

We further analyze the encoding rules of big_value area. According to the relevant provisions of the MP3 standard document , index no.0-15 (no. 4 and 14 unused) are different coding tables; Index no. 16-23 tables come from the same code book, no.17-23 tables have the same coding rules as the no.16 table, but different linbits values; no. 24-31 tables come from the same code book, no.25-31 tables have the same encoding rules as the no.24 table, just different linbits values. Maximum code value formula is shown as follows:

$$\text{Maximum code value} = 15 + 2^{\text{linbits}} \quad (1)$$

Big_value is further subdivided into three regions, region_0、region_1 and region_2, energy arrangement are roughly from high to low, The Huffman tables for the three regions are not necessarily the same. Region_0 has larger energy，so generally using no.16-23 or 24 to 31 code tables (only use the code tables from the same code book); the energy of region_1 and region_2 is relatively small, generally using no.0-15 code tables. Which tables are used depending on the maximum value of the region and coding principle of the minimum number of bits after encoding in the region. After encoding , information of the selected code tables was written into edges..

Taking into consideration the characteristics of big_value area, this article adopted the method of table mapping to hide information. Firstly, code tables was classified, classification standard is shown as following :(1) The maximum word that code tables can encode must be the same, in order to prevent the change the quantization step or appear coding error;(2)The code tables must come from the same code book, prevent the decoding error.

According to the above rules, the code tables are divided into two categories, one of which was used for imbedding information 0, the other for imbedding information 1. Grouped is shown in figure 4, embedding rules is shown in table 2.

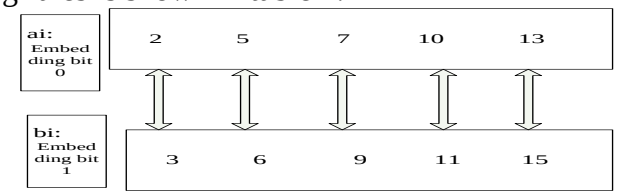

Fig.4 Grouping rule

Although there aren't many available tables, but the using frequency are highest. What's more, the method is simple and quick (only need to modify the edge information), so it will be no change the characteristic of statistics.

Table2 Embedding rules

Table index	Table index after embedding		Table index	Table index after embedding	
	Embedding 0	Embedding 1		Embedding 0	Embedding 1
2	No change	3	9	7	No change
3	2	No change	10	No change	11
5	No change	6	11	10	No change
6	5	No change	13	No change	15
7	No change	9	15	13	No change

Embedding and extracting process of the algorithm

Embedded process

In order to improve the algorithm's security, the three-layer DES[12] of cryptography was adopted to encrype information that will be embeded ,which were further then embed into count_1 areas by using matrix algorithm; the key of three-layer DES was imbedded in big_value areas by modifying the edge information. Of course, if ciphertext is too long, it could be imbedded in the big_value areas also. Specific algorithm is shown as follows:

1) Convert the message into binary steam and select the key to encrypt binary steam into ciphertext by using three-layer DES encryption;

2) Search count_1 area in a particle, eight bits as a group to be carrier;

3) Take 3 bits of ciphertext and modify one bit of the first seven carrier' bits in accordance with the rules of matrix algorithm to realize ciphertext embedding, the last bit of carrier not change.

4) Repeat steps 2 and 3 until all bits in the particle's count_1 area being finished up;

5) Search region_1 and region_2 in big_value area. No information imbedded in region_0. region_1 and region_2 select encoding tables according to their rules, if the tables within the scope of the index in the table 2, take out the key 1 or 2 bits (one matches condition taking1 bit, both matches taking 2 bits) and modify the corresponding frame edge information according to table 2 to realize the key embedding. If both encoding table' indexes don't belong to table 2, don't do anything.

6) Make Huffman coding for the particle, then search the next particle, until all the ciphertext and the key are embedded.

Extracting process

The steps of the extraction are as follows:

1) Extract the particle's frame edge information, get the encoding tables of region_1 and region_2;

2)Check if the encoding tables whether belong to table 2. if belong to it, extract part of the key according to the rule that shows in figure 4;

3) Decode the particle, and extract the count_1 area's sampling points, 8 bits of them as a group;

4) Extract 3 bits ciphertext in the first 7 sampling points of 8 according to the matrix decoding rules;

5) Repeat step 4, until complete all the ciphertext in the count_1 area;

6)Take the same operating for the next particle until all of the cipher key and ciphertext extraction completed;

7) Decrypt the ciphertext with the key, restore the message.

Experimental

This article used pop, classical, violin music clips as audio carriers and the sampling frequency was 44.1 Khz, quantization was 16 bits. changed the embedding bit rate per second. With gaussian white noise interference simulated the channel noise. The following , experiment focused on the four perspectives ， namely，embedding capacity、 imperceptibility、extraction accuracy and undetectability.

Embedding capacity

For MP3 format audio files, the duration time of each individual frame data is 20 ms, so there are 50 frames per second , namely 100 particles, about 200 bits of data per second can be hidden in big_value area. The Bits in count_1 areas is about 100-200 of a particle, 8 bits as a group, each group embedding 3 bits, so each particle can be embedded with 38-75 bits. The embedding capacity increased by about 1.5 times than the that of literature [6], nearly 10 times than literature [5].

Subjective imperceptibility

The most common audio subjective quality evaluation index is the MOS (Mean Opinion Score) Score [13] proposed by the international standard ITU p. 830. It uses 5 level criteria, as shown in table 6. Firstly, Testers listen to the audio samples. Then score them according to the rating in table 3.The average score of all testers is the MOS. This paper randomly selected 50 audiences, scored for the three types of loads under different embedding quantity music clips, the scoring results are shown in table 4.

Table3 Level criteria

MOS	Distortion description
5	No Distortion
4	A little distortion，but not be tiresome
3	Can feel distortion, tiresome a little
2	Obvious distortion, feel tiresome
1	Can’t stand it

Table4 Scoring results

Embedding rate Bit/s	POP MOS	CLASSICAL MOS	VIOLIN MOS
200	5.0	5.0	4.9
300	5.0	4.9	4.8
400	4.8	4.7	4.7
500	4.7	4.5	4.4
600	4.5	4.4	4.3

Experimental results showed that, even when the embedding rate is higher, this algorithm can still keep good subjective imperceptibility.

The objective imperceptibility and extraction accuracy

The objective imperceptibility and extraction accuracy’s formula as shown in (2), (3).

$$SNR(\mathrm{db}) = 10\lg\frac{\sum x^2(n)}{\sum (x(n)-y(n))^2} \qquad (2)$$

x(n)is original audio sampling ,y(n) is audio sampling after embedding message

$$Q = \frac{\text{The exactness bits of extracting}}{The\ \text{totality bits of secret message}} \times 100\% \qquad (3)$$

This article selected three types of audio and each of them had 50 clips. Recorded SNRs and extraction accuracy before and after embedding secret message, the average value are shown in as following table 5.

The experimental results showed that the algorithm may achieve higher imperceptibility and extraction accuracy than the literature [5], [6] .

Table 5 SNRs of the Algorithm in the article、Literature[5] and Literature[6]

Embedding rate Bit/s	Algorithm in the article			Literature[5]			Literature[6]		
	POP SNR	CLASSICAL (db)	VIOLIN (db)	POP (db)	CLASSICAL (db)	VIOLIN (db)	POP (db)	CLASSICAL (db)	VIOLIN (db)
200	92.9	89.5	86.1	84.4	78.2	70.3	90.0	84.3	79.6
300	90.2	87.3	83.7	79.5	73.6	67.7	87.5	79.2	75.5
400	86.9	82.1	81.2	73.5	69.3	65.2	81.8	72.6	72.3
500	81.8	78.1	75.3	68.4	64.2	60.5	74.2	68.9	66.9
600	77.6	74.2	72.3	63.2	60.3	56.6	70.6	64.2	61.7

Table6 Q values of the Algorithm in the article、Literature[5] and Literature[6]

Embedding rate Bit/s	Algorithm in the article			Literature[5]			Literature[6]		
	POP Q(%)	CLASSICAL (%)	VIOLIN (%)	POP (%)	CLASSICAL (%)	VIOLIN (%)	POP (%)	CLASSICAL (%)	VIOLIN (%)
200	98.3	98.6	99.8	96.1	96.9	98.3	92.6	93.2	93.9
300	97.3	98.6	98.9	95.3	95.3	97.2	90.3	90.1	91.1
400	96.8	96.4	97.8	94.2	94.5	94.3	88.2	88.3	88.2
500	95.9	95.6	96.5	92.1	92.3	92.9	84.2	84.9	85.1
600	92.7	93.5	94.0	90.3	91.5	91.5	80.1	80.9	81.9

Undetectability

This algorithm steganography used testing strategy mentioned in literature [14] to test the undetectability. In which R>1 indicates is MP3Stego steganography audio, and R<1 indicates common audio, the experimental data are shown in table 7.

Table7 R value of the Algorithm in the article

Embedding rate Bit/s	POP R	CLASSICAL R	VIOLIN R
200	0.8925	0.8952	0.8986
300	0.8995	0.8998	0.9008
400	0.9012	0.9024	0.9042
500	0.9112	0.9245	0.9337
600	0.9458	0.9587	0.9621

Experimental results showed that the method in literature [14] failed to detect our algorithm. The reason was that during the process of embedding, this method only changes the frame edge information when imbedding information in big_value area and and thus the imbedding process was done before coding in count_1 area. Therefore this algorithm has smaller influence on the characteristics of the code table distribution.

Summary

In view of the current international security situation and the wide application of digital steganography in engineering、business, this article designed a large-capacity steganography algorithm which combined both compression of the pre-type algorithm and inlaid-type algorithm of MP3 audio steganography, making full use of big_value area and count_1 area. Experimental results showed that this algorithm has larger capacity 、stronger robustness、better imperceptibility

and higher embedding efficiency. Basically meet the need of the secure communications future work would be to further study the characteristics of MP3 encoding to design a new better algorithm.

References

[1] WANG C T, CHEN T S, CHAO W H. A new audio watermarking based on modified discrete cosine transform of MPEG/Audio Layer III[C]. Proceedings of the 2004 IEEE International Conference on Network, Sensing & Control, 2004: 984-989.

[2]ZHU Kui-long, HOU lin-yuan.MP3 Compression Based Watermarking Capable of Resisting Decoding/Re-coding Attack[J]. JOURNAL OF SHANGHAI UNIVERSITY(NATURAL SCIENCE EDITION) ,2008, 14(4): 331-335.

[3] PETITCOLAS A P. MP3Stego 1.1.16[ED/OL]. Cambridge: Cambridge Computer Laboratory, 1997[2002-03-19]. http://www.cl.cam.ac.uk/-fapp2/steganography/mp3stego.

[4]ZHONG Shang-ping, CHEN Tie-run.Approaches of Improving Performance of MP3Stego and their Implementations[J]. Computer Engineering and Application.2006.42(35):95-100

[5]GAO Hai-ying. steganography algorithm based on Huffman code[J]. ACTA SCIENTIARUM NATURALIUM UNIVERSITATIS2007,46(4):32-35

[6]LIU Xiu-juan,GUO Li. High Capacity Audio Steganography in MP3 Bitstreams [J], Computer Simulation.2007,24(5):110-113

[7]KEN C Pohlmann. Principles of Digital Audio[M].Sixth Edition.BEIJING:POSTS& TELECOM PRESS,2013:421

[8] ZHANG Li-guang WANG Rang-ding.Psychoacoustic Model and its Application to MP3 Encoding[J].JOURNAL OF NINGBO UNIVERSITY(NATURAL SCIENCE & ENGINEERING EDITION) .2010,23(3):27—30.

[9] PHIlip C Ritchey, VERNON J Rego .Hiding Secret Messages In Huffman Trees[C]. 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE Conference Publications,2012

[10] J Fridrich.Feature—based steganalysis for JPEG images and its implications for future design of steganographic schemes[C].In:6th Information Hiding Workshop,Toronto,ON,Canada,2004

[11] LIU F L,LIU J F,LUO X Y. Digital Image Steganalysis[M] .BeiJing:China Machine Press,2010:32-38

[12]YANG Xiao-yuan,WEI Li-xian.Computer cryptology[M]. XI'AN JIAOTONG UNIVERSITY. 2007:49-50

[13] ITU. Subjective performance assessment of telephone-band and wideband digital codecs[S]. ITU-T Rec.1996,P830

[14]WAN Wei,ZHAO Xian-feng,HUANG Wei,et a1.Steganalysis of MP3Stegobased on Huffman table distribution and recoding[J].Journal of Graduate University of Chinese Academy of Sciences,2012,29(1):118-124.