IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 24, No. 2, April, 2019, pp. 161-165
Journal homepage: http://iieta.org/journals/isi

Extended Statistical Analysis on Multimedia Concealed Data Detections

Rupa Chiramdasu

V R Siddhartha Engineering College, Vijayawada, India

Corresponding Author Email: rupamtech@gmail.com

https://doi.org/10.18280/isi.240205

Received: 4 January 2019
Accepted: 10 March 2019

Keywords:
statistical analysis, classifier, extended statistical analysis, RS analysis, filter groups

ABSTRACT

Now- a- days, providing privacy and protection to illicit materials like videos and web pages became a major issue to the law enforcement authorities. Multimedia information transactions over external network became a threat as those files consist of stego-payload. Current new era is, cyber wars, also playing a vital role along with the physical wars in the world. Along with the people, Government organizations also want to maintain data secrecy in certain sensitive communication areas like Department of Military, Department of Air force, and other defense related areas to protect the data against opponent parties. One of the technique to protect the information steganograpgy. The art of hiding the data into the multimedia files as carriers referred as steganography. In order to this, identifying messages by steganography referred as steganalysis. These techniques can possible to use by unauthorized persons along with the authorized persons or organizations. Hence design and development of an application whether the selected multimedia holds any payload or not is highly recommended, in this new era. This paper work shows the functioning of various methods to predict the hidden data in the multimedia files. Performance analysis of the steganalysis methods by considering different key parameters is the main strength of this work.

1. INTRODUCTION

In the digital world, design and development and utilization of digital applications are being increase day by day. These applications can add serenity and easy connectivity in human life every activity [1]. By considering and following the security policies, standards and compliances in the phases of implementation of digital application can reduce the threats, attacks and data leakages by the eves. In many cases, sensitive data shared among various stakeholders such as employees working outside from organization's premises. This increases the risks ratio that confidential information will fall into unauthorized persons. They can misuse the data and exploit the system vulnerabilities by entering into the system through embed the bots or malwares into the sharing data (Multimedia). These risks can predict and detect by using advanced and Intelligent security measures, standard security measures, access control, designated attacks prevention and detection system tools, Cryptography/Steganography analysis. Cryptography transforms the plaintext into cipher text steganography hides the plain text into a carrier files like image, audio, video and text file [3].

In today's world internet has become the basic need for living. In such living for researcher's Steganography challenges and its applications has become a hot topic. From the resources of internet, steganography searched for 12 times in 1995, 500 times in 1996 and 1,000 times in 1998. Our Google search engine says that the keyword steganography is searched for 2,200,000 times [2]. Internet is the one which is making the technology much easier to use and hide the data in images. Some of the researchers proposed an attack which can based statistical as well as some are focused on types of data formats [3]. This statistical attack can apply on any fixed set of Pairs of Values (PoVs) based steganography technique. This technique calculates the Probability Distribution Function (PDF) and Chi-Square test. Probability Distribution Function uses the integration function from 0 to chi-square value [4]. Chi-square is the ratio of the difference between the expected frequencies to the measured frequencies to the square of the measured frequency with k-1 degree. Statistical analysis in stego analysis will evaluates by considering the input image and cover image as parameters. Let us denote PDF as shown in equation 1 [5].

$$P = 1- [(1/2\ \Gamma\ (2k-1)) – (ex0\ (k-1)] \quad (1)$$

RS Steganalysis is the other new steganographic method which is used to estimate the length of the message embedded into the Stego-image. By the previous technique i.e., Statistical analysis we can detect whether the image is Steg image or not but not the length of the message which is suspected to be embedded into it. RS technique estimate the length of the message embedded using four curves. The image is divided into several blocks. And then the flipping concept is applied. Positive flipping, negative flipping and zero flipping [5]. With the help of relative frequencies of these groups the image obtained from the original image of the LSB flipped and randomized and then predict eh level of embedding. Average flipping percentage is calculated, and the error is obtained. Error is maintained 0.5 % to 1 % till the end of the message i.e., Detected Probability Error (PE)=0.5 % - 1 % [6]. To further develop the detection, rate a new technique has been introduced i.e., Enriched Statistical Analysis (ESA). This technique overcomes the disadvantage of Statistical analysis.

The speediness of detection also increases.

Rest of the paper consists of like section 2 describes the details on prerequisites to the work. Existing techniques such as Statistical Analysis and RS steganalysis discussed in Section 3 and Section 4. Section 5 covers the methodology of Extended Statistical analysis. Performance analysis by considering key factors were covered in Section 6.

2. PRELIMINARIES

Lai et al. introduced different approaches for sharing images invertible. In this work the images were loss-less and distorted [2]. This author used notational system for secret pixels. If the array of the notational changes then by calculating threshold value we can say it as a stego image.

Wang et al. has developed a method that can able to performs identification of complex computation of sharing of secret images [8]. The author designed deterministic and probability schemes for performing on gray scale images.

In 2012 and 2015, released an article on how terrorists have been using steganography. They were used this method to transmit data through porn videos, Reddit, eBay which hides 141 files [8].

Forensic of multimedia files important for detecting the hidden payload. Some research has been done on this as motivation. In World War II, microdot technology has used to hide the sensitive data. In this method, image acted as cover medium and period was the size of that medium [9].

Chen et al. proposed a method by using DWT based securing the secret images [9]. Comparatively this technique has more tolerance over the data leakage and data loss and corruption then other image protection methods. Here some functions have performed priory that arithmetic and Huffman coding calculations.

An attack has happened on Instegogram and Apple Mac OS X in December 2016. Attacker used Image steganography for this attack. Stealth Malware Ops used for this attack. In 2017, McAfee who has working on against host based attack has identified that attacker utilization of steganography increased. Attackers used this method to pass malicious data through security protection system without detection [10].

So many methods have been proposing by the researcher for secret sharing of data. Some of the related methods discussed here.

Neha saxena in this paper the author discusses about the attacks on RS Stegaanalysis added some of the concepts related to data hiding in digital images [11]. This author explains how steganography is used for hiding the data in digital image. Main limitation of this work is stego analysis part has missed. It extends to that some other methods also using to enhance the capacity of payload embedding into images like LSB method as well as attacks on stegaanalysis [12-13].

3. STATISTICAL ANALYSIS

Pfitzman and Westfeld [14] proposed an attack which can able to on any type of steganographic method. This method designed and developed based on statistical preliminaries where Pair of Values (PoVs) fixed set are used in the evaluation process. At the time of embedding a message into the image these PoVs flipped with each other due to their flipping nature. Generally, every image palette consists of 8 bits, which is generated 256 colors powered with 2 i.e., 2^8 = 256. After embedding the message into the image, each pair has distinguished with a tendency to match the length of the message. In carrier image, two values from each pair can distributed unevenly. In this process, exchanging of one value with the other does not affects the sum of the occurrence of the colors in the image. Measured frequency for the original image and Expected frequency for the cover image is calculated.

The Statistical Analysis technique now applied to the original image and the cover image. This technique can calculate Chi–squared test (CST) and Probability Distribution Function (PDF). Depends on PDF values data can be analyzed to detect whether the forensic file is stego file or not.

Definition 2.1: [Chi-squared Test]. The ratio of the difference between the Measured Frequency and Occurred Frequency

The Chi – Square Statistics is denoted as x^2

$$\chi^2_{p-1} = \sum_{j=1}^{p} \frac{(n_j - n_j^1)^2}{n_j^1} \tag{2}$$

where $n_j^1 = \frac{n_j + m_j}{2}$

where nj = frequency of 2k

mj = frequency of 2k+1, 0 < k < 127

Definition 2.2: [Probability Distribution Function]. It refers that probability of embedding evaluated by integrating with density function and chi-squared upper limit.

$$P = 1 - \frac{1}{2k - \frac{1}{2}\,\Gamma\,(k - 1/2)} \int_0^{\chi 2} e^{-x/2}\, x^{\frac{k-1}{2}-1} dx \tag{3}$$

4. RS STEGANALYSIS

Image pixels categorized as two clusters such as Singular Groups (SG) and Regular Groups (RG). These groups depend upon the image properties. Basically, the image divides into several blocks for flipping. Positive flipping (f), negative flipping (f -1) and zero flipping (f0) can applied to the database. The difference between the positive and negative regular groups (RG and RG-1) should be equal to the difference between the positive and negative singular groups (SG and SG-1).

$$RG \cong RG_{-1} \text{ and } SG \cong SG_{-1} \tag{4}$$

The RS Steganalysis will consider only stego images as input and calculate the length of the message embedded into it that is depending on curves and minimum detected PE value [11], [15].

$$PE = \min PFA \tfrac{1}{2} (PFA + PMD (PFA)) \tag{5}$$

where,

PFA=Probability of False Alarms,

PMD=Probabilities of Missed Alarms.

5. EXTENDED STATISTICAL ANALYSIS

Extended Statistical analysis can reduce the limitations of the statistical analysis technique as well as in RS analysis. Extended Statistical Analysis announces which image is stego image or frame and which image is not a stego image or frame by considering given images or videos as input. In this method, Inverse Integration operation applies to the Chi-Square test to detect the stego data. The statistical function of this is:

$$P = Q\ (f, 0, xu) \tag{6}$$

The equation 6, describes the probability function which is quadruple (Q) to the function with the concern parameters are as input. Where, 'f' is exponential function of the Chi-Square. 'xu' is maintained as 127. The Probability Distribution Function of the extended statistical analysis defined as below:

$$PDF = \int_{-1}^{f} 1 - (1/p) \tag{7}$$

Equation 7, is the inverse integral of the Chi-Square function which varies from -1 to chi-square value which can be maintained 'k' as (k-1) degree. It effects on the accuracy of performance of stego detection as well as reduces the time consumption by less computation operations like Inverse instead of power operator. Extended statistical analysis tested on more than 9000 images and videos which have collected from the standard databases resources like cpntagiodump blog database, BOSS database.

6. RESULTS AND ANALYSIS

Figure 1 shows that performance of various stegoanalysis methods. These results have evaluated by considering various statistical analysis methods along with the proposed method extended statistical analysis methods. Performance of the extended statistical analysis method has evaluated by considering different image data bases and video databases, Where the time and accuracy of the method has considered as key parameters.

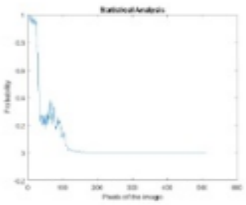
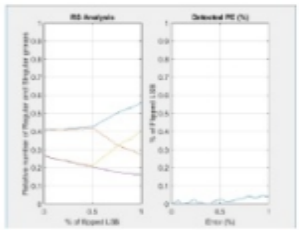
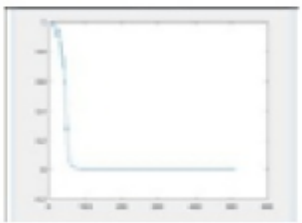
Schemes	Test Images	Stegoanalysis Results
Statistical Analysis	 (a)	 (a). Stego File
RS Analysis	 (b)	 (b). Stego File
Extended Statistical Analysis	 (c)	 (c). Stego File
Extended Statistical Analysis	 (d)	 (d). Non Stego File

Figure 1. Stego analysis on image data

Figure 1.a is showing that detection of stego image from a chosen database through statistical analysis method. Stego image detected using RS stegoanalysis method result is shown as Figure 1.b. Stego image detection by extended statistical analysis method result is shown as Figure 1.c. As well as, Figure 1. d shows that the finding of non stego image result from a chosen database using Extended statistical analysis. If we can observe that non stego images pixel frequency variants positions differ with the pixel frequency variants positions of stego images.

Video integrity performance has shown in Figure 2 and Figure 3. Figure 2 represents that an input video which can considered for testing whether the video holds stego payload or not. Fig 3 shows that frame by frame verification results by extended statistical analysis method. These graphs represent the variations among the stego frames and non stego frames. On comparison of these techniques can observe that ESA technique computation complexity is reducing by using the inversion integration operations with dot operator instead modulus and summation.

Figure 2. Input video for ESA based stego analysis

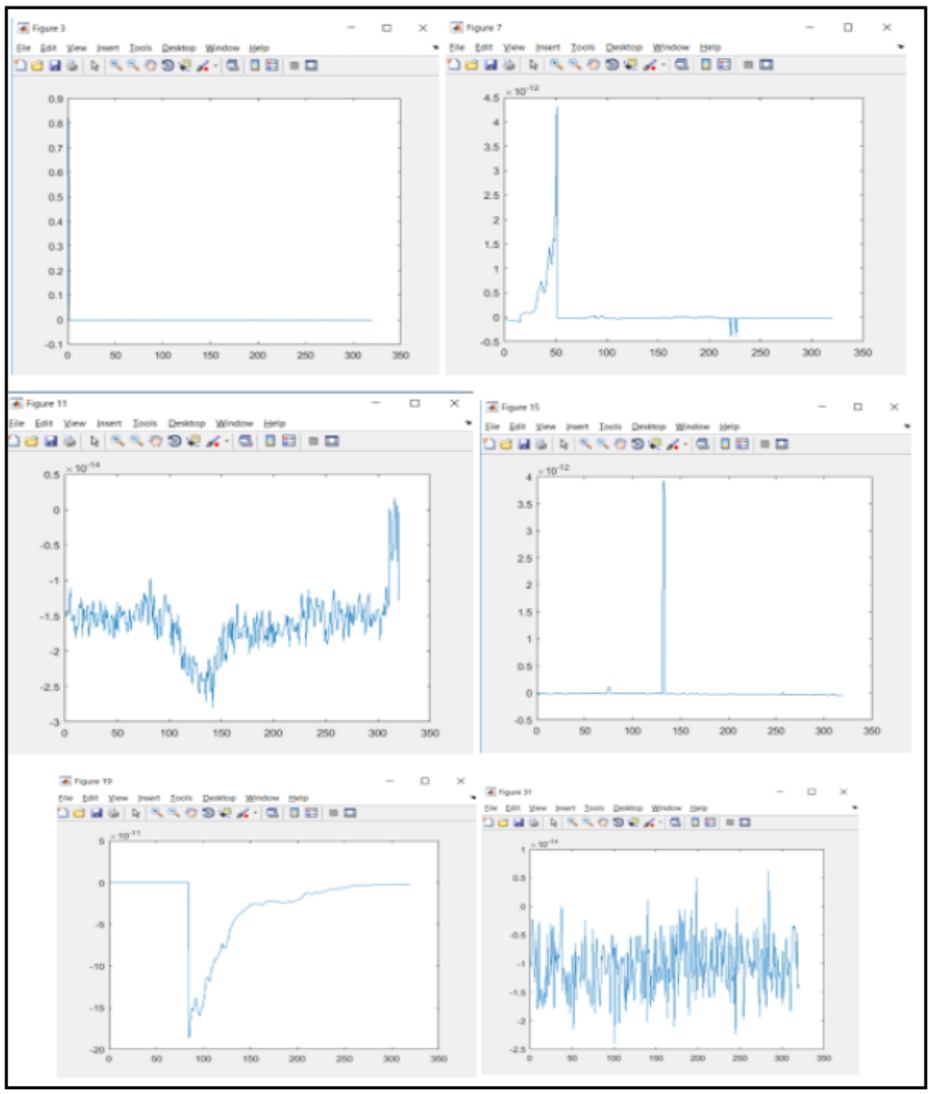


Figure 3. Frame by Frame Stego analysis of Video data

The performance results of Statistical Analysis (SA), RS stegoanalysis and Extended Statistical Analysis (ESA) have shown in Table 1 and Table 2. Reconnainece the standard data from different sources like Koggle, BOSS and cpntagiodump blog for integrity checking of the existed methods and proposed method. Here we considered more than 9000 images with different file formats as input. Table 1 shows the results analysis of statistical and RS stego analysis and Table 2 shows the result of extended statistical analysis.

Table 1. Performance anlysis of SA & RS

Input Size	Statistical Analysis (SA)		RS Steganalysis	
	Time (Sec)	Accuracy (%)	Time (Sec)	Accuracy (%)
253 K	0.01	0.65	0.01	0.65
30 M	1.01	0.65	1.40	0.65
130 M	8.45	0.65	10.45	0.65
200 M	13.13	0.65	12.65	0.65
340 M	32.02	0.65	17.50	0.65
600 M	63.57	0.65	18.56	0.65
1.73 M	132.13	0.65	21.06	0.65
2.55 G	429.47	0.65	26.65	0.65

Table 2. Performance Anlysis of ESA

Input Size	Extended Statistical Analysis (ESA)	
	Time(Sec)	Accuracy (%)
253 KB	0.01	0.4
30 MB	1.11	0.4
130 MB	5.21	0.4
200 MB	9.32	0.4
340 MB	13.0	0.4
600 MB	63.57	0.4
2.55 GB	28.54	0.4

7. CONCLUSIONS

Analyzing Forensic Multimedia is a crucial in this digitalization world due to increasing of innovative attacks and threats day by day. Forensic crime rate has been increasing rapidly. To reduce these kinds of attacks, as a first phase, needs to create awareness about various cybercrimes and their exploitation possibilities to the users of applications. Later, by following approved innovative standards and compliances in the development also can reduce victim ratio. One more approach to minimize the attack rate is adaptation of advanced technology tools in the digital applications. These can useful for prediction of threat, detection of attack and retrieve/extract the effected data by the unauthorized persons. The developed application is easy to use and is a new approach for detection of effected Multimedia. Further work will carry out on Audios.

REFERENCES

[1] Debbar, F., Ayad, B. (2017). A new steganalysis method to detect information hiding in speech. 13th International Wireless communications and Mobile Computing Conference (IWCMC), IEEE. https://doi.org/10.1109/IWCMC.2017.7986570

[2] Lai, B.,Chang, L. (2006). Adaptive data hiding for images based on HAAR discrete wavelet transform. Lecture Notes in Computer Science, 4319: 1085-1093. https://doi.org/10.1007/11949534_109

[3] Ing, X., Huang, W., Zhang, M., Zhao, I. (2016). A topography structure used in audio steganography. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 2134-2138. https://doi.org/10.1109/ICASSP.2016.7472054

[4] Cheddad, A., Condell, J., Curran, K., Kevitt, M.P. (2010). Digital image steganography: Survey and analyses of current methods. Signal Processing, 90(3). https://doi.org/10.1016/j.sigpro.2009.08.010

[5] Anusha, P., Medhane, S.P. (2017). Trapping of stego images on the basis of statistical evidences. International Journal of Advanced Research in Computer Science, 8(5): 2003-2006. https://doi.org/10.26483/ijarcs.v8i5.3681

[6] Thomas, P. (2013). Literature survey of various steganography. International Journal of Engineering Research & Technology (IJERT), 2(5). http://dx.doi.org/10.17577/IJERTV7IS02000

[7] Curran, K., Devitt, J.M. (2008). Image analysis for online dynamic steganography detection. Computer and Information Science, 1(3): 32-41. https://doi.org/10.5539/cis.v1n3p32

[8] Wang, R.Z., Su, C.H., (2006). Secret image sharing with smaller shadow images. Pattern Recognition Lett., 27(6): 55l–555. https://doi.org/10.1016/j.patrec.2005. 09. 021

[9] Chen, P., Lin, H., (2006). A DWT based approach for image steganography. International Journal of Applied Science and Engineering, 4(3): 275-290. https://doi.org/10.15439/2016F521

[10] Aljamea, M., Athar, T., Iliopoulos, C., Msamiruzzaman, (2017). Detection of hidden encrypted URL in image steganography. The Ninth International Conferences on Pervasive Patterns and Applications, Patterns, pp. 3-8. https://doi.org/10.1145/2896387.2896408

[11] Saxena, N. (2017). Steganography scheme against RS attack enriched with evolutionary programming (AGA) and OPAP. International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), 3(9): 64-68.

[12] Swadhin, K., Chirag, (2017). Video steganography using encrypted payload for satellite communication. IEEE Conferences Aerospace Conference. https://doi.org/10.1109/AERO.2017.7943978

[13] Kasapbasi M.C., Elmasry, W. (2018). New LSB-based colour image steganography method to enhance the efficiency in payload capacity, security and integrity check. Journal of Indian academy of sciences, 43(68): 1-14. https://doi.org/10.1007/s1204

[14] Westfeld, A., Pfitzmann, A. (1999). Attacks on steganographic systems. Proceedings of Third International Workshop on Information Hiding, pp. 61-76. https://doi.org/10.1007/10719724_5

[15] Djebbar, F., Ayad, B., (2017). A new steganalysis method to detect information hiding in speech. IEEE International Wireless Communications and Mobile Computing Conference (IWCMC). https://doi.org/10.1109/IWCMC.2017.7986570