Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Steganography: A Data Hiding Technique

Palak Mahajan
Department of Computer Science & IT,
University of Jammu, J&K, India

Abstract- The proliferation of the Internet has made it possible for anybody to send and receive information instantly from anywhere around the world. Communication of secret information is a major challenge and its complexity increases with the levels of sophistication. Steganography is a form of data hiding technique that provides mechanism for securing data over insecure channel by concealing information within information. It is based on invisible communication and this technique strives to hide the very existence of the secret message from the observer. As a result it is very commonly used by Intelligence Agencies for securely broadcasting and communicating information over the internet by hiding secret information inside images and text. Imperceptibility, robustness and capacity of the hidden data are the main characteristics of steganography. This paper provides a state-of-the-art review of the different existing methods of steganography along with some common standards and guidelines drawn from the literature.

Keywords- Data Hiding, Steganography, LSB Technique, Encryption

I. INTRODUCTION

Throughout time, confidentiality of information is a topic of concern. Information hiding plays a very important role in today's world that enables confidentiality and integrity. For many years information hiding has captured the attraction of researchers. Cryptography, watermarking and steganography are the most basic data hiding techniques that are being used to address copyright management, protect information, and conceal secrets [5]. It is a universal acknowledged truth that "a picture is worth a thousand words". The emergence of steganography has made it possible to hide not only 1000, but thousands of thousands of words in a normal image. For decades people strove to develop innovative methods for secret communication.

Steganography is derived from the Greek words "stego" means "covered" or "secret" and "graphy" means "to write" which means "to hide in plain sight". The secret information to be hidden is embedded into the cover object (text, image, or audio /video) in such a way that the very existence of the message remains undetected thereby maintaining the appearance of the resultant stego object. The main goal of steganography is to hide the fact that anything is present inside the transmitting data. As defined by Cachin [1] steganography is the art and science of communicating in such a way that the presence of a message cannot be detected. The goal of steganography is to hide messages inside other harmless messages in a way that does not allow any enemy to even detect that there is a second secret message present. One of the major weaknesses of cryptosystems is that even though the message has been encrypted, it still exists [6]. Therefore, the purpose of steganography is to complement cryptography and to avoid raising the suspicion of system attackers but not to replace cryptography Steganography has developed a lot in recent years because of advancements in the digital techniques being used to hide data. Almost any plain text, cipher text, image and any other media (audio, video) can be hidden using steganography.

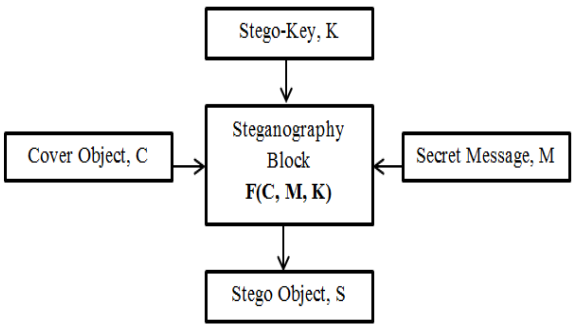

Fig 1. Block diagram of Steganographic system

Let C denote the cover carrier, and S the Stego object. Let K represent an optional key (a seed used to encrypt the message or to generate a pseudo random noise which can be set to {ϕ} for simplicity) and let M be the secret message we want to communicate. E_m is an acronym for embedding and E_x for Extraction. Therefore, the steganography process can be express as follows:

E_m: C ⊕ K ⊕ M = S (1)

$E_x(E_m(c, k, m)) \approx m, \forall c \in C, k \in K, m \in M$ (2)

Steganography modifies the carrier in an imperceptible way so that it reveals nothing neither the embedding of a message nor the embedded message itself. In general steganographic algorithms rely on the replacement of some noise component of a digital object with a pseudorandom secret message. The essence of steganography lies in the fact that the image bearing data should be statistically and visually identical to the original image so that an intruder cannot detect the presence of the hidden message. The secret message needs to be communicated to the destination is embedded into the cover image to derive the stego image. Steganalysis is the process of identifying steganography by inspecting various parameter of a stego media [4]. The primary step of this process is to identify a suspected stego media. After that steganalysis process determines whether that media contains hidden message or not and then try to recover the message from it.

In the realm of this digital world, steganography has created an atmosphere of corporate vigilance that has spawned various interesting applications, thus its continuing evolution is guaranteed. Steganography is employed in various useful applications, such as strong watermarks, military agencies, intelligence agencies, document tracking tools, document authentication, electronic money, radar systems remote sensing etc. Individual's details are also embedded in their photographs in smart IDs and identity cards.

The paper is organised as follows: Section II presents the measurement criteria of steganography followed by different categories of steganography in section III. Finally section IV discuss about the conclusion.

II. STEGANOGRAPHY TRIANGLE

Several important issues need to be considered while dealing with steganographic systems. They are steganographic robustness, capacity, and security. The association between them can be expressed by the steganography triangle shown in Figure 2. It represents balance of the desired characteristics associated with steganographic method. The main aim of steganography is to increase the steganographic capacity and enhance the imperceptibility while maintaining the robustness [17].

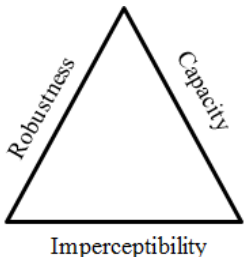

Fig 2. Measurement triangle of steganography

A. Capacity

Capacity is the maximum amount of secret information that can be embedded in a file. Capacity either can be defined as an absolute value in term of number of bits for particular cover or as a relative number regarding necessary bits to save final stego file. Capacity value depends on both embedding function and cover properties ($x^{(0)}$). P is a metric which shows proportion of length of the secret message relative to the maximum length of message can be embedded in cover. P value would be 0<p<1 and can be calculated by following formula:

$P=m/n, \quad x^{(0)} \in x^n$ (3)

Steganographic systems, mainly used for secret communication, aim to maximise the steganographic capacity and minimise the perception of hidden messages in stego images.

B. Imperceptibility

A steganographic system is perfectly secure if the statistics of the cover file and that of the stego file are identical. The higher fidelity of stego object, will give the better imperceptibility. This property would be satisfied if difference of resultant stego file be not distinguishable from original cover. Peak Signal to Noise Ratio (PSNR) is a metric to evaluate the ratio between possible maximum signal and influence of modifying noise to fidelity of its representation [3]. This metric can be calculated as follow:

$PSNR=10 \times \log (255^2 / MSE)$ (4)

$MSE=1/mn\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[f(x,y)-g(x,y)]^2$ (5)

Here f(x, y) and g(x, y) represent the pixel value at the position (x, y) in the cover-image and the stego-image respectively. The goal of the stego system is to achieve high PSNR value inorder to make steganography successful.

C. Robustness

Robustness is property of harness of eliminating secret information from stego file. While detection of embedded secret data has much higher importance than its removal, but property of robustness talks about resisting against intentional distortion of communication channel by means of systematic interface or channel noise aiming to ban use of steganography techniques [25]. Robustness metrics of steganographic algorithms are classified in distortion classes like geometric transformations or additive noise.

III. TYPES OF STEGANOGRAPHY

Steganography is differentiated on the basis of the media in which we hide the data. These are: text, image, audio, etc.

A. Text Steganography

Text steganography uses the text media to hide the data using different schemes like use of selected characters, extra white spaces of the cover text etc. [22]. There are different techniques to embed the secret data in text files. Format based method modifies the existing text to hide the data in such a manner that it involves the insertion of spaces, resizing the text, change the style of text. In random and statistical method characters are hidden that appeared in random sequence [1]. Statistical method determines the statistics such as mean, variance and chi square text which measure the amount of redundant information to be hidden within the text. Linguistics method is the combination of syntax and semantics. Syntactic steganalysis ensure the correct structure as the text is generated from grammar. In semantic method value is assigned to synonyms and data can be encoded to the actual word of text.

B. Image Steganography

In this method, images are used as cover object. There are two types of domains in which image steganography is categorized i.e. spatial domain & frequency domain [20]. In spatial domain, processing is applied directly on the pixel values of the image whereas in frequency domain, pixel values are transformed and then processing is applied on the transformed coefficients.

1) Spatial Domain Steganography: In spatial domain methods, a steganographer modifies the secret data and the cover medium in the spatial domain, which involves encoding at the level of the least significance bits (LSBs). To the human eye, changes in the value of the LSB are imperceptible, thus making it an ideal place for hiding information without any perceptual change in the cover object [12]. As LSB embedding takes place on noise, it is likely to be modified, and destroyed, by further compression, filtering, or a less than perfect format or size conversion. Hence, it is often necessary to employ sophisticated techniques to improve embedding reliability. Monica Adriana et.al [13] presented a spatial domain technique that applies LSB insertion to embed payload inside the color image. The algorithm is processed serially as well as in parallel by making use of threads to speed up the process of data hiding.

Masud Karim et.al [15] proposed a new method where the payload is stored into different position of LSB of image depending on the secret key. This method provides high security as attackers cannot extract information using general steganalysis tool, one needs the secret key to retrieve data. C.C. Chang et.al [5] proposed an adaptive technique applied to the LSB substitution method. They presented an approach that exploits the correlation between neighbouring pixels to estimate the degree of smoothness. Smoothness degree determines the number of LSBs to be used for embedding.

The current steganography tools that are based on the LSB algorithms include S-Tools, Hide and Seek, Hide4PGP and Secure Engine Professional [1]. These tools support BMP, GIF, PNG images and WAV audio files as the carriers. Each of these tools has unique features. S-Tools reduce the number of colors in the image to only 32 colors. Hide and Seek makes all the palette entries divisible by four. In addition, it forces the images sizes to be 320x200, 320x400, 320x480, 640x400 or 1024x768 pixels.Hide4PGP embeds the message in every LSB of an 8-bit BMP images, and in every fourth LSB of a 24- bit BMP image. These applications are flawed because they do not analyse the image file after it has been embedded with data to see how vulnerable it is to steganalysis. Furuta et.al [16] proposed a new spatial domain technique called Bit Plane Complexity Segmentation (BPCS) which is based on Most Significant Bit technique. BPCS provides the ability to hide data in higher bit planes of the cover image but it has low embedding capacity as changing most significant bits can cause significant changes in perception.

2) Frequency Domain Steganography: The transform domain based steganography tools embed the message in the transform coefficients of the image. DCT & DWT technique are implemented in frequency domain [7][11]. The discrete cosine transforms (DCT) and discrete wavelet transform (DWT) are mathematical function that transforms digital image data from the spatial to the frequency domain. In DCT, after transforming the image in frequency domain, the data is embedded in the least significant bits of the medium frequency components and is specified for lossy compression while In DWT, secret messages are embedded in the high frequency coefficients resulted from Discrete Wavelet Transform and provide maximum robustness. Abdelwahab and Hassan [19] propose a data hiding technique in the DWT domain. Both secret and cover images are decomposed using DWT (1st level). Each of which is divided into disjoint 4x4 blocks. Blocks of the secret image fit into the cover blocks to determine the best match.

Ali Al-Ataby et.al [8] proposed a modified high-capacity image steganography technique that depends on wavelet transform with acceptable levels of imperceptibility and distortion in the cover image and high level of overall security. Neda Raftari et.al [9] proposed a novel image steganography technique that combines the Integer Wavelet Transform (IWT) and Discrete Cosine Transform (DCT) is proposed which embeds secret image in frequency domain of cover image with high matching quality.

As for steganography in DCT domain, JSteg [1] is the classical JPEG steganographic tools utilizing the LSB embedding technique. JSteg embeds secret data into a cover image by successively replacing the LSBs of non-zero quantized DCT coefficients with secret message bits. Westfeld [10] introduced F5 algorithm. In this, the absolute value of the coefficient is decreased by one if it is needed to be modified instead of re-placing the LSBs of quantized DCT coefficients with the message bits. The advantage of F5 algorithm is that it minimizes the necessary number of changes to hide a message of certain length. OutGuess proposed a better technique which includes the ability to preserve statistical properties. It used pseudo random generator to select the DCT coefficients.

3) Adaptive Steganography: Adaptive steganography is a form of improved steganography which is a special case of the two former methods. It is also known as ''Statistics-aware embedding'', ''Masking'' or ''Model-Based'' Adaptive steganography considers statistical global features of an image before interacting with its LSB/DCT coefficients [18]. The statistics decides where to make the changes. It is characterized by a random adaptive selection of pixels depending on the cover image and the selection of pixels in a block with large local STD (standard deviation).

C. Audio Steganography

When secret data is embedded into digital sound, the technique is known as audio steganography. This method embeds the secret message in WAV, AU and MP3 sound files. There are different methods through which audio steganography explored:

1) LSB coding: Using the least-significant bit is possible, as modifications will usually not create audible changes to the sounds. Another method involves taking advantage of human limitations. It is possible to encode messages using frequencies that are inaudible to the human ear. Using any frequencies above 20.000 Hz, messages can be hidden inside sound files and will not be detected by human checks. In [26] Babul uses combined approach of DCT and LSB to implement audio steganography and provides high quality of steganography process in terms of Peak Signal–to-Noise Ratio (PSNR).

2) Parity coding: Instead of breaking a signal down into individual samples, the parity coding method breaks a signal down into separate regions of samples and encodes each bit from the secret message in a sample region's parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process flips the LSB of one of the samples in the region [23]. Thus, the sender has more of a choice in encoding the secret bit, and the signal can be changed in a more unobtrusive fashion.

3) Phase coding: Phase coding addresses the disadvantages of the noise inducing methods of audio steganography. Phase coding [28] relies on the fact that the phase components of sound are not as perceptible to the human ear as noise is. Rather than introducing perturbations, the technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to-perceived noise ratio.

4) Spread spectrum: In the context of audio steganography, the basic spread spectrum (SS) method attempts to spread secret information across the audio signal's frequency spectrum as much as possible [21]. This is analogous to a system using an implementation of the LSB coding that randomly spreads the message bits over the entire sound file. However, unlike LSB coding, the SS method spreads the secret message over the sound file's frequency spectrum, using a code that is independent of the actual signal. As a result, the final signal occupies a bandwidth in excess of what is actually required for transmission.

Nehru et.al in [24] has introduced a robust method of imperceptible audio data hiding. The system does not change the size of the file even after encoding and is also suitable for any type of audio file format. In [27] Tanmaiy et.al provided a combination of both models steganography as well as cryptography that provides better protection of the data from the intruders.

D. Protocol Steganography

The term protocol steganography refers to the technique of embedding information within messages and network control protocols used in network transmission. In the layers of the OSI network model there exist covert channels where steganography can be used. Ahsan and Kundur in [14] provided an example of where information can be hidden is in the header of a TCP/ IP packet in some fields that are either optional or are never used.

IV. CONCLUSION

Steganography is a powerful data hiding technique that provides a level of privacy while communicating data with others. This paper discusses several methods related with steganography. The research to device strong steganographic techniques is a continuous process. New techniques are developed inorder to implement steganography as well as to steganalysis. A steganographer needs to carefully select the cover object as well as a suitable steganographic technique in a particular domain. To provide a high level of security steganography-the concealment of data may not give the best result always. So one can incorporated steganography with cryptography.

REFERENCES

[1] N. Provos and P. Honeyman, "Hide and seek: An introduction to Steganography", *IEEE Conference on Security and Privacy*, pp. 32-44, 2003.

[2] C. Cachin, "An Information-Theoretic Model for Steganography", *Proceedings of 2nd Workshop on Information Hiding, MIT Laboratory for Computer Science*, May 1998

[3] "Peak Signal to Noise Ratio (PSNR)" from Wikipedia http://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio.

[4] Jessica Fridrich and Jan Kodovsky, "Rich Models for Steganalysis of Digital Images", IEEE *Transactions on Information Forensics and Security*, Vol. 7, No. 3, pp. 868-882, June 2012.

[5] C.C. Chang, P. Tsai, and M.H. Lin, "An Adaptive Steganography for Index- based images using Codeword Grouping", *Advances in Multimedia Information Processing-PCM, Springer*, Vol. 3333, pp. 731–738, 2004.

[6] "Cryptography" from Wikipedia, http://en.wikipedia.org/wiki/Encryption

[7] Moh Zan and Nyein Aye," A Modified High Capacity Image Steganography using Discrete Wavelet Transform", *International Journal of Engineering Research & Technology* (IJERT), Vol. 2,No. 8, pp. 2712-2715, August - 2013.

[8] Ali Al-Ataby and Fawzi Al-Naima, "A Modified High Capacity Image Steganography Technique Based on Wavelet Transform" *The International Arab Journal of Information Technology*, Vol. 7, No. 4, October 2010.

[9] Neda Raftari and Amir MasoudEftekhariMoghadam, "Digital Image Steganography Based on Assignment Algorithm and Combination of DCT-IWT", *Fourth International Conference on Computational Intelligence, Communication Systems and Networks*, 2012.

[10] Andrew Westfeld, "F5-a steganographic algorithm: high capacity despite better steganalysis", *Proc.of the 4th Information Hiding Work-shop*, vol. 2137, pp. 289-302, Springer, 2001

[11] T. Narasimmalou, Allen Joseph .R, "Optimized Discrete Wavelet Transform based Steganography", *IEEE International Conference on Advanced Communication Control and Computing Technologies* (ICACCCT), 2012.

[12] Rajbir kaur, Surbhi Gupta, and Parvinder S. Sandhu, "Randomized Steganography using Ycbcr Color Model Characteristics", *International Conference on Computer and Communication Technologies* (ICCCT' 2012) May 2012, Phuket.

[13] Monica Adriana Dagadita, Emil Ioan Slusanschi, and Razvan Dobre, "Data Hiding using Steganography", *IEEE 12th International Symposium on Parallel and Distributed Computing*, 2013.

[14] Ahsan, K. & Kundur, D., "Practical Data hiding in TCP/IP", *Proceedings of the Workshop on Multimedia Security at ACM Multimedia*, 2002.

[15] S. M. Masud Karim, Md. Saifur Rahman, and Md. Ismail Hossain, "A New Approach for LSB Based Image Steganography using Secret Key", *Proceedings of 14th International Conference on Computer and Information Technology* (ICCIT 2011) , December 2011.

[16] Furuta T., Noda H., Niimi M., and Kawaguchi E, "Bit-plane decomposition steganography using wavelet compressed video", *Joint Conference of the Fourth International Conference*, pp. 970 - 974, 2003.

[17] Manoj Kumar Ramaiya, Naveen Hemrajani, and Anil Kishore Saxena, "Improvisation of Security aspect in Steganography applying DES", *International Conference on Communication Systems and Network Technologies*, 2013.

[18] Raja, K. B., C. R. Chowdary, K. R. Venugopal, and L. M. Patnaik, "A secure Image Steganography using LSB, DCT and Compression techniques on raw images", *Third International Conference on Intelligent Sensing and Information Processing*, (ICISIP 2005), pp. 170-176, 2005.

[19] A.A. Abdelwahab and L.A. Hassan, "A discrete Wavelet transform based technique for image data hiding", *Proceedings of 25th National Radio Science Conference*, pp.1-9, March 2008.

[20] Moh Moh Zan and Nyein Aye, "A Modified High Capacity Image Steganography using Discrete Wavelet Transform", *International Journal of Engineering Research & Technology* (IJERT), Vol. 2, August 2013.

[21] Abdulaleem Z. Al-Othmani, Azizah Abdul Manaf and Akram M. Zeki, "A Survey on Steganography Techniques in Real Time Audio Signals and Evaluation", International Journal of Computer Science Issues, Vol. 9, No 1, pp. 30-37, January 2012.

[22] R Praveen Kumar and V Hemanth, Mshareef, "Securing Information Using Steganography", *International Conference on Circuits, Power and Computing Technologies*, 2013.

[23] K. Gopalan. , "Audio steganography using bit modification", IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03), Vol. 2, pp. 6-10, April 2003.

[24] Gunjan Nehru, Puja Dhar, "A Detailed look of Audio Steganography Techniques using LSB and Genetic Algorithm Approach", *International Journal of Computer Science Issues*, Vol. 9, No 2, January 2012.

[25] D. Sandipan, A. Ajith, and S. Sugata," An LSB Data Hiding Technique using Prime Numbers", *Third International Symposium on Information Assurance and Security, Manchester, UK, IEEE CS press*, 2007.

[26] Linu Babu1, Jais John, Parameshachari, Muruganantham, H. S. Divakara Murthy, "Steganographic Method for Data Hiding in Audio Signals with LSB & DCT", *International Journal of Computer Science and Mobile Computing*, Vol. 2, No. 8, pp. 54-62, August- 2013,

[27] Tanmaiy G. Verma, Zohaib Hasan, Dr. Girish Verma, "A Unique Approach for Data Hiding Using Audio Steganography", *International Journal of Modern Engineering Research* (IJMER), Vol. 3, No. 4, pp-2098-2101, Jul - Aug. 2013.

