ISSN 2229-5518

A study on steganography and steganalysis.

Habiba Sultana
Department of Computer Science and Engineering,
Jatiya Kabi Kazi Nazrul Islam University, Mymensingh-2220, Bangladesh.
srity.cse@gmail.com

Abstract— Steganography means hiding information into a media. Media may be a text, image, audio, video or network protocol. Embedding methods use different methods to implant information. Unauthorized persons, applications or devices may apply steganalysis to either detect the existence of hidden data in the carrier medium or explore the hidden information after that. In this paper, steganography, its classifications, applications, evaluation parameters, and related steganalysis are described briefly. That discussion will help the new researcher in finding a basic idea on steganography as well as on steganalysis.

Keywords—Steganography, cover image, stego image, LSB, PVD, steganalysis.

1 Introduction

Information plays a vital role in our daily life. Information needs to be secured. Every information security system tries to keep the data or information secured. Figure 1 depicts the commonly applied security systems. A famous securing technique is to hide the information in a media or to modify the information to an unrecognized format [1]. In Cryptography, a sender converts a message name plaintext into cipher text using an encryption algorithm and secret key. The receiver extracts that plaintext from cipher text using a decryption algorithm and the secret key [2,6]. The cryptography system is incapable of avoiding attentions of intruders. Invisible communication could come as a solution of the affairs. That is why information hiding mechanism becomes an attractive area in the field of information security [3]. Information hiding policy protects the data from all unauthorized party. The main area of application of information hiding systems is the military, intelligence agencies, online elections, internet banking, medical imaging and so on [1].

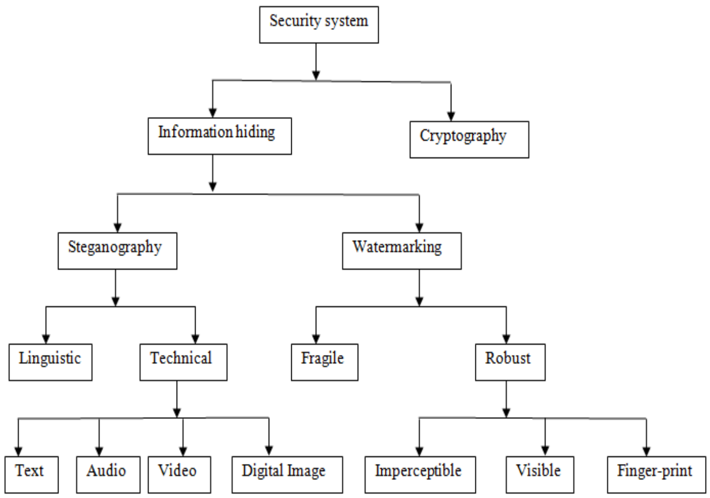

Figure 1: security system [1].

Information hiding consists of two branches - steganography and watermarking [1]. The watermarking works to add a mark using a key in a public data [5]. The signs need to be undetectable and robust to techniques of information processing [5].

It protects the integrity of secret data with or without concealing the existence of communication [3].

The purpose of watermarking is to protect the intellectual property of the contents [3]. On the other hand, the steganography policy conceals the secret information in the media.

Steganography is a Greek word. It means "covert writing" The steganography system hides the presence of messages within a media. In steganography, would not detect the presence of the message because the decoding algorithm is unknown. Important terminologies, used in the field of steganography, [7,8] are:

Cover image: Image steganography system implants its secret in an image known as a cover image.

Message: Original information that the system hides into the cover image.

Stego-image: Stego-image contains the implanted secret information.

Stego-key: For embedding or extracting the message in cover-images and from stego images, the algorithm uses a key, called stego key.

Embedding procedure: A set of rules which implants a secret message in an image.

Extracting procedure: The process of recovering a secret message from the stego image.

In steganography, if intruders want to find the secret message, he/she needs some secret information [8], such as detection of the existence of an embedded message in a given image, identification of the steganographic method, an approximation of the message length, and extraction procedure of the secret message. For this, steganography is a robust data hiding technique.

This article presents a study on steganography, its classification, evaluation parameters, and related steganalysis methods.

ISSN 2229-5518

The remaining of this paper is organized as follows: Section 2 describes the classification of steganography. Section 3 provides a view of the application of steganography. Section 4 narrates the evaluation parameters of steganography. Section 5 explains the processes of steganalysis. Finally, section 6 concludes the article.

2. Classification of Steganography in Various Aspects

Steganography can be classified in different ways, such that depending on carrier or key or embedding domain. Steganography mainly classified into two subsections, one is technical steganography, and another is linguistic steganography. This classification can be shown in the following:

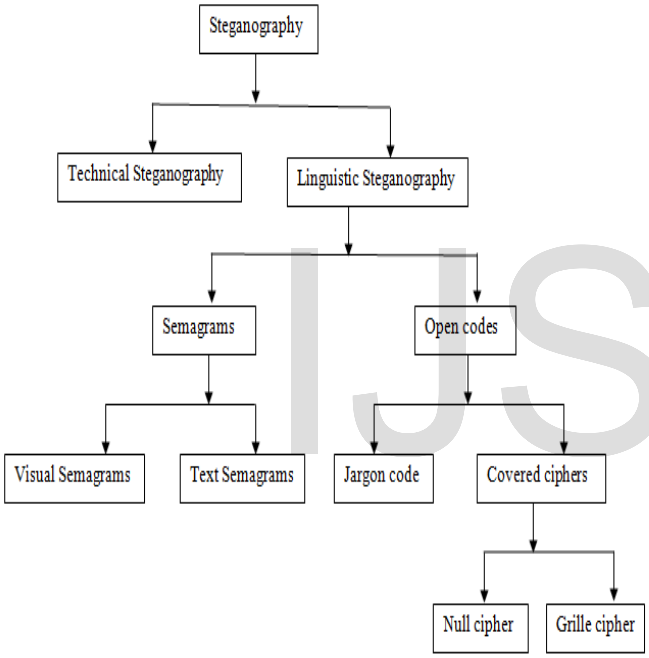

Figure 2: classification of steganography [4].

2.1.1 Technical Steganography

To conceal the existence of a message using physical or chemical means is called technical steganography. Examples of this type of steganography include invisible ink or microdots and other size-reduction methods [4].

2.1.2 Linguistic Steganography

Linguistic steganography is a branch of steganography in which embed the messages into cover and make stego. In the resulting stego, without actual recipient, no one easily finds the presence of the real information. In this steganography, the idea is to hide the presence of the actual messages. It hides the messages in natural language [9]. Linguistic steganography is further classified as semagrams and open codes [4].

Semagrams hide information by the use of symbols or signs in visually or text. Semagrams is classified into visual semagrams and text semagrams.

In visual semagrams, the physical object is used to embed secret data in such a way that only actual recipient can understand not another one. To hide the message semagrams used different symbol and sign of the particular language [10].

Text semagrams, modifying the appearance of text to embed the secret information. Modifying the appearance means changes in the font size, color, height and width of letters, add extra space, add extra letters, etc. [10].

Open codes are used to hide the secret message invisibly to the outsider. Open code is categorized in two ways: jargon codes and covered cipher.

In jargon code, information is embedded using the properties of the particular language. The people who are known with that particular language, only they can understand the embedded information.

In covered ciphers, secret information is embedded openly. So, if anyone knows the method of embedding, then it can be possible to recover the information. Covered ciphers can classify as a null cipher and grille cipher.

In grille cipher, secret information is embedded into a template. The words that appear in the openings of the template are the hidden message.

In null cipher, some predefined set of rules are used to implant the secret information. "Read the second character in every word" is one kind of predefined rule [4].

2.2 Classification of Steganography Based on the Key:

Key is an important element in steganography. Because of key increases more security. So, according to the use of the key, steganography is classified into three types as shown in fig .3.

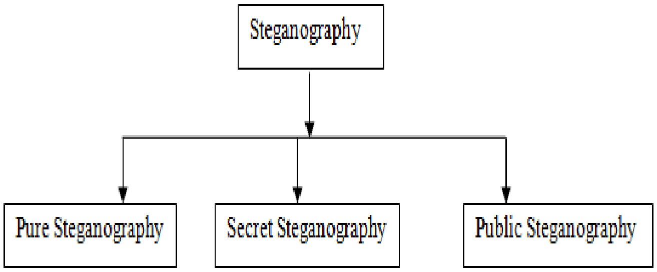

Figure 3: Classification of steganography based on key [1,10].

2.2.1 Pure Steganography

Pure steganography does not follow the concept of key. So, it does not require any prior exchange of data before sending the actual secret message. It is based on the assumption that no other party is aware of the communication.

2.2.2 Secret key steganography

In Secret key steganography, one secret key is used to embed the data. This secret key should be exchanged before communication so that after embedding the data by the sender, the receiver can extract the original message with the key.

2.2.3 Public key steganography

In public key steganography, two types of keys are used. One is the public key which is known to both the sender and the receiver, and another is the private key which is only known to the receiver. Using the public key, sender Implant the message. On the other hand, the receiver extracts the message using the private key.

2.3 Classification of steganography based on carrier file:

Carrier file means the file which is used to carry the secret message. To obtained the security, a various cover file is used, i.e. text, audio, video, network, and digital image. It is shown in the following:

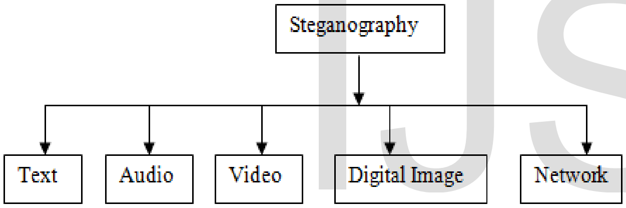

Figure 4: classification of steganography based on carrier [7,10].

2.3.1Text steganography

Text steganography is a technique where the text file is used as a carrier. In this methods, hiding the message of one text into another one for secure communication. Text steganography methods [11] are classified into three parts which consist of subparts as shown in the following:

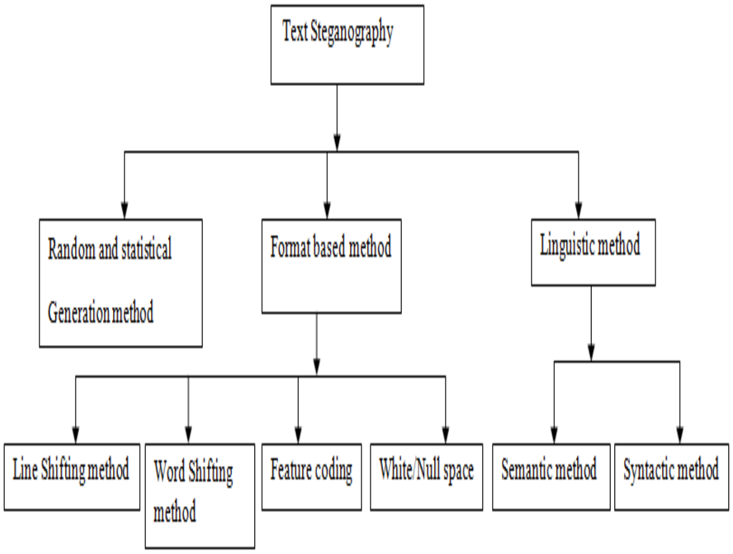

Figure 5: classification of text steganography

Format based methods used physical text formatting of text as a place in which to hide information. They do not change any word or sentence. Format based methods are line shifting, word shifting, feature coding, and white/ null space.

In the line shifting method, the lines of the text are vertically shifted to some degree and information is hidden by creating a unique shape of the text.

In word shifting method, the horizontal alignment of some words are shifted by changing distances between words to embed information in the text, and is acceptable for text where the distance between words is varied.

The feature coding method, some of the features (character height or font) of the text are chosen and altered depending on the message to be inserted.

White/ null space method is used to implant a small amount of data in a document. In this method, extra white/ null spaces are added at the end of each line or the end of each paragraph or sentence or between the words.

Random and statistical generation methods are used to generate cover-text automatically according to the statistical properties (grammar, probabilistic context-free grammar, etc.) of language.

Linguistic methods as a technique of data hiding that embeds the secret message within text based on some linguistic knowledge. They are a semantic and a syntactic method.

In a semantic method, the synonyms of certain words are used for hiding information in the text.

In a syntactic method, the syntactic structure (punctuation signs like a comma, full stop, etc.) of the text is used to embed the secret message into a text file.

ISSN 2229-5518

2.3.2 NETWORK STEGANOGRAPHY

Network steganography allows users to communicate secretly by embedding information within other messages. Steganophony, Transcoding steganography, lost audio packets steganographic method, Stream Control Transmission Protocol, etc. are network steganography.

2.3.3 AUDIO STEGANOGRAPHY

When an audio file is used as a carrier to implant a message, then this is called audio steganography. Audio files may be modified in such a way that they contain hidden information in such a way that it is difficult to remove the hidden data without destroying the original signal.

2.3.4 VIDEO STEGANOGRAPHY

A video consists both of images and audio. When this video file is used as a carrier, then this steganography is called video steganography. The video steganography technique is almost similar to image and audio steganographic techniques. Using a video file, a large amount of data can be implanted, but the problem is the size of such file and also the time computation.

2.3.3 IMAGE STEGANOGRAPHY

When taking the cover file as the image in steganography is called image steganography. Images are the most popular and useful cover image for steganography [17-23]. Image steganography is divided into two categories as spatial domain and frequency domain and is shown in the following:

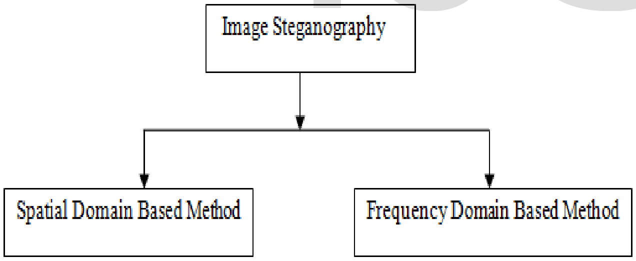

Figure 6: Classification of image steganography method

2.3.3.1 SPATIAL DOMAIN BASED METHOD

Spatial domain based method implants the secret data by modifying the spatial domain of the cover image. These methods use the gray pixel levels and their color values directly for encoding the message bits. Some spatial domain methods [7] are following.

- Least Significant Bit (LSB)
- Pixel Value Differencing (PVD)
- Edge-Based Data Embedding Method (EBE)
- Random Pixel Embedding Method (RPE)
- Mapping pixel to hidden data method
- Labeling or connectivity method
- Pixel intensity based method
- Texture-based method
- Histogram shifting methods.

2.3.3.2 FREQUENCY DOMAIN BASED METHOD

To implant, the message bits in the transform domain coefficients of the image is called frequency domain methods. Frequency domain is also known as transform domain. Some of frequency domain methods [7] are following:

- Discrete Fourier Transformation Technique (DFT)
- Discrete Cosine Transformation Technique (DCT)
- Discrete Wavelet Transformation Technique (DWT)
- Lossless or Reversible method
- Embedding in coefficient bits.

3. APPLICATION OF STEGANOGRAPHY

Steganography is very popular for its security. Steganography is implemented in a different field in our daily activities. Such as:

- Medical safety
- Terrorism
- Hacking
- Intellectual property offenses
- Corporate espionage
- Watermarking
- Indexing of video mail
- Military application
- Automatic monitoring of radio advertisements
- Protection against malware

4. EVALUATION OF STEGANOGRAPHY

An image steganography system can be evaluated based on some criteria or parameters. Those are visual quality, payload, robustness, undetected ability or security, and some computational complexity. Now briefly describe in following:

4.1 VISUAL QUALITY:

Visual quality is an important parameter to evaluate the steganography system. When a human visual system is unable to detect the differences between the original image and data embedded image, then the steganographic technique is successful. Sometimes visual quality of a cover image may be maintained or destroyed depends on the steganographic methods. The matrices which are used to measure the visual quality of an image are MSE, PSNR, Q, SIMM, cross-correlation, etc.

4.1.1 MSE:

MSE stands for a mean square error. MSE is obtained from the cover image C, and stego image S. The formula of MSE is given following:

$$MSE = \frac{1}{(M \times N)^2} \sum_{i=1}^{M} \sum_{j=1}^{N} (C_{i,j} - S_{i,j})^2$$

Where M and N are the height and width of cover and stego image,
If the value of MSE is small, then the quality of stego image is maintained. Then it can say that it is less detect ability.

4.1.2 PSNR:

PSNR stands for peak-signal-to-ratio. The PSNR is the ratio between a signal's maximum power and the power of the signal's noise. The PSNR is calculated using the following formula:

$$PSNR = 10\log_{10}\frac{Max^2}{MSE}dB$$

Where, Max denotes the maximum pixel value of the image. A higher PSNR value indicates the better quality of the stego image.

4.1.3 Cross-correlation:

Cross-correlation is a measure of similarity of cover image C and stego image S. Cross-correlation is measured as:

$$CC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(C_{i,j}-\mu_1\right)\left(S_{i,j}-\mu_2\right)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(C_{i,j}-\mu_1\right)^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(S_{i,j}-\mu_2\right)^2}}$$

Where μ_1, μ_2 is the mean pixel values of the cover image C, and stego image S.

4.1.4 UIQI:

UIQI stands for Universal Image Quality Index. UIQI is used to measure the similarity between cover image C and stego image S into three comparisons: correlation or structural information, luminance and contrast. UIQI is measured by,

$$Q \quad = \frac{4\times\sigma_{cs}\times c^{-}\times s^{-}}{\left(\sigma_c{}^2+\sigma_s{}^2\right)\times\left(c^{-2}+s^{-2}\right)} \quad \ldots..(1)$$

Where c^{-}, s^{-}, $\sigma_c{}^2$, $\sigma_s{}^2$ and σ_{cs} are given as

$$c^{-} = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(C(i,j)\right)$$

$$s^{-} = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(S(i,j)\right)$$

$$\sigma_c{}^2 = \frac{1}{M\times N-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(C(i,j)-c^{-}\right)^2$$

$$\sigma_s{}^2 = \frac{1}{M\times N-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(S(i,j)-s^{-}\right)^2$$

$$\sigma_{cs} = \frac{1}{M\times N-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(C(i,j)-c^{-}\right)\left(S(i,j)-s^{-}\right)$$

Equation (1) can be written as a product of three components

$$Q = Q_1\times Q_2\times Q_3$$

$$where\ Q_1 = \frac{\sigma_{cs}}{\sigma_{cs}\times\ \sigma_{cs}}$$

$$Q_2 = \frac{2\times c^{-}\times s^{-}}{\left(c^{-2}+s^{-2}\right)}$$

$$Q_3 = \frac{2\times\sigma_c\times\sigma_s}{\left(\sigma_c{}^2+\sigma_s{}^2\right)}$$

Where Q_1 measures the degree of linear correlation between C and S.

Q_2 Measures the luminance of C and S

Q_3 Measure the similarities between the contrasts of the images.

4.1.5 SSIM:

SSIM stands for Structural Similarity Index Measure. SSIM is used to compute the important information about the structure of the objects in the visual scence. It is the combination of the values obtained for the average intensity of the brightness, the variations in the contrast and the structure of the cross-correlation between the original and stego image. SSIM [3]is measured by

$$SSIM = \left(\frac{\left(2\times c^{-}\times s^{-}+K1\right)\left(2\times M_{cs}+K2\right)}{\left(M_c{}^2+M_s{}^2+K2\right)\left(c^{-2}+s^{-2}+K1\right)}\right)$$

Where, c^{-} and s^{-} are the means of pixels in image C and S. M_c and M_s are the computed variance of all pixels in both C and S images, M_{cs} is the co-variance between both C and S and K1 and K2 are the constants.
The UIQI and SSIM are considered more consistent and accurate than MSE and PSNR. The value of UIQI and SSIM varies between 1 and -1. If the value is one, then two images are identical and if shows -1 then images are mismatched [14].

4.2 Payload/ Embedding capacity:

Another steganographic evaluation parameter is payload or embedding capacity. It can be defined as some secret bits embedded per pixel in the cover image. The measurement of payload or embedding capacity is

$$EC = \frac{No.of\ embedding\ bits}{M\times N}bpp$$

Where, M and N are the height and width of a cover image.
4.3 Robustness:
Robustness protects the embedding data in stego image even though perform different image processing operations

on it.

4.4 Security:

Security is an important evaluation parameter in steganography. Different steganalysis attack the steganographic approaches. So protect the embedding information from attackers, security is vital.

4.5 Computational complexity:

Computational complexity refers to the efficiency of embedding and extraction methods concerning time and operation. If computational complexity is low then considered the steganographic system is ideal.

5. Steganalysis

Steganalysis exposed the existence of hidden information. It discriminates between stego objects and cover objects. Steganalysis is successful when it can detect the presence of a message. The goals of steganalysis are given below:

- To detect the existence of hidden information in a cover media.
- To identify the type of steganographic method that is used to create the stego object.
- To evaluate techniques that can be used to distinguish between cover object and stego object.
- Not only detect the presence of hidden information but also try to recover the hidden information.
- To estimate the secret message length.
- To break the security of its carriers.

5.1 different types of attacks

Steganography and steganalysis are reversed one another. Steganalysis discover the hidden information. But steganography is used to hide the information. There exist several attacks to destroy the purpose of steganography. Attacks may be in several forms, i.e., detecting, extracting, disabling or destroying secret information. An attack may depend on steganalyst (one who performs steganalysis). Several types of attack [13] are given below:

Stego-only attack:
Only stego object is obtainable for analysis.
Known-carrier attack:
Both of cover and stego object is obtainable for analysis.
Known-message attack:
The secret message is known and can be compared with the stego object.
Chosen-stego attack:
Both stego object and steganographic technique are available for analysis.
Chosen-message attack:
Approximate message and steganographic technique are used to create stego object for future analysis and comparison.
Known-steganography attack:
The secret message, stego object, and the steganographic technique are known and available for analysis.

Destroy everything attack:
Attacker destroys the implanted message.

Random tweaking attacks:
Attacker added small changes in the stgo-file so that the secret message is unreadable.
Add new information:
Attacker adds the new message into stego-file using the same steganographic method.
Reformat attack:
By changing the file format of stego-file, the attacker may destroy the secret information.
Compression attack:
When an attacker compresses the stego file, then embedded secret information may be a loss.

5.2 Classification of Image Steganalysis

Image steganalysis can be classified as targeted, and blind steganalysis and this is shown in the following:

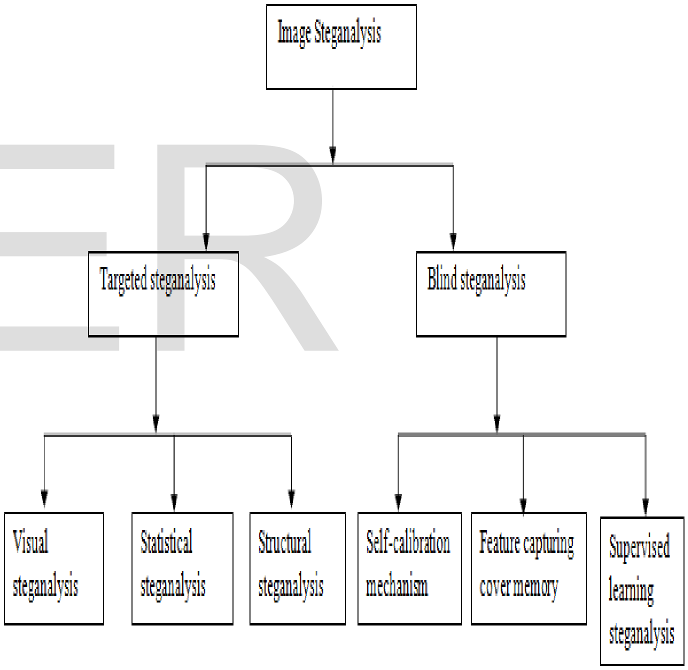

Figure 7: Classification of image steganalysis

5.2.1 Targeted steganalysis

Targeted steganalysis method is dependent on the specific steganographic algorithm that is being used for hiding secret data into the image. It works on a specific type of stego system. The results of this steganalysis are very accurate. But there is no way to extend one technique to another. So these techniques are inflexible [11]. A targeted steganalysis is classified into three parts as: visual, statistical and structural attacks.

Visual steganalysis
Maintain the visual quality is one of the most important

requirements of any steganographic method. Steganalyst applied different visual quality metrics such as MSE, PSNR, SSIM, Q index, etc., to evaluate the visual quality. So that steganalyst can able to find any inconsistency to classify between stego image and normal image. The visual attack can be a useful tool for known cover attacks and finding the sequential order of embedded message. This method is inefficient because it cannot be automated.

Statistical steganalysis

There are various types of statistical analysis, i.e., histogram analysis, bit plane analysis, sample pair analysis as chi-square and RS analysis methods [3]. These methods can discover the existence of the secret message and estimate message length.

Structural steganalysis

Structural steganalysis depends on particular steganographic methods and image. In this method, a more detailed investigation is done so that it can be able to make a difference between cover image and stego image. This method is not used to discover the existence of steganography.

5.2.2 Blind or universal or generic:

Blind steganalysis is preferred than targeted steganalysis. Blind steganalysis is independent of the underlaying steganographic algorithm [16]. These methods are used to differentiate between cover image and stego image based on image feature. Some blind steganalysis methods are self-calibration, feature capturing cover memory and supervised learning steganalysis.

6. Conclusion

This paper has presented a comprehensive survey of steganography and its classification. The evaluation parameter of steganography is also discussed. Here steganalysis also described. In future, different steganographic techniques and steganalysis methods will be implemented and compared.

REFERENCES

1. C.P. Sumathi, T. Santanam, G. Umamaheswari, “A Study on Various Steganographic Techniques Used for Information Hiding”, International Journal of Computer Science and Engineering Survey(IJCSES) vol.4, No. 6, December 2013.
2. Chin-Chen Chang, Min-Shian Hwang, Tung-Shou Chen, “A new encryption algorithm for image cryptosystems”, The Journal of Systems and Software, 58,(83-91), 2001.
3. Mehdi Hussain, Ainuddin Wahid Abdul Wahab, Yamani Idna Bin Idris, Anthony T.S. Ho, Ki-Hyun Jung, “Image Steganography in Spatial Domain: A Survey”, DOI: 10.1016/j. image 2018.03.012.
4. Gary C. Kessler, “An Overview of Steganography for the Computer Forensics Examiner ”, Forensic Science Communications, 2015.
5. Patrick Bas, Jean-Marc Chassery, Benoit Macq, “Image Watermarking: an evolution to content based approaches”, Pattern Recognition, Elsevier, 2002, pp.545-561. <hal-00166587>
6. Yuan-Hui Yu, Chin-Chen Chang, Iuon-Chang Lin, “A new steganographic method for color and grayscale image hiding”, Computer Vision and Image Understanding, (2006), doi: 10.1016/j.cviu.2006.11.002.
7. Abha Sharma, Prof Shreechnadra Upadhaya, Prof Rajkumar Paul, “A Survey of Image Steganography Techniques”, International Journal of Computer Science and Engineering, vol. 5, Issue 3, 2014.
8. Hayat Al-Dmour, Ahmed Al-Ani, “A steganography embedding method based on edge identification and XOR coding”, Expert Systems With Applications, 46(2016) 293-306, 2016.
9. Zhili Chen, Liusheng Huang, Zhenshan Yu, Wei Yang, Lingjun Li, Xueling Zheng, Xinxin Zhao, “Lingustic Steganography Detection Using Statistical Characteristics of Correlations between Words” International Workshop on Information Hiding, IH 2008, Lecture Notes in Computer Science, vol 5284, Springer, Berlin, Heidelberg.
10. Sunita Chaudhary, Dr. Meenu Dave, Dr. Amit Sanghi, “Review of Linguistic Text Steganographic Methods”, International Journal on Recent and Innovation Trends in computing and communication, volume 4, Issue 7, 2016.
11. Souvik Bhattacharyya, Indradip Banerjee, Gautam Sanyal, “A Survey of Steganography and Steganalysis Technique in Image, Text, Audio and Video as Cover Carrier”, Journal of Global Research in computer science, volume 2, No. 4, April 2011.
12. Khurrum Rahim Rashid, Aqsa Rashid, Nadeem Salamat, Saad Missen, “EXPERIMENTAL ANALYSIS OF MATCHING TECHNIQUE OF STEGANOGRAPHY FOR GRAYSCALE AND COLOUR IMAGE”, International Journal of Computer Science & Information Technology (IJCSIT), vol 6, No. 6, December 2014.
13. Katzenbeisser, F.A.P. Petitolas, “Information Hiding Techniques for Steganography and Digital Watermarking”, Artech House, Boston, USA,2000.
14. Shodhganga.inflibnet.ac.in/bitstream.
15. Manveer Kaur, Gagandeep Kaur, “Review of Various Steganalysis Techniques”, International Journal of Computer Science and Information Technologies, vol. 5(2),2014.
16. Sruthi Das N, Rasmi P S, “A Survey on Different Image Steganalysis Techniques”, International Journal of Modern Trends in Engineering and Research, volume 2, Issue 4, April 2015.
17. Kamal, A. H. M., and Mohammad M. Islam. "Enhancing embedding capacity and stego image

quality by employing multi predictors." Journal of Information Security and Applications 32 (2017): 59-74

18. Kamal, A.H.M. & Islam, M.M. " An image distortion-based enhanced embedding scheme", Iran J Comput Sci, 1.3 (2018): 175-186
19. Kamal A. H. M. and Islam M. M., "Enhancing the performance of the data embedment process through encoding errors", Journal of Electronics, 5.4 (2016): 79-95
20. Kamal, A H M and Islam, M. M.,. Boosting up the data hiding rate multi cycle embedment process, J. Vis. Commun. Image R., 40(2016): 574-588
21. Kamal A. H. M. and Islam M. M., “Capacity improvement of reversible data hiding scheme through better prediction and double cycle embedding process”, in Proceedings of IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Kolkata, India, 16-18 December 2015
22. Kamal A. H. M. and Islam M. M., “Enhancing the embedding payload by handling the affair of association and mapping of block pixels through prediction errors histogram”, in Proceedings of International Conference on Networking, Systems and Security (NSysS), BUET, Dhaka, 5-8 January, 2016
23. Habiba S., Kamal A. H. M. and Islam M. M., "Enhancing the robustness of visual degradation based HAM reversible data hiding", Journal of Computer Science, 12.2 (2016): 88-97