Indian Journal of Science and Technology, Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/90457, July 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Extensive Survey of Data Hiding Techniques

A. Rasmi and M. Mohanapriya*

Computer Science and Engineering, Karpagam University, Coimbatore - 641021, Tamil Nadu, India;
rzwsa@yahoo.com

Abstract

Background/Objectives: In the digital age because of the widespread use of internet, data hiding in digital imagery plays a vital role to ensure copyright protection and robustness from malicious attacks. **Methods/Statistical Analysis:** This research work explores an overview of the recently existing data hiding techniques used for the hidden exchange of secret information, from its earliest instances through potential future application. We enlighten different data hiding methods termed as watermarking, cryptography and steganography which have been proposed in last few years, some of them are flexible and simple for data hiding purpose. **Findings:** The potency of these techniques is based on certain features like robustness, imperceptibility and data embedding capacity which enables the users to evaluate the performance and efficiency of these techniques. **Application/Improvements:** It has great importance in the nearby future especially in the fields like intelligence agencies, military agencies, and cyber crime, so it will be more beneficial for researchers to develop innovative techniques.

Keywords: Cryptography, Data Hiding, Steganography, Techniques, Watermarking

1. Introduction

Data hiding defines a class of procedures used to embed data, such as text, image, audio or video, into various forms of media such as image, audio, or text. Since the ancient times secret hiding was used by spies and military intelligence operatives, or agents of companies to hide secret message inside written letters and speech. Compared with the traditional analog methods making seamless alteration is much easier on digital media. Nowadays with the advent of web technology the digital data such as text, image, audio and video can be transmitted by fast internet facility at high speeds. But the data transfer through internet is not secure, leakage of data content may lead to serious impact on social and personal life. So there is a need to hide the secret data inside other types of digital data, because security and integrity of data is crucial. Several methods have been proposed and used for protecting information from hackers. Existing data hiding methods are watermarking, steganography and cryptography and each has its own merits and demerits, radically all data hiding should follow the properties like high capacity, robustness, security, payload and reliability. But based on the degree of security of data transfer certain method can be applied, thus it can be stated as the techniques used for data inserting may vary depending on the quantity of information to be hidden[1]. Figure (1) depicts the existing data hiding techniques.

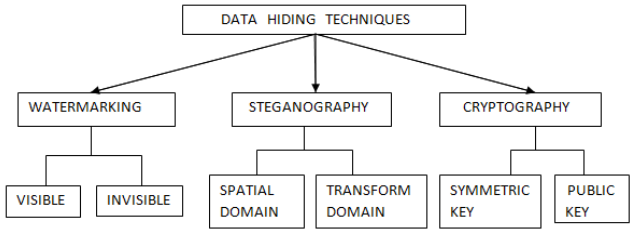

Figure 1. Existing data hiding techniques.

2. Watermarking

Watermark is normally a small amount of data that is used to indicate the ownership the particular object or data file. The watermark may be a signature of the author placed in the document for pride of authorship. More than 800 years ago, water marks were used in Italy to

*Author for correspondence

indicate the paper brand and the mill that produced it. Digital water marking provides security, authentication, transaction tracking and copy right protection for data when the image has been changed by image processing operation. Watermarking is an important application for the publishing and broadcast industries like TV, audio etc. A water mark may be a text or image imprinted onto the images, it tells us that who is the actual owner of the object. It hides a few bits of information, so the original and modified signal should be perceptually similar. Water marking is the method of imperceptibly change a cover to embed a data about that cover, without modifying the background of the original image, the data has to be very robust against all attacks and damage of the image. A typical example is a map, which takes time, and effort to create a road map of a town. Once the map is digitized, it becomes easy for someone to copy it, fabricate minor alterations, and put on the market as the original one. So in this case the hidden watermark may help the original owner, notice any attempts to take this work. Different available applications of watermarking are copy right protection, content authentication, forensics and piracy deterrence, and broadcast monitoring. The data to be embedded contains binary bits and content textures, so these factors are appropriate for application such as image authentication[2]. The required properties of digital water marks are robustness, security, perceptual invisibility and restrictions on computational complexity of embedding and extraction operation. Content based properties are more useful in watermark authentication in smooth areas .Water marking is classified into two types, visible watermarking and invisible watermarking. Visible watermarks are visible on the image typically logos or text. Invisible watermarks are not visible or perceivable. Watermarking enables the user to place an indelible mark on an image[3].

3. Steganography

The word steganography is originally derived from Greek words which literally mean covered writing. Steganography is an ancient art of conveying information in a secret way, and is the technique of concealing data within seemingly innocuous carrier which cannot be detached without drastically varying the data in which it is embedded. Early in the World War II Steganographic technology consisted of invisible inks, such as milk, vinegar and fruit juice. It is the art and science of writing hidden messages in a secure way such that only the sender and intended receiver can understand the existence of data. One of the first documents describing steganography is from the histories of Herodotus. In ancient Greece, text was written on wax covered tablets, the tablets appeared to be blank and unused so they passed inspection without any question. The advantage of steganography is that it can be used to transmit data securely without being discovered by eavesdropper, that is the quality of image is not so much altered, so the possibility of attacking by third party is less[4]. The goal of steganography is to hide message inside other harmless message in a way that does not allow any spy to even detect that there is a secret data present; it provides imperceptibility, so the most adaptive idea is to embed the data by minimizing the number of changes caused in the stego. Thus the main challenge in data hiding is how to embed message efficiently in a cover file, it can be implemented by considering certain criteria such as the embedded data should never exceed the size of the cover, because it might be noticed by intruders or attackers, so for implementing this identifying a cover file's redundant bits, then replacing these bits with data from the hidden information[5]. With the advancement of digital signal processing and information technology steganography moved from analog to digital. Steganographic embedding in dark colour images could be applied with a high degree of undetectability. In the branch of this digial world steganography has created various interesting applications. This digital revolution will surely help the intelligence agencies, military agencies, and cyber crime etc. The main objective of steganography is to hide the fact of communication undetectable from an unauthorized access, thus it ensures security. In this the sender embeds hidden content in unremarkable cover media where only the receiver can extract and understand the message, almost all digital file formats can be used in this. The main terminologies used in steganography are cover image, message and stego image. Original image which is used as a carrier for hiding information is known as cover. It may be a text, image, audio, video or protocol. Actual information that the sender wishes to remain confidential is the message, and it may be text, image audio, video or any other type of data that can be represented by bits. Embedded data is known as payload. After embedding message into cover image is known as stego image. Steganography technique is similar to sending letter from one person to other, here the envelope without letter acts

as cover, so after inserting letter into envelop it becomes stego. The letter is the payload or message.

In the Steganographic system, the sender should select the appropriate data carrier like image, video, text etc before inserting the data. The message embedding technique mainly depends on the structure of the cover, it cloak the data so it cannot be seen. The detection of tampering relies on both the embedding mechanism and the embedded data. For the secure data transfer using steganographic techniques, the foe should not be familiar with the cover image, otherwise if it is a well known one to them, then divide and embed the message bits in a random manner, thus it ensures security to an extent. Generally steganography can be defined as the compilation of creating a stego image and extracting the secret data from the stego image. The stego quality can be determined by using Peak Signal to Noise Ratio (PSNR)[6,7].

Based on the type of cover object used different categories are derived, they are image steganography, network steganography, video steganography, audio steganography and text steganography. In image steganography cover object is the image, it is the commonly used steganography method. When taking cover medium as network protocol, such as TCP, IP, and UDP etc is known as network steganography. Video steganography is the method of concealing some secret data inside a video file. Audio steganography is the process of hiding some secret information inside an audio file. Audio steganography uses audio formats like WAVE, MIDI etc for steganography process. If the data are embedded in a text file, and the outcome is a stego text is known as text steganography. Images are the most trendy files employed for data hiding, these image formats on the web are Graphics Interchange Format (GIF), Joint Photographic Experts Group (JPEG), Portable Network Graphics (PNG) and Bit Map Format (BMP). Joint photographic experts group(JPEG) is the most popular file format for image processing, so embedding data into JPEG files offer better masking, its data inserting rate is measured in bit per non zero coefficient. The strength and weakness of steganographic methods can be measured by using certain factors like embedding capacity, perceptual transparency, robustness, undetectability.

3.1 Robustness

This is the measure of the ability of the embedded data to retain intact even if the stego image may undergo transformations such as linear and non linear filtering, scaling, cropping, and lossy Compressions. Robustness means resistance to blind attacks and common image modifications.

3.2 Embedding Capacity

It defines the maximum amount of data that can be hidden in a cover image compared to the cover size. It is measured in bit per bit (bpb). Embedding capacity is known as payload, which is used to hide the data during the exchange of data.

3.3 Perceptual Transparency

The strength of perceptual transparency lies in its ability to be unnotified by the human visual systems, so after embedding data into cover image, perceptual transparency of the stego will be degraded as compared to cover. It can be defined as the invisibility of hidden data in stego image.

3.4 Undetectability

It is required for secure covert data transmission. The embedded information is undetectable if stego image is consistent with a model of the source from which images are drawn. If a particular image is examined and founds larger distortion than the original one, it may cause suspicion. Thus a good steganographic method should not change the features of the cover media. This property is termed as undetectability.

The existing steganographic techniques are spatial domain techniques and transform domain techniques. These methods are based on whether the pixels of the image are modified directly or indirectly, the former one is spatial while the latter is transform. An image can be represented as an array of pixels that represent the light intensities at various parts of the image. Each pixel is generally described as 8 bit or 24 bit. Grey scale images use 8 bits for each pixel, so it forms 2^8 combinations, and its value ranging from 0 to 255, 256 grey scale values for black and white images. In grey scale images the intensity of the pixel can be represented by using the value ranging from 0 to 255, and able to display 256 different shades of grey. For piercing the secret data into cover image, we have to decompose the grey scale of each pixel into eight bits, so the plane formed by the grey scale image is called a bit plane[8,9].

Colour images use 24 bit pixel and has 2^{24}colour combinations. In colour image each byte represents red, green and blue respectively. Each byte can have a value ranging

from 0 to 255, the darkest colour value is 0 and the brightest value is 255. Thus one pixel can have 256 shades of R, G and B respectively. Consider a colour pixel made up of 3 bytes as 00000000, 11111111, 11111111. The first 8 bits represent red, the second 8 bits represent green and the third group represent blue. In the above mentioned case red byte has the lowest value whereas in green and blue bytes have the highest value. Hiding of data can be done firstly by choosing redundant bits in a cover image, then select the redundant bits that can be customized without corrupting the integrity of the cover file. So embedding capacity can be measured by how much data can be hidden in a carrier file before it becomes noticeable by trespasser[10].

4. Spatial Domain Techniques

In spatial domain techniques the secret messages are embedding directly into the coverfile, it is also known as image domain. Spatial domain based methods embed data in the intensity of pixels of the image directly and in which the data hiding is performed directly on the pixel values of the cover image in such a way that the effect of the message is not visible on the cover image. Compared to frequency domain spatial domain methods are simple and computationally fast. In this domain the most common and simplest method is the Least Significant Bit (LSB) based steganographic method, it conceals the data in the least significant bit without any perceptible distortions. In this method the message is firstly preprocess, it will reduce the data size, after that it will be embedded into the least significant bits of the image based on pixel intensity.

The data embedding capacity can be increased by choosing two or more bits in each pixel. Embedding of bits can be performed either simply or randomly the amount of embedding data into an image is based on the size of the cover image. If the size of payload exceeds that of cover image then that will produce some distortions in the stego image, thus it will be perceptible by the intruders. So for embedding data into the regions which produce least distortions are selected, the common regions of interest are edge pixel, skin pixel, corner pixel. In this cover image is first decomposed into bit planes and then the least significant bit of the images are replaced with secret data bits. Different available methods in spatial domain are: A common and simple approach of hiding data within an image is LSB substitution

- LSB(least significant bit)
- PVD(Pixel value differencing)
- EBE(Edge based data embedding)
- RPE(Random based embedding)
- GLM(Gray level modification method)

A common and very popular methodology of hiding data within an image is LSB substitution.LSB based steganography is divided into LSB substitution and LSB matching technique. In LSB substitution method as the name implies replaces the least significant bit in some bytes of the cover image pixel with the information to be hidden. Suppose we have the following binary representation for the image.

10010101 00001101
10010110 00001111

And we want to embed the 4 bits of data: 1011 into the cover, so we get the following after applying LSB technique that is the stego image, here we are replacing the least significant bit of the cover with the secret data bits so we get the resultant stego.

1001010**1** 0000110**1**
1001011**0** 0000111**1**

In an 8 bit image the least significant bit of the image is replaced with the secret data, whereas in 24 bit colour image the least significant values of red, blue and green are altered. In this technique the data storing capacity can be increased by using two or more configuration of LSB for concealing the data, but it will affect the stego quality. So better security can be ensured by embedding secret data into specific regions within an image, these regions provide least distortions as compared to other parts of the image. Generally we are considering edges for data hiding, because these parts are less sensitive to visual distortions. Maintaining stego image qualiy is an important issue for the protection of secret data, so the image obtained after embedding data is almost similar to that of the original one, because the changes in the LSB pixel don't produce too much difference in the image. During the transmission of stego image if any suspicion is raised, then the hackers will try to know the hidden information inside the message.

Among the proposed data hiding techniques, LSB is one of the simplest techniques to insert data into cover image. The main advantages of LSB embedding are its simplicity and changing rapidly property provides high perceptual transparency. The LSB technique has some drawbacks; it is more predictable, because of the statistical difference between the altered and unaltered regions of the stego image. The distortion in stego image increases exponentially, so embedding more bits is detectable. There are

two types of LSB insertion methods, fixed-size and variable size, fixed one embeds the same number of data bits, in the cover image whereas in variable size embedding data rate is variable. In LSB method if one embeds information in the LSB plane, so the chance of detecting the existence of secret data is very high, it can be avoided by using embedding data in 2LSB plane of the cover image. It means that each of the least two significant bits hold one bit of the message .It is easy to implement, and has higher embedding capacity and it provides imperceptibility. An independent 2LSB is known as I2LSB, here changes occur in the least and second LSB planes independently. The payload embedding capacity of 2LSB and I2LSB are twice that of LSB. Capacity of hiding data and quality of stego image are the two benchmarks used by the steganography to evaluate the hiding performance.[11-13]

Pixel value differencing also known as PVD method is based on human vision sensitivity; in this the size of the hidden data bits can be calculated by the difference between two consecutive pixels in the cover image. It provides good imperceptibility, in this the number of bits can be embedded is determined by obtaining the difference between two consecutive pixels .Based on the obtained pixel difference we can determine region, if the difference is small it is in smooth area, otherwise it is in the edge region. The difference value is mapped into a range table this table is separated into different ranges of widths. So the width of a range table mentions the number of bits that can be embedded. It provides moderate capacity and high security. The embedding capacity in the smooth region is less compared to edge region that is more data can be embedded into the edge area and hence ensures security, whereas embedding in smooth regions may introduce artifacts. In this method cover image is divided into non overlapping blocks comprised with two consecutive pixels, a difference value can be found out from the nearby pixels. The difference value ranges from 0 to 255, and if the difference value is small, then the pixels are in smoother region, otherwise it in edge area. In the edge area pixel seems to be more noisy than their neighboring pixels, so edge regions formulate a better option to conceal secret information than any other parts of the image .Modified pixels in the normal regions are much more noticeable than the high texture regions. So the best way for embedding payload is in the noisy cover pixels or in high textured edge regions because they are imperceptible. In PVD next phase is to design a range table with contiguous range, and it ranges from 0 to 255. The upper and lower boundaries are marked by 'u' and 'l' respectively. This technique referring to only one direction for data hiding, whereas in tri-way PVD, three different edges are considered for hiding purpose, thus the embedding capacity is three times as that of PVD method.[14,15]

5. Transform Domain Techniques

In this domain method, images are first transformed and then the message is embedded into the image. It is also known as frequency domain. In this method secret bits are embedded indirectly, it means that the data is embedded in the transform or frequency domain of the cover image pixels. Compared to spatial domain techniques, transform domain techniques appear to be more complex and hence slow, if features of the cover file could not be utilized. Transform domain based methods conceal the data into the areas that are less exposed to external factors. This method embeds the data into the cosine or Fourier coefficients of an image indirectly, it means that firstly the input image file is divided into various number of coefficients then only data embedding takes place into the coefficients. In Transform domain mode data hiding is performed in significant regions of the cover image. Spatial domain techniques are the easiest way to embed informations, but they are highly vulnerable to minor variations in the cover file, which leads to destroy the secret data completely by the intruder. So the safest way is to embed the data into the frequency domain of a signal, it is much more robust than the spatial domain methods but it provides less payload capacity. Different types the transform domain embeddings are existing one of them uses discrete cosine transformation to insert the data into the image file, another one uses wavelet transforms. In this domain the sender first transforms the cover file into frequency domain coefficients, and then only embedding is performed. In transform domain the embedding operation can be done by using different frequency bands of the cover image. Its high embedding properties and fragility are beneficial to ensure proper authentication.

Frequency domain techniques are much more stronger than spatial domain, nowadays most of the systems prefer transform domain because they hide secret message in the parts of the cover file that are less revealed to transformations like cropping, compression and filtering. Frequency Domain techniques are broadly classified such as, Discrete Fourier Transformation technique (DFT), Discrete Cosine Transformation technique (DCT),

Table 1. Comparison of watermarking, steganography and cryptography

CRITERION	WATER MARKING	STEGANOGRAPHY	CRYPTOGRAPHY
Input files	At least two	At least two	One
Key	Optional	Optional	Mandatory
Carrier	Almost image	Digital media	Mostly text based
Visibility	Frequently	Never	Always
Flexibility	Restriction on cover selection	Free to choose any cover file	Not mentioned
Secret data	Watermark	Payload	Plain text
Detection	Informative	Blind	Blind
Concern	Robustness	Capacity	Robustness
Authentication	Achieved by cross correlation	Full retrieval of information	Full retrieval of information
Type of attack	Image processing	Steganalysis	Cryptanalysis
Objective	Copyright protection	Secret communication	Data protection
Fails when	It is replaced	It is detected	De-cipherer
Result	Watermarked file	Stego file	Cipher text
History	Modern era	Ancient	Modern era

Discrete Wavelet Transformation technique (DWT), lossless or reversible method, embedding in coefficients. This transform divides the image into different frequency bands like high, middle and low levels, so it is more helpful for the users to find out an apt embedding position for inserting data based on their needs. This technique uses various algorithms and transformations to embed the data into the image. A large number of the security based steganographic systems today prefer frequency domain techniques because of its stronger embedding properties, that is embedding data in frequency domain is much stronger than embedding data in time domain. High imperceptibility and high robustness are the advantages of frequency domain techniques. Low payload capacity is the drawback of transform domain technique.[16-19]

6. Cryptography

Cryptography is a widely used overt secret writing procedure that makes modifications on the structure of data in such a way that only its intended recipient can receive it, and manipulate the data in order to hide their presence in text file. The main objective of Cryptography is to protect user's integrity and confidentiality from unauthorized access and securing the secrecy of communication using methods like encryption and decryption. The key idea behind the secure transfer of message in Cryptography is based on the principle of message scrambling, so it cannot be easily detectable by malicious people, thus it ensures multiple layers of shield. Cryptography is one of the most commonly used technique to provide secure data transmission between sender and receiver, for this it encrypts the plain text and generates the cipher text. Plain text may a text document, a bank account number, a password or any other information. The original unscrambled data before applying encryption is known as plain text, after the encryption process we get the encrypted plain text, it is known as cipher text. Main components of cryptography are plain text, key, cipher text, encryption and decryption algorithms. Encryption algorithm converts plain text to cipher text, whereas decryption algorithm converts cipher text to plain text. Most commonly used cryptographic schemes are symmetric key cryptography, public key cryptography. For encryption and decryption process symmetric key cryptography uses a single key. Public key cryptography uses a pair of keys one for public is known as public key for encryption and other for private is known as private key for decryption pro-

cess. Example for symmetric key algorithm is DES (Data Encryption Standard). The digital signature algorithm, RSA algorithm, key exchange algorithms are examples of public key cryptography. The comparisons of the three techniques are shown in Table 1.[20,21].

7. Conclusion

The initial aim of this survey is to investigate the recent existing data hiding techniques based on certain features like robustness, imperceptibility and data embedding capacity and how it is implemented for its future scope. The review conducted with the aid of comparison table enables the users to evaluate the performance and efficiency of these techniques and choose it based on their application program. It is more helpful for beginners to understand the features, background, and history of the existing schemes by combining factors like high payload and data security.

8. References

1. Thanikaiselvan V, Bansal T, Jain P, Shastri S. 9/7 IWT Domain data hiding in image using adaptive and non adaptive methods. Indian Journal of Science and Technology. 2016 Feb; 9(5). Doi: 10.17485/ijst/2016/v9i5/87189.
2. Kaur G, Kaur K. Digital watermarking and other data hiding techniques. IJITEE. 2013; 2(5):181. ISSN: 2278-3075.
3. Sharma PR, Mishra J. A comprehensive survey on data hiding technique. IRJET e. 2015 Jul; 2(4):1–5. ISSN: 2395 -0056.
4. Shastri S, Thanikaiselvan V. PVO based reversible data hiding with improved embedding capacity and security. Indian Journal of Science and Technology. 2016 Feb; 9(5). Doi: 10.17485/ijst/2016/v9i5/87191.
5. Anandpara D, Kothari AD. Working and comparative analysis of various spatial based image steganography techniques. International Journal of Computer Applications. 2015 Mar; 113(12):1–5. (0975 – 8887).
6. Patel PR, Patel Y. Survey on different methods of image steganography. IJIRCCE. 2014 Dec; 2(12):1–5.
7. Halder T, Karforma S, Mandal R. A novel data hiding approach by pixel-value-difference steganography and optimal adjustment to secure e-governance documents. Indian Journal of Science and Technology. 2015 Jul; 8(16). Doi: 10.17485/ijst/2015/v8i16/51269.
8. Vidya G, HemaPreetha R, Shilpa GS, Kalpana V. Image steganography using ken ken puzzle for secure data hiding. Indian Journal of Science and Technology. 2014 Jan; 7(9). Doi: 10.17485/ijst/2014/v7i9/48564.
9. Babloosaha B, Sharma S. Steganographic Techniques of Data Hiding using Digital Images Defence Journal. 2012 Jan; 62(1):11–8.
10. Rabara V, Shah V. Image based steganography review of LSB and HASH-LSB techniques. IJAERD. 2014; 2014.
11. Thasneem Salim PT, Vigneswaran T. FPGA Implementation of Hiding Information using Cryptography. Indian Journal of Science and Technology. 2015 Aug; 8(18). Doi: 10.17485/ijst/2015/v8i19/76853.
12. Rahma AMS, Abdulmunim ME, Al-Janabi RJS. New spatial domain steganography method based on similarity technique. International Journal of Engineering and Technology. 2015 Jan; 5(1):1–4.
13. Islam S, Modi MR, Gupta P. Edge-based image steganography. EURASIP Journal on Information Security. 2014.
14. Sonaniya AK, Rai RK. A review on comparison between different image steganography methods. IJAECE. 2014 Nov; 3(8):355–58.
15. Shelke FM, Dongre AA, Soni PD. Comparison of different techniques for Steganography in images. IJAIEM. 2014 Feb; 3(2):1–6.
16. Suri S, Joshi H, Mincoha V, Tyagi A. Comparative analysis of steganography for coloured images. International Journal of Computer Sciences and Engineering Research Paper. 2014; 2(4):1–5.
17. Dash S, Das M, Jena KC. Region based data hiding for high payload. International Journal of Computer Science and Information Technologies. 2015; 6 (1):913–19.
18. Tiwari A, Yadav SR, Mittal NK. A review on different image steganography techniques certified. International Journal of Engineering and Innovative Technology (IJEIT). 2014 Jan; 3(7):1–4.
19. Kundra S, Madaan N. A comparative study of image steganography techniques. IJSR. 2014 Apr; 3(4):1–5. ISSN (Online): 2319-7064.
20. Mehndiratta A. Data hiding system using cryptography and steganography: A comprehensive modern investigation. IRJET. 2015 Apr; 2(1):1–7.
21. Abikoye Oluwakemi C, Adewole Kayode S, Oladipupo Ayotunde J. Efficient data hiding system using cryptography and steganography. IJAIS. 2012 Dec; 11(4):1–6. ISSN: 2249-0868.