ISSN:0975-9646

Information Hiding Systems

Meenakshi Kaul
School of Computers and Electronics, IPS Academy, Indore, India

***Abstract*- The art and science of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message is Steganography. This differs from cryptography, the art of secret writing, which is intended to make a message unreadable by a third party by encrypting the data. If a piece of text looks suspicious , it is easy to suspect that someone wants to hide something for the reader and several decryption methods can be employed to assure an attack .The opportunity to suspect that there is something hidden is not there in case of steganography , thus increasing the level of security by at least one step. The paper is organized into parts; in the first ones an overview and brief technical introduction to steganography is made. However the rest of parts are consecrated for its emphasis on digital applications, focusing on hiding information in image or audio files.**

***KEYWORDS*- Steganography, Cryptography, Audio-Video streaming, Modulations.**

I. INTRODUCTION

Some authors categorize steganography as a form of cryptography since hidden communication is a form of secret writing. The advantage of steganography over cryptography alone is that messages do not attract attention to themselves, to messengers, or to recipients. An unbidden coded message, no matter how unbreakable it is, will arouse suspicion and may in itself be incriminating. A steganographic message appears to be something else: a picture, an article, a shopping list, or some other message -the cover text. Steganographic message is often first encrypted by some traditional means, and then a cover text is modified in some way to contain the encrypted message, resulting in stegotext. [1][2]

The steganography process generally involves placing a hidden message in some transport medium, called the carrier. The secret message is embedded in the carrier to form the steganography medium [3][2].

II. DIGITAL IMAGE STEGANOGRAPHY

The carrier used as Steganography medium here are graphical images or audio files .Most digital image applications today support 24-bit true color, where each picture element (pixel) is encoded in 24 bits, comprising the three RGB bytes .The choice of color encoding affects image size. A 640 x 480 pixel image using 8-bit color would occupy approximately 307 KB (640 x 480 = 307,200 bytes), whereas a 1400 x 1050 pix image using 24-bit true color would require 4.4 MB (1400 x 1050 x 3 = 4,410,000 bytes).Color palettes and 8-bit color are commonly used with Graphics Interchange Format (GIF) and Bitmap (BMP) image formats which are generally considered to offer loss less compression because the image recovered after encoding and compression is bit-for-bit identical to the original image. Graphical image is a collection of numbers that constitute different light intensities in different areas of image [4].A color value is normally a three-component vector in a color space; A well known color space is RGB. Since the colors red, green, and blue are additive primaries, every color can be specified as a weighted sum of a red, green, and a blue component. A vector in RGB space describes the intensities of these components. Another space, known as YCbCr, distinguishes between a luminance (Y) and two chrominance (Cb, Cr) components. Whereas the *Y* component accounts for the brightness of a color, Cb and Cr distinguish between the color grades. A color vector in RGB can be converted to YCbCr using the well known transform:

$$Y = 0.299R + 0.587g + 0.114B$$
$$Cb = 0.5 + (B - Y)/2$$
$$Cr = 0.5 + (R - Y)/1.6$$

Number of bits in a colour scheme called bit depth refers to the number of bits used for each pixel. [5]

A. Least significant bit substitution or overwriting method

The most common steganography method for audio and image files uses some type of least significant bit substitution for overwriting. The least significant bit term comes from the numeric significance of the bits in a byte [4][6]**.**As far as reviews of concerned literature, Least significant bit substitution is a simple, and yet common, technique for steganography. Its use, however, is not necessarily as simplistic as the method sounds. Only the most naïve steganography software would merely overwrite every least significant bit with hidden data. A difference between adjacent pixels xi and xi+ 1 is calculated and fed into a quantizer Q which outputs a discrete approximation of the difference signal. Thus, in each quantization step a quantization error is introduced. For highly correlated signals we can expect it to be close to zero, so an entropy coder which tries to create a minimum redundancy code given a stochastic model of the data to be transmitted, is considered to be efficient. At the receiver side the difference signal is dequantized and added to the last signal sample in order to construct an estimate for the sequence. For steganographic purposes the quantization error in a predictive coding scheme can be utilized, say we can adjust the difference signal so that

it transmits additional information. Let, the stego-key consists of a table which assigns a specific bit to every possible value of error. In order to store a specific message bit in the cover-signal, the quantized difference signal is computed. If this error value does not match (according to the secret table) with the secret bit to be encoded, error is replaced by the nearest value where the associated bit equals the secret message bit. The resulting error values are then fed into the entropy coder. At the receiver side, the message is decoded according to the difference signal and the stego-key.
Though LSB modification techniques are easy ways to embed information, but they are highly vulnerable to even small cover modifications. An attacker can simply apply signal processing techniques in order to destroy the secret information entirely. In many cases even the small changes resulting out of lossy compression systems yield to total information loss.
Incase, the sender uses all cover-elements for information transfer, starting at the first element. Since the secret message will normally have less bits, the embedding process will be finished long before the end of the cover. In this case, the sender can leave all other cover elements unchanged. This can, however, lead to a serious security problem: the first part of the cover will have different statistical properties than the second part, where no modifications have been made. To overcome this problem, the secret message needs to be enlarged with random bits so as to create an equal change in randomness at the beginning and the end of the cover. The embedding process thus changes far more elements than the transmission of the secret would require. But here the probability that an attacker will suspect secret communication increases. A more sophisticated approach is the use of a pseudorandom number generator to spread the secret message over the cover in a rather random manner**.** If both communication partners share a stego-key k usable as a seed for a random number generator, they can create a random sequence and use the elements for information transfer. Thus, the distance between two embedded bits is determined pseudo randomly. Since the receiver has access to the seed *k* and knowledge of the pseudorandom number generator, he can reconstruct the entire sequence of element.

B. Transform Domain Techniques

Image steganography can be divided into Image domain and Transform domain. Image or spatial domain embeds messages in the intensity of pixels directly and in transform or frequency domain, images are first transformed and then message is embedded in the image.[7]Transform domain involves the manipulation of algorithms and image transform [8]methods hide messages in significant areas of the cover image which makes them more robust to attacks, such as compression [9], cropping, and some image processing, than the LSB approach. However, while they are more robust to various kinds of signal processing, they remain imperceptible to the human sensory system. One method is to use the discrete cosine transformation (DCT) and other is wavelet transforms. The two-dimensional DCT is the "heart" of the most popular lossy digital image compression system used today: the JPEG system which converts the image to be compressed into the YCbCr color space [10] and breaks up each color plane into 8x8 blocks of pixels. Then, all blocks are DCT transformed. In a quantization step all DCT coefficients are divided by some predefined quantization values and rounded to the nearest integer .The purpose of this process is to modulate the influence of the different spectral components on the image. The influence of the highest DCT coefficients is reduced as they are likely to be dominated by noise. The resulting quantized DCT coefficients are compressed using an entropy coder by arithmetic coding. In the JPEG decoding step **[10]**, all DCT coefficients are dequantized and an inverse DCT is performed to reconstruct the data. The restored picture will be close to (but not identical with) the original one; but if the quantization values were set properly, there should be no noticeable difference for a human observer.
One popular method of encoding secret information in the frequency domain is modulating the relative size of two (or more) DCT coefficients within one image block.[11][4] During the encoding process, the sender splits the cover-image in 8x8 pixel blocks; each block encodes exactly one secret message bit. The embedding process starts with selecting a pseudorandom block *bi* which will be used to code the *i*th message bit. Before the communication starts, both sender and receiver have to agree on the location of two DCT coefficients, which will be used in the embedding process; The two coefficients should correspond to cosine functions with middle frequencies; this ensures that the information is stored in significant parts of the signal (hence the embedded information will not be completely damaged by JPEG compression).We choose the DCT coefficients in such a way that the quantization values associated with them in the JPEG compression algorithm are equal.[12] The sender then performs an inverse DCT to map the coefficients back into the space domain. To decode the picture, all available blocks are DCT -transformed. By comparing the two coefficients of every block, the information can be restored.

C. Paletted image method

Here, the order of the colors in the palette is altered [8] or least significant bit encoding is done on the palette colors rather than on the image data. In a palette-based image only a subset of colors from a specific color space can be used to colorize the image. Every palette-based image format consists of two parts: a palette specifying *N* colors as a list of indexed pairs (i, ci), assigning a color vector ci to every index *i*, and the actual image data which assigns a palette index to every pixel rather than the color value itself. If only a small number of color values are used throughout the image, this approach greatly reduces the file size. There are

two ways to encode information in a palette-based image: either the palette or the image data can be manipulated. The LSB of the color vectors could be used for information transfer. Alternatively, since the palette does not need to be sorted in any way, information can be encoded in the way the colors are stored in the palette. [13][8] However, all methods which use the order of a palette to store information are not robust, since an attacker can simply sort the entries in a different way and destroy the secret message. Information can be encoded in the image data. Color values can, for instance, be stored according to their Euclidian distance in RGB space: Since the human visual system is more sensitive to changes in the luminance of a color, another (probably better) approach would be sorting the palette entries according to their luminance component. After the palette is sorted, the LSB of color indices can safely be altered for every pixel, the set of closest colors (in the Euclidian norm) is calculated. Starting with the closest color, the sender proceeds to find the next-closest color until a color is found where its parity (R + G + B mod 2) matches with the secret bit to encode. Once such a color is found, the pixel is changed to this new color.

III. Hiding Information in Digital Sound

Audio encoding involves converting an analog signal to a bit stream. Analog sound-voice and music-is represented by sine waves of different frequencies. Storing the sound digitally requires that the continuous sound wave be converted to a set of samples that can be represented by a sequence of zeros and ones, using Pulse Code Modulation technique. Analog signals need to be sampled at a rate of twice the highest frequency component of the signal so that the original can be correctly reproduced from the samples alone.

Embedding secret messages in digital sound is generally more difficult than embedding information in digital images. Moore noted that the human auditory system is extremely sensitive; perturbations in a sound file can be detected as low as one part in 10 million. Although the limit of perceptible noise increases as the noise level of the cover increases, the maximum allowable noise level is generally quite low. It is however known that the human auditory system is much less sensitive to the phase components of sound; this fact has been exploited in numerous digital audio compression systems. [14]In phase coding, a digital datum is represented by a phase shift in the phase spectrum of the carrier signal; the carrier signal is split into a series of *N* short sequences, DFT is applied, and a matrix of the phases and Fourier transform magnitudes is created. Since phase shifts between consecutive signal segments can easily be detected, their phase differences need to be preserved in the stego-signal. The embedding process thus inserts a secret message only in the phase vector of the first signal segment and creates a new phase matrix using the original phase differences. The sender then uses the new phase matrix and the original matrix of Fourier transform magnitudes to construct the stego-signal using the inverse Fourier transform. Given the knowledge of the sequence length, the receiver is able to calculate the DFT and to detect the phases.

A. Echo Hiding

Echo hiding attempts to hide information in a discrete signal f(t) by introducing an echo f(t – Δt) in the stegosignal c(t),c(t) = f(t) + af(t -Δt)

Information is encoded in the signal by modifying the delay Δ*t* between the signal and the echo in a way that the echo signal is not audible for a human observer. The basic echo hiding scheme can only embed one bit in a signal; therefore a cover signal is divided into l(m) blocks prior to the encoding process. Consecutive blocks are separated by a random number of unused samples so that the detection and extraction of the secret message bits is harder. Before the secret message can be extracted out of the stego-signal, some sort of synchronization must take place; the receiver must be able to reconstruct the l(m) signal blocks the sender used to embed one secret message bit. Each signal segment can then be decoded via the autocorrelation function of the signal's spectrum. [14]

B. Spread Spectrum and Information Hiding

Spread spectrum techniques are a means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information; the band spread is accomplished by means of a code which is independent of the data, and a synchronized reception with the code at the receiver is used for despreading and subsequent data recovery. Although the power of the signal to be transmitted can be large, the signal-to-noise ratio in every frequency band will be small. Even if parts of the signal could be removed in several frequency bands, enough information should be present in the other bands to recover the signal. Thus, Spread Spectrum makes it difficult to detect and/or remove a signal. This situation is very similar to a steganography system which tries to spread a secret message over a cover in order to make it impossible to perceive. Since spreaded signals tend to be difficult to remove, embedding methods based on SS should provide a considerable level of robustness. The two special variants of SS generally used are direct-sequence and frequency-hopping schemes. In direct-sequence schemes, the secret signal is spread by a constant called chip rate, modulated with a pseudorandom signal and added to the cover. On the other hand, in frequency- hopping schemes the frequency of the carrier signal is altered in a way that it hops rapidly from one frequency to another. [15]

C. Audio steganography

In a perceptual audio coder, the codec does not attempt to retain the input signal exactly after encoding and decoding, rather its goal is to ensure that the output signal sounds the same to a human listener. The psychoacoustic model uses

auditory masking to analyze the input signals within consecutive time blocks and determines for each block the spectral components of the input audio signal by applying a frequency transform. Then it models the masking properties of the human auditory system, and estimates the just noticeable noise-level, sometimes called the threshold of masking. In its quantization and coding stage, the encoder tries to allocate the available number of data bits in a way that meets both the bit rate and masking requirements. This can be framed as, that the audio signal contains a significant portion of information that can be discarded without the average listener noticing a change. For the purpose of understanding, it can be considered that two nested iteration loops are utilized here; the inner iteration loop (rate loop) and the outer iteration loop (noise control/distortion loop).The system will hide .0information in audio files during the compression process. The data is first compressed, encrypted and then hidden in the audio bit stream. The hiding process takes place under encoding process namely in the inner loop. The inner loop quantizes the input data and increases the quantizer step size until the quantized data can be coded with the available number of bits. Another loop checks that the distortions introduced by the quantization do not exceed the threshold defined by the psychoacoustic model. One-time pad (or key stream generator) can be used to select a set of pixels, and embed the cipher text bit as their parity .This way, the information can be hidden by changing whichever of the pixels can be changed least obtrusively. There is an interesting tradeoff: the more bits in the selection channel, the more bits we can hide in the cover text.SHA-l can be employed to generate pseudorandom bits for use in the hiding process.

D. Steganalysis

The detection of steganographically encoded packages is called steganalysis. Steganalysis broadly follows the way in which the steganography algorithm works. One simple approach is to visually inspect the carrier and steganography media. Many simple steganography tools work in the image domain and choose message bits in the carrier independently of the content of the carrier. A second approach is to look for structural oddities that suggest manipulation as structural changes often create a signature of the steganography algorithm that was employed. Least significant bit insertion in a palette-based image often causes a large number of duplicate colors, where identical colors appear twice in the palette and differ only in the least significant bit. Steganographic techniques generally alter the statistics of the carrier and, obviously, longer hidden messages will alter the carrier more than shorter ones.

IV. Some Applications of Information Hiding

Unobtrusive communications are required by military and intelligence agencies: even if the content is encrypted the detection of a signal on a modern battlefield may lead rapidly to an attack on the signaler. For this reason, military communications use techniques such as spread spectrum modulation making signals hard for the enemy to detect or jam. Information hiding techniques can also be used in situations where plausible deniability is required as may be when the two communicating parties are engaged in an activity which is somehow illicit, and they wish to avoid being caught. The healthcare industry and especially medical imaging systems may benefit from information hiding techniques. They use standards such as DICOM (digital imaging and communications in medicine) which separates image data from the caption, such as the name of the patient, the date, and the physician. Embedding the name of the patient in the image could be a useful safety measure in case of any information loss. Another emerging technique related to the healthcare industry is hiding messages in DNA sequences. This could be used to protect intellectual property in medicine, molecular biology or genetics.

A. Rumored Usage in Terrorism

The rumors about terrorists using steganography started first in the daily newspaper USA Today on February 5th 2001. The articles available online and were titled "Terrorist instructions hidden online", and the same day, "Terror groups hide behind Web encryption". In July of the same year, the information looked even more precise: "Militants wire Web with links to jihad". A citation: "Lately, al-Qaeda operatives have been sending hundreds of encrypted messages that have been hidden in files on digital photographs on the auction site eBay.com". These rumors were cited many times - without ever showing any actual proof- by other media worldwide, especially after the terrorist attack of 9/11.Unless the people are under active investigation, it is unclear that anyone will notice this activity.

V. Conclusions

An investigation about how information can be hidden and detected in static media like audio and video files is a vast matter for study. It has been observed that the whole idea with steganography is to avoid suspicion using a flow of ordinary information as a cover .Images when used for hiding information increase the file size. But the decision of use of cover depends on the purpose of hiding and on the size of information to be hidden. How a changed packet length affects the capacity, ease of detection, performance etc can always be considered under further investigations .It is impossible to know how widespread the use of steganography is by criminals and terrorists. The use of steganography is certain to increase and will be a growing hurdle for law enforcement and counterterrorism activities. Ignoring the significance of steganography because of the lack of statistics is "security through denial" and not a good strategy.

REFERENCES

[1] Provos, N., Honeyman, P.,"Detecting Steganographic Content on the Internet", ISOCNDSS'02, California, 2002.

[2] Wang, H & Wang, S, "Cyber warfare: Steganography vs. Steganalysis", *Communications of the ACM*, 47:10, October 2004

[3] Korjik ,V.,Morales-Luna,G.,"Information Hiding Through Noisy Channels", Lecture Notes in Computer Science,vol.2137,pp.42-50,Springer-Verlag,2001

[4] Johnson, N.F. & Jajodia, S., "Exploring Steganography: Seeing the Unseen", *Computer Journal*, February 1998

[5] Owens, M., "A discussion of covert channels and steganography", *SANS Institute*, 2002

[6] Westfeld, A., Pfitzmann,A.,"Attacks on Steganographic Systems", In Proceedings of Information Hiding-Third International Workshop, Springer Verlag,pp.61-76, September 1999.

[7] Lee, Y.K. & Chen, L.H., "High capacity image steganographic model", *Visual Image Signal Processing*, 147:03, June 2000

[8] Johnson, N.F. & Jajodia, S., "Steganalysis of Images Created Using Current Steganography Software", *Proceedings of the 2nd Information Hiding Workshop*, April 1998

[9] Dunbar, B., "Steganographic techniques and their use in an Open-Systems environment", *SANS Institute*, January 2002

[10] Moerland, T., "Steganography and Steganalysis", *Leiden Institute of Advanced Computing Science*, www.liacs.nl/home/ tmoerl/privtech.pdf

[11] Currie, D.L. & Irvine, C.E., "Surmounting the effects of lossy compression on Steganography", *19th National Information Systems Security Conference*, 1996

[12] Fridrich, J.Golyan, M.,DU,R., "Steganalysis Based on JPEG Compatibility " Special Session on Theoretical and Practical Issues in Digital Watermarking and Data Hiding, SPIE Multimedia Systems and Applications IV, Denver, CO, August 20-24,2001.

[13] "Reference guide: Graphics Technical Options and Decisions", http://www.devx.com/projectcool/Article/19997

[14] Bender,W., Gruhl,D., Morimoto,N.,Lu,A., "Techniques for Data Hiding ", IBM Systems Journal, vol.35, Nos 3&4, 1996.

[15] Manamalkav, M.,"Audio File Steganography", available at <www.cise.ufl.edu/~smanamal/steganography.htm> 9 April 2002.

Meenakshi Kaul holds a B.E degree in Electronics and Telecommunication from IET, Devi Ahilya University, Indore-INDIA. She is M.E in Computer Science from RGPV University Bhopal-INDIA. Her research interests include Network Security and Mobile Computing.