Journal of Information Security, 2013, 4, 12-18
http://dx.doi.org/10.4236/jis.2013.41003 Published Online January 2013 (http://www.scirp.org/journal/jis)

Peak-Shaped-Based Steganographic Technique for MP3 Audio

Raffaele Pinardi[1], Fabio Garzia[1,2], Roberto Cusani[1]
[1]Department of Information, Electronics and Telecommunications Engineering Sapienza University of Rome, Rome, Italy
[2]Wessex Institute of Technology, Southampton, UK
Email: fabio.garzia@uniroma1.it

Received September 4, 2012; revised October 12, 2012; accepted October 22, 2012

ABSTRACT

The aim of this work is the development of a steganographic technique for the MP3 audio format, which is based on the Peak Shaped Model algorithm used for JPEG images. The proposed method relies on the statistical properties of MP3 samples, which are compressed by a Modified Discrete Cosine Transform (MDCT). After the conversion of MP3, it's possible to hide some secret information by replacing the least significant bit of the MDCT coefficients. Those coefficients are chosen according to the statistical relevance of each coefficient within the distribution. The performance analysis has been made by calculating three steganographic parameters: the Embedding Capacity, the Embedding Efficiency and the PSNR. It has been also simulated an attack with the Chi-Square test and the results have been used to plot the ROC curve, in order to calculate the error probability. Performances have been compared with performances of other existing techniques, showing interesting results.

Keywords: Peak-Shaped Steganography; MP3 Steganography

1. Introduction

Steganography techniques are used to hide secret information in the most common audio/video formats. There are three main different kinds of audio/video steganography [1]:

1) insertion steganography, where the secret message is inserted in the cover object;

2) substitution steganography, where some bits of the cover object are substituted with the bits of the secret message;

3) constructing steganography, where an ad hoc cover object is generated to contain the secret message.

The developed technique is based on LSB steganography, a substitution steganography, that replaces the least significant bit of the audio/video file with the secret message bit. This method is very simple to implement and does not allow the human eye/hear to perceive significant changes in the stego object. In **Figure 1** an example is shown: the letter "A" is embedded in the audio samples replacing the least significative bit.

However, this technique has lower resistance to the statistical attacks since with a proper steganalysis it is possible to detect the secret information. To solve this problem, Model Based Steganography can be used [2]. The cover object is divided into two parts, x_a and x_β, to embed the secret information. The first part is the most relevant, and it will not be modified. The second one is less relevant with respect the other and it will contain the secret message. The division is based on the statistical model of the cover object.

After the embedding process, x_β contains the secret information, called x_β. The union between this part and x_a is the stego object.

The purpose of this paper is to present a new steganographic algorithm for the MP3 [3-5] format based on the change, in the Peak Shaped Based for the JPEG [6], of the discrepancy equation, adapting it to vectors and studying the statistical distribution of the MDCT [3-5] coefficients. In the following the analysis of the performance of the proposed algorithm is shown and it is demonstrated that this method does not introduce audible distortion when the signal audio is reproduced. Further, it is demonstrated that this method does not create relevant statistical differences in the samples distribution, showing its suitability for steganographic applications and its robustness to steganographic attacks.

2. The MP3 Format and Compression

The MP3 format was born to have good audio quality and low file size [3,4]. An audio file is first converted into a digital format, with a sampling, and then it is processed with the human psychoacoustic model [3-5].

Sampled Audio Stream (16 bit)	A in binary	Audio stream with encoded message
1001 1000 0011 1100	0	1001 1000 0011 1100
1101 1011 0011 1100	1	1101 1011 0011 1001
1011 1100 0011 1101	1	1011 1100 0011 1101
1011 1111 0011 1100	0	1011 1111 0011 1100
1011 1010 0111 1111	0	1011 1010 0111 1110
1111 1000 0011 1100	1	1111 1000 0011 1101
1101 1100 0111 1000	0	1101 1100 0111 1000
1000 1000 0001 1111	1	1000 1000 0001 1111

Figure 1. LSB example: the letter "A" is inserted into an audio file by replacing the least significant bit.

With this model it is possible to delete the frequency that the human ear can't hear; an algorithm characterized by this properties is called "lossy" because it deletes some information. This happens with a compression, that uses the MDCT [3,5], the Modified Discrete Cosine Transform, described in the following equation:

$$X_k = \sum_{n=0}^{2M-1} x(n) h_k(n) \tag{1}$$

that is another version of the DCT-II used in the JPEG format, where:

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}\cos\left[\frac{(2n+M+1)\cdot(2k+1)\cdot\pi}{4M}\right] \tag{2}$$

and $w(n)$ is a window (it is possible to choose different kinds of windows).

The audio samples are processed in a MDCT filterbank. The audio sequence is divided in "frame", each frame contains M samples and is processed as in **Figure 2**.

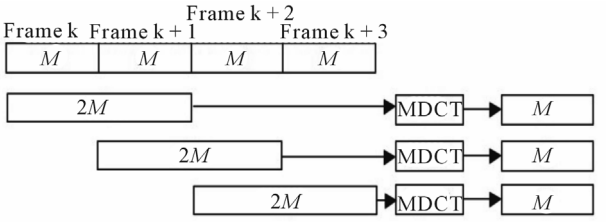

Figure 2. MDCT filterbank.

3. Peak Shaped Based Steganography and MP3

3.1. Peak Shaped Based Steganography

The Peak Shaped Based (PSB) [6] steganography has been used for the JPEG format. This is a method based on:

- LSB steganography;
- Least significant bit;
- Model Based steganography [2].

The JPEG coefficient are, for first, divided by group, indicated with $g(b)$ [6]:

$$g(b) = \text{sign}(b)\cdot\left\lfloor\frac{b}{2}\right\rfloor \tag{3}$$

and by offset, indicated with $O(b)$ [6]:

$$O(b) = |b - 2\cdot g(b)| + 1 \tag{4}$$

where b is the JPEG coefficient and $|b| > 1$.

The PSB algorithm is based on an assumption, from the properties of the JPEG coefficients statistical distribution processed by the algorithm F5, which is [6]:

$$h(b) > h(b+1) \tag{5}$$

$$h(b) + h(b+1) > h(b+1) - h(b+2) \tag{6}$$

where $h(b)$ indicates the histogram of the coefficient b. With this assumption is possible to calculate a probability, called "offset probability".

Subsequently the coefficients are processed with the "discrepancy", an operator that allows to calculate the statistical dependence between two closer coefficients. This is defined as follows:

$$S_0 = \frac{\sum_{j=1}^{4}\sum_{i=1}^{64} q^i \cdot \left|\hat{b}_0^j - \hat{b}_i^j\right|}{4} \tag{7}$$

where [6]:

$$\hat{b}_i^j = \begin{Bmatrix} b_i^j, b_i^j \in x_\alpha \\ 2\cdot g\left(b_i^j\right), b_i^j \in x_\beta \end{Bmatrix} \tag{8}$$

where, as shown in **Figure 3**, the blocks from 1 to 4 are the neighbors of the 0th. Each block contains 64 DCT-II coefficient, 8∗8, and their sequences composing the JPEG image.

3.2. Differences between JPEG and MP3

To apply the PSB algorithm to the MP3 format it is necessary to study the differences between this format and the JPEG standard, in order to identify possible changes. These differences are:

- the JPEG uses the DCT-II while the MP3 uses the MDCT;
- the JPEG works on blocks; each blocks, or matrix, size is 8∗8. Instead, the MP3 works on frame; each frame has dimension equal to 1∗1152, that are vectors;
- the PSB is based on an assumption from the F5 algorithm, that is used for the JPEG format.

Concerning the first point, it is possible to notice that the PSB works on the coefficients. It is therefore necessary to demonstrate that the DCT-II and the MDCT statistical distributions have the same properties.

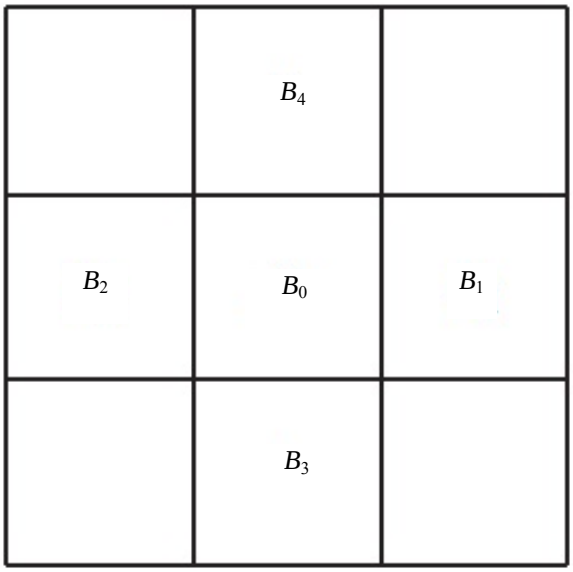

Figure 3. Adjacent blocks to the block B_0.

Concerning the second point, it is necessary to detect the operations that in the PSB works on matrix and, to apply it on the MP3 format, transforming them in operations that works on vectors.

Concerning the third point, it should be studied the statistical distribution of the MP3 coefficients after the F5-algorithm in order to identify if this distribution has the same properties than the JPEG distribution processed by F5.

3.2.1. MP3 Discrepancy

The JPEG discrepancy works on matrix, as described in paragraph 3.1. It is necessary to modify this operation to enable it to operate on vectors. Considering the MP3 format, and the MP3 frames, it is possible to call one of them as k_0. The discrepancy works on the previous frame and the subsequent. The frames are showed in **Figure 4**. The mean is done only for 2 frames because 2 frames are taken and each frame contains 576 coefficients.

$$S_0 = \frac{\sum_{j=1}^{2}\sum_{i=1}^{576} q^i \left|\hat{b}_0^j - \hat{b}_i^j\right|}{2} \quad (9)$$

where $\hat{b}_i^j$ is the same as in the JPEG discrepancy.

3.2.2. Statistical Distribution of the MP3 Coefficients

The statistical distribution of the MP3 coefficients, after the compression, is Peak Shaped [7]. This trend can be modeled by the Generalized Gaussian (GG). The GG varies with a parameter, called r, and it is possible to choose different values of *r*, *i.e.* $r = 2$ is a Gaussian, $r = 1$ is a Laplacian etc. A good approximation for the distribution of the MP3 coefficient has the value *r* set to 0 or 1,

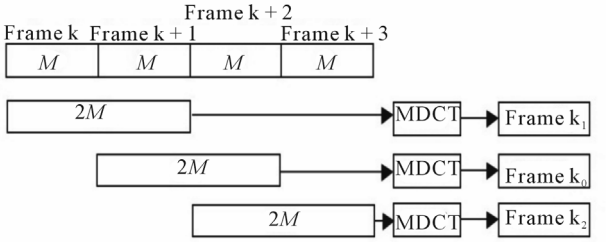

Figure 4. The frames on which works the MP3 discrepancy.

as shown in **Figure 5**.

With the parameter *r* set to 0.5, it is possible to have the best approximation of the statistical distribution. It is possible to choose the value $r = 1$, the Laplacian distribution, to have a good approximation with low complexity.

The Laplacian distribution is used to approximate the statistical distribution of the JPEG coefficients, as well [8-10].

3.3. The PSB over MP3

Using the previous considerations, it is possible to apply the Peak Shaped Based steganography to the MP3 format. In fact it is possible to utilize the assumption used by this algorithm from the F5 [11] steganography because the MP3 coefficients and the JPEG coefficients have the same statistical distribution. F5 modifies the coefficients, without considering their source.

Having both formats, namely both transformed, the same statistical distribution of the coefficients, the use of different transformed becomes irrelevant to the development of the algorithm.

3.4. The Embedding Process

In the following, the list of steps of the embedding process is reported:

- the first step is represented by the analysis of the MP3 statistical distribution;
- successively the value of Hg vector that contains the histograms of the MP3 is calculated;
- with the Hg values it is possible to exclude the samples that are statistical most significant;
- the Hg values allow the calculus, with the algorithm shown in **Figure 6**, of the offset probability vector, called ***P***;
- each frame of the MP3 file, that contains 576 coefficients, is taken and analyzed;
- the coefficients *b* are divided by group, with the $g(b)$ (4), and by offset, with the $O(b)$ (5);
- the discrepancy is calculated by means of Equation (9);
- with the discrepancy, the vector ***P*** and a PRNG, according with the secret key, it is possible to determi-

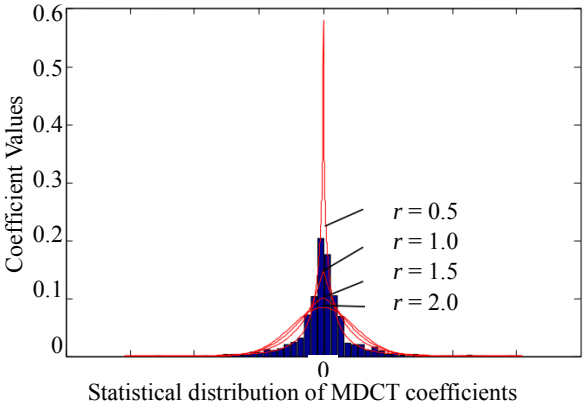

Figure 5. Statistical distribution of the MP3 coefficients approximated with the GG [7].

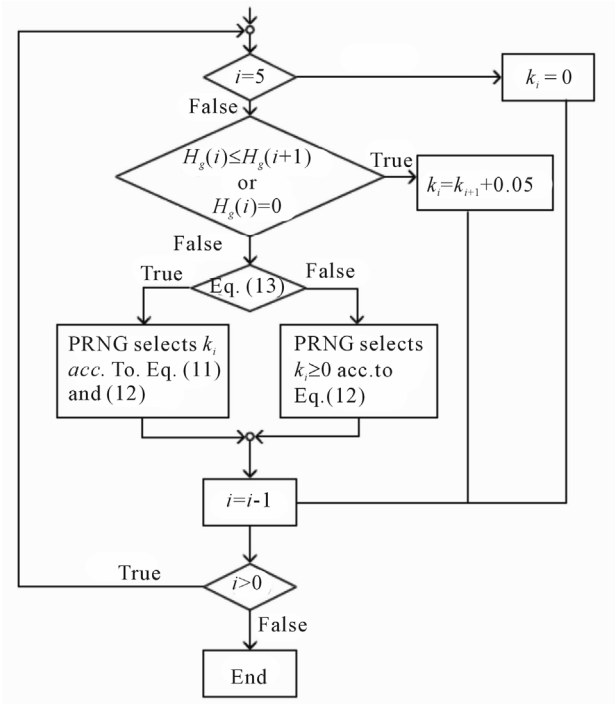

Figure 6. Block diagram of the calculus of the offset probability vector *P*.

nate the coefficients that contain the stego message;

- the offset of each coefficient that is possible to modify is changed according to the value of the bit of the secret message, as showed in **Table 1**.

To extract the secret message the embedding process must be repeated to determinate the coefficients that were modified. The analysis of each offset allows the reconstruction of the secret message.

4. Results

4.1. Performance Parameters

To analyze the performance of a steganographic technique three parameters are used:

- Embedding capacity;
- PSNR;
- Embedding efficiency.

Table 1. Offset of each coefficient that is changed.

	Even	Odd
0	No change	Decrease
1	Increase	No change

4.1.1. Embedding Capacity

The embedding capacity (EC) [7,10] indicates the maximum data size that it is possible to hide in the cover object. It is defined as follows:

$$\mathrm{EC} = \frac{\langle \text{secret message size} \rangle}{\langle \text{cover object size} \rangle} \tag{10}$$

4.1.2. PSNR

The Peak to Noise Ratio [10] provides the similarity between the cover object, the original file, and the stego object, the file where the secret message has been hidden. It is defined through the Mean Square Error (MSE):

$$\mathrm{MSE} = \left[\frac{1}{M \cdot N}\right]^2 \sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(I(i,j) - I^i(i,j)\right)^2 \tag{11}$$

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{12}$$

4.1.3. Embedding Efficiency

Through the Embedding Efficiency (EE) [2,12] it is indicated the average number of bits inserted for each change. It is defined as follows:

$$\mathrm{EE} = \frac{\langle \text{secret message size} \rangle}{\langle \text{number of changes} \rangle} \tag{13}$$

4.2. Peak Shaped Based for MP3 Performance

To evaluate the PSB for MP3 performance it is necessary to compare the Efficiency, the Capacity and the PSNR with different kinds of steganographyc algorithms and with the original PSB, that it was implemented for the JPEG format.

4.2.1. JPEG Performance

The embedding capacity and efficiency for the Model Based Steganography for the JPEG format are shown in **Table 2**.

The PSNR is evaluated for the PSB algorithm for the JPEG format, as shown in **Table 3**. MB1 and MB2 are two model based techniques used in [2], PSB is the Peak

Table 2. Model based steganography performance [2].

Image Name	File Size	Message Size	Capacity	Efficiency
Barb	48,459	6573	13.56%	2.06
Boat	41,192	5185	12.59%	2.94
Bridge	55,689	7022	12.61%	2.07
Goldhill	48,169	6607	13.72%	2.11
Lena	37,678	4707	12.49%	2.16
Mandrill	78,316	10,902	13.92%	2.07

Table 3. PSNR for the JPEG-PSB algorithm [6].

PSNR	Min	Mean	Max
MB1	34.2 dB	40 dB	43.6 dB
MB2	35.2 dB	39.9 dB	44.3 dB
PSB	34.2 dB	40.4 dB	46.6 dB

Shaped Steganography used for the JPEG format.

4.2.2. PSB for MP3

To calculate the Embedding Capacity for the MP3-PSB steganography it is necessary to calculate the file size as follows:

$$L = \frac{\text{bitrate} \cdot M}{\langle \text{sampling frequency} \rangle} \tag{14}$$

A new variable is defined and it is called L_s that is the secret message length. Then the Capacity is:

$$\mathrm{C} = \frac{L_s \cdot 100}{L} \tag{15}$$

and the results are shown in **Table 4**.

The PSNR is evaluated as described in Equation (11) and the results are showed in **Table 5**.

The Efficiency is calculated by analyzing the number of changes to insert the secret message in the cover object. It is possible to see the performance in **Table 6**.

4.2.3. MP3 Steganography Comparison

In **Tables 7** and **8** the PSNR and the Capacity are illustrated for different kinds of steganographyc algorithms for MP3 format.

4.3. Steganalysys

With the steganalysis [9] it is possible to have a better analysis of the PSB-MP3 performance. This method calculated two probabilities, the False Alarm probability, when a cover object is classified as stego, and the Missed Detection probability, when a stego object is classified as cover. The first one is indicated with the symbol $\mathrm{P_{fa}}$

Table 4. PSB-MP3 capacity.

Capacity	Min	Mean	Max
PSB-MP3	4.45%	12.75%	22.05%

Table 5. PSNR for PSB-MP3.

PSNR	Min	Mean	Max
PSB-MP3	55.67 dB	58.21 dB	62.90 B

Table 6. Efficiency for the PSB-MP3.

Efficiency	Min	Mean	Max
PSB-MP3	1.9969	1.9995	2.0012

Table 7. Capacity for different algorithm.

Steganographyc Technique	Embedding Capacity
Peak Shaped for MP3	12.75%
Generic LSB	34%
Tone Insertion	0.006%
Phase Coding	0.02%
Spread Spectrum	0.003%
Echo Data Hiding	0.012%
SVD	0.08%
VAS	10%

Table 8. PSNR for different algorithms.

Steganographyc Technique	PSNR
Peak shaped for MP3	58.21 dB
Phase Coding	69.5 dB
Spread Spectrum	44 dB
SVD	41 dB

and the second one with the symbol $\mathrm{P_{md}}$. The values of these probabilities depend on a threshold, called τ, that modify the steganalysis system accuracy.

With these probabilities it is possible to calculate other parameters, like the detection probability, $\mathrm{P_{det}} = 1 - \mathrm{P_{md}}$, and the error probability $\mathrm{P_{err}}$:

$$\mathrm{P_{err}} = \frac{1}{2} \cdot \left(\mathrm{P_{fa}} + \mathrm{P_{md}} \right) \tag{16}$$

4.3.1. Chi-Square Test

A steganalytic technique that is possible to use for the PSB-MP3 is the Chi-Square test [12]. Some parameters are calculated with the histograms of the MDCT coefficients probability distribution and the results of the

Chi-Square test are compared with the threshold.

One method to calculated the chi-square test is the Zhang-Ping attack [13] that evaluates two variables:

$$f_0 = \sum_{i>0} h_{2i} + \sum_{i<0} h_{2i+1} \quad (17)$$

$$f_1 = \sum_{i<0} h_{2i} + \sum_{i>0} h_{2i+1} \quad (18)$$

and if $f_1 > f_0$ it will calculate the chi-square values:

$$\text{chi}^2 = \frac{(f_0 - f_1)^2}{f_0 + f_1} \quad (19)$$

to compare with the threshold.

4.3.2. ROC Curve

This analysis to evaluate the PSB performances is done on a random set of MP3 files. With the comparison with the threshold and the Chi-Square value it is possible to calculate the two probabilities of false alarm and missed detection, shown in **Table 9**.

With these two probabilities it is possible to graph the ROC curve. This curve indicates the efficiency of the steganalytic method. If the curve is near the first quadrant bisector the steganographyc algorithm is very strong, otherwise the steganalytic method is efficient.

In **Figure 7** it is possible to see the ROC curves (solid line) of the PSB algorithm for the MP3. This curve is very smash on the first quadrant bisector (dashed line). This indicates that the steganographyc method is very robust when this steganalytic algorithm is used. When the ROC curve is far from the bisector the steganographic algorithm isn't very robust or else the steganalytic technique is very efficacious. Instead when this curve is very close to the bisector the technique is very secure, since there is perfect security when the ROC curve is exactly equal to the bisector.

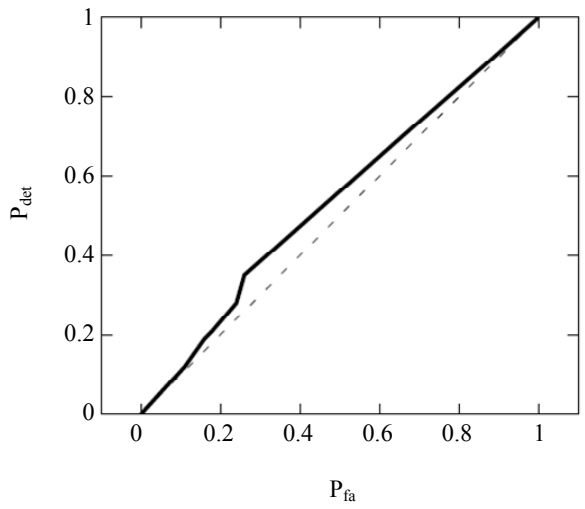

Figure 7. ROC curve for the PSB-MP3.

Table 9. False alarm, missed detection and detection probabilities when the threshold takes different values.

Tau	P_{fa}	P_{md}	$P_d = 1 - P_{md}$
0.01	0.26	0.65	0.35
0.1	0.24	0.72	0.28
0.5	0.18	0.79	0.21
1	0.16	0.81	0.19
10	0.11	0.88	0.12

5. Conclusions

A new steganographic algorithm for the MP3 format has been developed by changing, in the Peak Shaped Based for the JPEG, the discrepancy equation, adapting it to vectors and studying the statistical distribution of the MDCT coefficients. The analysis of the performance of this algorithm showed that this method does not introduce audible distortion when the signal audio is reproduced. Further, this method does not create relevant statistical differences in the samples distribution.

The Peak Shaped Based for the MP3 has an high capacity compared to the other algorithms and a good PSNR. In fact the mean Embedding Capacity is equal to 12.75%, higher than the most relevant techniques used for MP3 steganography; the PSNR is equal to 58.21 dB, higher than the PSB for the JPEG.

A steganalytic attack has been simulated to evaluate the robustness of the algorithm. This attack is implemented on the Zhang Ping analysis and on the Chi-Square test. This attack has been adapted to the PSB-MP3 since it was created for the JSteg steganography. By calculating the false alarm probability and the missed detection probability it is possible to draw the ROC curve. The analysis of this curve shows that this attack is not suitable for this steganographic method because the ROC is crushed on the bisector. The error probability, calculated with the ROC curve, tends to 0.5 when the threshold increases; when it takes this values the choice is completely random.

The steganographyc algorithm implemented, as assumed, is resistant to the statistical attacks.

REFERENCES

[1] G. Kipper, "Investigator's Guide to Steganography," Auerbach Publications, Boca Raton, 2003. doi:10.1201/9780203504765

[2] P. Sallee, "Model-based Steganography," Springer Verlang, Berlin, 2004.

[3] A. Spanias, T. Painter and V. Atti, "Audio Signal Process And Coding," John Wiley and Sons, Hoboken, 2007. doi:10.1002/0470041978

[4] U. Zolzer, "Digital Audio Signal Processing," John Wiley and Sons, Hoboken, 2008. doi:10.1002/9780470680018

[5] J. S. Jacaba, "Audio Compression Using Modified Discrete Cosine Transform: The MP3 Coding Standard," University of the Philippines, Manila, 2001.

[6] L. Rossi, F. Garzia and R. Cusani, "Peak-Shaped-Based Steganographic Technique for JPEG Images," *EURASIP Journal on Information Security*, 2009, Article ID: 382310. doi:10.1155/2009/382310

[7] R. Yu, X. Lin, S. Rahardja and C. C. Ko, "A Statistic Study of the MDCT Coefficient Distribution for Audio," *IEEE International Conference on Multimedia and Expo*, Taipei, 30-30 June 2004, pp. 1483-1486.

[8] D. Liu, H. Zhang, M. Polycaropou, C. Alippi and H. He, "Advances in Neural Networks," Springer Verlag, Berlin, 2011.

[9] R. Bohme, "Advanced Statistical Steganalysis," Springer, Berlin, 2010.

[10] K. B. Shiva Kumar, K. B. Raja and R. K. Chhotaray, "Sabyasachi Pattanaik, Bit Length Replacement Steganography Based on DCT Coefficients," *International Journal of Engineering Science and Technology*, Vol. 2, No. 8, 2010, pp. 3561-3570.

[11] A. Westfeld, "F5-A Steganographic Algorithm," Springer Verlang, Berlin, 2001.

[12] K. Lee, A. Westfeld and S. Lee, "Category Attack for LSB Steganalysis of JPEG Images," Springer Verlang, Berlin, 2006.

[13] T. Zhang and X. Ping, "A Fast and Effective Steganalytic Technique against Jsteg-Like Algorithms," *Proceedings of the* 2003 *ACM Symposium on Applied Computing* (*SAC*), Melbourne, 9-12 March 2003, pp. 307-311.

