A Stego-analysis Techniques by SOD Using Statistical Measurements Based on FPGA

Elaf Sabah Abbas[1], Dhamyaa H. Mohammed[2], Thamir Rashed Saeed[3] and Sabah A. Gitaffa[4]

[1,2] Dept. of Computer Communication Eng., Al-Mansour University College, Baghdad, Iraq

[3,4] Dept. of Electrical Eng.- University of Technology, Iraq

(FPGA circuit design for DSP Research Group, Ciphering Research Group)

e-mail: thamir_rashed@yahoo.com

Received: 17/02/2014

Accepted: 04/06/2014

Abstract – Steganalysis is the technique of analyzing a stego-image to determine whether it has embedded data or not. More deliberately steganalysis, it can be achieved by coding a program that examines the stego-image structure and measures its statistical properties. This paper presents a novel steganalysis algorithm by detecting the sequence occurrence distribution (SOD) of cover/setgo-image using three types of statistical randomness properties tests: Frequency, Serial and Poker. Where hidden a $2.4X10^{-7}$% distortion of covering image in multiple-LSB (MLSB), the difference achieved detection between cover-stage images as; frequency is 0.91362828; serial is 3.45887 and poker is 160.6455. Also, this proposed algorithm can point to the occurrences of the sequence which is affected by the embedded message, then implemented it by using 8-bit pair code and made by Xilinx-spartan-3A XC3S700AFPGA, with 50 MHz internal clock.

Keywords: – Steganalysis, SOD, Statistical properties.

1. Introduction

The steganalysis refers to the body of techniques that are designed to distinguish between cover and stego-objects. This discrimination between a stego and cover-objects can be achieved with or without the knowledge of the steganographic algorithm that is used for embedding the secret message [1]. Most steganographic techniques involve changing properties of the covered source, and there are several ways of detecting these changes [2]. Steganalysis techniques are divided into two broad categories: active and passive [3], [4]. The biggest challenges in active steganalysis are the identification of reliable, feature detects stego-images [4], and it is an inherently difficult problem [5]. While passive steganalysis attempts to destroy any trace of secret communication, without detecting the secret data [6]. More deliberately, it can be achieved by coding a program that examines the stego-image structure and measures its statistical properties [6], and it can be considered as successful, and the respective steganographic system as 'broken', if the steganalysis decision can be solved with higher trustworthy than random guessing[7], [8]. Overall, there is still no universal, "one size fits all" detection solutions; thus, steganalysis methods must be adjusted precisely to the specific information hiding technique [9].

2. Related work

Several steganalysis techniques have been proposed in the literature. These methods can be classified into two general categories: specific and universal. A universal distortion design called universal wavelet relative distortion (UNIWARD) for embedding in an arbitrary domain that was proposed by Vojt¿ch Holub et al [10]. The Edge Detection Filter Technique for detecting the hidden message in the edges of the image was presented by Nitin Jain et al [5]. While Ch. Demudu Naidu et al [11] presented a detection method by using Functional Link Artificial Neural Network for detecting the coded content of Steganography. Also, Xianyang Luo et al [12] proposed a method for combining appropriate trace sets to estimate the modification ratio of each natural binary bit-plane. Quantization index modulation (QIM) based Steganography, Hafiz Malik et al [13] presented a nonparametric steganalysis method using irregularity (or randomness) in the test-image to distinguish between the cover-image and the stego-image [13]. Tu-Thach Quach et al and Yoan Mi che [4], [14] presented a steganalysis method using features calculated from a measure that is invariant for cover images and is altered for stego images by modeling the distribution of the DCT coefficients as a Laplacian. The fusion of the histogram of running length and histogram characteristic function for detection LSB matching was proposed by [15]. In the other words, the detection of JPEG Steganography by 2-D arrays formed from the magnitudes of JPEG quantized block DCT coefficients was presented in [16]. A multi-scale, multi-orientation image decompositions, first- and higher-order magnitude and phase statistics are relatively in consistent across a broad range of images, but are disturbed by the presence of embedded hidden messages as described in [17]. While Jessica Fridrich et al [8] presented a detection method by using a linear classifier trained on feature vectors correspond to cover and stego images. Whereas, the relation between the length of embedded massage and gradient energy was analyzed by [18], and then Gradient Energy-Flipping Rate detection was proposed. Ismail Avcıbas et al [19] used a statistical feature for stego-analysis, while Hany Farid [20] also used

the same technique but with wavelet-like decomposition.

3. Sequence Occurrence Distribution (SOD) Statistics

The tests are designed for a random number, while we are doing a modification to the input parameters to serve our proposed algorithm. The aim of the SOD [21] statistical tests is to measure the quality of randomness for cover and stego-image to identify certain kinds of weakness, which represents the hidden message that it may have. Three kinds of tests will be made; Let the binary sequence of Input Image Pixels IIP = IIP_0, IIP_1, IIP_2, …., IIP_{n-1} of length n , where n is the number of image pixels, and IIP_n is 8-bit for BW image and 24-bit of color one. The three tests are;

3-1 Frequency test (mono-bit test):

The frequency tests can be determined by counting the numbers of pixels which have LSB 0's and 1's. Where, n_0 denotes the number of pixels with LSB equal 0's and n_1 represent the number of pixels with LSB equal 1's and the test defined by;

$$X_1 = (n_0 - n_1)^2 / n \qquad (1)$$

3-2 Serial tests (two-bit-test):

This test can be used to determine the occurrence number of 00, 01, 10, and 11 of LS2B (Least significant two bits). Where, n_{00}, n_{01}, n_{10}, and n_{11} denote the number of occurrences in IIP, respectively, the test is calculated by;

$$X_2 = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2\right) - \frac{2}{n}\left(n_0^2 + n_1^2\right) + 1 \qquad (2)$$

3-3 Poker test:

The Poker tests can determine whether the occurrences of each part of the length m (least significant m bits (LSmB)). Then, the sequence IIP is divided into k non-overlapping parts of a length m, and n_i be the number of occurrences of the i^{th} part. This test is defined by;

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k \qquad (3)$$

Where k=1024000, and m=4. Since, the Statistical steganalysis is more powerful than signature steganalysis, because mathematical techniques are more sensitive than visual perception[3].

4. Proposed Work

For detecting the MLSB Steganography [22], [23], a steganalysis method was proposed based on randomness statistical analysis of the cover/stego-image. This detection is based on the SOD of the image pixel bits. Where, after embedded a 35 bits as a massage in LSB of 409600 pixels as a cover image Figure (1-a) which represents 2.4X10-7 % and 1.2X10-7 % of LS2B distortion of the cover image (stego-image) Figure (1-b).

a) Cover- image

b) Stego- image

Figure (1) a) Cover image b) Stego-image

The detection algorithm was based on the process of the stego-image by the algorithm in Figure (2) by calculating the statistical properties of SOD (X_{1s}, X_{2s} and X_{3s}) for receiving stego-image after that we do the same for cover-image (X_{1c}, X_{2c} and X_{3c}) which stored in the database of the algorithm, and then make comparison between them (C_1, C_2 and C_3) .

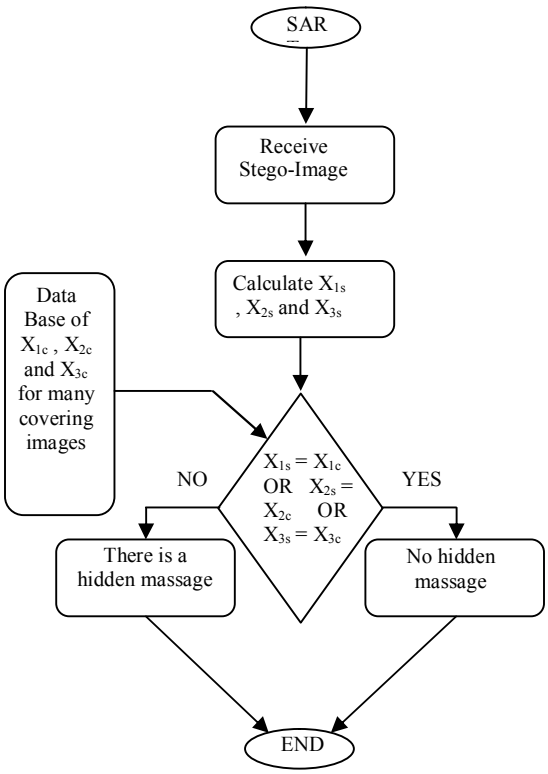


Figure (2) Proposed SOD stego-analysis Algorithm

After satisfying the proposed algorithm operation, we implement it by using 8-bit pair code which is made by Xilinx-spartan-3A XC3S700AFPGA, with 50 MHz internal clock as shown in Figure (6). The operation is based on four stages, first one counting the number of pixels (n), number of pixels to LSB equal one and zero (n_1, n_0), number of pixels with LS2B 00,….,11 (n_{00}, n_{01}, n_{10}, and n_{11}) and number of pixels with LS4B 0000…..1111 (n_{0000},… n_{1111}) at the same time. The second stage is calculated using equations (1), (2), and (3) by making a VHDL core for each equation. The third stage compares the output of the second stage (X_1, X_2, and X_3) with the same variables of the original image (Cover image). The last stage is the decision stage (Target) which is satisfied by using OR gate between the output of the comparator (C_1, C_2, and C_3). Where, target equal to 1 or equal to zero, which indicates for a hidden text message or not respectively.

5. Results and discussion

From the Figure (1), it is clear that the invisibility of distortion detection because the percent of distortion that be added is $2.4X10^{-7}$ % for LSB and $1.2X10^{-7}$ % for LS2B. Then, after the SOD treatment of LSB to return the same as before the distortion was added as in Table (1).

The proposed algorithm detected the distortion of stego- image by calculating X2 and X3 where its value is 756 and 453 respectively. While, for LS2B the detection of $1.2X10^{-7}$ % distortion is clear in the Table (2) where, the difference of X1, X2 and X3 are 0.91362828, 3.45887 and 160.6455 respectively

Figure (3) represents the SOD of LS4B after hiding the test (red bar) and a blue one before it, and the invisibility is clear. Whereas, the position which effected by hiding the text are in Table (3).

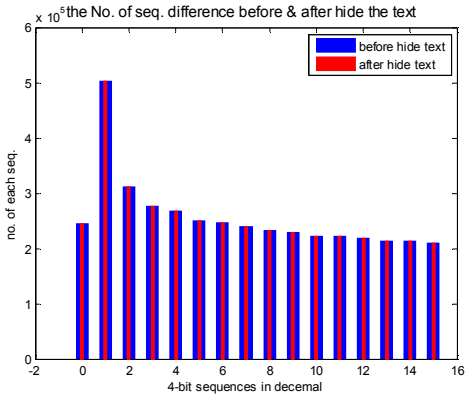


Figure (3) No. of sequences before and after adding a message

Table(1) Cover/stego-Images and Message Occurrence Sequences of 1's and 0's

[1] Before padding a text		**[2] After padding a text**		**[3] After the treatment site go- Image according to ref. [21,22]**	
[4] Image pixels		**[5] Image pixels**		**[6] Image pixels**	
[7] 4096000		[8] 4096000		[9] 4096000	
[10] 1's in LSB	**[11] 0's in LSB**	**[12] 1's in LSB**	**[13] 0's in LSB**	**[14] 1's in LSB**	**[15] 0's in LSB**
[16] 2141558	[17] 1954442	[18] 2141553	[19] 1954447	[20] 2141558	[21] 1954442
[22] **Text bits** = 35					
[23] No. of 1's	**[24] No. of 0's**				
[25] 13	[26] 22				
[27] Text % of covering the image					
[28] $85.44X10^{-7}$					

Table(2) Statistical Tests Results

[29] Frequency test (X_1)		**[30] Serial test (X_2)**		**[31] Poker test (X_3)**	
[32] Before hid a text	*[33] After hiding a text*	*[34] Before hid a text*	*[35] After hiding a text*	*[36] Before hid a text*	*[37] After hiding a text*
[38] 8547.948598	[39] 8547.03497	[40] 30683.73179	[41] 30687.19066	[42] 1.7471e+008	[43] 1.7471e+008
[44] Difference		**[45] Difference**		**[46] Difference**	
[47] 0.91362828		[48] 3.45887		[49] 160.6455 [50]	

Table(3) Effected Sequences

Effected sequence	Effected value
0000	-1
0001	1
0010	-1
0011	2
0110	1
0111	-1
1000	3
1001	-3
1100	-1
1101	1

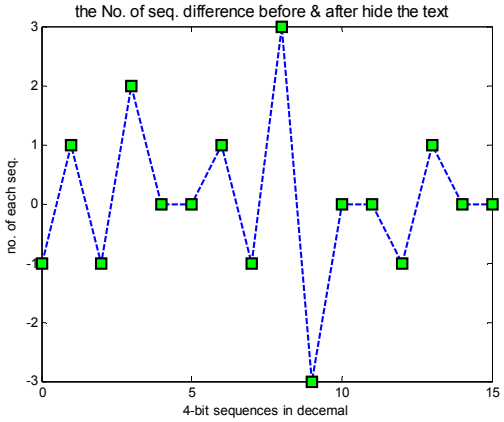


Figure (4) Effected Values of Sequences

While Figure (4) represents the difference between the SOD before and after hiding the text in the effected sequences and the correspondence of these two figures are in Figure (5) for clarifying the sequences which effect the distortion and the quantity of these effects.

This means, the proposed algorithm can identify the sequence which is affected by the embedded text message. Figure (6) represents the four stage algorithm which is implemented by Xilinx-spartan-3A XC3S700AFPGA.

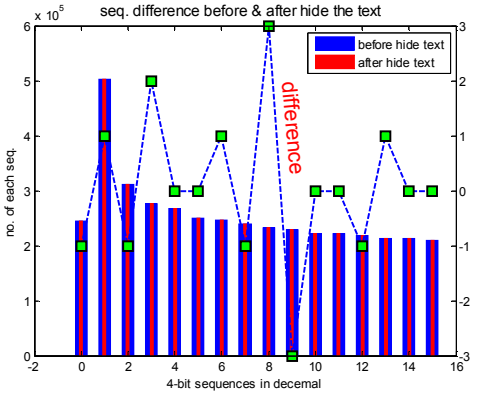
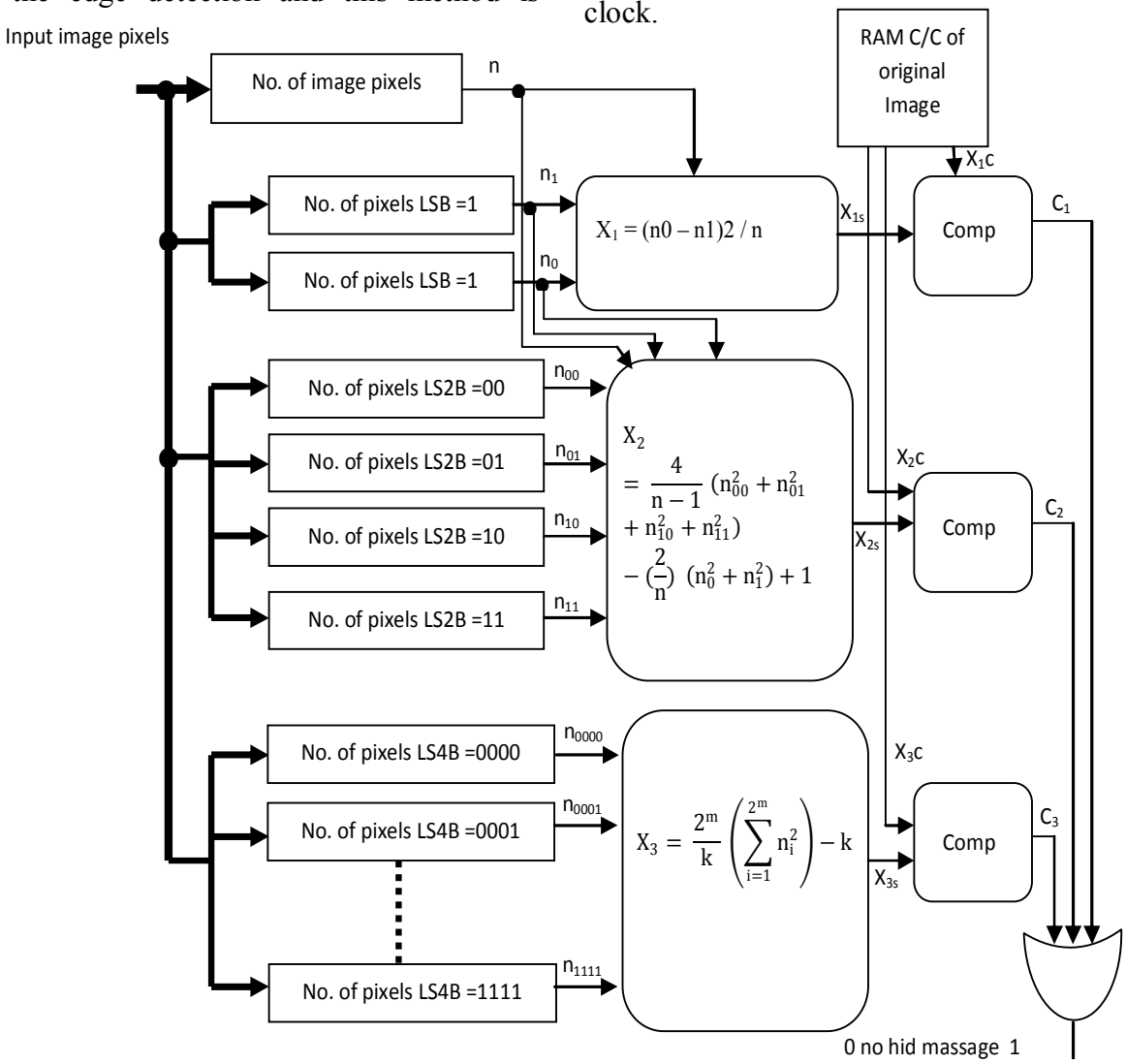


Figure (5) Correspondence the Effected Values with Occurrence Sequences

Its output is shown in Figures (7) and (8), where, the waveform of the proposed algorithm represents the decision state with and without text message. Whereas, after calculating the SOD LSB, LS2B and LS4B and X_1, X_2 and X_3 before and after hiding the text. Then, the Target as in figure (6) shows HIGH where C_1, C_2, and C_3 are not at LOW. While, Figure (8) represents another case where after Target signal appeared, the hidden message is removed, then, the Target is at LOW because C_1, C_2, and C_3 are at LOW. Therefore, the result was significant difference between the cover and stego-images. While the work of [5] depends on the edge detection and this method is turned off when the hidden message are not at the edges.

6. Conclusion

The most obvious finding to emerge from this study is that, the detection of the hidden text message, whatever their percent of covering the image. In this work is $1.2X10^{-7}$ %, also, can identify the sequence occurrence which is affected by this hidden text message. While the simplicity and ease to implement, its clear by using the proposed algorithm with low cost based on Xilinx-spartan-3A XC3S700AFPGA, with 50 MHz internal clock.



Figure(6) Block diagram of the proposed algorithm

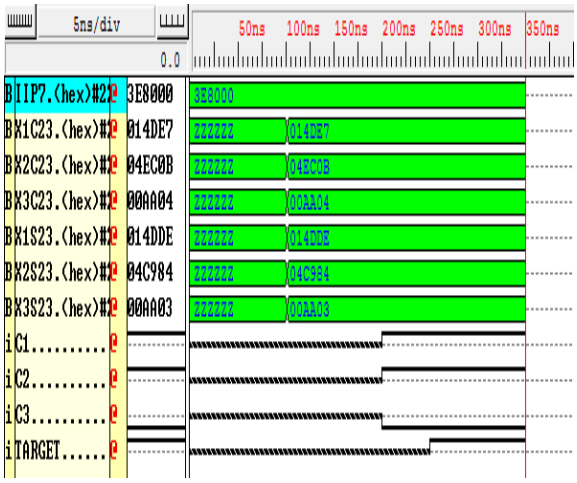


Figure (7) Waveform of the Proposed Algorithm with Hidden Message

Figure (8) Waveform of the Proposed Algorithm with and without Hidden Message

References

[1] El-Sayed M. El-Alfy and Azzat A. Al-Sadi, " Pixel-Value Differencing Steganography: Attacks and Improvements", 2nd International Conference on Communication and Information Technology, Feb. 2012.

[2] Rainer B¨ohme and Andreas Westfeld, "Breaking Cauchy Model-based JPEG Steganography with First Order Statistics", Computer Scurity ESORICS, Vo. 3193, pag. 125-140, 2004.

[3] Yambem Jina Chanu, Manglem Singh, and Themrichon Tuithung, "Image Steganography and Steganalysis: A Survey", Computer Applications, Volume 52–No.2, August 2012.

[4] Tu-Thach Quach, Fernando Perez-Gonzalez, and Gregory L. Heileman, "Model-Based Steganalysis Using Invariant Features", SPIE, Media Forensics Security Proc., Vol. 7254, San Jose, 2009.

[5] Nitin Jain, Sachin Meshram, and Shikha Dubey, "Image teganography Using LSB and Edge – Detection Technique", International Journal of Soft Computing and Engineering (IJSCE) Vol.2, Issue-3, July 2012.

[6] Abbas Cheddad, "Steganoflage: A New Image Steganography Algorithm", PhD thesis, University of Ulster, Sept. 2009.

[7] Rainer Bohme, " Advanced Statistical Steganalysis", Springer-Verlag Berlin Heidelberg, 2010.

[8] Jessica Fridrich, " Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes", Information Hiding-Springer Link, Vol. 3200, Pag. 67-81, 2005.

[9] Wojciech Mazurczyk, " VoIP Steganography and Its Detection – A Survey", Cornell University Library, computer Science, arxiv.org, Cryptography and security, 2013.

[10] Vojt¿ch Holub, Jessica Fridrich and Tomá¿, Denemark, "Universal distortion function for steganography in an arbitrary domain", EURASIP Journal on Information Security 2014.

[11] Ch. Demudu Naidu, S. Pallam Setty, M. James Stephen, S.K.Prashanth, and Ch. Suresh, "Steganography Detection using Functional Link Artificial Neural Networks", International Journal of Computer Applications, Volume 47, No.5, June 2012.

[12] Xiangyang Luo, Fenlin Liu, Chunfang Yang,Shiguo Lian and Ying Zeng, " Steganalysis of adaptive image steganography in multiple gray code bit-planes", Springer Science and Business Media, Vol. 57, Pag. 651– 667, 2012.

[13] Hafiz Malik, K. P. Subbalakshmi and R. Chandramouli, " Nonparametric Steganalysis of QIM Steganography using Approximate Entropy", IEEE Transactions on Information Forensics and Security, Vol. 7, Issue 2, 2011.

[14] Yoan Mi che, Patrick Bas, Amaury Lendasse, Christian Jutten and Olli Simula, " Reliable Steganalysis Using A Minimum Set of Samples and Features", Hindawi Publishing Corporation EURASIP Journal on Information Security, 13 pages, Volume 2009.

[15] Xiaoyi Yu and Noboru Babaguchi, "Run Length Based Steganalysis for LSB Matching Steganography", IEEE Conference on Digital Object Identification, Pag. 353- 356, 2008.

[16] Yun Q. Shi, Chunhua Chen, and Wen Chen, "A Markov Process Based Approach to Effective Attacking JPEG Steganography", IH'06 Preceding of the 8th International Conference on Information Hiding, Springer-Verlag, Pag. 249-264, 2007.

[17] Siwei Lyu, and Hany Farid, "Steganalysis Using Higher-Order Image Statistics", Information Forensics and security IEEE Transaction, Vol. 1 Issue 1, Pag. 111-119, 2006.

[18] Li Zhi and Sui Ai Fen, " Detection of Random LSB Image Steganography", IEEE Vehicular Technology Conference, ISSN 1090-2038, Vol. 3 Pag. 2113-2117, 2004.

[19] Ismail Avcıbas, Nasir Memon, and Bülent Sankur, "Steganalysis Using Image Quality Metrics", IEEE Trans. On Image Processing, VOL. 12, NO. 2, Feb. 2003.

[20] Hany Farid, " Detecting Hidden Massage using Higher-Order Statistical Models", International Conference on Image Processing Vol. 2, Pag. 905-908, 2002.

[21] Yas A. Alsultanny, and Hashim J. Jarrar, " Generating and Testing Random Key for Image Encryption using ECB and CBC Modes", J. J. Appl. Sci., Vol. 8, No. 1, Pag. 1-11, 2006.

[22] Thamir Rashed Saeed and Shaymaa Abd-Elghany, " Efficient Adaptive Setganography Algorithm", International Journal of Pure and Applied Research in Engineering and Technology, Vo. 2, No. 1, 2013.

[23] Thamir Rashed Saeed, "A Novel Steganography with Preserving Statistical Properties", International Journal of Computer Science Issues, Vol. 10, Issue 5, No.2, September 2013.