IMPLEMENTATION OF LSB STEGANOGRAPHY ALGORITHM IN MP3 AUDIO FILE

C. AQUINO[1] , I. RIVERA[2] , S. A. GARCÍA[1]

[1] ESIME CULHUACAN, National Polytechnic Institute. Av. Santa Ana San Francisco Culhuacan, Deleg. Coyoacan, C.P. 04430, México D.F.[2] Center for Innovation and Development in Computer Technology, National Polytechnic Institute. Unidad Profesional Adolfo López Mateos, Edificio del CIDETEC, México, D. F., 07700

E-mail:{caquino; irivera; silvestregarcia }@ipn.mx

Abstract: The steganography is a discipline that studies the techniques and algorithms to hide information, their main objective is to transform a hidden message through unusual means called carriers.

It is based in two basic principles: The first one, is to select very well the way or via that is best referring to the file or archive that is going to be hidden even though it looses quality it would not be perceptible. The second one is taking in mind the limitation of men if of perception we are talking, just as it is, the range of colors that even if they vary a little the human eye is not capable to decode it. In this work is presented a first approximation of the Bit Algorithm less significative (LSB) in audio files using C Language for its implementation.

Key Words: Steganography in audio, LSB codification, hidden information

1.- INTRODUCTION

Criptography and Steganography are two independent fields within informatics security, both complement each other, the first one hides the meaning of the message thus the second hides its existence. Each by separate cannot assure the secret, but if both techniques are applied to code and hide a message, the security level is increased and with it the possibilities of exit.

The steganography is the discipline that studies the assembly of techniques that have in common hiding sensible information, messages or objects within the container files, commonly multimedia: digital images, videos or audio files with the objective that the information may pass unaware to third ones and may be recovered only by the original or legitimate user[4].

Although multiple forms exist to hide an information, commonly are based on two principles:

A) To take advantage of stegomedia with superfluous information nonuseful that can be modified without raising suspicions, for example by means of the technique LSB which will be explained ahead.

B) To take advantage of the reordering of the elements that define stegonomedia, for example, to rearrange the pixels of the trowel of colors in a GIF file. Fig.1 The Monalisa is one of the images most used to exemplify the steganography in images.

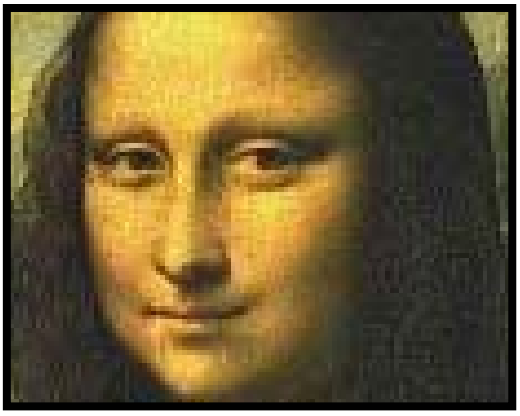


Figure 1. The Monalisa is one of the images most used to exemplify the steganography in images.

Steganography is classified in two ways; the one with protection against erasure and with protection against detection. The one that we will use as a study base (that is here first) enters the LSB algorithm. Next will be explained what each one of them does.

1.- The steganography of protection against detection:

One is based on hiding binary data in the maze of bits that supposes a file. The bits that compose the message to hide are introduced (either by adding them or by conducting arithmetical operations with the original ones) in the already existing file trying that the resulting file after realizing the changes, seems the original one, that is; without changes. For example we may find this type of steganography in images, sound and feasibles.

2.-Steganography against erasure protection:

Inside this division we may find two branches: water marks the information is hidden related to the object within in such a way that it may be extracted and valid afterwards by an specific computer. It adds data of copy rights, finger prints, besides containing the information of the owner it contains information of the original buyer or of the one that obtains the rights of use [4].

2.- RELATED WORK

At present, diverse methods and algorithms exist and are used to hide information within the multimedia archives: [5]

1) Masking and Filtrate: in this case the information is hidden within a digital image using water marks that include that information, just like the author rights, the property and licenses. The objective is to add an attribute to the image that acts like the cover.

2) Algorithms and Transformations: this method hides the message in bits of less important data.

3) Insertion in the less significant bit (LSB insertion): It consists of making use of the least significant bit of the pixels of an image and to alter it, the best results are obtained in images with color format RGB (Three bytes, components of color, by pixel) The same technique may be applied to audio and video.

4) Cetel technique: The use of the steganography in documents may work just by adding a blank space and the cards to the end lines of a document. This type of steganography is very effective, since the use of the blank spaces in targets and abates or tabs are not visible to the human eye, at least in the majority of the text editors, and they take place in a natural way in the documents (and that is why) it is very difficult that it raises suspicions.

5) Audio archives techniques: When information within audio files or archives is hidden, the most used technique is the of low bit encoding [5] that is similar to the LSB that is usually used in images [5] the problem that appears with the low bit encoding is that it may be appreciated by the human ear, so it is a risky method for those that use it specially if they are trying to hide information within an audio file.

6) Spread Spectrum: it works (at random) by means of the addition of random noises at the signal that the information is hidden inside an airline company and the propagation in all the frequency spectrum [5].

7) Echo Data Hiding: it uses the echoes in files or archives of sound with the aim of hiding information. What this method obtains best is that it may really improve the sound of audio within the file of audio.

8) Video techniques: it is common to use DCT method (Discrete Cosine Transform) and DCT functions by changing slightly each one of the images in the videos, the information is hidden within each video

photogram in such a way that it would not be perceptible by the human eye.

Fig.2 general scheme of the steganography

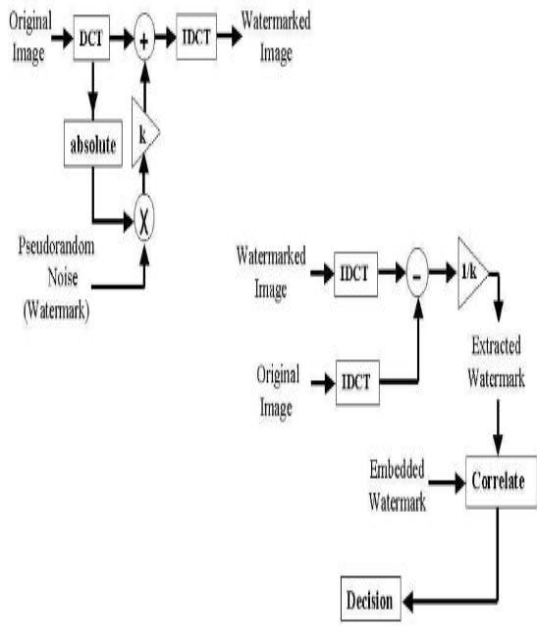
3.- OUR CONTRIBUTION

The human ear is extremely sensible to the changes in the audio patterns but not as much as to perceive changes within the same frequency. At the time of hiding a message in audio it is important to know the mean by which the message is going to be transmitted since it is not the same between digital-digital means (between computers) than between air-digital ways (microphone). That is why an LBS algorithm is developed.

3.1.- METHODOLOGY

The following figure shows how the steganography works.

This is one of the most used algorithms in steganography since it is easy to apply to an image and audio. A great amount of information may be hidden with just a small amount of video time or in a



very small image. The LBS process consists of choosing a subgroup $\{j_1,..,j_m\}$ of elements of the cover in execution of an operation of substitution C_{ji}<->m_i within the information since it changes the LBS of C_{ji} by some O or 1 bit, one could also image a substitution operation that changes more than a chunk of cover, for example storing two pieces of message in both less significant chunks of an element of the cover.

In the extraction process, the LSB of the selected elements of the cover are extracted and bordered until the secret message is reconstructed [1].

Next the algorithms of process of fixation and process extraction are shown:

Algorithm 1.- Process of fixation: the substitution of less significant chunk

Para $i = 1,...,lc$ hasta

$si \leftarrow ci$

Fin para

Para $i = 1 ...,lm$ hasta

Cálculo del índice ji donde almacenar i el mensaje del bit

$sji \leftarrow cji\ mi$

Fin para

Algorithm 2.- Process of extraction: the substitution of less significant chunk

Para $i = 1,...,lM$ hasta

Calculo de índice *ji cuando el i del mensaje almacena el bit más significativo*

$mi \leftarrow LSBC_{ij}$

Fin para

Algorithm 3.- Process of fixation: method of arbitrary interval

Para $i = 1...,lc$ hasta

$si \leftarrow ci$

Fin para

Generar secuencia aleatoria *ki* utilización de la semilla *k*

$n \leftarrow k1$

Para $i = 1,...,lm$ hasta

$sn \leftarrow cn\ mi$

$n \leftarrow n + ki$

Algorithm 4.- Process of extraction: method of arbitrary interval

Generar secuencia aleatoria *ki* utilización de la semilla *k*

$n \leftarrow k1$

Para $i = 1,...,lm$ hasta

$mi \leftarrow \text{LSB}cn$

$n \leftarrow n + ki$

Fin para

4.- DEVELOPMENT

The LSB process consists of choosing a subgroup $\{j_1,..,j_m\}$ of elements of the cover and the execution of an operation of substitution of C_{ji}<->m_i within the information since it changes the LSB of C_{ji} by some bit of O or 1.

One could also imagine an operation of substitution that changes more than a piece of the cover, for example storing two pieces of the message in both less significant pieces of an element of the cover.

In the extraction process the LSB of the selected elements of the cover are extracted and bordered until the secret message is reconstructed. [1]

5.- EXPERIMENTAL RESULTS

The results that appear next, were the product of the development that is described beforehand.

The program is in charge to compress the sound tracks of format MPGE III since it offers us a quality compression from 11 to 1 (128 kilobits per second) this proportions a very good opportunity of the concealment or hiding of information.

Although WMA has the best quality in general, I did not have the access code and only counted with an accomplishment for MP3Stego. The concealment process happens in the heart of the cover Layer III. Process of codification that is know in the cycle of codification. The inner process quantifies the input data

And increases the size until the quantified data can be based with the number available of filegrams.

Following is an interface of how the stenganography must become or should be done.

Fig.3 output of based filegrams

In the order to present or display these results we leaned in a software called audacity which serves to analyze the audio archives. The first that we did was to load in the program both audio archives, the original one and the other archive with steganogrphy and at first sight difference is not seen at the time of reproducing and the sound is absolutely not altered and the reproduction time is the same.

Figures 4 and 5 show us the comparison of the waves of both archives, in the right part is the original file and in the left part the composite file. We will thus call it composite file. The one which was applied the steganography for convenience during the explanation.

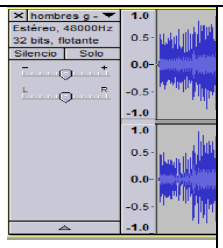


Fig 4.Original Image

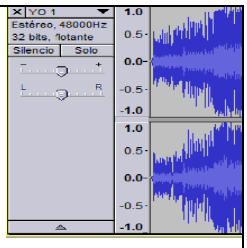


Fig 5. Image with LBS

5.1.- ANALYSIS OF DECIBELS

As you can appreciate in the previous images the significant changes are not shown and on the basis of it a comparison is realized in decibels and frequencies that are handled. In the figure 6 the results are shown.

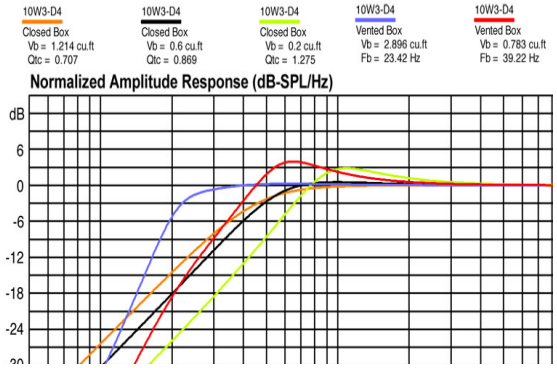


Fig. 6 Decibel analysis

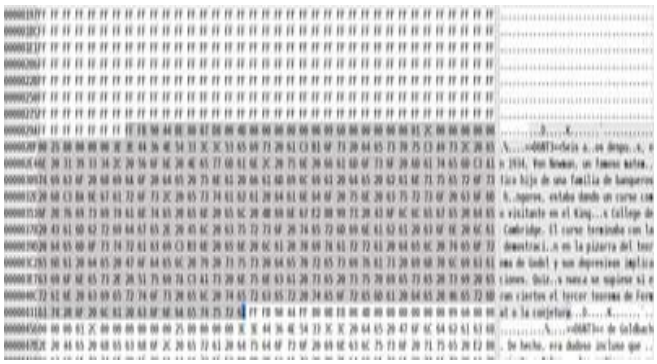


Fig. 7 Code of the obtained signal

6.-CONCLUSIONS

Due to the special characteristics of this work, the fundamental properties of the Human Auditory System has been studied to find “hollows” in order to develop a stego-system on codes of MP3 audio.

The realized study allow to obtain a quite ample vision and with an average (middle) level of depth from the present landscapes and composes a solid base from which it may be possible to continue researching in specialized methods, furthermore it has allowed us to establish guides from which a new stego-system can be designed and created. The study of basic stegoanalitic techniques allow us to know better the weakness of the own steganographic algorithms, fundamental knowledge to develop affective algorithms and with the desired quality.

Of the created and implemented stego-system, in no way was tried to create an “unbreakable” system but rather a departure point to create a complete system. Thus although the fact that, as it is commented, further on it is possible to be considered statically steganoanalyzed, this is something that is not simple even for basic stego-system, this should not be considered only as a defect, otherwise like the recognition of the weakness of the algorithm in the mentioned aspects and the procedure that marks the next steps to follow in order to reach a safe system.

Fig. 8 Codified signal

7.- REFERENCES

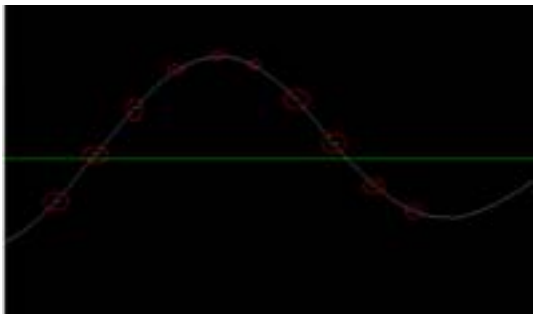
[1] Stefan Katzenbeisser, Fabien A.P Petitcolas, Information Hiding techiques for steganography and digital watermarking, pp.15 2000, Artech House, Inc

[2] ARTZ, Donovan. Digital Steganography: Hiding Data within Data. p. 77 ss. Junio 2001. Los Alamos National Laboratory. Spotlight.

[3]SIMMONS, G. J. "The prisioners' Problem and the subliminal channel", in Advanced in Criptology, Proceedings of CRYPTO 83, Plenum Press,1984.

[4] Roberto Gómez Cárdenas, "La esteganografía". Artículo publicado en la revista Bsecure de marzo del 2004.

[5] Greg Kipper , "Investigator's Guide to Steganography", Intellectual Property Protection Systems , Auerbach Publications 2004 .



8.- AUTHORS PROFILE

Ing. Carlos Aquino Ruiz Born in Mexico on April 2, 1973 is in Communications and Electronics Engineering from the National Polytechnic Institute, a research professor and areas of development are applied computing, data networking and security.

Master´s Degree Israel Rivera Zarate. Born in Mexico DF on Sep 11, 1970, holds a degree in Communications and Electronics from the National Polytechnic Institute in and earned his Master of Science in Engineering with specialization in Digital System by the National Polytechnic Institute in 2005.

Master´s Degree Silvestre Ascención García Sánchez. Born in Mexico DF on Jul 11, 1975, holds a degree in Communications and Electronics from the National Polytechnic Institute in and earned his Master of Science in Engineering with specialization in Microelectronics Digital Signal Processing by the National Polytechnic Institute in 2005.