Embedding Text in Audio Steganography System using Advanced Encryption Standard, Text Compression and Spread Spectrum Techniques in Mp3 and Mp4 File Formats

Adeboje Olawale Timothy
Federal University of Technology, Akure, Nigeria

Adetunmbi Adebayo Olusola
Federal University of Technology, Akure, Nigeria

Gabriel Arome Junior
Federal University of Technology, Akure, Nigeria

ABSTRACT

Hiding text in a digital audio file format has been a major challenge because of the Human Auditory System (HAS) and how the digital audio will be converted into analog form for text to be embedded into it. Several techniques that include but are not limited to Least Significant Bit, Echo Hiding, Phase Coding, Parity Coding and so on have been proposed in both research communities and the academia. Audio steganography hides data in selected audio files. Several audio steganography works exist, but their major limitations include their inability to embed information in multiple audio file formats, high distortion rate and low level of robustness of their resultant stego files. This research attempts to proffer solution to the obvious challenges of the previous works by developing an efficient and robust audio steganography system for the security of information whether in store or on transit across the Internet. Results of performance evaluation of the developed system shows that it has very low level of distortion as revealed by the Signal to Noise Ratio (SNR). The compression ratio obtained is also equal to one (1), which shows that the cover audio file is identical to the resultant stego file.

Keywords

Stego file, Cryptography, Encryption, Decryption, Private key, Steganography, Spread Spectrum, Discrete Cosine Transform and Cipher Text.

1. INTRODUCTION

Secret digital music MP3 and MP4 files, are important and popular audio compression standard in the Internet. MP3 and MP4 use the destructive compression technologies to achieve high compression rate and the original file is shrunk to a very small size. MP3and MP4 compression always consumes time. To protect digital media files, researchers propose and improve many data-hiding algorithms, which are known as steganographic algorithms.

Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity. The main goal of steganography is to communicate securely in a completely undetectable manner and to avoid drawing suspicion to the transmission of a hidden data.

The most crucial parameters of the data-hiding applications are: security, reliability, invisibility, complexity, and data-hiding capacity, these parameters are mostly related to each other [14].

This research work combined both Cryptography and Steganography to hide text file into MP3 and Mp4 digital audio signal. The use of Huffman algorithm and Lempel-Ziv-Welch algorithm is used for text compression in order to reduce the distortion in the audio file and Frequency Hopping Spread Spectrum technique to increase the robustness of the stego file.

2. RELATED WORKS

In an attempt to address security issue, information hiding techniques like steganography have shown some promising solutions. However, there are some raising concerns when using this approach, for example, in research work presented by [1] and [2], the research works suffered a limitation of high distortion in the stego file. [12] introduce LSB method as a robust method of hiding data in audio signal. However, as the size of the secret message become larger, the embedding process becomes complicated. [2] explained a novel method for data embedding in the audio stream. But the system can only embed information in wav file.

[15] develop a tool to encrypt a message and hide the message into a digital object. The objectives of the research work are to combine various encryption and steganography techniques in other to provide more powerful message concealing method and to improve the quality of the resultant stego-file. Thereby, reducing suspicion of convert communication. The encryption method used in their research is ElGamal Encryption. The encrypted message is embedded in the homogenous frames of mp3 audio file. Before embedding the message into the file, the encrypted message is enhanced with spread spectrum method and XOR modulation to improve its randomness. The limitations of this research work are: a. the quality of the stego file depends on file size and message length. The research work was only carried out on mp3 audio format and not on other audio format.

3. METHODOLOGY

Audio file format is a container format for storing audio data and metadata on a computer system. This format is divided into three major groups namely: the uncompressed audio formats, (e.g., WAV), formats with lossless compression, (e.g., WMA) and formats with lossy compression (e.g., MP3). In this research work, the selected cover media which is the audio signal undergoes different stages; the selected audio file irrespective of its format will be converted into .Wav (Wave File Audio) format using a function in MATLAB.

Many types of audio file are available, such as Windows Audio Visual (WAV), Windows Media Audio (WMA), and MPEG (MP3). The type used in this study is WAV file format of type Pulse Code Modulation (PCM), because it is uncompressed audio format, which gives more flexibility for data hiding. A stego object (WAV file) with high sampling rate and sampling resolution may draw suspicion, because of its large size, especially if its subjective quality is not high. Usually, it is easy to hide more secret data in the high quality audio data (for example, the use of least significant bit encoding to embed one bit in each sample, consist of 16 bits, sample has less effect on the stego object than adding one bit in a sample consist of 8 bits). In the developed system the wave files, with 8-bit samples resolution, are used as cover media for hosting the secret data.

There are some basic parameters to understand when discussing on digital audio files. This parameter includes:

i. **Sampling rate (f_s)** is the number of samples of audio carried per seconds. It is measures in Hertz. In this research, 44100Hz is used in wav audio format.

ii. **Frame size:** The frame size is the amount of bits in each frame.

$$\text{Frame Size} = \frac{1}{Partitioned\ frame\ size} \qquad 1$$

iii. **Frequency Resolution (k_s):** It is the scaling factor, which ensures the cover medium is embedded below the audibility threshold. It is measured in Hertz. Human Auditory System (HAS) works dynamically in a wide range of frequencies between 20Hz-20000Hz. It is calculated using equation 2

$$k_s = \frac{f_s}{2\times\frac{N}{2}} \qquad 2$$

3.1 Embedding Module

Embedding module is a module at which the secret message (text file) is embedded into the selected cover medium (audio Signal). This is depicted in the figure 1.

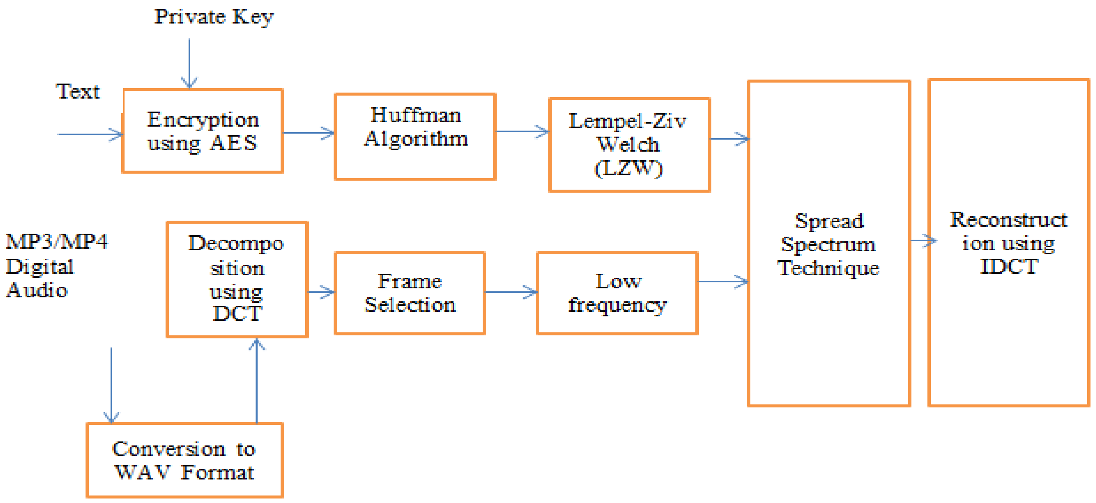

Figure 1: Architecture of the Embedding Module

The embedding module involves three processes which include: plain text (secret message) encryption and compression, audio signal decomposition as well as combined signal construction processes.

The secret message which is in plain text is encrypted in order to enhance the security of the secret message from the intruders by using a public key and private key to avoid unauthorized access of the text. This research work makes use of Advanced Encryption Standard (AES) over other encryption algorithms because it uses higher length key sizes such as 128, 192 and 256 bits for encryption, it uses 128-bit block size and also has 10, 12 or 14 rounds of bits depending on the key size used. Hence it makes AES algorithm more robust against hacking. The general encryption procedure is mathematically represented in equation

$$E(K,M) = \{C\}K \qquad 4$$

where E represents the Encryption function, M denotes the Plain text (Secret message), K stands for Encryption Key, and C is the cipher text.

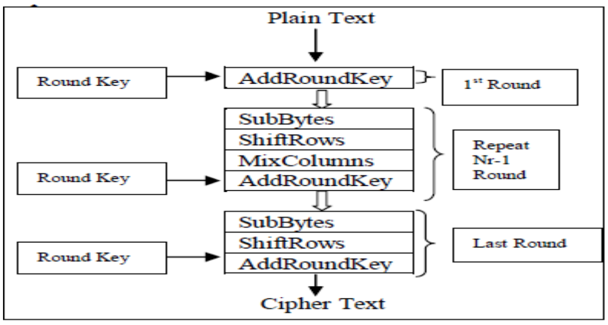

Figure 3: AES encryption algorithm

The encrypted text will be compressed using a two stage compression techniques which are Huffman algorithm and Lempel-Ziv Welch (LZW). Data compression is a method of encoding rules that allows substantial reduction in the total number of bits to store or transmit a file and this will help to reduce distortion in the stego file. The feature of both LZW (Lempel-Ziv-Welch) and Huffman algorithms are combined to improve the compression ratio. The Huffman algorithms is represented as

$$b_{Huff} = \sum_{a=1}^{n} f(a_i) L(\{C\}) \qquad 5$$

where $L(\{C\})$ denotes the length of the cipher text, $f(a_i)$ is the word character and n is the number of bit.

Lempel Ziv Welch is represented as

$$c_{lzw} = b_{Huff}(\log_2 b + \log_2 N + 2) \qquad 6$$

where b_{Huff} represents the compressed cipher text using Huffman model, N denotes the fixed codes and c_{lzw} denotes the LZW compression.

The main advantage of this combined algorithm is that the percentage of data reduction increases more compared to the existing text compression techniques.

The audio format irrespective of its audio file format is converted into .wav format. This is achieved using an audio write function in MATLAB.

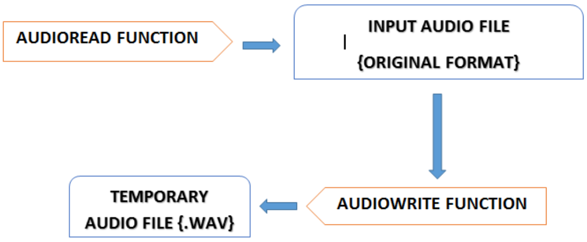

Figure 3: Audio format conversion to wav format

The reason for this is that it will be easier to decompose the digital audio signal into analog signal using the One Dimensional Discrete Cosine Transform (DCT).

The DCT is one of the powerful compact transforms. It relocates most of the signal energy into the first transform coefficients, lesser energy or information is relocated into other (i.e., high frequency) coefficients. The frames thus created are queued based on the energies of the frames. DCT is applied on the voiced blocks that have power less than the predefined second threshold value (T'). The block size was taken small to avoid the high computational complexity of DCT calculations which makes the system slow. Thereafter, frames will be selected using frequency frames with low frequency will be selected in order to ensure the secret message does not introduce audible distortion in the audio signal.

$$f_{dct}(x) = \sum_{u=1}^{N-1} \alpha(u)\ c\ (u) \cos\left[\frac{\pi(2x+1)u}{2N}\right], \qquad 7$$

for $x = 0,1, \dots\dots\dots. N-1$

$$\text{where } \alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} \text{ for } u = 0 \\ \sqrt{\frac{2}{N}} \text{for } u \neq 0 \end{cases}$$

where $f_{dct}(x)$ represents the original sequence of the audio signal, N denotes the Last frame in the audio file, x is the Number of frames in an audio file and u denotes the frame size.

Spread Spectrum technique was used to hide the encrypted and compressed secret message (Text file) into the digital Audio signal. Spread spectrum is a method by which energy generated in particular bandwidth is deliberately spread in the frequency domain, resulting in a signal with a wider bandwidth. Spread spectrum ($s_{sprectrum}$) systems encode data as a binary sequence which sounds like noise but which can be recognized by a receiver with the correct key. There are two types of spread spectrum techniques which are: Direct Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS). In Direct Sequence Spread Spectrum, data to be transmitted is divided into small pieces and each piece is allocated to a frequency channel across the spectrum. In this research work, Frequency Hopping Spread Spectrum is used. In Frequency-Hopping Spread Spectrum, the audio file's frequency spectrum is altered so that it hops rapidly between frequencies. Spread spectrum combined the compressed text file with the low frequencies of the audio signal using:

$$s_{sprectrum} = c_{lzw}.\ f_{dct}(low) \qquad 8$$

The embedded signal is added to the other frames of high frequency using:

$$f_{frame}(t) = s_{sprectrum} + f_{dct}(high) \qquad 9$$

The analog signal generated is then converted into digital signal using the Inverse Discrete Cosine Transform (IDCT) as given below:

$$c_{dct}(u) = \alpha(u) \sum_{x=1}^{N-1} f_{frame}\ (t) \cos\left[\frac{\pi(x+1)u}{2N}\right] \qquad 10$$

where $c_{dct}(u)$ is the new audio signal (stego file).

4. RESULT AND DISCUSSION

The system was implemented using MATLAB (R2017a version) programming language on Windows 8.1 Operating System platform with hardware configuration of 3GB RAM, 1.6Hz Intel processor speed and 250GB of hard disk. In this developed system, MP3 and MP4 digital audio file format were used as the cover media and different ranges of secret text were hidden into them for evaluation. The stego files (embedded audio file) were evaluated. The stego file (audio file) retains its initial size after evaluating the proposed approach and the amount of information that the developed system can hide is very high (500KB).

The research work was evaluated using the following performance metrics: computational time, bit per character, compression ratio and signal to noise ratio.

4.1 Compression Ratio

This is the ratio of the cover medium before and after the secret message is embedded into it to its ratio when the secret message is embedded into it.

$$\text{Compression Ratio} = \frac{\text{output file size}}{\text{input file size}} \qquad 1$$

From the result carried out in this research work, the system has a compression ratio of 1, which means the size of the audio file before the secret message is embedded into it is still the same size after the secret message is embedded into it.

4.2 Signal to Noise Ratio (SNR)

Signal to noise ratio is a parameter used to know the amount by which the signal is corrupted by the noise. It is defined as the ratio of the signal power to the noise power. Alternatively, it represents the ratio of desired signal (say a music file) to the background noise level. It is measured in decibel (db). SNR can be calculated by equation below.

$$\text{SNR (db)} = 10\log\frac{\sum_n I_n}{\sum n(E_n - I_n)^2} \qquad 2$$

where E_n= Stego file and I_n= Original Audio Signal

4.3 Computational Time

This is the time taken for the system to execute its function.

4.3.1 MP3 Audio Format

Mp3 file format, with size 4.58Nb and a length of 5 minutes was used as an experiment to perform the performance metrics. Table 1 shows the result of the evaluation.

Table 1: Mp3 File Evaluation Result

Text size (KB)	Audio file size(MB)	Audio Length (Minutes)	Compressed Text Size (KB)	Compression Ratio in Percentage	Computational Time (second)	Signal to Noise Ratio (Db)	Bit per character	Audio Size After Embedding (MB)	Extraction Time	Compression Ratio
50	4.58	5	43923.6	87.5	6	59.5	8	4.58	4	1
100	4.58	5	82741.5	82.7	6	54.7	8	4.58	4	1
150	4.58	5	128710	85.8	6	53.9	8	4.58	4	1
200	4.58	5	181829	90.9	6	51.3	8	4.58	4	1
250	4.58	5	211964	84.8	6	48.5	8	4.58	4	1
300	4.58	5	260484	85.8	6	32.4	8	4.58	4	1
350	4.58	5	293173	83.8	6	28.4	8	4.58	4	1
400	4.58	5	355485	88.9	6	20.4	8	4.58	4	1
450	4.58	5	376936	83.8	6	17.9	8	4.58	4	1
500	4.58	5	434141	89.8	6	10.7	8	4.58	4	1

From table 1, the values of the Signal to Noise ratio are more than 50db when the size of the text file to be embedded ranges from 50kb to 200kb, and this indicates that there will be no distortion in the audio. But from 250kb to 500kb, the values of the Signal to Noise Ratio began to decrease; making the values to be less than 50db and this implicates that there will be distortion as the value decreases from 50db. This can be represented graphically in figure 1.

From table 2, the values of the Signal to Noise ratio goes above 50db when the size of the text file to be embedded ranges from 50kb to 100kb, but the value of SNR for 150kb is 49.3 which van be approximately to 50db can also be accepted as good SNR vale. This indicates that there will be no distortion in the audio with a secrete message that's up to 150kb. But from 200kb to 500kb, the values of the SNR began to decrease; making the values to be less than 50db and this implies that there will be distortion as the values decreases from 50db. This is represented graphically in figure 2.

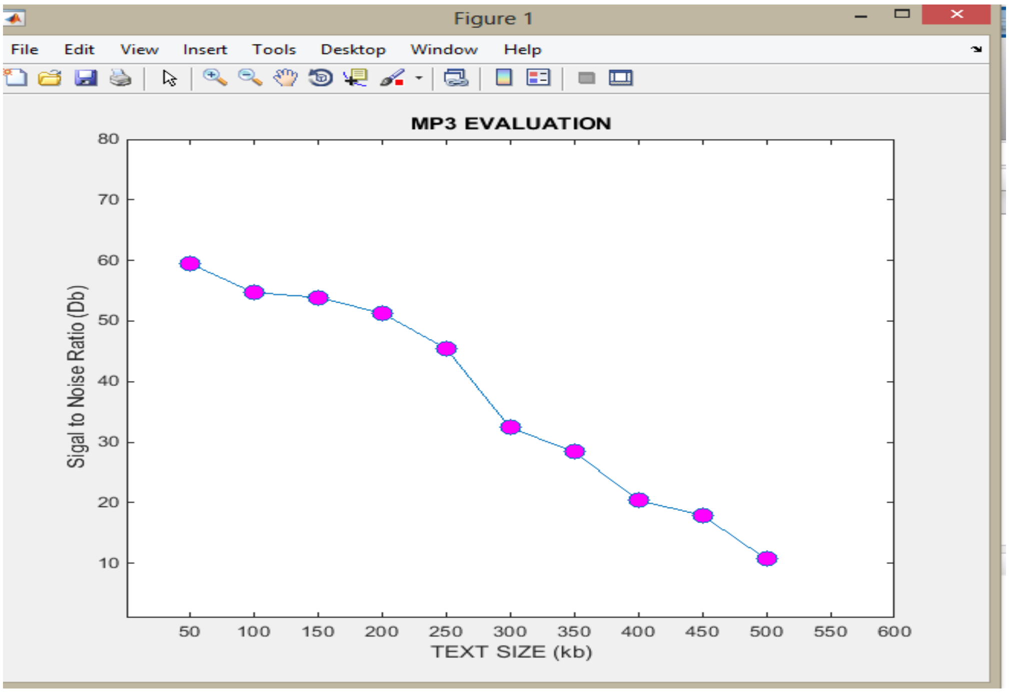

Figure 1: Graph of MP3 file format

4.3.2 MP4 Audio Format

MP4 file format, with size 4.70 Mb and a length of 5 minutes was used as an experiment to perform the performance metrics. Table 2 shows the result of the evaluation.

Table 2: .MP4 File Evaluation Result

Text size (KB)	Audio file size(MB)	Audio Length (Minutes)	Compressed Text Size (KB)	Compression Ratio in Percentage	Computational Time (second)	Signal to Noise Ratio (Db)	Bit per character	Audio Size After Embedding (MB)	Extraction Time (second)	Compression Ratio
50	4.81	5	42902.2	14.2	9	59.7	8	4.81	7	1
100	4.81	5	88870.5	11.1	8	55.9	8	4.81	6	1
150	4.81	5	131774	12.2	8	51.6	8	4.81	6	1
200	4.81	5	163442	18.3	8	49.8	8	4.81	6	1
250	4.81	5	224733	10.1	8	42.6	8	4.81	6	1
300	4.81	5	272742	9.1	8	37.6	8	4.81	6	1
350	4.81	5	307474	12.2	8	31.7	8	4.81	6	1
400	4.81	5	355485	11.2	8	28.7	8	4.81	6	1
450	4.81	5	399920	11.1	8	22.0	8	4.81	6	1
500	4.81	5	423926	15.2	8	18.6	8	4.81	6	1

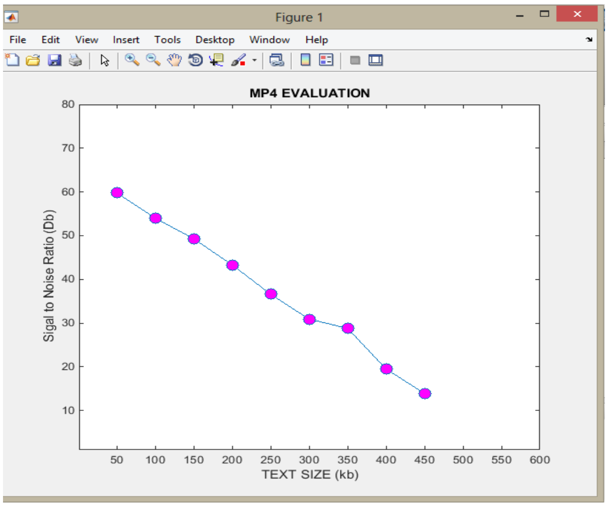

Figure 2: Graph of MP4 file format

5. CONCLUSION AND RECOMMENDATION

There are a number of proven methods for applying steganography to hide information within audio data. In this research work, an audio steganography system for MP3 and MP4 that uses using Discrete Cosine Transform (DCT) and spread spectrum techniques was developed. It was shown through implementation and subjective experimentation that the developed audio steganography system supports MP3 and MP4 digital audio format. The system developed has the ability to embed a secret message of size that is up to 500kb but the system has the ability to embed a text size of 250kb with respect to the digital audio length or size without any distortion and has the ability to retain the same size after embedding text into it.

The work has been able to develop a robust stenographic system that would be very useful in securing and sharing large amount of sensitive data or information without arousing suspicion. This system is therefore recommended for security agencies and other organization that consider information security as being of uttermost priority. This system is a useful means for transmitting covert battlefield information via an innocuous cover audio signal.

6. REFERENCES

[1] Wheeler, D., Johnson, D., Yuan, B., and Lutz, P. (2012). Audio Steganography Using High Frequency Noise Introduction.

[2] Ghanwat, D., and Rajan, R. S. (2013). Spread Spectrum based audio steganography in transformation domain. Global Journal of Advanced Engineering Technologies, Vol2, Issue4-2013.

[3] Bandyopadhyay, S. K., and Datta, B. (2011). Higher LSB layer based audio steganography technique. *IJECT*, *2*(4), 129-135.

[4] Bertino, E. (2013, August). Data security–challenges and research opportunities. In Workshop on Secure Data Management (pp. 9-13). Springer, Cham.

[5] Bhowal, K., Pal, A. J., Tomar, G. S., and Sarkar, P. P. (2010, November). Audio steganography using GA.

In Computational Intelligence and Communication Networks (CICN), 2010 International Conference on (pp. 449-453). IEEE.

[6] Can, Y. S., Alagoz, F., & Burus, M. E. (2014). A Novel Spread Spectrum Digital Audio Watermarking. Journal of Advances in Computer Networks, 2(1)

[7] Cheng, W. Q., Han, F., Tung, M. J., & Xu, K. (2007). Robust audio steganography using direct sequence spread spectrum technology. *Technology*, 1-6.

[8] Dhore, V., & Arfat, P. M. (2015). Secure Spread Spectrum Data Embedding and Extraction. *International Journal of Science and Research*, *4*(1), 743-747.

[9] Divya, S. S., & Reddy, M. R. M. (2012). Hiding text in audio using multiple LSB steganography and provide security using cryptography. International journal of scientific & technology research, 1(6), 68-70.

[10] Geetha, K., & Muthu, P. V. (2010). Implementation of ETAS (embedding text in audio signal) model to ensure secrecy. International Journal on Computer Science and Engineering, 2(04), 1308-1313.

[11] Olanrewaju, R. F., Othman Khalifa, H. A., and Suliman, R. (2013). Increasing the hiding capacity of low-bit encoding audio steganography using a novel embedding technique.

[12] Abdullah, A. M. (2017). Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data. *Cryptography and Network Security*.

[13] Ali Khayam, S. (2003) .The Discrete Cosine Transform (DCT): Theory and Application. Information Theory and Coding, Seminar. The Discrete Cosine Transform: Theory and Application.

[14] Atoum, Mohammed Salem, O. A. A. Rababah, and Alaa Ismat Al-Attili. "New technique for hiding data in audio files." *Journal of Computer Science* 11.4 (2011): 173-177.

[15] Kresnha, P. E., and Mukaromah, A. (2014). A Robust Method of Encryption and Steganography using ElGamal and Spread Spectrum Technique Based on MP3 Audio File. In Proceeding Conference on Application and of Electromagnetic Technology, 3(9):11-15.