International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Security through Steganography - A Survey

Manu Narula[1], Chahat Goyal[2], Nikhil Sinha[3]

[1]Student of M.Tech, University School of Information, Communication & Technology, Guru Gobind Singh Indraprastha University, Delhi, India
[2]Application Development Analyst, Accenture, Gurgaon, India
[3]Application Development Analyst, Accenture, Gurgaon, India

[1]manunarula1996[at]gmail.com
[2]chahatgoyal2206[at]gmail.com
[3]nikhilsinha931[at]gmail.com

Abstract: ***In today's world, digitization has become a popular concept. However, it has also led to vast new ways of compromising privacy and accessing sensitive information giving rise to the need for high standard security in this field. Three major players surfaced in response to this problem; Cryptography, Water marking and Steganography. Steganography hides information in visual actions or media such as hand signs, visual Morse code, image pixels or frame pixels for streamed flow like videos, making it invisible to a wide variety of attacks as well as simple visual detection from humans making it almost impenetrable layer of security. Here we will be discussing the very roots of emergence of Steganography, their importance at their inception as well as their role in shaping the modern day landscape for data hiding schemes, along with the modern take and techniques in persuasion today.***

Keywords: Steganography, Transform Domain, Spatial Domain, L.S.B., Quantization, Histogram

1. Introduction

In the rapidly evolving and expanding world, digitization has become a necessity for everyone. However, it has also led to vast new ways of stealing and eve's dropping on sensitive information giving rise to the need for high standard security in this field. Three major domains came into existence in response to this problem; Cryptography, Water marking and Steganography. While cryptography and steganography deals with hiding/protecting the secret message, Water marking focusses on protection of cover image or property to revoke unauthorized access. Cryptography mathematically hides plain textual data in complex cypher texts that are of no meaning without the assigned secret key. Steganography takes one step ahead and hides information in visual actions or media such as hand signs, visual Morse code, image pixels or frame pixels for streamed flow like videos, making it invisible to a wide variety of attacks as well as simple visual detection from humans. Here we will be discussing the very roots of emergence of Steganography, their importance at their inception as well as their role in shaping the modern day landscape for data hiding schemes, along with the modern take and techniques in persuasion today. Some of the examples being LSB, DCT, DWT etc.

2. Aims and Objectives

The document primarily aims to give a brief overview to study and catalogue different prominent algorithms and procedures involved in steganography particularly taking its focus on image steganography. This field was made our focus because of its capability of hiding information in plain sight as well as it forms the base for the video based steganography. Though steganography provides an array of advantages, it often has a tradeoff between data hiding capacity and quality of image produced, steganography attracts interest due to its versatile nature of implementations, challenges and efficiency, it concerns security, image processing and in some scenarios, involves IoT also.

A. In this particular document, the progress of our work so far is discussed in detail, we will be highlighting the origins of Steganography from the time of world war 2 till the late 1980's when it became popular among masses.
B. The various techniques present in the field today will also be introduced at a small level, leaving space for future improvements.
C. The associated advantages, tradeoffs, and challenges are also debated upon in the subsequent sections.

3. Origins of Steganography

The first ever recorded implementation of steganography techniques can be effectively traced back to the year 440 BC where Herodotus quotes two examples in his writings.[8] Histiaeus sent a secret message to his vassal, Aristagoras, by shaving the head of his most trusted servant, "marking" the intended message onto his scalp, then waited till his hair returned to send him on his way, with the instruction, "When thou art come to Miletus, bid Aristagoras shave thy head, and look thereon." Second example quoted being: Demaratus warned Greece about a devastating attack by writing it directly on the wooden back panel of a wax tablet before applying the wax on it.

A. Origin of Techniques of Steganography

a. Physical Steganography

Steganography has found its usage for centuries implementing physical involvement such as morse codes, etc. Some examples are:

- Messages hidden in plain sight on paper written in secret inks or inks invisible in normal conditions.

- Messages encrypted in Morse code on yarn that is then made into a piece of clothing worn by a courier.
- Messages written on envelopes in a way that they are covered up by the postage stamps.
- In the initial era of the printing press, it was a common practice to use different typefaces on a printed page simply because the printer did not have enough copies of some letters in a single typeface. Thus, a message could be easily hidden by using different typefaces, such as normal or italic or bold.
- During and after World War II, agents used photographically-created microdots to send sensitive information back and forth. Microdots were natively minute even less than the size of a 'full stop' made by a typewriter. Microdots were tactically placed in the paper and covered with an adhesive detectable by viewing against bright light.
- During World War II, a spy of Japan hiding as a doll seller used order transcripts and records discussing amount and specifications of dolls to exchange sensitive and vital information. Her case later got famous as the Doll lady case.
- Jeremiah Denton tactically blinked his eyes in Morse code during a televised press conference in mid-1966 that he was forced into as an American prisoner-of-war by his captors, translating to "T-O-R-T-U-R-E".

b. Digital Messages

Modern steganography came in to existence in the year 1985 with the power of personal computing being applied to classical steganography problems. [9] A slow progress has taken place since then, but has been successful in terms of vast variety of methods and techniques available to us now:

- Hiding messages within the bits of noise in images and/or audio files.
- Hiding data within cypher data or within some arbitrary random data. The message to hide is cyphered, then used to replace part of a much larger block of encrypted data or a block of arbitrarily selected data.
- Hiding messages in adulterated executable files, exploiting the redundancy in the aimed instruction set.
- Pictures embedded inside streaming media (optionally played at slower or faster speed).
- Altering arrangements of elements in a set.
- Content-Aware Steganography is a special type of steganography that hides the data in the semantics that a human user assigns to a datagram. These systems provide security against a nonhuman entity like a spyware or such bot.

c. Digital Text

- Making text color same as the color of the background of document in word processor documents, e-mails etc.
- One can hide information by using Unicode symbols in place of ASCII symbols which are identical, although there is no visual difference in the 2 sets, some systems may show the fonts in a slightly different style, and the extra information would then be easily spotted.
- Use of hidden control characters and repetitive use of markup tags such as bold, underline or italics to embed data within HTML, this usage is easily visible if one examines the document source making it less secure. HTML pages can also house code for extra blank spaces and tabs at the end of lines, and colors, fonts and sizes, which are not identifiably visible when displayed in the browser or text editor.
- Use of non-printing characters in the Unicode character set like Zero-Width Joiner and Zero-Width Non-Joiner. [11] [12] These characters are generally used for joining and disjoining letters in Arabic and Persian languages, but can also be used in Roman alphabets for hiding information as they have no literal meaning in Roman alphabets. Being zero-width they are not displayed on screen. ZWJ and ZWNJ can represent "1" and "0".

These and many other formats of Steganography like Printed Steganography, Echo Steganography, Social Steganography, Network Steganography, etc. emerged at different points in history

I. Modern Steganography

Steganography covers a wide range of applications in modern day scenario including Special intelligence services, everyday products and much more, some examples of it are as follows:

A. Modern Printers

Some modern computer printers use steganography to place data onto the printed documents to keep track and mark it for identification, including Hewlett-Packard and Xerox brand color laser printers. These printers insert minute yellow dots in every page. The hardly-visible dots contain encoded printer serial numbers along with date and time stamps

B. Use by Intelligence Agencies

In the year 2010, the Federal Bureau of Investigation put up an allegation on the Russian foreign intelligence service that they make use of tailored steganography software for embedding cypher text messages inside image files for certain interactions with "illegal agents" stationed in different nations.

C. Distributed Steganography

The Distributed steganography methods split the message into many smaller units to be assembled after it is received by the intended person. The disassembled or distributed message in different cover images or media makes it even more difficult to be caught as no part is complete by itself and is certainly meaningless without all the other splits.

D. Spatial Domain Steganography

Spatial Domain Steganography gives rise to the most computationally inexpensive and simpler methods for information embedding. Here, the embedding rate is measured in bits per pixel or bpp. Spatial Domain

Steganography can be very broadly categorized into direct and indirect methods, while the former manipulates the pixel information of cover image directly, hiding the message in the very image itself by means of substitution, addition or adjustments in data. The latter however, changes the notations of the cover media and hides the message in the introduced noise bits. Some of the commonly used Spatial Domain Methods are discussed below.

a. L.S.B. Methods

LSB flipping or Replacement: LSB Flipping simply arranges the message bits onto the cover image. It does so by changing the LSB of pixel values with the required bits. For example: if a cover image having pixel values ranging from 0-255 (8-bit image) and the intended message has n bits, then after embedding, the value of the ith message bit is equal to stego image's ith pixel's LSB. Similarly, the value of (i+1) th message bit will be the (i+1) th pixel's LSB and so on till the entire message bits have been traversed.

LSB Matching: LSB Matching tries to minimize the substitutions by selecting the start point of embedding such that the number or pixels whose LSB match with the message bit at that position so that no substitution is required at that point. After selection whenever a non-matching value is encountered, the value is then randomly increased or decreased by 1 to make it match with the message bit. This technique is also sometimes referred to as +- 1 embedding. A general trend followed while increasing or decreasing pixel value is that even pixel values are increased by 1 while odd pixel values are decreased by 1.

b. Grey Level Modifications

Introduced by Potdar in 2004. This technique works by mapping the grey levels of pixels with secret message. A sequence of cover pixels pre-defined by certain algorithm is selected and making use of the notion of odd and even numbers for mapping. The image is taken as a 2-D matrix, a function of x and y of size M x N. Grey Level Modification alters the light intensity intensity of the grey level scalar image at coordinates given by (x,y) where intensity belongs to the range of 0-255 by n units using arbitrary function. The sequence of locations of bits of cover image denoted by p should be predetermined by some function as follows.

$p=\{g(x,y)|0\leq x\leq M,0\leq y\leq N\}$

The one necessary condition in order to embed the value of L for any Pi={0,(MxN)} is odd, then L is incremented by 1 else it remains unchanged. The value of L for every Pi is examined again for mapping the message bit Mi to Pi. If l is even at Pi and Mi is 0, then L is not changed else it is decremented by 1. Like LSB methods, this method also exhibits low complexity with high payload embedding capacity.

c. Pixel Indicator Technique

This is a modification of LSB techniques which aims to enhance security. It works by dividing the three channels of RGB image into an indicator channel and data channels. The indicator channel decides which data channel to use for data hiding [33]. The least two significant bits of a specific color indicator channel (I-channel) is used to reveal the presence of the secret message in the least significant two bits of the other channel. The randomization in embedding provides enhanced security to complicate the detection of secret data. The secret information and I-plane are divided into four sub-blocks and each of the message blocks is embedded into a specific image block by using "Magic LSB" to make the data extraction even more challenging than the LSB techniques.

d. Pixel Value Differencing

This technique is based on the visual effect of human visual perception capacity. The differences of grey values in two- pixels blocks of the cover image are clustered according to different contrast and smoothness properties and the secret message is embedded into these cluster blocks. This technique yields better result than LSB embedding techniques. We take two consecutive pixels Pi and Pi+1 in the cover image having grey values gi and gi+1. The difference in intensity is calculated as$d=|gi+1-gi|,d\in[-225,255]$. The extremely smooth regions in the image had d value close to zero (li =0) and d close to ui= - 255 to 255 representing edged region. The possible values of |d| are classified into range Ri, i = 1..n where the bounds are (li, ui) . The n message bits which can be embedded is formulated as n=log2(li ,-ui+1) and the next distance is calculated using
$d'=\langle lk+b,d\geq 0\rangle$and $d'=-\langle lk+b),d<0\rangle$ where the value of b is $b\in[0,uk-lk]$ which is the value of sub stream S from n bit message. An enhanced variant of PVD called Tri-way PVD works by partitioning the cover image into non-overlapping 2x2 blocks of 4 pixels

e. Pixel Pair Matching

The technique engages two pixels as an embedding component to hide a message digit Sn in N-notational system. Here, the LSB of the first pixel is used for embedding one message bit and a binary function carries another bit. The improvised PPM techniques are exploiting modification direction (EMD) in which one pixel in the pixel pair is changed by one grey-scale unit and the message digit is represented in 5-notational system.

f. Predictive Coding

The technique embeds the secret message into the compression codes during image compression. The technique works in two-stages. The first stage is a prediction stage and the second is an entropy coding stage. The secret data is concealed into the difference value of the cover image after the prediction stage using prediction error values. This technique integrates image compression and steganography to overcome the bandwidth limited in

the network which results in increased data hiding efficiency. The various existing coding techniques are Gradient-adjusted prediction predictor, the median edge detector. If $\bar{x}$is the predicted value of x, then the predictive rule of the predictor is as follows.

$\bar{x}$=min(x,y), $if\ c\geq$max(x,y)
$\bar{x}$=max(x,y), $if\ c\leq$min(x,y)
$\bar{x}$=x+y−c, $otherwise$

Where min(x,y) and max(x,y) function computes the minimum and maximum values of x and y respectively. Gap Predictor estimates the intensities in horizontal and vertical directions represented by gh and gv. Finally, embedding is carried

g. Multibit Plane Steganography

The LSB embedding is extended to hide data in multiple bit planes. It doesn't consider the local property while embedding which decreases its performance in terms of imperceptibility as it follows a non-adaptive embedding approach. It's variant bit plane complexity segmentation (BPCS) represents the raw image in pure binary coding (PBC) and converts it to canonical Grey coding (CGC) system. According to the bit- plane, the image is decomposed into a set of binary images and divided into non- overlapping consecutive blocks of 2x2. The parameter should be chosen with careful selection. The complexity of the image block is computed as below.

$\alpha=k/2\times 2L\times(2L-1)$

Where $\alpha>\alpha°$ is the predefined threshold, which help select the image block for data embedding. The message bits are also grouped into 2Lx2L. If the message block complexity is less than threshold $\alpha°$, then the block is processed using conjugation operator, whose complexity will be $(1-\alpha)$ larger than $\alpha°$, the message block will be replaced with the image block. The embedding rate is observed to be as high as 4 bpp without causing severe distortion.

h. Quantization Based Steganography

The embedding technique used in digital watermarking known as Quantization Index Modulation(QIM) is extended to image steganography by quantizing an input signal x to y using a set of quantizers Qm(.). The choice of quantizer is depended on the secret message bit m. QIM is generally employed to transform domain coefficients before quantization because it produces signs of discreetness in first order statistics, the histogram in spatial domain when quantization scales greater than two.

$yi=Qm(xi)=\Delta[(xi/\Delta)+(12)]$, if $mi=0$
$yi=Qm(xi)=\Delta[(xi/\Delta)+(\Delta 2)]$, if $mi=1$

A variant of QIM is called Dither modulation (DM). It adds a dither signal to the input signal before quantization stage and subtracted after quantization covering all the values of the input signal and di is the dithering signal which is uniformly distributed.

$yi=Qm(xi+di)-di$ Over $[-\Delta/4,\Delta/4]$

i. Histogram Shifting

The process of embedding a MxN grey scale image with values on l=[0,255]. The objective is to find the minimum zero-point h(z), where zϵ [0,255] and a maximum point h(m), where mϵ [0,255]. If the h(z)>0 then coordinates (x,y) is recorded and h(z) is set to zero. The part of the histogram H(i) with iϵ [z, m] is shifted by 1 unit towards right. If the message bit is 1, the grey scale pixel value is changed to m+1. IF the message but is 0, m remains unchanged. The embedding capacity denoted by C=h(m)-0, indicates the amount of overhead information.

E. Transform Domain Steganography

Spatial domain techniques performance in terms of embedding capacity is high. However, it is not robust against statistical attacks. This prompted the need for enhanced and secure transform domain steganography. The techniques to implement digital image steganography by converting the image from spatial to transform domain is termed as transform domain steganography. It started by adopting the transform domain techniques used in robust watermarking to image steganography for designing large capacity embedding steganography. There are many candidates' transforms such as Discrete Cosine Transform, Discreet Wavelet Transform, Discrete Fourier Transform, Haar Transform, Integer Wavelet Transform, Contourlet Transform etc. These techniques are popular because transform coefficients embedding occurs in more robust areas, spreads across the entire image, there by producing better resistance against attacks than spatial domain techniques. We shall discuss the most common transform used in the subsequent sections.

a. Discrete Cosine Transform

JPG uses DCT to convert from spatial to transform domain during compression. The DCT based steganography system integrates with the image-compression algorithm to design various JPEG steganographic schemes. The Non-zero AC DCT Coefficients are used for data embedding algorithm. DCT breaks an image into low frequency (FL), mid frequency (FM) and high frequency (FH) and embeds in mid frequency range. Given 2-D image f(x,y) of NxN size, the 2-D Discrete cosine transform C(u,v) of image f(x,y) is defined as below.

$C(u,v)=\alpha(u)\alpha(v)\Sigma\Sigma f(x,y)N-1y=0cos[2x+12N]N-1x=0\times cos[2y+12N]$

Where $C(u,v)=\{\sqrt{1}N,for\ u=0\sqrt{2}N,for\ u=1,2,..N-1\}$

The process begins by dividing the cover image into non-overlapping blocks and 2D-DCT is applied on each block following which the DCT coefficients are quantized according to the quantization table and the secret message bits are then embedded into the quantized DCT Coefficients which are coded using run length coding and Huffman coding. The High frequency components of DCT are better suited for embedding because they are visually

more resistant to noise than their low frequency components. The major steganographic tools based on DCT are JPEG/JP Hide which uses LSB embedding to replace LSBs of non-zero quantized AC coefficients which are randomly selected with the secret bits. YASS (Yet another Steganographic Scheme) divides the image into fixed size of six blocks and within each block; sub-blocks called Host blocks (H-Block) are randomly selected. The secret message bits are embedded into the DCT coefficients of the H-Block using QIM DCT based steganography received widespread acceptance with F5 algorithm which embeds by decreasing the absolute value of the coefficient by 1 using matrix encoding. In Outguess, LSB replacement is used for embedding in DCT coefficients with values not equal to 0 or 1. Outguess and Steg-Hide are statistics preserving steganography which uses statistical restoration but are highly detectable. Steg-Hide avoids changing the histogram as DCT coefficients are swapped. In F5, if absolute value becomes 0 i.e. when coefficient absolute value is -1 or 1 then it is said to have shrinkage and message bit is embedded in next coefficient. YASS resists blind steganalysis and embeds at randomized locations.

b. Discrete Wavelet Transform

The DWT hides data in regions that are less sensitive to the HVS at the high-resolution detail bands (HL, LH and HH). These techniques increase the robustness while maintaining high imperceptibility. The higher sub band represents the finer details of the image and can be used for embedding while the lowest sub band has the most important and relevant information [18]. DWT provides more compression ratios and avoids interference from artifacts [6] High Frequency contains edge information, low frequency contains signal information. DWT is performed in vertical direction followed by Horizontal direction and follows these representations LL → for approximation coefficients, LH → Vertical details and HH →Diagonal details. Secret images are disintegrated into LH, HL, HH and LL is processed to set the next wavelet coefficient values. Along with DWT, several other transforms such as curvelet transform, Slantlet transform, Integer Transform, Contourlet transform etc., are used to generate robust stego systems; each having certain advantage than the other.

c. Contourlet Transform

Contourlet transform has wavelet features and the sub bands at each scale are decomposed into different directions. It solves wavelet band mixing problem. Contourlet transform proves to be more powerful in characterizing images rich in directional details and smooth contours. It provides multi-scale and multi directional representation of cover image using Laplacian Pyramid. Hence, they are more effective than DWT to capture smooth contours and geometric structures. On decomposition of the cover image by applying contourlet Transform, a low pan and high pan sub bands of the image is obtained. Data is embedded in the high pan sub band. **Integer Wavelet Transform (IWT)** maps an integer dataset. The floating point coefficients of DWT wavelet filters are the location of embed data embedding. However, truncation of floating point values to integer causes loss of the hidden information leading to data hiding failure. To avoid, floating point precision of wavelet filters. The LL sub band of IWT is a closer copy than LL of DWT of the original image. Disadvantages of Transforms Domain Steganography are Low hiding capacity and Complex computations.

F. Adaptive Steganography

Adaptive steganography is presently the most secure technique for empirical cover sources for secret data communication. The embedding method embeds the secret message while considering costs of modifying the cover image pixel to embed the payload which should minimize a distortion function designed to capture statistical detectability. There exist two general frameworks for adaptive embedding paradigm; firstly, by establishing empirical payload–distortion bound for additive distortions which uses near-optimal practical coding schemes known as non-additive distortion. It accounts for the inter-pixel correlations and interactions among the embedding modifications. Examples of adaptive non-additive distortion steganography are HUGO, HILL and UNIWARD variants. However, adaptive non-additive distortion function modelling steganography is more challenging because there is no coding technique capable of minimizing an arbitrary distortion function and hence, the non-additive distortion functions is approximated to an additive form such as CMD which uses side informed precover to embed; Secondly, by designing cover image model noise residuals during image acquisition which takes the difference between a pixel and its estimated value, such as MiPOD additive steganography [65]. HUGO is the first method based on distortion minimization, it reflects on the smallest impact that modified distortions of a cover pixel groups have on statistical distribution in SPAM feature space. The distortion minimization factor is further extended in Wavelet Obtained Weights (WOW) by using correlations between pixels in the predicted pixels. In recent times, the most successful heuristic additive distortion bound adaptive approach is UNIWARD, which is the generalization of WOW to be used in any domain (S-UNIWARD, SI-UNIWARD and J-UNIWARD). UNIWARD distortion function is the relative sum of changes rates of cover and stego wavelet coefficients by evaluating their smoothness in multiple direction using 8-tap wavelet directional filter bank consisting of LH, HL and HH directional high pass filter in spatial domain.

$$D(X,Y)\triangleq(\Sigma\Sigma|Wuv(X)\ k-Wuv(Y)\ kuv3k=1)+|Wuv(X)\ k|$$

where Wuv(X) k and Wuv(Y) k are the wavelet coefficients for cover-stego image pair (X,Y) and (u, v) is taken over all sub band and stability constant is > 0. In case of J-UNIWARD, first image should be decomposed to spatial domain. It follows the same procedure as S-UNIWARD. In side informed SI-UNIWARD, the distortion is represented as $DSI(X,Y)\triangleq D(P,Y)-D(P,X)$, where P is the precover used as side information for both binary and ternary embedding with syndrome trellis

coding. MiPOD is an optimal content adaptive steganographic method whose performance is like empirical detectors built with real image model classifier. The image is modelled by mutually independent embedding using LSB-M. The cover image x={x1,..xn} which generates the stego image y={y1,..yn} to which we apply the probabilistic rule P(yn=xn+1)=βn ,P(yn=xn-1)=βn and P(yn=xn)=1- 2βn where the change rate βn=[0,1/3]. The distortion cost of MiPOD is pn =ln(1/ βn-2). Steganalysis is performed with maxSRMd2 feature set. The most commonly used coding scheme for near optimal payload-distortion bound embedding is Syndrome Trellis Codes (STC).

The embedding additive distortion function $D(x,y)=\Sigma d(x,y)ni$ where d(x,y) is the distortion cause by a single letter distortion. STC is a variant of matrix encoding based on convolutional codes such as binary convolutions code denoted by Hϵ{0,1}mxn and message m ϵ{0,1}n. The embedding and extraction process is determined as below.

Emb: $y=argmin\ D(x,w)$, $w\epsilon\{w|\ Hw=m,\ w\ \epsilon\{0,1\}n\}$ *and* ***Ext***:$m=Hy$.

G. Spread Spectrum Based Steganography

The spread spectrum image steganography (SSIS) is key-based blind steganographic scheme, in which the original cover image is not exactly needed to extract the secret message. Extraction is based on the encryption key. The message is encrypted, encoded with a low rate error control coding method and embedded in a randomly generated noise using a key with the help of a modulator. SSIS is usually modeled as additive white Gaussian noise. The embedded message is passed over to the quantizer to generate the stego image, which is send over a secure channel along with the keys. The decoder produces the cover image estimate by means of image restoration technique using reverse-encoding process. The imperceptibility factor performance of SSIS scheme is measures with signal to noise ratio (PSNR). A chaos based SSIS scheme has been developed where chaotic encryption based on chaotic shift keying is used to encrypt the secret message signal instead of the pseudorandom noise generator to generate white noise as the embedding channel. At the receiver side, chaotic modulation is applied.

H. Side Informed Steganography

In side-informed steganography, the secret message is embedded in the higher quality cover images known as precover while converting to a lower quality cover image using quantization errors. It is capable of embedding the desired payload using binary or ternary embedding technique. Side Information can be included in any of the steganographic scheme to increase robustness and imperceptibility. We could also use image acquisition tools (such as camera) and multiple pictures of the same scene as side- information. However, they are labor intensive and not as computationally efficient as precover approach. The SI-steganography technique has been majorly used in JPEG compression using Perturbed Quantization, F5, and nsF5, MME, MMEx and BCHopt, all of these methods are centered on the rounding errors of unquantized DCT coefficients. In F5 algorithm, the impact of embedding or the embedding efficiency is equal for each coefficient so, the total distortion minimization for a given message payload corresponds to the impact of minimizing the individual coefficients which is modified. The distortion function is designed by computing the impact of embedding costs from quantized and unquantized image cover sources. The performance of F5 embedding was increased through matrix encoding. This enhanced F5 when incorporated with wet paper code (WPC) resulted in the improved version called nsF5 that took care of the shrinkage issue caused by F5 embedding originally. The modified matrix embedding (MME) used in JPEG steganography made use of the side information of the original uncompressed cover image to build the distortion function. Here, only those coefficients with minimum distortion are selected to be modified. Studies have shown that MME performs better in terms of embedding efficiency, i.e., the number of message bits embedded per embedding change. SI-steganography evolved with heuristic optimization algorithms such as BCHopt anf fast VCH-syndrome coding. These heuristic approaches exploits the rounding error, the quantization step for designing the distortion function, thus resulting in huge improvement in security against steganalysis attacks. The most attractive feature of the use of side-informed steganography is it can be easily integrated in any steganographic scheme designed to minimize distortion. The steganalysis involves testing the stego image by resizing, color-depth reduction and image color conversion such as color to grey scale conversion in the spatial domain and in the quantization stage during JPEG compression. The embedding distortion:

βij =exp(-aPij))/(1+exp(-aPij)),

where Pij is the cost of each cover element and a>0 is a parameter chosen so that denominator does not become zero.

4. Future Scope

Steganography is the art of hiding information in plain sight, this information may range from simple texts to complex executable codes. Such a type of field has numerous applications in security scenarios. The current research strives towards embedding information into watermarks for providing a legitimate proof of ownership, this method is attracting interest from the real estate sector as well as big companies that take copyrights seriously. Focus is also diverted on coming up with more sophisticated and accurate techniques to analyze and detect hidden information to check for illegal usage like virus implants and other harmful malicious entities. Both Steganography and Stegnalysis have a vast host of opportunities and techniques yet to be discovered.

Steganography also paves way for development via integration of machine learning algorithms to produce techniques with better data capacity while maintaining a minimal level of noise possible, being a technology

independent art of data hiding, steganography can combine with any and all recent technologies as well as those in development to come up with more productive ways.

5. Conclusion

Steganography is a technique of covering the data in such a way that the message could be transmitted secretly and only the sender and receiver knows the way of decrypting that secret text or message. Steganography increases the security of data to be transmitted and also ensures that only authorized personnel can have access to that message. This document aims to present a survey of steganography and techniques that are used for steganography. Focusing on both, the origins of Steganography as well as the modern approach in this digital world. We see the emergence of pictoral data hiding or visual data hiding from times as ancient as 400BC, moving through the world wars and into the modern era diversifying in nature and implementations harnessing the power of computer units. Various papers have been reviewed on steganography. It is studied that there are various types of steganography like text, audio, video, image, network or protocol steganography. This shows that text or data using steganography can be hidden in many ways.

Acknowledgment

We would like to thank the staff of UIRC, GGSIPU, Delhi for providing the resources and support pivotal for the research work of this document. We would also like to extend our thanks to the varied authors of the research papers used for the reference work on this document for providing valuable insights into various fields of Steganography and its implementations.

References

[1] Amritpal Singh, “An Improved LSB based Image Steganography Technique for RGB Images”, Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on. IEEE, 2015., pp 1-4
[2] Mehdi Hussain, “A Survey of Image Steganography Techniques”, International Journal of Advanced Science and Technology Vol. 54, May, 2013, pp 113-124
[3] T. Morkel, “AN OVERVIEW OF IMAGE STEGANOGRAPHY”, ISSA. 2005, pp 1-11
[4] R.Poornima, “AN OVERVIEW OF DIGITAL IMAGE STEGANOGRAPHY”, (IJCSES) Vol.4, No.1,February 2013, pp 23-31
[5] Anil Kumar, “A Secure Image Steganography Based on RSA Algorithm and Hash-LSB Technique”, IJARCSSE, Volume 3, Issue 7, July 2013, pp 363-372
[6] Shaveta Mahajan, “A Review of Methods and Approach for Secure Stegnography”, IJARCSSE, Volume 2, Issue 10, October 2012, pp 67-70
[7] Jasleen Kour, “Steganography Techniques –A Review Paper”, International Journal of Emerging Research in Management &Technology, Volume-3, Issue- 5, May 2014, pp 132-135
[8] Petitcolas, FAP; Anderson RJ; Kuhn MG (1999). "Information Hiding: A survey" (PDF). Proceedings of the IEEE. 87 (7): 1062–78.
[9] CiteSeerX 10.1.1.333.9397. doi:10.1109/5.771065. Retrieved 2008-09-02.
[10] http://www.mikebarney.net/stego.html
[11] Cheddad, Abbas; Condell, Joan; Curran, Kevin; Mc Kevitt, Paul (2009). "A skin tone detection algorithm for an adaptive approach to steganography".
[12] Signal Processing. 89 (12): 2465–2478. doi:10.1016/j.sigpro.2009.04.022
[13] Akbas E. Ali (2010). "A New Text Steganography Method By Using Non-Printing Unicode Characters" (PDF). Eng. & Tech. Journal. 28 (1).
[14] Aysan, Zach (December 30, 2017). "Zero-Width Characters". Retrieved January 2, 2018. In early 2016 I realized that it was possible to use zero-width characters, like zero-width non-joiner or other zero-width characters like the zero-width space to fingerprint text. Even with just a single type of zero-width character the presence or non-presence of the non-visible character is enough bits to fingerprint even the shortest text.
[15] T. Y. Liu and W. H. Tsai, "A New Steganographic Method for Data Hiding in Microsoft Word Documents by a Change Tracking Technique," in IEEE Transactions on Information Forensics and Security, vol. 2, no. 1, pp. 24-30, March 2007. doi: 10.1109/TIFS.2006.890310
[16] "Criminal complaint by Special Agent Ricci against alleged Russian agents" (PDF). United States Department of Justice.
[17] "Distributed Steganography". IEEE. October 2011.
[18] Wenbo Zhou, Weiming Zhang & Nenghai Yu. (2017). A New rule for cost reassignment in adaptive steganography. IEEE transactions on information forensics and security.
[19] Antonio Tasheva, Zhaneta Tasheva & Plamen Nakov. (2017). Image based steganography using modified LSB insertion mwthod with contrast stretching.
[20] Proccedings of the ACM 18th International Conference on Computer Systems and Technologies, Bulgaria.
[21] Qingqing Shen, Guangjie Liu, Weiwei Liu & Yuewei Dai. (2015). Adaptive image steganography based on pixel selection. IEEE International Conference on Progress in Informatics and Computing (PIC), Nanjing, China.
[22] Baby Della, Jitha Thomas, Gisny Augustine, Elsa George & Neenu Rosia Michael. (2015). A Novel DWT Based Image Securing Method Using Steganography. Procedia Computer Science.
[23] Bin Li, Ming Wang, Xiaolong Li, Shunquan Tan & Jiwu Huang. (2015). A Strategy of Clustering Modification Directions in Spatial Image Steganography.
[24] IEEE Transactions on Information Forensics and Security.
[25] Li, Jun, Xiaoyuan Yang, Xin Liao, Feng Pan & Minqing Zhang. (2017). A game-theoretic method for designing distortion function in spatial steganography.
[26] Multimedia Tools and Applications.

[27] Kavitha, C. T & C. Chellamuthu. (2010). Multimodal medical image fusion based on Integer Wavelet Transform and Neuro- Fuzzy. International Conference on Signal and Image, Chennai, India.
[28] Luo, Xiangyang, Xiaofeng Song, Xiaolong Li, Weiming Zhang, Jicang Lu, Chunfang Yang & Fenlin Liu. (2016). Steganalysis of HUGO steganography based on parameter recognition of syndrome-trellis-codes. Multimedia Tools.
[29] Winkler, Antj. (2012). Advances in Syndrome Coding based on Stochastic and Deterministic Matrices for Steganography. Saechsische Landesbibliothek- Staats- und Universitaetsbibliothek, Dresden.
[30] R. O. El Safy, H. H. Zayed & A. El Dessouki. (2009). An adaptive steganographic technique based on integer wavelet transform. International Conference on Networking and Media Convergence, Cario, Egypt.
[31] Feng Gu & Lu J. (2006). A new composite implicit iterative process for a finite family of nonexpansive mappings in Banach spaces. Fixed Point Theory and Applications.
[32] Thomas Mittelholzer. (2000). An Information-Theoretic Approach to Steganography and Watermarking. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg.
[33] Y. Qi, Y. Wang & J. Yuan. (2009). Audio Steganalysis Based on Co-occurrence Matrix and PCA. International Conference on Measuring Technology and Mechatronics Automation, 433-436, Zhangjiajie, Hunan.
[34] Vojtěch Holub & Jessica Fridrich. (2013). Digital image steganography using universal distortion. ACM workshop on Information hiding and multimedia security, Montpellier, France.
[35] Mohan, Malini, & P.R Anurenjan. (2011). A new algorithm for data hiding in images using contourlet transform. Recent Advances in Intelligent Computational Systems (RAICS) IEEE, Trivandrum, India.
[36] Hayat Al-Dmour & Ahmed Al-Ani. (2016). A steganography embedding method based on edge identification and XOR coding. Expert Systems with Applications.