A Review of Steganographic Methods and Techniques

Abdesselam Beroual, Imad Fakhri Al-Shaikhli
Computer Science, International Islamic University Malaysia, Malaysia.
berslem@hotmail.fr
Computer Science, International Islamic University Malaysia, Malaysia.
imadf@iium.edu.my

Abstract— The frequent and intensive transmission of information on the internet requires protection techniques to secure sensitive information. Steganography is one of many security techniques that consist of concealing information inside an appropriate multimedia (e.g. image, video, audio…etc.) with the objective to hide the embedded information. Many challenges confront steganography systems like capacity (size of data embedded) and imperceptibility (level of undetectability). These two aspects are inversely proportional to each other which produces a data hiding dilemma. This paper presents a review of the steganography methods used in recent years and a critical analysis based mainly on capacity measurement and, secondly, on imperceptibility measurement and steganography domain.

Keywords— Steganography, capacity, properties of steganography, types of steganography.

I. Introduction

Recent years have seen a surge in the use of intensive transmission of information via public communication channels like the internet. When the information is considered sensitive, it is very important to protect it from unauthorised parties. Hence, a strong security system is involved. Steganography is one of these security systems used to secure transmission between the sender and receiver.

A. Steganography

Steganography is a science of hiding information inside an appropriate container (that can be as a text file, image, video, audio, etc.) in a way that prevents detection of secret data by unintended recipients [1]. A successful steganography system is considered if the secret message is not revealed; otherwise, it is broken. The most important aspects of any steganographic method are the high level of imperceptibility (also, called undetectability) and the amount of secret data to embed in the multimedia carrier, (also called capacity) [2]. In the popular form of digital steganography called container modification, the carrier is altered in the embedding process and then sent to the receiver which extracts the data. The problem is that every modification of the container changes its statistics slightly (e.g., histogram). If an adversary (warden) gets access to the carrier, he will try to check it for the presence of the secret message. The actions taken by the warden to find hidden information are called steganalysis and in many cases are based on statistical distribution [3].

B. Terminology

- *Cover Object/Container/Carrier/Digital Medium: These terms refer to the input files such as image, video, audio or text files in which concealment of secret data is to be performed.*
- *Stego-Object: The result should be after the embedding of secret data into the cover object, the cover object becomes a stego-object.*
- *Embedding: Is the process of concealment of a secret message inside a digital medium.*
- *Extraction: Is considered the reverse process of embedding, so the embedded message is recovered from the stego-object to be read.*
- *Message/data: Refers to the secret information that is to be embedded in the cover object for secure transmission from sender to receiver.*

C. Types of steganography

- *Text Steganography: It comprises concealing data inside the text files. In this*

technique, the secret information is hidden in the back of each nth letter of every word of a text message. Many methods are available for concealing data in the text file. These methods are Format Based Method, Random and Statistical Method, Linguistics Method.

- *Image Steganography: Is the most popular method used. The secret message is embedded in a digital image as a cover by using an embedding algorithm. The result of this process is a stego-image which will be sent on the transmission channel. The receiver will extract the message using a specific extraction algorithm. All along the transmission of the stego-image, the unauthorised parties can only notice the transmission of an image but cannot expect the existence of a hidden message.*
- *Audio Steganography: It consists of hiding data in audio files. This method conceals the data in WAV, MP3 and AU audio files. Many different methods of audio steganography exist like Low Bit Encoding, Phase Coding and Spread Spectrum.*
- *Video Steganography: It is a technique of hiding data or files in a digital video file. In this method wherein the video is considered a "combination of pictures" is used as a carrier for hiding the secret data. Also, many techniques are used in this field.*

D. Techniques of steganography

There are many techniques in the field of steganography, are presented. Their classification or categorisations are different, and there is no claim of a unique classification. In this proposal, the focus is on the techniques widely used by the four steganographic types mentioned above. Here, the classification is divided into three main groups depending on the domain used for embedding. These three categories are spatial domain, transform domain, and adaptive domain.

1) Spatial domain:

It encodes the secret information in significant parts of the cover image and shares it with the receiver only, and thus the secrecy of data is increased as the attacker must have knowledge about the pixels where data has been hidden. It is one of the simplest techniques that creates a covert channel in the original image in which changes are likely to be scant when compared to the human visual system. The spatial domain techniques are classified into different categories and are mostly used is the LSB technique. Details of spatial techniques can be referenced in [4] and [5].

2) Transform domain

In the transform technique, the secret message is embedded in the frequency coefficients of the cover image. It is a more complex way of hiding messages in an image compared to other steganographic techniques of hiding. It can hide a larger amount of data with high security, invisibility, lossless transmission of data compared with the spatial technique. An extra work added to this technique is that after changing the values to binary frequency DCT coefficients are also calculated followed by embedding the secret information. In this technique, the most used types are DWT, DCT and DFT. More details can be referenced in [4].

3) Adaptive domain

The adaptive domain is a recent domain that uses techniques from both spatial and transforms domains.

II. DATA EMBEDDING ALGORITHMS

This section reviews data embedding algorithms presented in the literature. The focus of this paper is on algorithms and techniques used to enhance the capacity and imperceptibility. Algorithms focusing on the enhancement of security and robustness fall outside the scope of this paper, however, many works have been done concerning the enhancement of these last two features in different types of steganography. In the audio type (like in [6], [7] and [8]) the authors proposed techniques prioritising robustness of data embedding in audio covers. In the image type, like in ([9], [10], [11] and [12]) they produced a more secure system in comparison with previous works about image steganography, and in the video type such as works in ([13], [14] and [15]) they showed amelioration in terms of security of the embedded message and robustness against attacks.

Algorithms proposed to achieve a high capacity are reviewed next, followed by those aimed at providing good imperceptibility. Table 1 summarises the reviewed works that recently proposed for hiding secret data.

In [16] the author tried to provide a clear performance measurement on diverse models of concealing image and/or text data in the image steganography by constructing four models. The

first two are based on the existing security enhanced LSB based approaches for hiding image and text separately. The second two methods proposed (third and fourth) which present a novelty, are hybrid models that allow hiding both the image and text together in the carrier image. The third hybrid model is a direct implementation of manipulating image and text data combined, and the fourth hybrid model uses a method that is optimised for manipulating the image and text bit streams simultaneously.

In [17], the authors proposed a method in the spatial domain that improves the capacity and security, it consists of dividing the image into two parts; one is reserved for embedding the secret data, the other is used to indicate which change is applied to each pixel exist in the first part.

In [18], the authors proposed a scheme about reversible steganography in encrypted images based on feature mining with increased capacity in plaintext domain. The authors used two multi-granularity encryption and residual histogram shifting. First of all, the cover image is encrypted on fine-grained level and coarse-grained level together with content-owner key. Thereafter, additional data can be hidden into the encrypted image by scouting both the similarity of neighbouring pixels in local level and residual histogram in global level with a data-hiding key. If both content-owner and data-hiding keys are adopted at the same time, the cover can be restored with faultless during the data extraction.

In [19], the authors used a spatial domain method to introduce an LSB technique that consists of hiding the secret message by using a dynamic substitution in the LSB of pixels of an image instead of hiding the message sequentially.

[20] presented a new concept of multi-secret and false digital image steganography. The idea consists of embedding in a single carrier more than one message. One (or more) of the hidden secrets is/are a real message and the others are false messages. The real message contains essential data intended to be securely transmitted between authorised parties; the false is considered bait for focusing attention on an unimportant message.

In [21], the authors proposed a new embedding technique of audio steganography in the spatial domain to increase the carrier medium capacity for substitution additional hidden message. In comparison with the previous works that use four LSBs as a maximum number of bits without significant effect on host audio signal for LSB audio steganography, this paper used seven LSBs to embed the secret message into variable and multiple LSBs layers.

In [22], the author proposed a video steganography in the spatial domain that consists of extracting video frames using video reader function. These frames have been collected for extraction of three different colour regions, namely red, green and blue. These true colours have been used for hiding the secret information. The multiple least significant bits have been computed from three different colour regions. Secret information that has to be embedded in the cover object has been converted into the binary format, and the XOR operation has been used for embedding the information behind the cover object. These frames have been recombined after embedding secret information behind the cover object. The frame that has been recombined provides stego-video that can be transmitted to the receiver side for extraction of secret information.

In [23], the authors used a frequency domain method based on integer lifting wavelet transform (LWT) where the cover image is partitioned into non-overlapping 8x8 blocks, and 2D Integer LWT is applied to each block. To identify proper location of the secret message, a Tree Scan Order (TSO) is performed in transformed blocks. To enhance the security, the authors used an encryption method for the secret message before embedding it.

In [24], the author used a frequency domain method based on irreducible polynomial mathematics. The technique can be represented with two schemes: Data Hiding Approach based on Mix Column Transform (DHAMCT) and Enhanced Data Hiding Approach based on Mix Column Transform (EDHAMCT). After dividing a colour image into blocks, it applies the proposed transformed method to the specify block and hides the secret message therein.

In [25], the author proposed a new method of MP3 steganography in the frequency domain. The objective is to emphasise increasing the robustness of data embedding based on the MP3 statistical properties using the Modified Discrete Cosine Transform (MDCT) technique to compress these properties. For this purpose, selected MDCT coefficients are modified through MPEG audio compression procedure. In this algorithm, the secret message is embedded directly after achieving the MDCT coefficient and before quantization in the MP3 encoder.

[26] introduces a scheme for hiding data in audio files. Its principle is making modifications in the amplitude of the cover audio file and uses a

key to increasing the security purpose to hide the secret message.

In [27], the authors presented a new steganography algorithm for video based on the multiple object tracking algorithm and Hamming codes. This algorithm includes four stages: Secret Message Pre-Processing Stage: the secret message is a text file, after converting all characters into ASCII code in a binary array, the author used a key for encryption. The encrypted array is divided into 11-bit blocks. Then, every block is encoded by the Hamming codes. Motion-Based Multiple Object Tracking Stage: This stage detects each moving object within an individual frame, and then associates these detections throughout all of the video frames. The background subtraction method is applied to detect the moving objects. Then, the Kalman filter is used to predict estimation trajectory of each moving object. Data Embedding Stage: The motion regions are identified and tracked through the video frames. The region of interest changes in each frame based on the size and the number of the moving objects. The algorithm is applied to predict trajectories of all moving objects. Data Extraction Stage: The stego-video is divided into frames through the receiver, and then two keys are extracted from the non-motion region of the first frame.

The author in [28] proposed a method in video steganography that transforms the video frames to YCbCr colour space then embed a secreted message in a particular region of interest. The particular region of interest is selected using the skin detection algorithm. The secret message is embedded in the selected region of interest of a frame having the least MSE.

In [29], the authors proposed a general multiple-base data embedding (GMB) algorithm, which carries a secret M-ary digit in a pixel-cluster consisting of n pixels. The proposed algorithm presented many advantages. First, it provides a multiple-purpose style producing a high quality of images or providing a large payload for feasible applications. Second, it is possible to accurately predict the overall capacity and the stego-image quality using mathematical expressions without conducting real message embedding. Third, a coefficient mapping technique increases the security and resists steganalysis attacks from visualising the histogram of stego-images. Finally, according to the experiment of this work, when offering the capacity of 1.0 bpp, the proposed GMB algorithm can resist steganalytic algorithms including RS steganalysis, and the SPAM steganalyzer can be used in steganographic applications.

In [30], the author used an adaptive domain method based on multilevel simple embedding technique. The main idea in this approach is to use three methods of audio steganography on a single audio file instead of using a single method. This paper proposed hiding three messages in one audio file through three levels. The first level is LSB coding; the second is parity bit coding followed by the level of spread spectrum method.

In [31], the authors proposed an adaptive data hiding algorithm for video files. The main process is creating a skin map for each frame using an adaptive skin detection algorithm with reduces the number of false positives. Then, the skin map is converted to a skin-block-map to eliminate the error-prone skin pixels that can result in inefficient retrieval of the hidden data. To increase the robustness, the authors conducted embedding process by using a wavelet quantization technique over the red and blue channels of the host frames.

The work in [32] focused on the capacity performance by using three methods namely: Changing in Alphabet Letter Patterns (CALP), Vertical Straight Line (VERT) and Quadruple Categorization (QUAD). These methods use text files containing hidden text, and this hidden text is converted to binary bits before being applied in the embedding process. The CALP method tries to manipulate English letters by mapping the binary sequence of the hidden text through pattern changes of several letters of the cover text during the embedding process. These pattern changes have been incorporated using some unused symbols of the ASCII number system. The author used the VERT method by selecting English letters divided into two groups. The letters contain one vertical straight line identified as a G1 group which will hide 1 bit hidden data. Whereas, a letter containing more than one single line or does not contain a vertical straight line is identified as a G2 group and will hide 0 bit hidden data.

In [33], a text steganography technique is used to enhance some steganographic properties. To enhance the imperceptibility, Zero Distortion Technique is applied on the text. For achieving a large amount of data to be embedded, the abbreviation method is used, and for the security purpose, the author used the Indexed Based Chaotic Sequence encryption.

[34] proposes a new text steganography method for hiding text messages into SQL

carriers. This technique is considered a generation-based method that generates SQL queries out of the secret message using a dictionary of words organised into 65 categories with no common words. These categories represent 65 different characters including the 26 letters of the English language, the 10 digits of the decimal system, and a set of 29 special characters. The word to generate is chosen from the dictionary, so, when changing the dictionary content, this would result in a different SQL query.

In [35], a text steganography method was proposed using Huffman coding. The secret data is hidden in the forward email platform. The authors used the characters' number used in the email id to refer to the secret data bits, and they took from the processed secret data, the characters will be added to the email id for an optimal utilisation of a number of characters. The new characters are supplemented before the symbol of '@'.

- Discussion of the literature

Through reading and analysing the different papers from this literature and others, we observe that most of the mentioned techniques and methods have improved (or tried to improve) one property at the expense of other properties, even though the other properties have showed some improvement, only that this improvement is clearly modest in comparison with the improvement of the main property focused by the paper. The majority of these reviewed papers were aimed to ameliorate the capacity property; however, almost all of them have achieved a size of capacity less than 20%, except three papers where they increased the capacity up to 50% as in [19], 52% in [17] and 55% as in [21]. From other papers, some have slightly ameliorated the capacity in comparison with their previous works as in [16, 27, 28, 32, 35], some papers have slightly balanced between the capacity and imperceptibility like in [18, 24, 33]. [25, 31, 34, 26, 29] have showed balanced performance in terms of capacity, security and robustness. In [20, 22, 23, 30], the authors achieved good imperceptibility.

As this research interests in the property of capacity, we found in these reviewed papers that their maximum amount of embedded data is less than 56%, where the two biggest sizes are 55% and 52% in the audio type and image type respectively. Moreover, almost all these papers have common weakness in terms of capacity which is the embedding of the whole size of secret data in the cover media. Hence, this paper tries to conduct a novel technique for overcome this weakness as well as the trade-off between the two properties (capacity and imperceptibility).

III. Conclusions

This paper highlights concepts regarding popular steganography techniques and algorithms. There are different types of steganography methods available. The number of methods is not restricted. Here the importance is given to methods focusing on the capacity property, and secondly on imperceptibility. Other properties of a steganogrpahic system "robustness and security" are mentioned briefly in the literature section. According to this paper, the maximum size of embedded data reached by the previous works in this literature review is 55% and 52% in [21] and [17] respectively. This technique consists of transforming the secret data on the base of the cover into a small part considered a sequence of references. This sequence is the subject to be embedded instead of the original secret data. From the preliminary work, the original size of secret data is reduced significantly, and embedded without distortion.

References

[1] Subhedar, Mansi S., and Vijay H. Mankar. *"Current status and key issues in image steganography: A survey."* Computer science review 13 (2014): 95-113.

[2] Barni, Mauro. *"Steganography in Digital Media: Principles, Algorithms, and Applications (Fridrich, J. 2010) [Book Reviews]."* IEEE Signal Processing Magazine 28.5 (2011): 142-144.

[3] Cheddad, Abbas. *Digital Image Steganography: Concepts, Algorithms, and Applications*. VDM Publishing, 2009.

[4] Solanki, Roshni, Monika Chuahan, and Madhavi Desai. *"SURVEY OF IMAGE STEGANOGRAPHY TECHNIQUES."*, IJARESM, ISSN: 2394-1766.

[5] Anandpara, Dimple, and Amit Kothari. *"Working and comparative analysis of various spatial based image steganography techniques."* International Journal of Computer Applications 113.12 (2015).

[6] Bazyar, Mohsen, and Rubita Sudirman. *"A Robust Data Embedding Method for MPEG Layer III Audio Steganography."* International Journal of Security and Its Applications 9.12 (2015): 317-327.

[7] Kamalpreet Kaur and Er. Deepankar Verma, *"Multi-Level Steganographic Algorithm for Audio Steganography using LSB, Parity Encoding and Advanced LSB Technique"*, International Journal of Advanced Research in Computer Science, Volume 4, No. 9, July-August 2013

[8] Dieu, Huynh Ba, and Nguyen Xuan Huy. *"An improved technique for hiding data in audio."* Digital Information and Communication Technology and its Applications (DICTAP), 2014 Fourth International Conference on. IEEE, 2014.

[9] Lwin, Thandar, and SUWAI PHYO. *"Information Hiding System Using Text and Image Steganography."* International Journal of Scientific Engineering and Technology Research 3.4 (2014): 1972-1977.

[10] Manjula, Y., and K. B. Shivakumar. *"Enhanced secure image steganography using double encryption algorithms."* Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016.

[11] Mohan Megha, and Anitha Sandeep. *"Multiple security enhancements for image steganography."* Inventive Computation Technologies (ICICT), International Conference on. Vol. 1. IEEE, 2016.

[12] Verma, Vaidehi, and Trapti Ozha. *"Enhancing the Security and Quality of Image Steganography Using a Novel Hybrid Technique."* International Conference on Smart Trends for Information Technology and Computer Communications. Springer, Singapore, 2016.

[13] Ramalingam, Mritha, and Nor Ashidi Mat Isa. *"A data-hiding technique using scene-change detection for video steganography."* Computers & Electrical Engineering 54 (2016): 423-434.

[14] Zhang, Hong, Yun Cao, and Xianfeng Zhao. *"Motion vector-based video steganography with preserved local optimality."* Multimedia Tools and Applications 75.21 (2016): 13503-13519.

[15] Sharma, Shikha, and Devendra Somwanshi. *"A DWT based Attack Resistant Video Steganography."* Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. ACM, 2016.

[16] Ouyang, Linqiang, Jin H. Park, and Harbhinder Kaur. *"Performance of Efficient Steganographic Methods for Image and Text."* Journal of Advances in Information Technology Vol 7.1 (2016).

[17] Marghny H. Mohamed, Loay M. Mohamed, *"High Capacity Image Steganography Technique based on LSB Substitution Method"*, Applied Mathematics & Information Sciences 10(1):259-266• January 2016.

[18] Zhaoxia Yin, Wien Hong, Jin Tang and Bin Luo *"High capacity reversible steganography in encrypted images based on feature mining in plaintext domain"*, Int. J. Embedded Systems, Vol. 8, Nos. 2/3, 2016.

[19] Rashid, Aqsa, and Muhammad Khurrum Rahim. *"Critical Analysis of Steganography "An Art of Hidden Writing"."* International Journal of Security and Its Applications 10.3 (2016): 259-281.

[20] Ogiela, Marek R., and Katarzyna Koptyra. *"False and multi-secret steganography in digital images."* Soft Computing 19.11 (2015): 3331-3339.

[21] Bazyar, Mohsen, and Rubita Sudirman. *"A New Method to Increase the Capacity of Audio Steganography Based on the LSB Algorithm."* Jurnal Teknologi 74.6 (2015): 49-53.

[22] Kaur, Jaspreet, and Mrs Naveen Kumari. *"Steganography Over Video Files Using Multiple Least Significant Bits."* International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) 4.8 (2015): PP-127.

[23] Seyyedi, Seyyed Amin, Vasili Sadau, and Nick Ivanov. *"A Secure Steganography Method Based on Integer Lifting Wavelet Transform."* IJ Network Security 18.1 (2016): 124-132.

[24] Wafaa Mustafa Abdullah, Abdul Monem S. Rahma, Al-Sakib Khan Pathan,"*Mix column transform based on irreducible polynomial mathematics for color image steganography: A new approach*", ELSIVIER Computers and Electrical engineering 40(2014)1390-1404.

[25] Bazyar, Mohsen, and Rubita Sudirman. *"A Robust Data Embedding Method for MPEG Layer III Audio Steganography."* International Journal of Security and Its Applications 9.12 (2015): 317-327.

[26] Dieu, Huynh Ba, and Nguyen Xuan Huy. *"An improved technique for hiding data in audio."* Digital Information and Communication Technology and its Applications (DICTAP), 2014 Fourth International Conference on. IEEE, 2014.

[27] Mstafa, Ramadhan J., and Khaled M. Elleithy. *"A New Video Steganography Algorithm Based on the Multiple Object Tracking and Hamming Codes."* Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on. IEEE, 2015.

[28] Khupse, Sneha, and Nitin N. Patil. *"An adaptive steganography technique for videos using Steganoflage."* Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on. IEEE, 2014.

[29] Chen, Wei-Sung, et al. *"A novel general multiple-base data embedding algorithm."* Information Sciences 358 (2016): 164-190.

[30] Kaur, Ramandeep, et al. "Enhanced Steganographic Method Preserving Base Quality of Information Using LSB, Parity and Spread Spectrum Technique." Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on. IEEE, 2015.

[31] Sadek, Mennatallah M., Amal S. Khalifa, and Mostafa GM Mostafa. *"Robust video steganography algorithm using adaptive skin-tone detection."* Multimedia Tools and Applications 76.2 (2017): 3065-3085.

[32] Osman, Baharudin, Roshidi Din, and Mohd Rushdi Idrus. *"Capacity performance of steganography method in text based domain."* ARPN Journal of Engineering and Applied Sciences 10.3 (2015): 1345-1351.

[33] Yadav, Virendra Kumar, and Saumya Batham. *"A novel approach of bulk data hiding using text steganography."* Procedia Computer Science 57 (2015): 1401-1410.

[34] Bassil, Youssef. *"A Generation-based Text Steganography Method using SQL Queries."* arXiv preprint arXiv: 1212.2067 (2012).

[35] Kumar, Rajeev, et al. *"A high capacity email based text steganography scheme using Huffman compression."* Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference on. IEEE, 2016.