Mathematical Model that using scrambling and Message Integrity Methods in Audio Steganography

Mohammed Salem Atoum
Department of Computer Sciences
Irbid National University
Moh_atoom1979@yahoo.com

Abstract: The success of audio steganography is to ensure imperceptibility of the embedded message in stego file and withstand any form of intentional or un-intentional degradation of message (robustness). Audio steganographic that utilized LSB of audio stream to embed message gain a lot of popularity over the years in meeting the perceptual transparency, robustness and capacity. This research proposes an XLSB technique in order to circumvent the weakness observed in LSB technique. Scrambling technique is introduce in two steps; partitioning the message into blocks followed by permutation each blocks in order to confuse the contents of the message. The message is embedded in the MP3 audio sample. After extracting the message, the permutation code book is used to re-order it into its original form. Md5sum and SHA-256 are use to verify whether the message is altered or not during transmission. Experimental result shows that the XLSB performs better than LSB.

Keywords: Mathematical Model, Scrambling, Integrity

Introduction:

Information is share globally through the Internet, in digital form (Fricker and Rand, 2002). There are issues and challenges regarding the security of information in transit from senders to receivers. The major issue is the protection of digital data against any form intrusion, penetration, and theft. The major challenge is developing a solution to protect information and ensure their security during transmission (Feruza and Kim, 2007). Three components of information security are confidentially, integrity, and availability (Feruza and Kim, 2007). . Confidentiality ensures that information is kept secret from any unauthorized access. This could be done through information hiding techniques, namely cryptography and steganography (Lenti, 2000).

Cryptography involves the act of encryption and decryption of a digital data. The major weaknesses of such techniques are that even though the message has been encrypted, it still exists. Steganography dwells on concealing any digital data in a in an innocuous digital carrier, the word steganography is derived from an old Greek word which means covered writing (Katzenbeisser and Petitcolas, 2000). Steganography has been used as a mean of concealing secret messages during ancient times (Rahim, Bhattacharjee and Aziz, 2014). It was used by Histiaeus, the tyrant of Miletus who, in 499 BC, tattooed the scalps of his slaves with a hidden message with a command for his men to attack the Persian (Ricardo *et al.,*1999; Huayin and Li, 2008; Emelia *et al.,* 2008; Yu *et al.,* 2010). The message became hidden when the slaves' hair grew back.

According to researchers, steganography can be described as a study of the means of hiding secondary information within primary information without affect the size of information nor the cause of any form of distortion which could be perceived (Francia *et al.,* 2006; Qiao, Sung and Liu, 2009b; Ganeshkumar and Koggalage, 2009; Petrovic and Yann, 2009; Liu, Qiao and Sung, 2009; Jangra, and Singh, 2014). The primary information, known as the carrier or host, was embedded within the secondary information, which is typically hidden and could be in the form of a file or message. The media with the embedded information is called stego

signal, file, bit stream or sequence (Basu and Bhoumik, 2010; Khairullah *et al.,* 2009; Alla, Parsad and Siva, 2009; Changder, Debnath and Ghosh, 2009; Qi, Ye and Liu, 2009; Farouk, 2014).

Steganography is the one of two techniques used covert communication. However, watermarking is the second techniques that can be embedded watermark into host cover to keep copyright for the hosts. Steganography typically establishes point-to-point data security (Mandala, Kotagiri and Kapala, 2013). The strength of steganographic technique in keeping the data in the carrier medium against attacks or alteration is weak during transmission, storage or format conversion is weak (Katzenbeisser and Petitcolas, 2000).

The process of embedding information in host media in steganography technique and watermarking are usually done transparently (Manimegalai *et al.*, 2014; Koziel, 2014). The difference between steganography and watermarking is that while steganography is a technique which hides the information, watermarking actually allows the third person to see the message (Cvejic and Sebbanen, 2004; Neeta, Snehal and Jacobs, 2006). Thus, in terms of watermarking, the process needs to ensure robustness so that any intentional attacks would not compromise, remove, or cause destruction of the information in any way in the marked media while at the same time preserving the quality of the signal (Scagliola, Berez and Guccione, 2009; Bhattacharyya and Sanyal, 2010). Watermarking is the most suitable technique in cases where knowledge of the hidden information could cause possible manipulations (Mitchell, 2003; Avcibas, Memon and Sankur, 2003; Yusnita and Othman, 2007; Naji *et al.*, 2009).

Proposed Scheme

This paper describes in detail the design of the proposed scheme, which is developed to enhance the security and robustness of the basic steganography model. The proposed scheme comprises of three phases: scrambling message, embedding and extraction algorithm and message integrity. Scrambling message and message integrity are the two phases added to the basic model of steganography in order to improve security and robustness. In addition XLSB is presented, this is scheme is the main part of the contribution of this research, because it extends the basic LSB technique, through incorporating high security level for the secret message before embedding and message integrity method after extraction. Figure 1 shows the proposed scheme

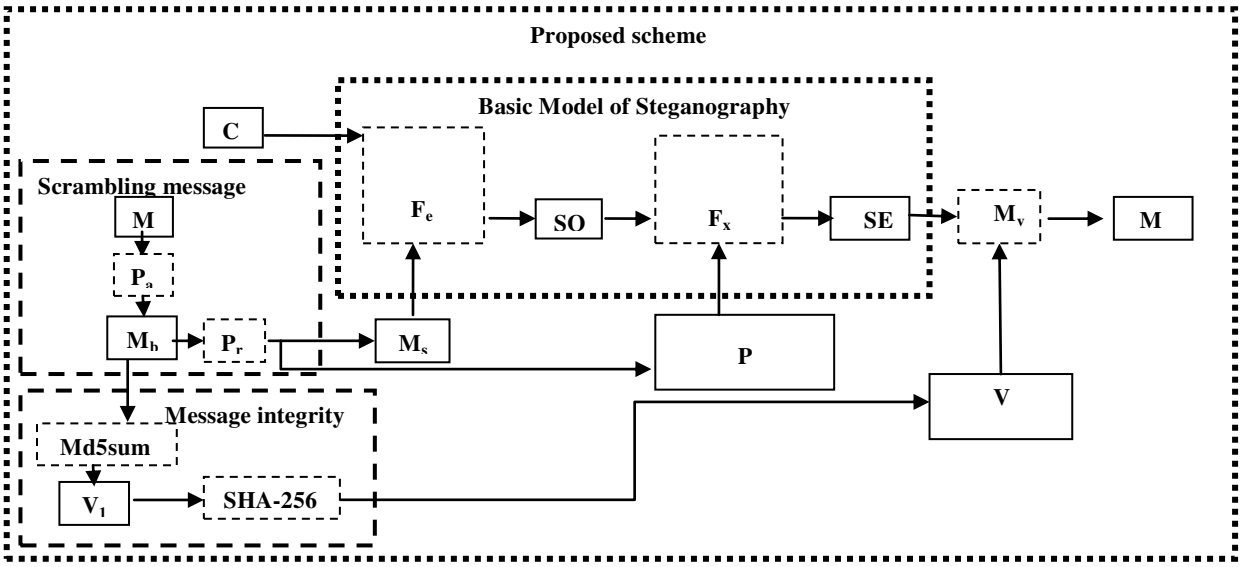

Figure.1 Proposed scheme

The model is presented in there different parts, which consist of the basic model of steganography, scrambling of message and message integrity respectively. The following variables were involved in the interrelationships between each section of the model. Thus, c represent the cover file, and C is the set of all possible cover files for which they belong to c, therefore mathematically it can be represented as $c \in C$. Before embedding, there is a need to prepare the secret message, thus if m represent the secret message, and M represents the set of all possible secret messages for which they belong to m, mathematically it can be represented as $m \in M$. The length of the cover file is given by L_c= length of c, whereas the length of the secret message is given by L_m= length of m. After preparing the dataset, the scrambling message M_s which is one of the most important aspect of the general models start with partitioning process P_a, where n_p represent the number of partitions within n_p=ceil($L_c \backslash L_m$), for each dataset, P_a is the Partition function of the message m, which produces M_b= message blocks after partitions M_b= {m_1, m_2, m_3, ……., m_{np}}, now the M_b= P_a(m) undergoes permutation step, where P_r is the Permutation function of the message block. P is the codebook of permutation, it is a random equation. Thus a

M_s is generated which is the secret message after scrambling. The message scrambled after permutated M_b are presented as M_s= P_r(M_b,P). The embedding of the of the secret message on the cover file, dwells on the F_e denoted as the Embedding Process, for which F_e(M_s,C)= SO, where SO is the Stego Object. After embedding the extraction process F_x for which F_x(SO) = M_s and C will be carried out, thereafter, the extraction function F_x(SO,P) = M_b makes it possible for the extraction of the secret message. The extracted message undergoes a process of validation, whereV_1 represent the validation value one, thus the validation procedure is presented below**:**

Md5sum= Checksum function using MD5 method

Md5sum(M_b)= V_1

SHA-256= Secure hashing algorithm 256 method

SHA-256(V_1)= V

V: Validation value to check message integrity.

M_t: Message integrity function.

V_1'= validation value generated from end user.

Md5sum(M_b)= V_1'

SHA-256(V_1') = V'

V': Validation value generated from end user.

If M_t(V') = = V, then the message is intact,

Else the message is being altered

In general the model explores the framework of the research and the steps requires to for the conceptualization of the different sections of the model. By implementation, consider P and V which are secret information required to send to the receiver through a secure channel. If message scrambling comprises of two processes: the partition process (P_a), which partition M into message blocks (M_b); and the permutation process (P_r) which re-orders Mb to construct $\mathbf{M_s}$ and generate the first part of secret information, P.

A new scheme was developed based on the basic steganography model, which was described in detail in section 2.2 of chapter two. The basic steganography model utilizes two processes, namely embedding (F_e), and extraction (F_x). In F_e both M and C are converted into bits stream, and one of the known embedding algorithms (XLSB) is used to embed M into C. F_x includes the same steps in F_e to extract the message but in an inverse way.

Scrambling Method

Algorithm 1 Scrambling Algorithm

```
Inputs: C is cover, M is message, Lc is length of C, Lm is length of M,
Outputs: Ms, P
        Parameters: np is a number of partitions, Mb is message
        blocks, S is the size of blocks, P is P_Codebook.
        // Partition process
                np= ceil(Lc\Lm)
                        S= Lm/ np
                        for i=1 to np
                                for j=1 to S
                        Mb[i]= M[j]
                                End for
                        End for
        // Permutation process
        // Select random permutate
                        RandomPerm(Mb[i], j)
                        For i= 1 to np
                            choose j uniformly at random from [i,...,n]
                                // for example can choose j= ( (np-i)+1) or any  P= j
                                swap Mb [i] and Mb [j]
                                Ms[i]= Mb[i]
                        End for
```

Message Integrity Method

Algorithm 2. Message integrity algorithm

```
Input: Mb is message blocks.
Outputs: True or False
Parameters: Md5sum is function to generate checksum, SHA-256 is a function to generate
verification value, V1 is first verification value, V is the second verification value.
// Md5sum process
        For i=1 to np
                V1= Md5sum( Mb(i))
        End for

// SHA-256 process
V= SHA-256(V1)

13: Validation Check
If V=V' then T else F
End
```

Conclusion

This research has explored and reviewed MP3 audio steganography, particularly with respect to MP3 files after compression. LSB in time domain has been extended and XLSB algorithm is formed. The new technique aims at meeting the three most important audio steganography requirements, which are imperceptibility, capacity, and robustness. Any technique tries to enhance the capacity or robustness should preserve imperceptibility. This research increased the capacity and robustness as well as improved the imperceptibility. Two algorithms, standard least significant bit (SLSB) and extended least significant bit algorithm (XLSB) were implemented; the first algorithm was implemented based on the general idea of LSB to

be a benchmark for the new extended algorithm. SLSB has sufficient information about cover format, and avoids manipulating those bits in samples, which cause larger error or distortion. Second approach was a proposed algorithm based on the concept secure least significant bit. XSLB has been proposed, implemented and tested.

References

[1] A.Z. Al-Othmani, A. A. Manaf and A. M. Zeki, "A Survey on Steganography Techniques in Real Time Audio Signals and Evaluation", *IJCSI International Journal of Computer Science Issues*, Vol. 9, no. 1, 2012

[2] S. B. Kumar, D. Bhattacharyya, P. Das, D. Ganguly and S. Mukherjee, "A tutorial review on Steganography", *International Conference on Contemporary Computing (IC3-2008),* Noida, India, 2008, pp. 105-114.

[3] P. Dutta, D. Bhattacharyya, and T. Kim, "Data Hiding in Audio Signal: A Review", *International Journal of Database Theory and Application*, Vol. 2, No. 2, June 2009

[4] W. Bender, W. Butera, D. Gruhl, R. Hwang, F. J. Paiz, and S. Pogreb, "Techniques for data hiding", *IBM Systems Journal*, Vol. 39, No. 3-4, pp. 547 – 568, 2000.

[5] V.N. Vapnik, "Statistical Learning Theory". John Wiley and Sons, New York, USA, 1998

[6] F. Djebbar, B. Ayad, K. A. Meraim and Ha. Hamam, "Comparative Study of Digital Audio Steganography Techniques", Survey paper, 2011.

[7] P. Jayaram, H. R. Ranganatha, and H. S. Anupama , "Information Hiding Using Audio Steganography-A Survey", *IJMA,* Vol.3, No.3, 2011

[8] A. Z. Al-Othmani, A. Abdul Manaf, and A. M. Zeki, "A Survey on Steganography Techniques in Real Time Audio Signals and Evaluation", *IJCSI*,Vol.9,No 1, 2012

[9]B. A. Usha, N. K. Srinath and N. K. Cauvery, "ANALYSIS OF IMAGE STEGANALYSIS TECHNIQUES TO DEFEND AGAINST STATISTICAL ATTACKS – A SURVEY", *International Journal of Research in Engineering and Technology*, pp. 148-151

[10] P. Chandrakar, M. Choudhary and C. Badgaiyan, "Enhancement in Security of LSB based Audio Steganography using Multiple Files", *International Journal of Computer Applications*, Vol. 73, No.7, pp. 21-24, 2013

[11] R. A. Al-Dallah, A. M. Al-Anani, R. I. Al-Khalid, and S. A. Massadeh, "An Efficient technique for data hiding in audio signals", *American Academic & Scholarly Research Journal*, Vol. 4, No. 5, 2012

[12] I. J. Kadhim, "A New Audio Steganography System Based on Auto-Key Generator", *Al-Khwarizmi Engineering Journal*, Vol. 8, No. 1, pp. 27 – 36, 2012

[13] U. Mehta and D. Sihag, "Multi-Part Data Hiding in Audio Steganography", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 3, No. 9, pp. 8037- 8039, 2014

[14] P. Praveen, and R. Arun, "Audio-video Crypto Steganography using LSB substitution and advanced chaotic algorithm", International Journal of Engineering Inventions, Vol. 4, No. 2, pp. 01-07, 2014

[15] T. Barbu, "Variational Image Denoising Approach with Diffusion Porous Media Flow" , *Abstract and Applied Analysis*, Vol. 2013, 2013.

[16] Gaussian waves, "Simulation and Analysis of White Noise in Matlab", [Online]: available at: www.gaussianwaves.com/2013/11/simulation-and-analysis-of-white-noise-in-matlab/

[17] Atoum, M. S. (2015). A Comparative Study of Combination with Different LSB Techniques in MP3 Steganography. In Information Science and Applications(pp. 551-560). Springer Berlin Heidelberg.

[18] Atoum, M. S., Ibrahim, S., Sulong, G., and Zamani, M. (2013). A New Method for Audio Steganography Using Message Integrity, *Journal of Convergence Information Technology*,8(September), 35–44.

[19] Atoum, M. S., Ibrahim, S., Sulong, G., and Ahmed, A. (2013).New Secure Scheme in Audio Steganography (SSAS). *Australian Journal of Basic and Applied Sciences*, 7(6), 250–256.

[20] Atoum, M. S., Ibrahim, S., Sulong,G. and Ahmed, A. (2012). MP3 Steganography: Review. *Journal of Computer Science issues*, 9(6).

[21] Atoum, M. S., Suleiman, M., Rababaa, A., Ibrahim, S., and Ahmed, A. (2011). A steganography Method Based on Hiding secretes data in MPEG / Audio Layer III. *International Journal of Computer Science and Network Security*, 11(5), 184-188.

[22] Atoum, M. S., Rababah, A. and Al-attili, A. I. (2011).New Technique for Hiding Data in Audio Files. *International Journal of Computer Science and Network Security*, 11(4), 173-177.

[23] Atoum, M. S., Ibrahim, S., Sulong, G., Zeki, A and Abubakar, A. (2013). Exploring the Challenges of MP3 Audio Steganography. *Proceding IEEE from 2nd International Conference on Advanced Computer Science Applications and Technologies (ACSAT)* ,Sarawak, Malaysia.

[24] Atoum, M. S. (2015, August). New MP3 Steganography Data Set. In IT Convergence and Security (ICITCS), 2015 5th International Conference on (pp. 1-7). IEEE.