International Journal of Cryptology Research 5(1): 77-93 (2015)

Analysis of Steganography Substitution System Methods Using New Testing Techniques Proposed for Strict Avalanche Criterion

[1]Abdul Alif Zakaria, [2]Nor Azeala Mohd Yusof,
[3]Wan Zariman Omar, [4]Nik Azura Nik Abdullah and
[5]Hazlin Abdul Rani
Cyber Technology Research Department
CyberSecurity Malaysia, Kuala Lumpur, Malaysia
Email: [1]alif@cybersecurity.my,[2]azeala@cybersecurity.my,
[3]wanzariman@cybersecurity.my, [4]azura@cybersecurity.my,
[5]hazlin@cybersecurity.my

ABSTRACT

In this research, we present the analysis on steganography substitution system methods; Least Significant Bit Substitution, Random Interval, Pseudorandom Permutation, Image Downgrading & Covert Channels, and Cover-regions & Parity Bits. New testing techniques proposed by Phyu Phyu Mar and Khin Maung Latt for strict avalanche criterion is implemented to analyze secret message bit distribution in all methods. One million bits of secret messages have been used for sampling in this testing. This analysis compares each methods results on three tests; total number of bit changes in each output, output values in each output, and total number of bit changes in each bit position. From our observation, Random Interval and Cover-regions & Parity Bits produced the best test results compared to the other three steganography substitution methods.

Keywords*: Steganography, Substitution System, Strict Avalanche Criterion, Hamming Weight*

1. INTRODUCTION

Steganography is the art and science of hiding messages. The word Steganography comes from the Greek words Steganós meaning covered and Graptos meaning writing. Steganography and cryptography have similarities where both are used to protect important information. However, steganography is different from cryptography because it involves hiding information without noticing any alteration made to the cover object (Cheddad et. al., 2007). Cover object or carrier is the file such as text, picture, image, audio or video in which secret message is hidden. The secret message can also be the same form as cover object. File containing secret message that has been hidden in cover object is called stego-object (Tariq et. al., 2013).

Steganalysis is the art and science of detecting a secret communication in steganography. Detectable traces in the cover medium

may exist if message is hidden. Changes in statistical properties of the cover may lead to steganalyst attempting to detect the existence of the secret communication. This attempting process to detect statistical traces is called statistical steganalysis.

Many developed embedding techniques in the last few years have been successfully attacked. Statistical properties of secure stego-systems should be protected and controlled because each time a new secure embedding algorithm is developed, steganalyst will find a new statistic to ensure the success of their attack (Leivaditis, 2010). It is important to inject a secret message into a carrier that no detectable changes are introduced for secure secret communication. The main objective is to avoid introducing statistically detectable modifications into the carrier and to not raise suspicion to the attacker. Size of the secret message, format and content of the carrier image may directly influence the ability of not being detected and lead to higher probability that the modification can be statistically detected (Fridrich and Du, 2000).

This paper illustrates comparison of five known steganography substitution system methods using new techniques proposed for Strict Avalanche Criterion (SAC) in search of the best substitution system method to be implemented. A short description of Least Significant Bit Substitution, Random Interval, Pseudorandom Permutation, Image Downgrading & Covert Channels, and Cover-regions & Parity Bits methods are described in **Section 2**. **Section 3** explains the purposes for all Hamming weight tests. In **Section 4** and **Section 5**, the experimental setup and analysis are discussed. Summary and conclusion of the research are lastly demonstrated in **Section 6**.

2. STEGANOGRAPHY SUBSTITUTION SYSTEM METHODS

Figure 1 shows an existed classification of steganography (Al-Ani et. al., 2010). Substitution system is one of six steganography classification method that is applied in steganography. This system substitutes redundant or least significant parts of a cover with a secret message. Receiver can extract the information if the position where secret information has been embedded is known. Below (section *A*, *B*, *C*, *D* and *E*) are the methods under substitution system that are publicly known and implemented in steganography tools (Johnson and Katzenbeisser, 2000).

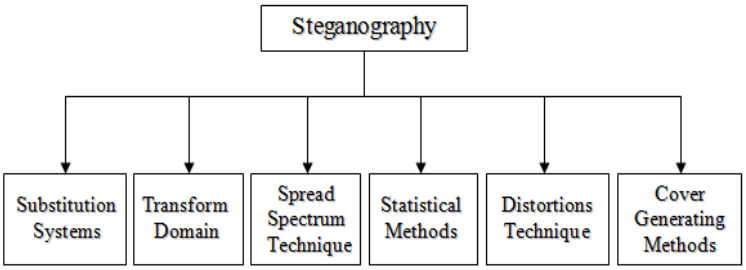

Figure 1: Steganography Classification

A. Least Significant Bit Substitution

This method is performed by substituting the least significant bit (LSB) of cover-element with the bit of a secret message (Rodrigues et. al. 2004) and is illustrated in Figure 2. To reconstruct the secret message, the LSB of the selected cover-elements are extracted and lined up. By using this method, information can be hidden with little impacts to the carriers. Because of the simplicity of applying it into image and audio, this method is now common to steganography.

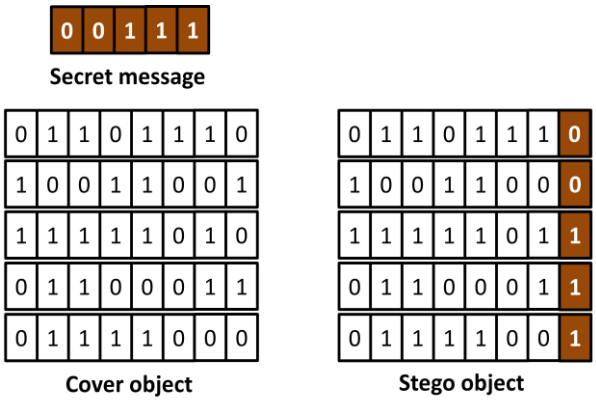

Figure 2: Least Significant Bit Substitution

B. Random Interval

A pseudorandom number generator is used in this method to spread the secret message over the cover-elements in a rather random manner (Johnson and Katzenbeisser, 2000). Both sender and receiver share a stego-key as a seed for a random number generator. A random sequence is created in which the interval between two embedded bits is determined

pseudorandomly. The secret message bits are stored according to the interval between two embedded bits. This method is illustrated in Figure 3.

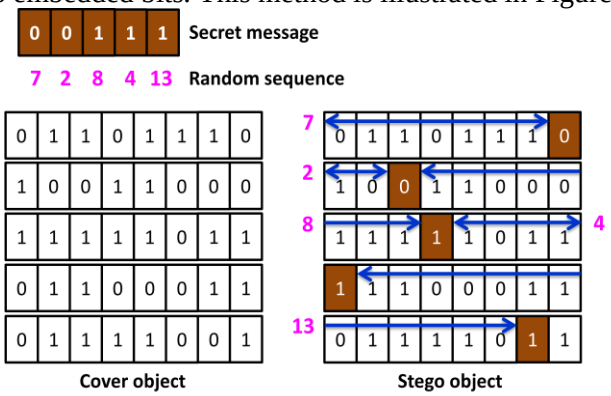

Figure 3: Random Interval

C. Pseudorandom Permutation

Distribution of the secret message is done in random manner over the whole cover-elements. The main goal is to increases the complexity for an attacker because there is no guarantee that subsequent message bits are embedded in the same order. A sequence is generated using a pseudorandom number generator. The secret message bits are stored according to bit position of cover-elements which is determined by the generated sequence (Hossain, 2014). This method is illustrated in Figure 4.

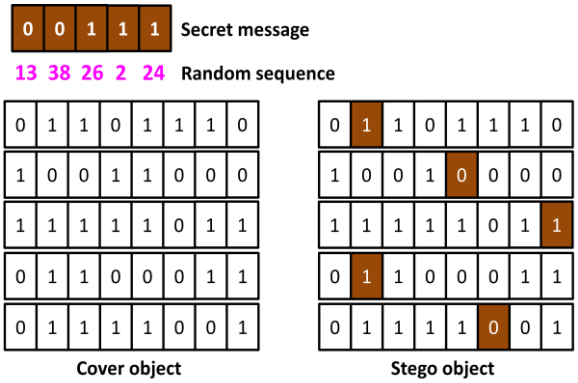

Figure 4: Pseudorandom Permutation

D. Image Downgrading and Covert Channels

Images can be exchanged covertly using this method (Zaidan et. al., 2009). This method is usually used for "leaking" information. Convert channels in operating systems allow processes to communicate "invisibly" and possibly across different security zones specified by a security policy. Both secret and cover messages are in form of images. Given a cover-image and secret image of equal dimensions, the sender exchange the four least significant bits of the cover's color values with the four most significant bits of the secret image. Access to the most significant bits of the secret image is gained when the receiver extracts the four least significant bits out of the stego-image. Four bits are sufficient to transmit a rough approximation of the secret image since the degradation of the cover is not visually noticeable in many cases. This process is illustrated in Figure 5.

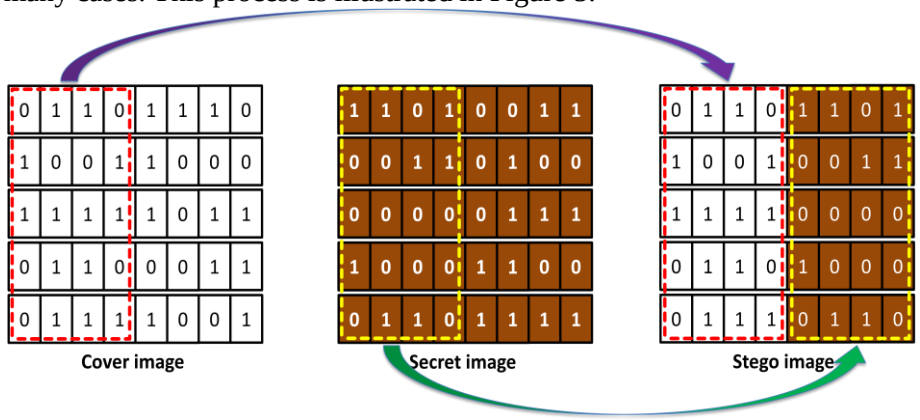

Figure 5: Image Downgrading and Covert Channels

E. Cover-regions and Parity Bits

A stego-key is used as the seed to generate pseudorandom sequence of disjoint cover-regions. Cover-region length is calculated by dividing the cover image length with the message length. Only one bit of the secret message is stored in a whole cover-region rather than in a single element. In the embedding process, disjoint cover-regions are selected, each encoding one secret bit in the parity bit. A parity bit of a region can be calculated by counting total number of '1's and modulo the value with two. One LSB of a random chosen cover-element is flipped if the parity bit of the cover-region does not match with the secret bit to encode. The parity bits of all the selected cover-regions are calculated and lined up to reconstruct the message at the receiver (Bandyopadhyay and Banik, 2012). This process is illustrated in Figure 6.

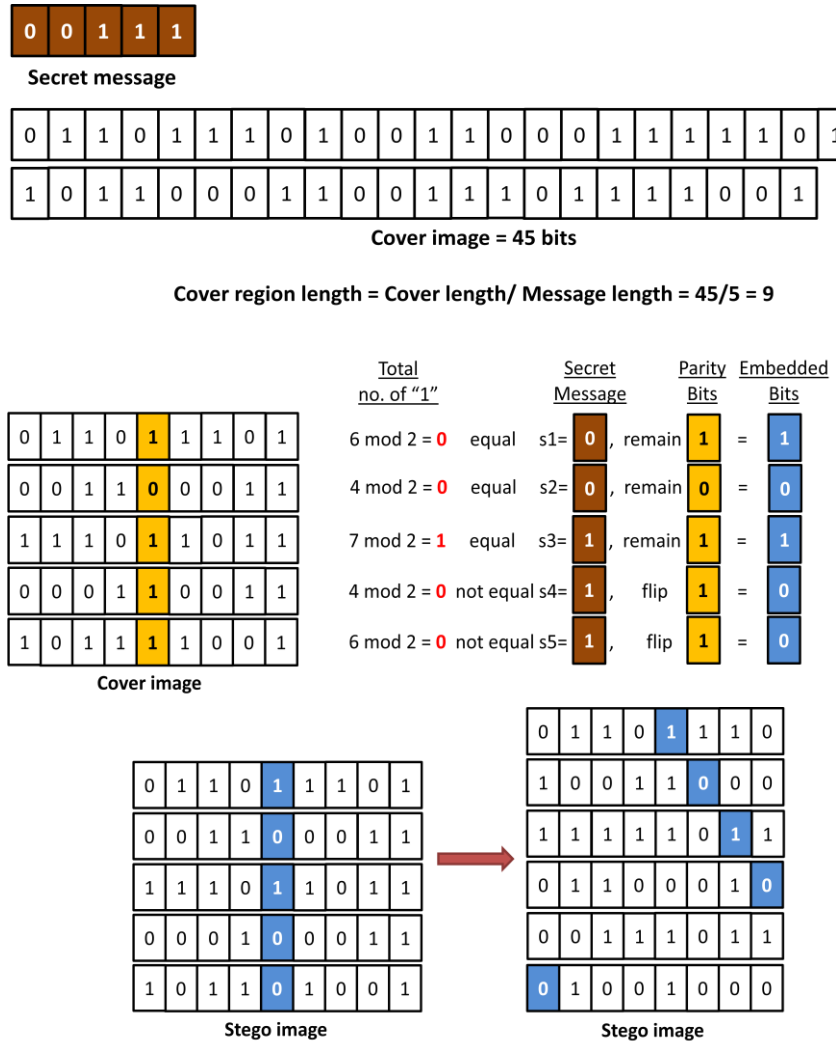

Figure 6: Cover-regions and Parity Bits

3. NEW TESTING TECHNIQUES

These testing techniques are based on new techniques for Strict Avalanche Criterion (SAC) that was proposed by Phyu Phyu Mar and Khin Maung Latt (Mar and Latt, 2008). The techniques were previously used to evaluate the strength of substitution box (S-Box) which is a component in cryptographic algorithm. Since this paper is analyzing the distribution of

secret message bits of substitution system methods, these techniques are suitable to be used in this analysis. The proposed techniques highlighted three main criteria which are avalanche effect, completeness, and strong function. Definition of each criteria are described as follows:-

- Avalanche effect; a function exhibits the avalanche effect if and only if an average of one half of the output bits change whenever a single input bit is complemented.
- Completeness; a function is complete if and only if each output bit depends on all of the input bits. Thus, if it is possible to find the simplest Boolean expression for each output bit in terms of the input bits, each of these expressions would have to contain all of the input bits if the function is complete.
- Strong function; a function is a strong function if and only if each of its output bits should change with a probability of one half whenever a single input bit is complemented.

This new testing techniques are simpler and easier because the existing techniques (Li and Cusick, 2007) use mathematical equations and requires test to be repeated. By using only simple calculation, the result can easily be evaluated from representation of bar graphs as it indicates whether the testing substitution system methods are good or poor.

4. III. EXPERIMENTAL SETUP

This research paper considers five substitution system algorithms (methods) to be tested namely Least Significant Bit Substitution, Random Interval, Pseudorandom Permutation, Image Downgrading & Covert Channels, and Cover-regions & Parity Bits. Each algorithm is tested by using 100 samples containing 10,000 bit message per sample. Outputs from each algorithm are analyzed using the new testing techniques proposed for Strict Avalanche Criterion.

A. Frequency Analysis Of Various Hamming Weight (Avalanche Effect)

Output values of the algorithm which correspond to two inputs were chosen. Apply XOR function to compute the differential value of these two

outputs and find the hamming weight in the differential value. For necessary count of testing, repeat above steps. Analyze the frequency of various differential values by counting the number of '1's in each output (byte). This method is to observe total number of bit changes in each output. This process is illustrated in Figure 7.

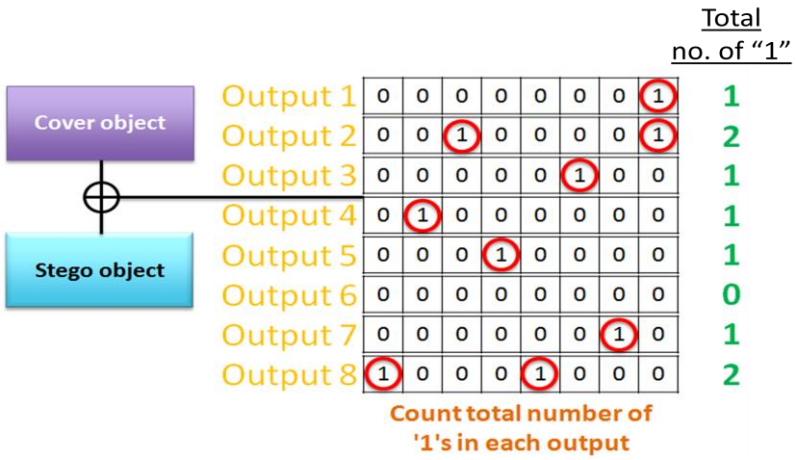

Figure 7: Avalanche Effect

B. Frequency Analysis Of Various Differential Value (Completeness)

First, two inputs with their corresponding output values of the algorithm were chosen. Next, the differential value of these two outputs was computed by applying XOR function. Then, the hamming weight in the differential byte value in decimal of the outputs (byte) was determined. Above steps were repeated for necessary count of testing. The frequencies of various differential values were analyzed by counting the total number of occurrence for each output. The objective of this method is to observe the byte value in decimal of each output. This process is illustrated in Figure 8.

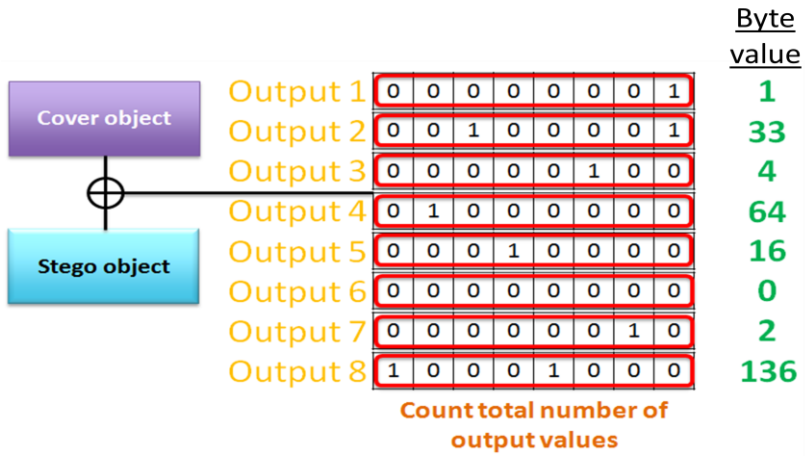

Figure 8: Completeness

C. Analysis Of Hamming Weight According To The Bit Position (Strong Function)

Choose two inputs and find their corresponding output value of the algorithm. Compute the differential value of these two outputs by applying XOR function. Repeat above steps for necessary count of testing. Analyze the hamming weight according to the bit position of resulting differential values by counting total number of '1's in each bit position (row). This method is to observe total number of bit changes in each bit position. This process is illustrated in Figure 9.

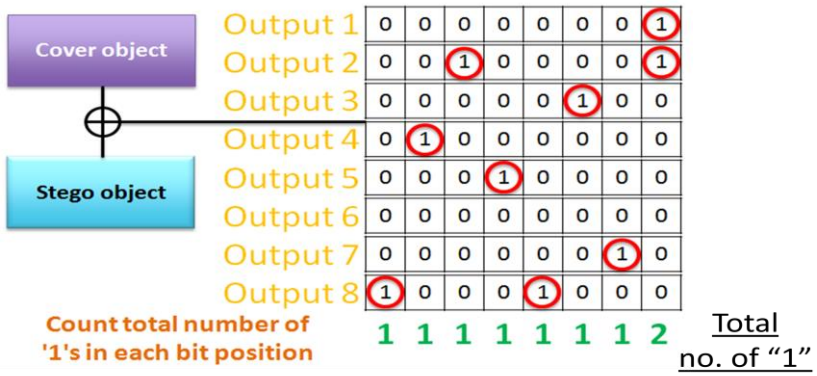

Figure 9: Strong Function

5. IV. RESULTS AND ANALYSIS

This section describes the results and analysis carried out on the three tests mentioned above namely Avalanche Effect, Completeness, and Strong Function. The objectives of these tests are to observe total number of bit changes in each output, to observe the byte value in decimal of each output, and to observe total number of bit changes in each bit position. All of the test results are displayed in the form of bar graph to facilitate the reader.

A. Frequency Analysis Of Various Hamming Weight (Avalanche Effect)

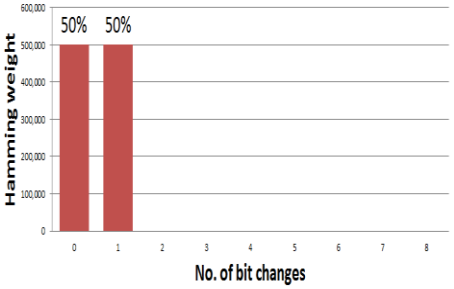

Figure 10: Least Significant Bit Substitution

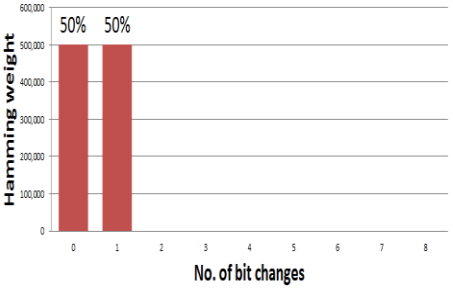

Figure 11: Random Interval

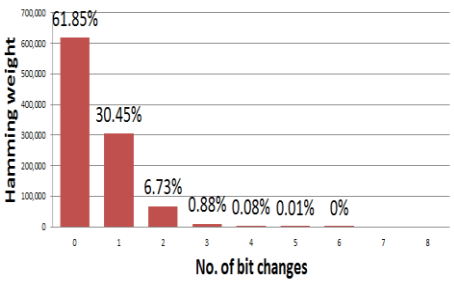

Figure 12: Pseudorandom Permutation

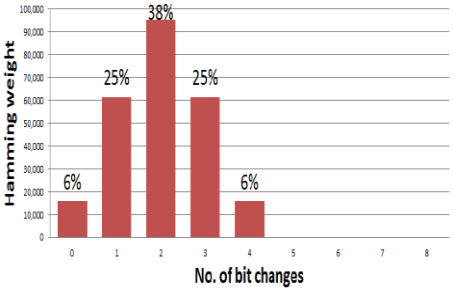

Figure 13: Image Downgrading & Covert Channels

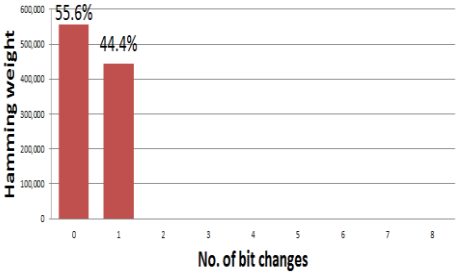

Figure 14: Cover-regions & Parity Bits

Test results for avalanche effect indicate that Least Significant Bit Substitution, Random Interval, and Cover-regions & Parity Bits methods changed up to one bit output, whereas Image Downgrading & Covert Channels method changed up to 4 bits output. Pseudorandom Permutation method changed up to 6 bits output. It is important to note that more bit changes will increase the suspicious level of the existence of secret messages. These results are referred from Figure 10,11,12,13 and 14.

B. Frequency Analysis Of Various Differential Value (Completeness)

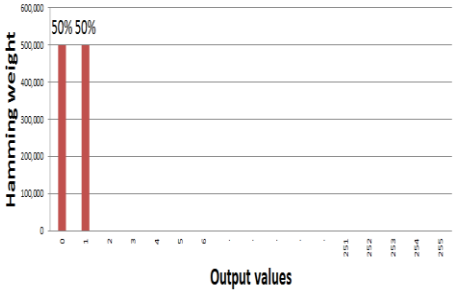

Figure 15: Least Significant Bit Substitution

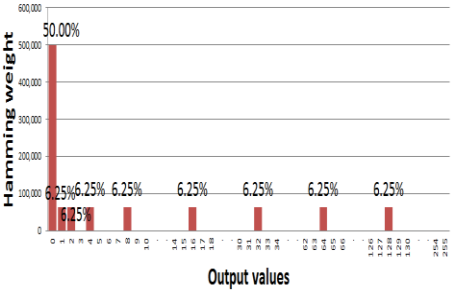

Figure 16: Random Interval

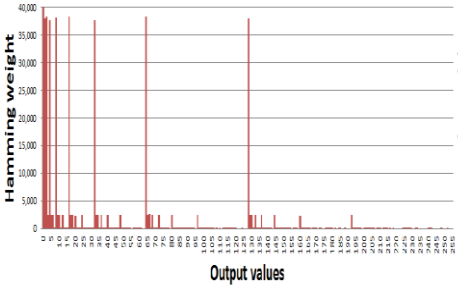

Figure 17: Pseudorandom Permutation

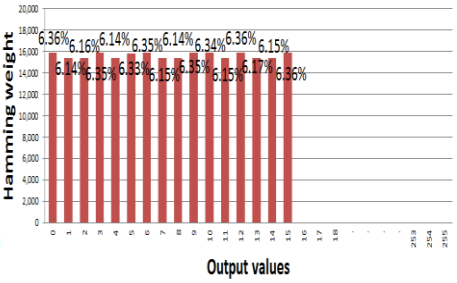

Figure 18: Image Downgrading & Covert Channels

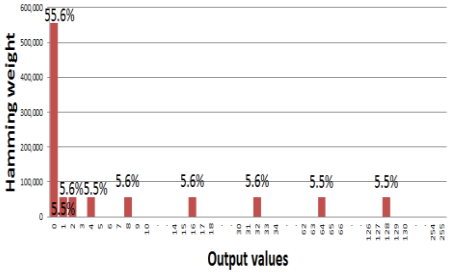

Figure 19: Cover-regions & Parity Bits

Completeness test results are showed in Figure 15,16,17,18 and 19. Least Significant Bit Substitution method produced 2 output values. Random Interval and Cover-regions & Parity Bits methods produced 9 output values. Image Downgrading & Covert Channels method produced 16 output values. Pseudorandom Permutation method produces various output values. Lesser output values will increase the ability to guess the secret message bit position.

C. Analysis Of Hamming Weight According To The Bit Position (Strong Function)

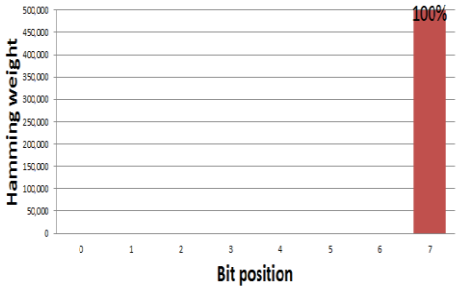

Figure 20: Least Significant Bit Substitution

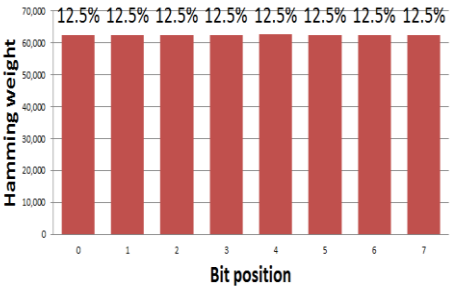

Figure 21: Random Interval

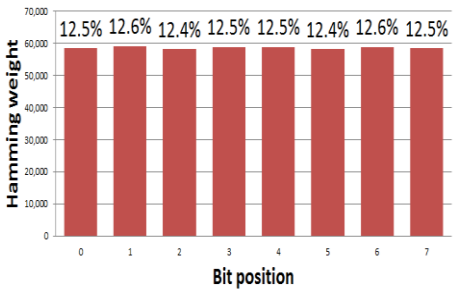

Figure 22: Pseudorandom Permutation

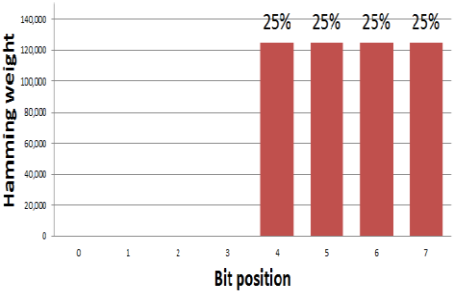

Figure 23: Image Downgrading & Covert Channels

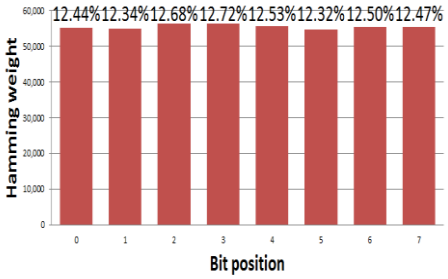

Figure 24: Cover-regions & Parity Bits

In the strong function test, results show that Least Significant Bit Substitution method embeds secret messages in 7th position. Image Downgrading & Covert Channels method embeds secret messages in 4th, 5th, 6th and 7th position. Random Interval, Pseudorandom Permutation, and Cover-regions & Parity Bits methods embed secret messages in all positions. Lesser bit position embeds will increase the ability to guess the secret message bit position. The results are showed in Figure 20,21,22,23 and 24.

6. CONCLUSIONS AND FUTURE WORKS

This research paper presented the analysis on steganography substitution methods using new testing techniques proposed for strict avalanche criterion. Overall results of the analysis are shown in Table 1. All methods are analyzed in detail to determine the best method to be implemented. Random Interval and Cover-regions & Parity Bits methods seem to be the two best methods out of the five substitution system methods because both produced good results in all three tests. Results of this analysis can be used to change or improve current embedding methods. Therefore an important consideration in steganography is to have a good randomization method which could improve the security. Motivated by the results derived, we will continue on this research to propose a better substitution system method which is not susceptible to steganalysis.

Frequency Analysis Of Various Hamming Weight (Avalanche Effect)	Methods	Total number of bit changes in each output							
		0	1	2	3	4	5	6	7	8
	Least Significant Bit Substitution	50%	50%	-	-	-	-	-	-	-
	Random Interval	50%	50%	-	-	-	-	-	-	-
	Pseudorandom Permutation	61.85%	30.45%	6.73%	0.88%	0.08%	0.01%	0.00%	-	-
	Image Downgrading & Covert Channels	6%	25%	38%	25%	6%	-	-	-	-
	Cover-regions & Parity Bits	55.6%	44.4%	-	-	-	-	-	-	-
Frequency Analysis Of Various Differential Value (Completeness)	Methods	Output values								
	Least Significant Bit Substitution	0 / 1								
	Random Interval	0 / 1 / 2 / 4 / 8 / 16 / 32 / 64 / 128								
	Pseudorandom Permutation	Various output values								
	Image Downgrading & Covert Channels	0 / 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10 / 11 / 12 / 13 / 14 / 15								
	Cover-regions & Parity Bits	0 / 1 / 2 / 4 / 8 / 16 / 32 / 64 / 128								
Analysis Of Hamming Weight According To The Bit Position (Strong Function)	Methods	Total number of bit changes in each bit position								
		0	1	2	3	4	5	6	7	
	Least Significant Bit Substitution	-	-	-	-	-	-	-	100%	
	Random Interval	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	
	Pseudorandom Permutation	12.5%	12.6%	12.4%	12.5%	12.5%	12.4%	12.6%	12.5%	
	Image Downgrading & Covert Channels	-	-	-	-	25%	25%	25%	25%	
	Cover-regions & Parity Bits	12.44%	12.34%	12.68%	12.72%	12.53%	12.32%	12.50%	12.47%	

Table 1: Overall results

7. ACKNOWLEDGMENT

The main author would like to acknowledge the effort of team members in producing this paper. Not forgetting Cyber Technology Research Department of CyberSecurity Malaysia for support and encouragement. A special thanks is also due to Dr Tuan Sabri b. Tuan Mat for introducing cryptography to the main author and guidance throughout involvement in this field.

REFERENCES

Al-Ani Z. K., Zaidan A. A., Zaidan B. B. and Alanazi H. O. 2010. Overview: Main fundamentals for Steganography. *Journal of Computer*. Vol. 2, No. 3: 158-165.

Bandyopadhyay S.K. and Banik B.G. 2012. Multi-Level Steganographic Algorithm for Audio Steganography using LSB Modification and Parity Encoding Technique. *International Journal of Emerging Trend & Technology in Computer Science (IJETTCS).* Vol. 1, Issue 2: 71-74.

Cheddad A., Condell J., Curran K. and McKevitt P. October 2007. A Comparative Analysis of Steganographic Tools. Proceedings of *The 7th Information Technology and Telecommuting Conference IT&T 2007*

Fridrich J. and Du R. 2000. Secure Steganographic Methods for Palette Images. *Pfitzmann A. (ed.): 2nd International Workshop on Information Hiding. Lecture Notes in Computer Science*. Vol. 1768: 47–60.

Hossain M.J. January – February 2014. Information Hiding using Image Steganography with Pseudorandom Permutation. *Bangladesh Research Publications Journal.* Vol. 9(3): 215-225.

Johnson and Katzenbeisser. 2000. A survey of Steganographic techniques. *Information hiding Techniques for Steganography and Digital Watermarking.* 43-78.

Leivaditis M. 2010. Statistical Steganalysis. *MSc thesis, University of Surrey.*

Li Y. and Cusick T.W. 2007. Strict Avalanche Criterion over Finite Fields. *J. Math. Crypt*. Vol. 1: 65-78.

Mar P.P. and Latt K.M. 2008. New analysis methods on strict avalanche criterion of Sboxes. *World Academy of Science, Engineering and Technology.* 48: 150-154.

Rodrigues J.M., Rios J.R. and Puech W. April 2004. SSB-4 System of Steganography using Bit. 5th *International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004).*

Tariq A., Falih E. and Shaker E. 2013. A New Approach for Hiding Data within Executable Computer Program Files Using an Improvement Cover Region. *Iraq Journal of Computers, Communication, Control and Systems Engineering (IJCCCE).* Vol. 13, No.1, 2013.

Zaidan B.B., Zaidan A.A., Taqa A. and Othman F. December 2009. Stego-Image Vs Stego-Analysis System. *International Journal of Computer and Electrical Engineering(IJCEE).* Vol. 1, No. 5: 572-578.