International Journal of Digital Crime and Forensics
Volume 13 • Issue 5 • September-October 2021

Multimedia Concealed Data Detection Using Quantitative Steganalysis

Rupa Ch., V. R. Siddhartha Engineering College (Autonomous), Vijayawada, India

Sumaiya Shaikh, V. R. Siddhartha Engineering College (Autonomous), Vijayawada, India

Mukesh Chinta, V. R. Siddhartha Engineering College (Autonomous), Vijayawada, India

ABSTRACT

In current days, there is a constant evolution in modern technology. The most predominant usage of technology by society is the internet. There are many ways and means on the internet through which data is transmitted. Having such rapid and fast growth of communicating media also increases the exposure to security threats, causing unintellectual information ingress. Steganography is the main aspect of communicating in an aspect that hides the extent of communication. Steganalysis is another essential concern in data concealing, which is the art of identifying the existence of steganography. A framework has been designed to identify the concealed data in the multimedia file in the proposed system. This work's main strength is analyzing concealed data images without embedding and extracting the image's payloads. A quantitative steganalysis approach was considered to accomplish the proposed objective. By using this approach, the results were achieved with 98% accuracy.

KEYWORDS

Classifier, Concealed Data, Images, Steganalysis, Steganography

1. INTRODUCTION

Exchanging data now-a-days is not a complex task. But exchanging the same data without the interruption of the third party is a critical task. Two most powerful and popular techniques to achieve that are introduced based on the concept of secret writing. They are Steganography and Cryptography. Mostly Steganography concepts has been used for hiding the data in text files or images or videos or audio in-short via Multimedia. Cryptographic tools are developed for exchanging concealed data without any interruption (Alazab, Shalaginoy, Mesleh, Awajan, 2020). At the same time, Steganography tools are developed to maintain the secrecy of the concealed data in Multimedia.

The main indicators of the steganographic system are Capacity, Security and Interoperability. These parameters are interdependent means that a parameter value effects the remaining values (Xintao, Daidou, Nao, Boxia, Mengxiao, Chuan, 2020)Steganography means "Covered Writing" in Greek. Steganography is the complement of Cryptography and its goal is to hide the presence of a message and to create a covert channel (Chunfang, Fenlin, Shuangkui, Jicang, Junwei, 2019). A prominent steganography instance is simmon's prisoner's problem (Feng, Zhang, Dawei, Zhanyang, Shaohua, Lianyong, 2019). They were considering the famous example of Bob and Alice. Depending on the model proposed in this paper, Bob is the sender and Alice is the receiver. Bob is willing to share some concealed data with Alice but not in a common way. He used this Steganography technique and hide the data in an image and transmitted to the Alice. At the receiver end Alice knowing about the technique can easily extract the data hidden in the image. Generally, any unauthorised persons (intruders) without prior knowledge of the technique used by the authorised persons can only download

DOI: 10.4018/IJDCF.20210901.oa6

the image but will not be able to extract the hidden information. Attackers are using some of these techniques to add the dangerous malicious data to disrupt the target systems or individuals (Sairam, Boopathybagan, 2019).

The main goal of steganography is defeating the identification of hidden data. However, this approach became a victim of several passive and active attacks. Later, people have used this method to do illegal activities like cyberwars, cyber-attacks, etc(Rupa, Thippa, M.H, 2020). Day by day, this approach's utilization is increasing by the industry and government for sharing the sensitive data among the authorized partners. Statistical and quantitative steganalysis have often used to identify the data within the multimedia (Rupa, 2020). Generally, steganography uses to conceal the data by authorized persons and deceive the data by the intended persons only. It is a type of covert communication and also helpful to work on complex communication models. It itself has some strengths and weaknesses (Alazab, Venkataraman, Watters, Alazab, 2013).

In olden days, at the sender side a technique involving writing crypto messages on an empty paper with wax making it invisible is used. Later at receiver side, heat is applied to the paper to extract the secret message. This technique is named as Secret Writing (Silman, 2001, Neha, 2015). During Second World War, a famous king used Micro dot Technology to transfer large amount of data. The dots' spaces are considered periods, and the crypted dot is the message (Cole, 2001) (Rupa, 2017). These incidents are motivated to develop an application system that is used to scan multiple images and displays whether the images have any hidden information. This type of applications will be best suited for forensic analysis (Imam, Vassilakis, 2019) (Samiksha, Mahip, Bartere, 2015).

In recent years, there have been reports about the use of Steganography in espionage, terrorist attacks, crimes and other activities. In 2001, some mainstream media in the Unites States, such as CNN and US Today, reported the news of secret communications between Al Qaeda members using Steganography. Reported that Bin Laden gang will attack the target map, pass instructions and other information hidden in pornography, sports chat and other websites. According to Die Zeit, an al-Qaida suspect was arrested in Berlin in May 2011 and police found him carrying a memory card. Later, after being cracked by experts in charge of computer criminal investigation in the German Federal Criminal Police Bureau, it was found that on the surface, only one pornographic video named "KickAss" appeared on the card, but in fact, 141 text documents were hidden in the video, including a large number of Al Qaeda action reports, future plans and so on. In June 2010, the Federal Bureau of Investigating (FBI) successfully arrested 10 Russian agents in New Jersey, which caused a great diplomatic shock between the United States and Russia.

This paper deals with more Quality and Quantitative analysis as it loads bulk number of images as input irrespective of its format and size. Then the powerful Quantitative Steganalysis is performed on all the images at once and graphical as well as textual outputs are obtained in the form of result. The first section of the paper deals with the introduction part explaining the true meaning of Steganography and Steganalysis and differences between them are also highlighted. The Section 2 explains the history research on Steganography and Steganalysis where many authors like (Cohen, 2020), (Fridrich, 2010) proposes many techniques and methodologies in this research area. The Section 3 deals with the methodology which we proposed i.e., Quantitative Steganalysis which on comparison gives better performance than the existing methodologies. This paper continues with explanation of the methodology used and conclusion carrying the future work scope.

2. LITERATURE SURVEY

(Rupa, 2016) presented a powerful Steganography system that uses merging scientific techniques and algorithms which will be efficient for communication. This systematic way uses RSA cryptography method for secret data encryption. In this paper, cryptography and steganography have been used to detect the message they have incorporated. First, they use some techniques to hide the message and apply the application of detecting the hidden information. Cryptography is used to encrypt the data which is to be converted plaintext into ciphertext. That ciphertext embedded into an image. After extracting the embedded data from the images at receiver side, he has applied decryption technique to decrypt the ciphertext. Then receiver can able to read the original message.

(Cohen, Nissima, Elovici, 2020) proposed a scanning application that detects malicious data in all image formats. These malicious programs can be very dangerous because they can steal all the data possessed in the laptop and charge you to relieve the content present in the laptop. Such kind of images will be harmful in many ways. So, Robert T. R, proposes such techniques that can easily scan and detect the images which are downloaded from the web browser. So many researchers carried their work in identifying the concealed data in images using various techniques like DWT based securing the secret images, micro dot technology used in olden ages mostly in World War II and many Steganalysis techniques.

(Fridrich, Golijian, Rui, 2010) proposes a technique named RS Steganalysis with the concepts of hiding the data in the images and retrieving the same data using some measures. But the author's paper has some limitations on the work are size of payload to embed and supported image sizes. The capacity of the payload is limited and also supporting for all image formats.

(Al-Jarrah,2017)proposes the RS Steganalysis method using the BossBase 1.01 gray scale dataset which consists of 10,000 PGM format images. The author confines to the single format of the images throughout the paper and carried out his analysis. This paper is not confined to single format of images. Any format of the image can be used in the application for analyzing. (Xintao, Daidou, Nao, Boxia, Mengxiao, Chuan, 2020) proposed a method using Discrete Cosine Transform (DCT) and elliptic curve cryptography (ECC). The authors have used SegNet Neural Network to improve the capacity of the proposed steganographic approach (Table 1).

3. PROPOSED METHOD

This system intends to develop a classy system that can help society from the dangerous malicious programs or concealed data present in the Multimedia. Figure 1 explains the Steganalysis system, this system will directly load the images from the web crawler with the permission access from the administrative. After accessing the images from the crawler wherever they are saved in the computer must go through the analysis session so that it is declared that the image is a clean and doesn't possess any hidden payload in it (Rupa, 2010). The developed system works so that the user must specify the location on the computer to which the downloaded images are saved. The user can also select a single image or a database like BOSS-Base Database which consists of 10,000 digital images of same size (512X512) and same format (.pgm). After the image is loaded, the image is given to a Less Complexity Linear classifier to classify the image. After the image has been classified the RS steganalysis (Tran, Alazab, Broadhurst, 2014) is chosen to scan the image and then the result will be displayed. An application has been developed to perform all these actions at a single visit. This project is developed on MATLAB 2017a, where an application is designed using **"MATLAB GUI".** A Graphical User Interface is designed to be very user-friendly to the user with all the guidelines and comments. The computational resources required to develop this application is 1TB of space and fast operating system with i7 7th generation core. The thousands of Input images, output images and graphs which are produced at the end of the system should be stored in the same location.

This application will directly lead you to browse in all format of images and apply the action based on the user interest.

3.1 Low Complexity Linear Classifier

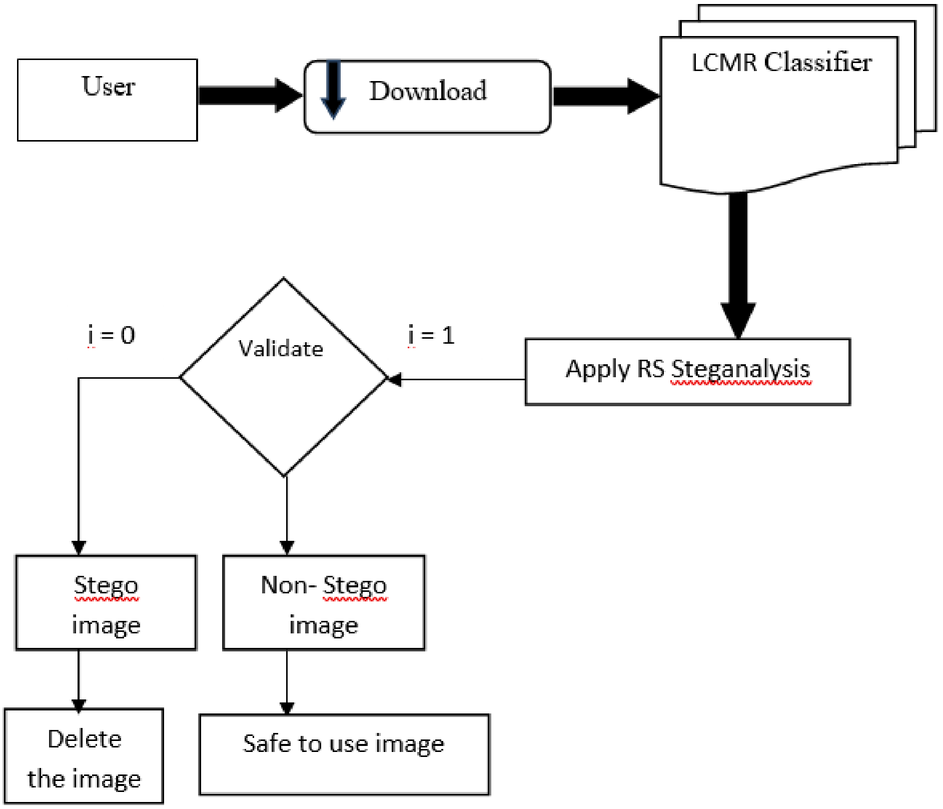
Classification plays a vital role in this system. Making the decision depending upon the features of the database is done by decision-based Classifier. The combination of characteristics obtains the linear value and the vector value. 997 samples among 10,000 digital images are trained and among them 548 are used for feature extraction. Optimal regularization parameter uses 5-folds for cross

Table 1. Summary of Related Work

Scheme	Analysis		Features			
	RS	PA	Payload Capacity	Robustness	Imperceptibility	Quality
Fridrich (2010; 2013; 2013)	No	Yes	Low	Partial	Moderate	High
Imam (2019)	Yes	No	Moderate	High	High	Low
Rupa (2016; 2017; 2019)	Yes	No	Moderate	High	High	Low
Swadhin (2017)	Yes	No	High	Low	Low	Low
Samikhsha (2015)	Yes	No	High	Moderate	Moderate	High
Rupa (2016; 2017; 2019)	Yes	No	High	Low	Moderate	High
Proposed	Yes	No	High	High	Moderate	High

*RS means Regular Singular Analysis *PA means Payload Analysis.

Figure 1. Architecture of Steganalysis System



validation. As BOSS-Base database 1.01 (Rupa, 2019)(Numan, Sunhan, Khan, Haka, Haider, Reddy, Alazab, 2020) grayscale is given as input database for low complexity linear classifier, classification

Table 2. Standard Data set of Images

Standard Boss-Base1.01 Data set		Statistical Analysis	
No. of images	Size	Time in secs	Accuracy (%)
1	257 KB	0.01	0.35
100	25 MB	1.10	0.37
500	125 MB	8.54	0.37
700	175 MB	14.23	0.37
1000	250 MB	33.03	0.38
2000	500 MB	64.67	0.39
5000	1.37 MB	137.23	0.39
10000	2.74 MB	430.57	0.45

is done on the concept of training the dataset and then testing the dataset. With this minimum PE value is calculated. The dataset with different formats of images and their sizes are being performed on Statistical Analysis. Its accuracy is compared as shown in Table 2, with the proposed methodology i.e., RS Steganalysis.

Figure 2 shows optimal regularization based on the K- fold Cross Validation (CV), where 'K' is the number of selections or number of folds, here named as folds as per this paper. The complete data set is divided into 5 folds where a s K=5 folds. In the first iteration, fold-1 is considered as the testing model and the rest are the training model. In the second iteration, fold-2 is the testing set and the remaining revert as the training set the same process is repeated until each fold of the 5 folds are tested and used as testing sets. Different colors are used for 5-folds for easy identification as shown in above Figure 2(b). The graph, Figure 3(b) is between tolerance vs P_E. The below graph and selected database fold 3 are selected, which is indicated with blue color line with minimum P_E value. The formula for calculation of P_E is encrypted below:

$P_E = \text{Min } P_{FA} + ½ [P_{FA} + P_{MD} (P_{FA})]$………….. (Fridrich, Golijian, Rui, 2010)

Where,

P_{FA} is the probability of False Alarms,
P_{MD} is the probability of Missed Alarms.

Figure 3 clearly shows that for all the 5-folds, fold-3 has minimum P_E value on validation set. On comparison with all the three images with their respective generated graphs presented in Figure 2 and Figure 3, the first image has minimum $P_E$value as per the rule of the RS Steganalysis after the point of 0.15 the graph is linear in its shape which describes that the considered image is not a Stego image with the minimum P_E value. Among all these images fold-3 of the first image has minimum validation. All the images are tested for Low Complexity Linear Classifier and its values are obtained for a validation of 5-Folds. Among all these, 0.15 is the minimum value obtained so far on performing the RS Steganalysis.

Figure 2. (a) Input Images from Dataset (b) 5-fold Cross Validation

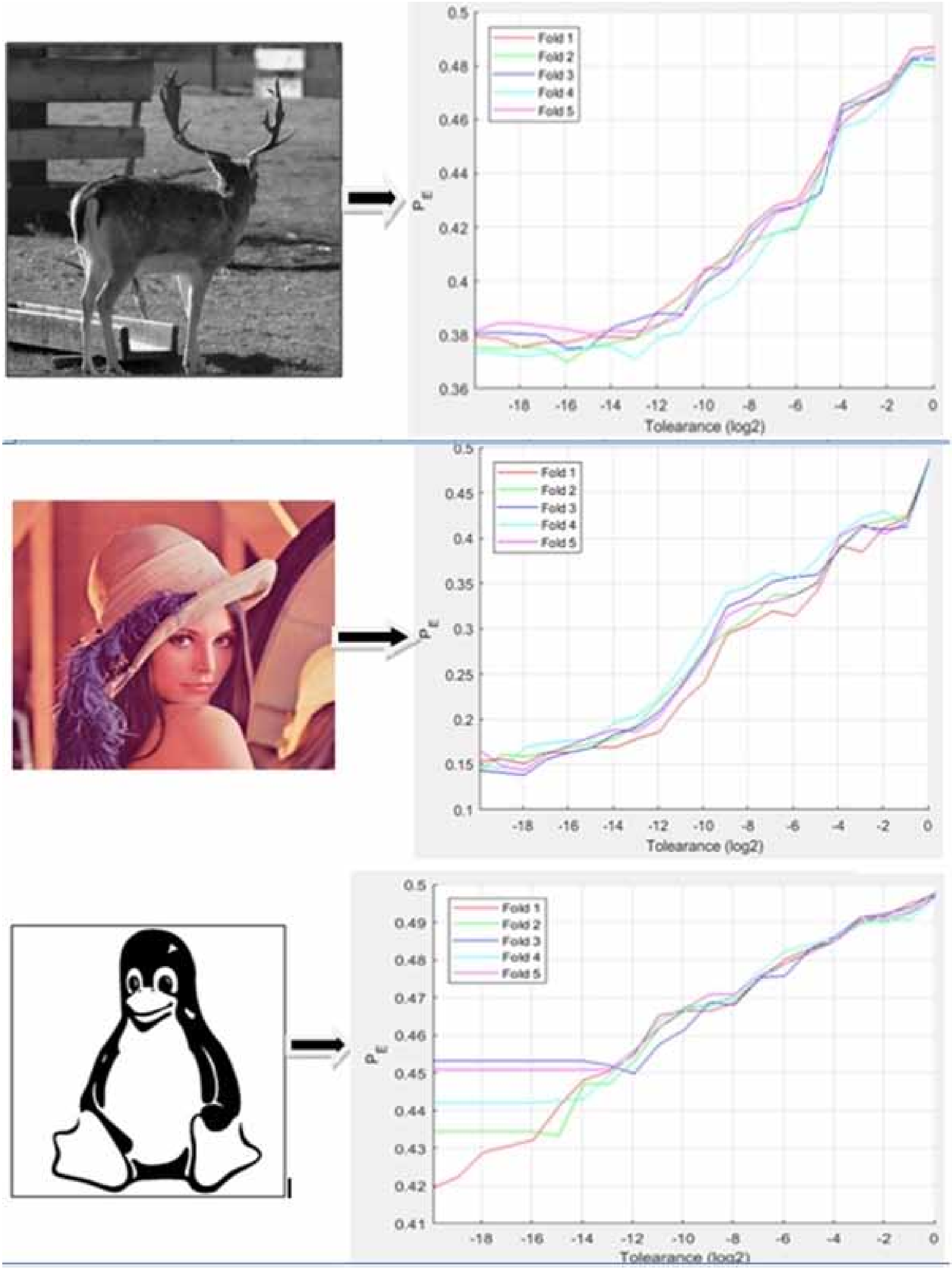
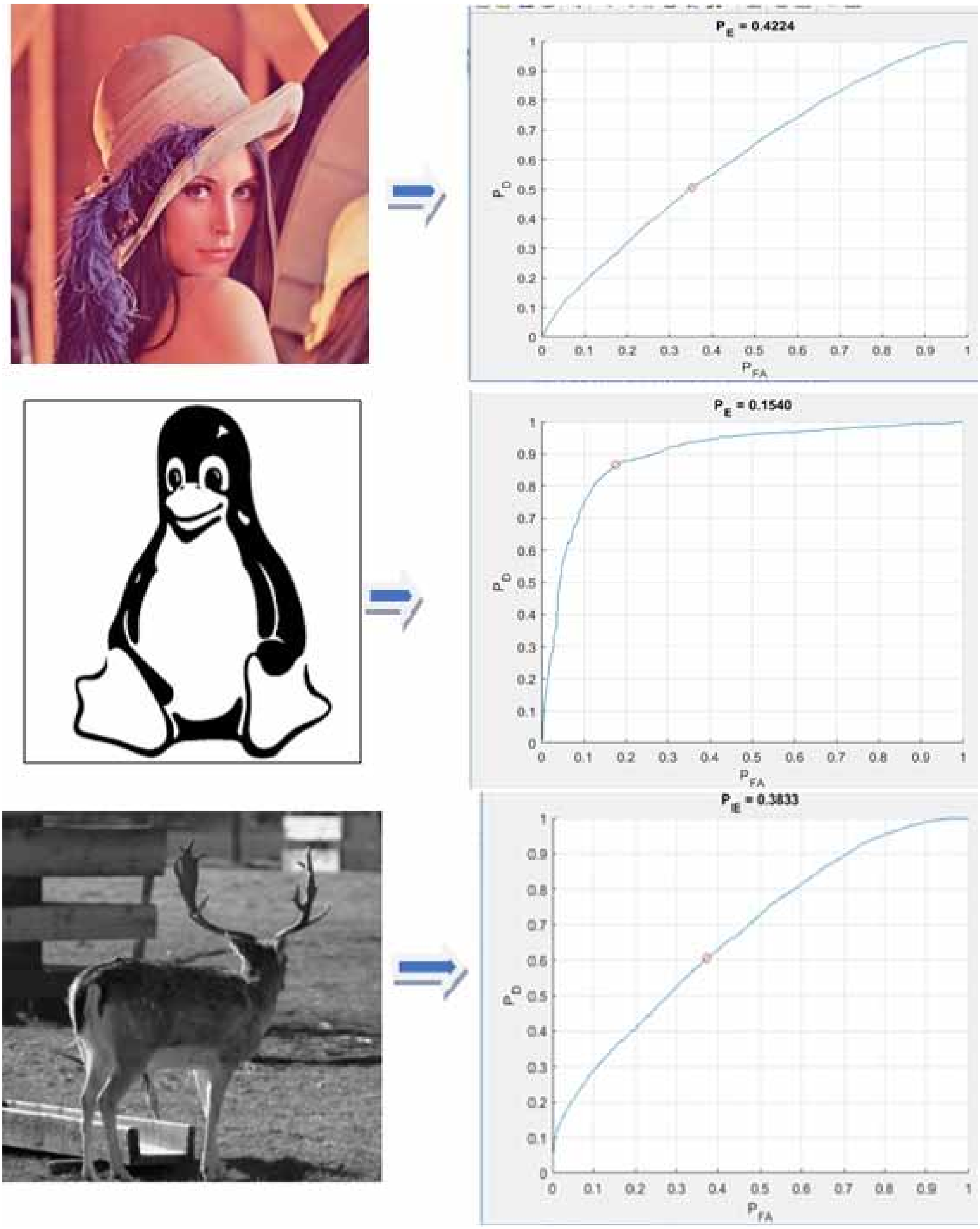


Figure 3. (a) Input Images (b) Calculated P_E Graph



4. METHODOLOGY

The main method used to develop this project is RS Steganalysis. Regular and Singular group will randomize the image depending upon the pixels of the image. The Randomization starts from the Least Significant Bits of the image. The total number of pixels in the image will be 255. They are given different colors for identification. Every 8-bit of pixel forms a frame. RS method is based on flipping concept. The 0th pixel flipped and formed as 1st bit, then 1st bit is flipped and formed as 2nd bit so on 255 bits is flipped to 266 bits. The negative flipping is same as positive flipping. Negative flipping starts from -1th bit and ends at 255th bit. Regular group is denoted as G_R and Singular group is denoted as G_S. The expected value of G_R is equal to G_{-R} and is true for G_S too (Table 3).

Table 3. Calculation of Runtime for RS Steganalysis

No. of images	Size	Time	Accuracy (%)
1	257 KB	0.01 sec	0.65
100	25 MB	1.50 sec	0.65
500	125 MB	10.56 sec	065
700	175 MB	13.56 sec	0.65
1000	250 MB	18.06 sec	0.65
2000	500 MB	19.56 sec	0.65
5000	1.37 MB	22.06 sec	0.65
10000	2.44 GB	27.56 sec	0.65

$G_R = G_{-R}$ and $G_S = G_{-S}$ …………(Fridrich, Golijian, Rui, 2010)

Where,

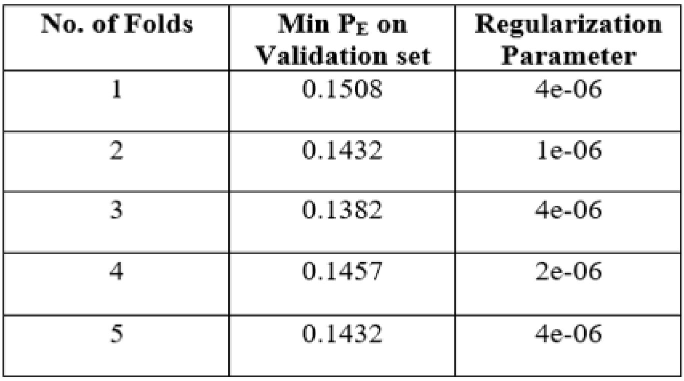
G_R is the Positive Regular Group,
G_{-R} is the Negative Regular Group,
G_S is the Positive Singular Group,
G_{-S} is the Negative Singular Group.

Percent of pixel *'p'* with length *'l'* is embedded randomly with scattered pixels which are arranged in LSB order. Regular and Singular groups correspond to the points with *p/2* as shown in below Figure 4(b) and 5. accuracy can be calculated as below:

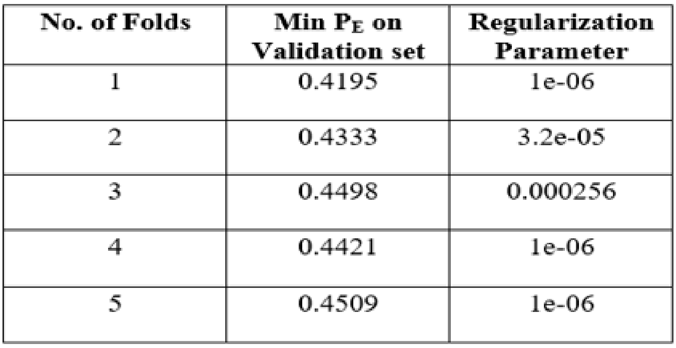
Accuracy = $[(N_i^* + Ni)/2] \times 100$…….. (Fridrich, Golijian, Rui, 2010)
Where,
$N_i^* = |\{ \text{color}|\text{SortedIndexOf (color)} \Sigma^{k}_{i=1} \{2i,2i+1\}|\}|$
2
$N_i = |\{\text{color}|\text{SortedIndexOf (color)} \Sigma^{k}_{i=1} 2i|\}|$
N = Number of indices
i = 1,2,3,4,…………………,K pairs
K = Categories of pairs.

The Figure 6 describes the RS analysis with the embed rate of 0.5. the lines of color red and blue indicate relative numbers of regular and singular groups that get deviates from 0.5% of flip. The error percentage is also greater than zero as calculated in Low Complexity Linear Classifier. This method works the same with the JPEG format images also as shown in Figure 7.

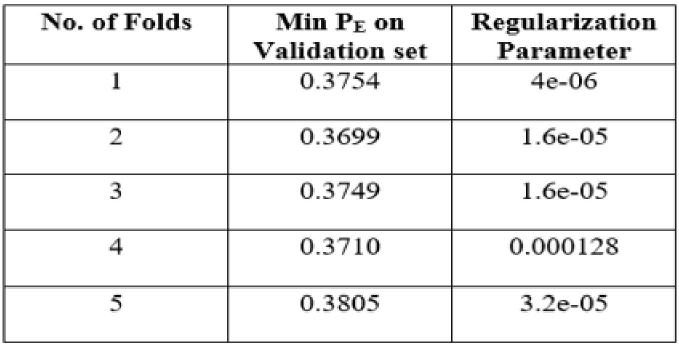
Figure 4. (a) Input Image (b) Calculation of P_E and Regularization parameter



No. of Folds	Min P_E on Validation set	Regularization Parameter
1	0.1508	4e-06
2	0.1432	1e-06
3	0.1382	4e-06
4	0.1457	2e-06
5	0.1432	4e-06



No. of Folds	Min P_E on Validation set	Regularization Parameter
1	0.4195	1e-06
2	0.4333	3.2e-05
3	0.4498	0.000256
4	0.4421	1e-06
5	0.4509	1e-06



No. of Folds	Min P_E on Validation set	Regularization Parameter
1	0.3754	4e-06
2	0.3699	1.6e-05
3	0.3749	1.6e-05
4	0.3710	0.000128
5	0.3805	3.2e-05

5. CONCLUSION AND FUTURE WORK

Technology is playing a vital role globally. In this computing world, every transaction required security from the unintended third parties. Authorised persons are using many kinds of security mechanisms to provide privacy for the sensitive data. Unfortunately, unauthorised persons such as attackers and intruders are also using the same security mechanisms to launch malwares on the intended target systems. The proposed method in this paper is used to analyse the multimedia data formats such as image and text to check whether it has the concealed data or not using RS Steganalysis method. Notify to the user with the result in a graphical way if the concealed data was detected. The main strength of the proposed work is can able to do steganalysis either one to one image scan or many

Figure 5. (a) BMP Format of Input Image (b) Output of RS Application

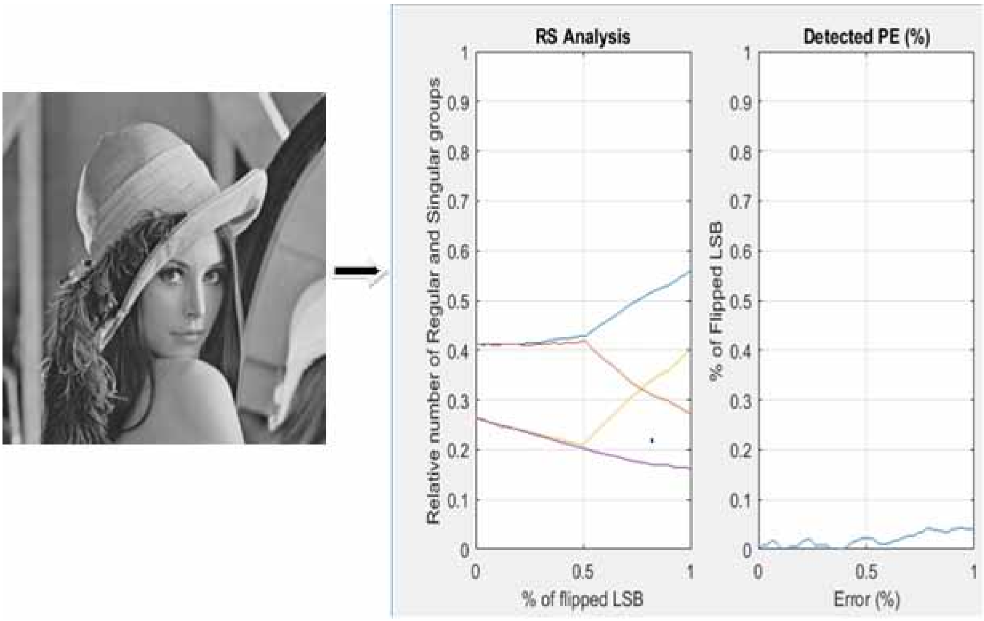
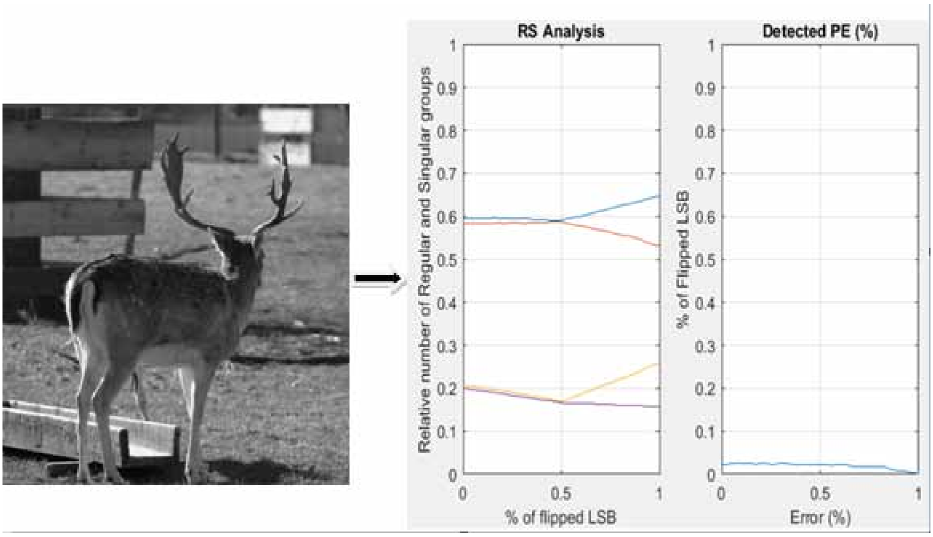
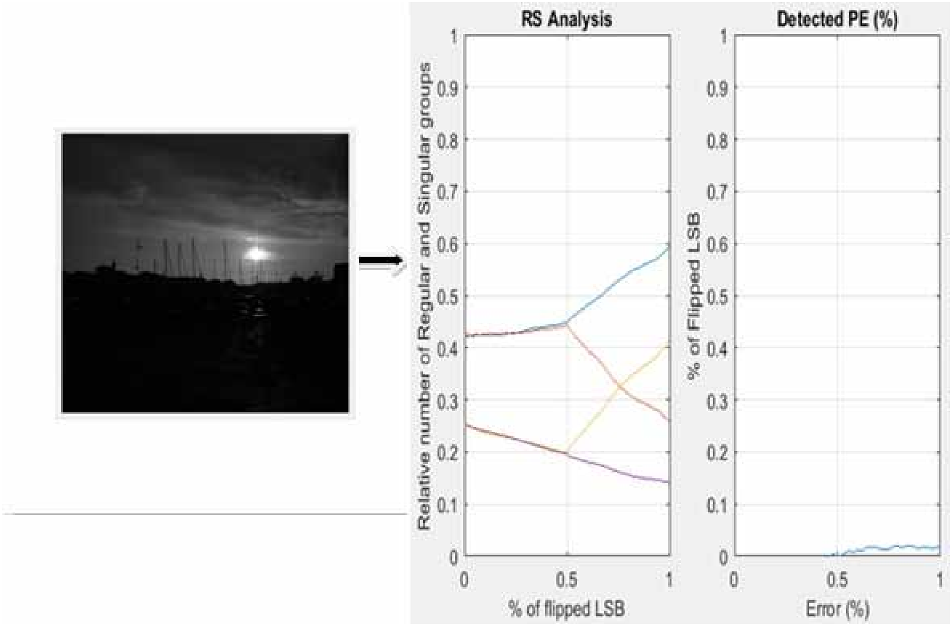


Figure 6. (a) Input image of .PGM format (b) Output of RS Steganalysis with 0.5 embed rate



to many image scans. In the future work, would like to apply the proposed mechanisms on videos to check their integrity. As well to detect and extract the concealed data if exists.

Figure 7. (a) Input image of .JPEG format (b) Output of RS Steganalysis with 0 embed rate



REFERENCES

Al-Jarrah. (2017). Steganalysis Using LSB-focused Statistical features. *International Conference on Future Networks and Distributed Systems, 54*, 1-5.

Alazab, M., Shalaginov, A., Mesleh, A., & Awajan, A. (2020). Intelligent Mobile Malware Detection using Permission requests and api calls. *Future Generation Computer Systems*, *107*, 509–521.

Alazab, M., Venkatraman, S., Watters, P., & Alazab, M. (2013). Information security governance: the art of detecting hidden malware. In *IT security governance innovations: theory and research* (pp. 293-315). IGI Global.

Bhattacharya, Banerjee, & Sanyal. (2011). A survey of Steganography and Steganalysis Technique in Image, Text, Audio and Video as cover Carrier. *Journal of Global Research in Computer Science, 2*(4).

Cohen, A., & Nissima, N. (2020). Machine Learning Based Solution for the Detection of Malicious JPEG Images. IEEE Access, 99.

Curran & McDevitt. (2008). Image Analysis for Online Dynamic Steganography Detection. *Computer and Information Science, 1*(3).

Duan. (2020). *A New High Capacity Image Steganography Method Combined With Image Elliptic Curve Cryptography and Deep Neural Network* (Vol. 8). IEEE Access.

Fridrich, J., Golijan, M., Hogea, D., & Soukal, D. (2013). Quantitative steganalysis of digital images: Estimating the secret message length. *Multimedia Systems*, 288–302.

Fridrich, J., & Goljan, M. (2002). Practical Steganalysis of Digital Images – State of the Art. SUNY Binghamton, Department of Electrical Engineering.

Fridrich, Golijan, & Du. (2010). Detecting LSB Steganography in color, and grey-scale images. *IEEE MultiMedia*, *8*(3), 22–28.

Hameed, Hassaballah, Aly, & Awadi. (2019). *An Adaptive Image Steganography Method Based on Histogram of Oriented Gradient and PVD-LSB Techniques*. IEEE.

Imam, N., & Vassilakis, V. (2019). Detecting Spam Images with Embedded Arabic Text in Twitter. *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 1-6.

Lee & Lee. (2020). New Approach on Steganalysis: ReverseEngineering based Steganography SW Analysis. *9th International Conference on Software and Computer Applications*.

Numan, M., Subhan, F., Khan, W. Z., Hakak, S., Reddy, G. T., & Alazab, M. (2020). A systematic review on clone node detection in static wireless sensor networks. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 65450–65461.

Ruan, F., Zhang, X., Zhu, D., Xu, Z., & Wan, S. (2019, October 4). Deep Learning for Real – Time Steganalysis: A Survey. *Journal of Real-Time Image Processing*.

Rupa. (2020). Security and privacy of UAV data using blockchain technology. *Journal of Information Security and Applications*.

Rupa, C. (2016). Squint Pixel Steganography: A Novel Approach to detect Digital Crimes and recovery of medical images. *Int. Journal of Digital Crimes and Forensics, 34*(4), 37-47.

Rupa, C. (2017). A Secure Information Framework with APRQ Properties. *Springer IEI*, *98*(4), 359–364.

Rupa, C. D. (2019). Privacy and Protection of Medical Images ROI Using SPLSB and Bit-plane based Watermarking. In *ACM International Conference on Cryptography, Security and Privacy 2019*. University of Malaya.

Rupa, C. (2019). Extended Statistical Analysis on Multimedia Concealed Data Detections. *Ingénierie des Systèmesd'Information*, *24*(2), 161–165.

Rupa, C., Thippa Reddy, G., Abidi, M. H., & Alahmari, A. (2020). Computational System-to-Classify Cyber Crime Offenses Using Machine Learning. *Journal of Sustainability*, *12*(10), 1–15.

Sairam, T. D., & Boopathybagan, K. (2019). Computational Intelligence based Steganalysis Comparison for RCM-DWT and PVA-MOD methods. *Journal for Control, Measurement, Electronics, Computing and Communications*, *60*(3), 2019.

Samikhsha, K. (2015). A Comparative Study of Various Image Steganographic Techniques used for Information Hiding. *Compusoft. International Journal of Advancements in Computing Technology*, *4*(5).

Saxena, N. (2015). Steganography scheme Against RS Attack Enriched with Evolutionary Programming (AGA) and OPAP. *International Journal of Mathematics and Science*, *3*(11), 64–74.

Shoniwa, R. T. R., & George, G. (2013). Design of Application to Detect Images Embedded with Malicious Programs. *International Journal of Scientific Research (Ahmedabad, India).*

Simmons. (1984). The Prisoners' Problem and the Subliminal Channel. In Proceedings of CRYPTO '83 (pp. 51-67). Plenum Press.

Swadhin, Ksustubh, & Chirag. (2017). Video Steganography Using Encrypted Payload for Satellite Communication. *IEEE Conference*.

Tran, Alazab, & Broadhurst. (2014). *Towards a feature rich model for predicting spam emails containing malicious attachments and urls*. Academic Press.

Yang, C., Liu, F., Ge, S., Lu, J., & Huang, J. (2019, May). Locating Secret Message Based On Quantitative Steganalysis. *Mathematical Biosciences and Engineering*, 30.

Zou, Y., Zhang, G., & Liu, L. (2019). Research on Image Steganograpghy Analysis Based on Deep Learning. *Journal of Visual Communication and Image Representation*. Advance online publication. doi:10.1016/j,jvcir.2019.02.034

Ch. Rupa (PhD) is working as a professor in VRSEC (A), Vijayawada. She was a senior member of IEEE and Life Member of CSI, ISTE, IAENG, IEI, IACSIT. She published more than 70 papers in various journals and conferences. JNTU Kakinada has awarded her as a Young Engineer of 2010. IEI awarded her as National young Engineer of 2011 Govt of A. P and IEI by combined awarded her as Young Engineer of 2012. Her main research interests includes information security, Image Processing, Security algorithms. She has received couple of awards from IETE, IEI(I) for her work.

Sumaiya Shaikh is an M.Tech grad who is interested in research at cyber forensics and cyber security.

Mukesh Chinta is currently working as an Assistant Professor in Department of CSE at V R Siddhartha Engineering College, Vijayawada. He has received B.Tech degree in Electronics and Communications Engineering and Masters degree in Distributed Systems and Networks from University of Hertfordshire, UK. He has 10 years of teaching experience. He has actively organized and participated in various workshops and technical events. His research interests include Adhoc Networks, Wireless Sensor Networks and Network Security. He is also a certified Cisco CCNA and CyberOps instructor.