

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 28-08-2012		2. REPORT TYPE Related Material		3. DATES COVERED (From - To) -	
4. TITLE AND SUBTITLE How do Statistical Detection Methods Compare to Entropy Measures			5a. CONTRACT NUMBER W911NF-11-1-0174		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 206022		
6. AUTHORS Jesus Velez Torres, Dr. Alfredo Cruz (Advisor), Dr. jeff Duffany (Mentor)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Polytechnic University of Puerto Rico 377 Ponce De Leon Hato Rey San Juan, PR 00918 -			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 58924-CS-REP.8		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT After the September 11, 2001 terrorist attack on the World Trade Center the United States declared war on terrorism. The Patriot Act (H.R. 3162) was passed, which authorizes law enforcement agencies to monitor and intercept the electronic communications of suspected terrorists. This meant that telephone communication and encrypted telephony would no longer be safe for terrorists to communicate with each other. This provoked them to search for alternate means to communicate sensitive information. The New York Times, USA Today, and the					
15. SUBJECT TERMS Statistical steganalysis; cryptography; steganography					
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT		15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	UU		Alfredo Cruz
					19b. TELEPHONE NUMBER 787-622-8000

## **Report Title**

How do Statistical Detection Methods Compare to Entropy Measures

### **ABSTRACT**

After the September 11, 2001 terrorist attack on the World Trade Center the United States declared war on terrorism. The Patriot Act (H.R. 3162) was passed, which authorizes law enforcement agencies to monitor and intercept the electronic communications of suspected terrorists. This meant that telephone communication and encrypted telephony would no longer be safe for terrorists to communicate with each other. This provoked them to search for alternate means to communicate sensitive information. The New York Times, USA Today, and the United States Institute of Peace have reported that terrorists may be using steganography and cryptography on the web as a means for covert communications [1, 22, 23]. These reports have been the basis of several studies into covert communication by terrorists. It is becoming increasingly important to detect the images that contain steganography such that we can reduce foul-play. This counter technique is known as steganalysis.

There are two problems in steganalysis: (1) detecting the existence of a hidden message and (2) decoding the message. This research is only concerned with the first problem, which is the detection of hidden messages using statistical steganalysis

***Title: How do Statistical Detection Methods Compare to Entropy Measures.***

*Jesús Vélez Torres*

*IT Management & Information Assurance*

*Dr. Alfredo Cruz*

*Department of Electrical & Computer Engineering*

*Polytechnic University of Puerto Rico*

**Abstract**

After the September 11, 2001 terrorist attack on the World Trade Center the United States declared war on terrorism. The Patriot Act (H.R. 3162) was passed, which authorizes law enforcement agencies to monitor and intercept the electronic communications of suspected terrorists. This meant that telephone communication and encrypted telephony would no longer be safe for terrorists to communicate with each other. This provoked them to search for alternate means to communicate sensitive information. The New York Times, USA Today, and the United States Institute of Peace have reported that terrorists may be using steganography and cryptography on the web as a means for covert communications [1, 22, 23]. These reports have been the basis of several studies into covert communication by terrorists. It is becoming increasingly important to detect the images that contain steganography such that we can reduce foul-play. This counter technique is known as steganalysis.

There are two problems in steganalysis: (1) detecting the existence of a hidden message and (2) decoding the message. This research is only concerned with the first problem, which is the detection of hidden messages using statistical steganalysis.

**Introduction**

Steganography is the art of covered or hidden writing. The purpose of steganography is covert communication to hide a message from a third party. This differs from cryptography, the art of secret writing, which is intended to make a message unreadable by a third party but does not hide the existence of the secret communication.

Steganography hides the covert message but not the fact that two parties are communicating with each other. 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. The use of a steganography key may be employed for encryption of the hidden message and/or for randomization in the steganography scheme. In summary:

$$\text{steganography\_medium} = \text{hidden\_message} + \text{carrier} + \text{steganography\_key}$$

As an increasing amount of data is stored on computers and transmitted over networks, it is not surprising that steganography has entered the digital age. On computers and networks, steganography applications allow for someone to hide any type of binary file in any other binary file, although image and audio files are today's most common carriers.

Steganography has a number of nefarious applications; most notably hiding records of illegal activity, financial fraud, industrial espionage, and communication among members of criminal or terrorist organizations.

The most common steganography method in image files employs some type of least significant bit substitution or overwriting. The least significant bit term comes from the numeric significance of the bits in a byte. The high-order or most significant bit is the one with the highest arithmetic value, whereas the low-order or least significant bit is the one with the lowest arithmetic value.

As a simple example of least significant bit substitution, imagine "hiding" the character 'G' across the following eight bytes of a carrier file (the least significant bits are underlined):

10010101 00001101 11001001 10010110

00001111 11001011 10011111 00010000

A 'G' is represented in the American Standard Code for Information Interchange (ASCII) as the binary string 01000111. These eight bits can be "written" to the least significant bit of each of the eight carrier bytes as follows:

```
10010100 00001101 11001000 10010110
00001110 11001011 10011111 00010001
```

In the sample above, only half of the least significant bits were actually changed. This makes some sense when one set of zeros and ones are being substituted with another set of zeros and ones.

Least significant bit substitution can be used to overwrite legitimate RGB color encodings or palette pointers in GIF and BMP files, coefficients in JPEG files, and pulse code modulation levels in audio files. By overwriting the least significant bit, the numeric value of the byte changes very little and is least likely to be detected by the human eye or ear.

Least significant bit substitution is a simple, albeit common, technique for steganography. Its use, however, is not necessarily as simplistic as the method sounds. Only the most naive steganography software would merely overwrite every least significant bit with hidden data. Almost all use some sort of means to randomize the actual bits in the carrier file that are modified. This is one of the factors that make steganography detection so difficult.

One other way to hide information in a paletted image is to alter the order of the colors in the palette or use least significant bit encoding on the palette colors rather than on the image data. These methods are potentially weak, however. Many graphics software tools order the palette colors by frequency, luminance, or other parameter, and a randomly ordered palette stands out under statistical analysis.

## **Problem**

Steganography is the art and science of communicating in a way which hides the existence of the communication. The events of 911 prompted significant discussion and speculation as to the use of Steganography by terrorists for clandestine and secured

communications. According to nameless “U.S. officials and experts” and “U.S. and foreign officials,” terrorists groups are “hiding maps and photographs of terrorists targets and posting instructions for terrorists activities on sport chat rooms, pornographic bulletin boards and other Web sites”[1]. Three years ago, FBI Director Freeh tried to convince government authorities that terrorists were using encryption and steganography to support their organizations. Freeh urged legislators to enact stricter Internet usage laws, emphasizing that ignoring the issue would not only cause harm to the United States government, and indeed, the world, is hastening to discover viable methods for the detection and prevention of this electronic subterfuge, in the hope that denying criminal networks (the use of the Internet as a communications medium) will render them incapable of mounting another attack that could cost more lives.

## **Goal**

The goal is to statistically analyze the least significant bit(s) of each color dimension of each pixel to look for some kind of a pattern. In the absence of a hidden message this should look like random noise. Addition of a hidden message will affect the entropy of the data. This difference should be detectable by comparing the entropy of unaltered picture files with the entropy of files with embedded steganography. Also I will be using several different statistical attacks and compare them to the entropy results.

## **Research Questions**

- Can the less significant byte method be a sure proof way of identifying images containing hidden information?
- Is entropy comparison a reliable way to identify steganography on images?
- Which method is more reliable, when analyzing images, entropy, Chi-Squared or RS attacks?
- Which method is more efficient, when analyzing images, entropy, Chi-Squared or RS attacks?

## **Relevance**

Steganography made the headlines when the New York Times [22] and USA Today [1] published articles on terrorists using steganography for covert communications. The New York Times published articles regarding the French defense ministry, which then reported that terrorists could be using steganography to communicate their plans to blow up the U.S. Embassy in Paris [22]. *“The terrorists were under instruction that all their communications were to be made through pictures posted on the Internet, the defense official said.”* Using steganography for malicious purposes has become a threat not only to individuals and businesses, but also to democratic governments.

Steganography and cryptography are technologies that can be used for secret and secure communications. Hypertext markup language (HTML) and the World Wide Web are technologies that can be used for open communications. When combined, steganography, cryptography, HTML and the Web can be very effective tools for covert communications among terrorist groups, drug traffickers and other nefarious groups. They can use the Internet as their global network to communicate and “hide information in plain sight” of law enforcement agencies.

The “Search for International Terrorist Entities” (SITE) Institute shows evidence that terrorists make effective use of the Web in their propaganda, recruitment, and fund raising efforts [23]. It is common knowledge that terrorist websites frequently publish and disseminate information in anonymity, and using various groupings. They could also be using the Web for covert communication. This would make it more difficult for law enforcement agencies to discover and intercept covert communications.

## **Review of the Literature**

Given time, it has been possible to break every steganographic system that has ever been published. As a result of this, new techniques are developed to improve upon the flaws of the predecessor. It is very much the same case in steganalysis where the algorithms are often tweaked or combined in order to attack the latest steganography algorithm. The two fields therefore operate in a cat and mouse style strategy with steganography aiming to be ahead of the field such that covert communications may

exists. There is then a call for steganalysis to catch up such that covert communications are minimized as much as possible.

One of the most basic steganographic techniques is achieved by manipulating the pixel values of the cover image in sequence such that they then turn into code that can be used to reconstruct the message when extracting. The most popular method for developing this concept left a pattern in the beginning of the images that steganalysts bought to light. This led to the development of a series of techniques that could detect the existence of steganography within any stego-program that followed this embedding strategy, and so steganographers set about developing a new system that would render the attacks useless.

This led to the development of randomized embedding in pixel values, and once again steganalysts found a method for detecting steganography for all images created using this strategy.

## **Approach**

The work plan would be to hide messages, i.e. pictures, text messages and other types of files. Given the proliferation of digital images, and given the high degree of redundancy present in a digital representation of an image (despite compression), there has been and increased interests in using digital images as cover-objects for the purpose of steganography. There should be noted that there has been much work on embedding techniques which make use of the transform domain or more specifically JPEG images due to their wide popularity.

Different open source programs, which use different methods of steganography, will be used to hide messages and files into pictures. The data from the pictures, with hidden messages, will be imported to a statistics program like R language, StegSecret, Vsl Studio, etc. There will be a two part analysis, first the less significant bit portion of the pictures will be analyzed and compared with the original picture, and second the pictures will be analyzed and compared in terms of statistical attacks.

## **Milestones**



The major steps necessary to develop and complete the project in a timely fashion would be to:

1. Learn how to use the programs steganography and steganalysis.
2. Familiarize myself with the different algorithms that are used to embed the data on the digital images.
3. Import the data acquired to the statistical programs and analyze, and compare it to the original data from the source material.
4. Create tables and graphs using the results from the analysis.
5. Redacting my project into a well written paper that would be used to present my results.

## Resources

For this project I will need mid range computer, that can process the data from the different open source programs, which are readily available, and will be using for the project. Some of these programs are:

- *Hide Seek*: this are two programs. Hide produces a file called outfile.gif, so the original gif is left untouched. Seek produces a file called whatever you tell it to, via the name you choose for <outfile.ext>. It works by using the Least Significant Bit of each pixel to encode. Also it uses dispersion to spread the data (and thus the picture quality degradation) out a bit throughout the image in a pseudo-random fashion
- *JP Hide Seek*: these are also two programs JPHIDE.EXE is a DOS program to hide a data file in a jpeg file. JPSEEK.EXE is a DOS program to recover a file hidden with JPHIDE.EXE
- *StegDetect*: is an automated tool for detecting steganographic content in images developed by Niels Provos. It is capable of detecting several different steganographic methods to embed hidden informations in JPEG images. Currently, the detectable schemes are:
  - jsteg
  - jphide

- invisible secrets
- outguess
- F5(header analysis)
- appendix and camouflage
- *Stegbreak*: is a tool used to launch dictionary attacks against Jsteg-Sheel, JPhide and Outguess.
- *Hide Reveal*: Composed of a Java steganographic library and GUI to use steganography. Primarily intended for research communities on security and steganography to implement new dissimulation and steganalysis algorithms.
- *StegSecret*: is a steganalysis open source project (GNU/GPL) that makes possible the detection of hidden information in different digital media. StegSecret is a java-based multiplatform steganalysis tool that allows the detection of hidden information by using the most known steganographic methods. It detects EOF, LSB, DCTs and other techniques.

Also I will use statistical programs and tools to compare and analyze my results some of these programs are:

- *Steganography Studio*: Steganography Studio software is a tool to learn, use and analyze key steganographic algorithms. It implements several algorithms highly configurable with a variety of filters. Also implements image analysis algorithms for the detection of hidden information. This software is developed in Java, allowing use in any operating system.
- *R Language*: R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.
- *VSL studio*: VSL is free image steganography and steganalysis software in form of graphical block diagramming tool. It allows complex using, testing and adjusting different steganographic techniques and provides simple GUI along with modular, plug-in architecture

It is important to have access to specialized books, which can be acquired online or at the university library. Lastly professor guidance will be needed, in regards to questions about some of the programs, tools, and the subject in general. I already talked to Dr. Duffany which recommended some of the tools and is an expert on R Language. Also Dr. Alfredo Cruz is guiding me through the process of developing the needed literature for the project.

### **Digital Image Format: JPEG**

JPEG stands for Joint Photographic Experts Group, named after the group that designed the original file format. JPEG file format is the industry standard for digital image delivery over the internet and for amateur photography. It is favored because its compression produces a reasonable image with a small file size.

JPEG compression reduces image files to about 5% their normal size with loss due to destructive or lossy compression. Lossy is used for reducing the number of pixels by simply removing the colors that the human eye cannot see. This action reduces the image's size but also lowers the image quality. JPEG can achieve compression ratios by the factor of 10 to 20 compared to TIFF format with a very little visible loss in image quality. This also means that image information is lost each time the files are manipulated and saved. This loss results of the decompression and recompression from subsequent saving of the image. JPEG achieves acceptable quality results for the internet and amateur photography by successfully taking advantage of human eye limitations. JPEG's generational loss of data means that it is not favored for professional quality printing or archival image delivery over the internet. Lastly JPEG files can be readily made from TIFF files.

### **Statistical Steganalysis**

The ideal of a perfectly secure steganography is that, that the presence of hidden message should be perceptually and statistically invisible. Since embedding a secret message in a cover media needs modification of cover, steganography inevitably leaves some traces in statistical properties of the medium [26]. This encourages an eavesdropper to discover

distortions in statistical properties of the cover signal, through statistical steganalysis, to detect whether it contains hidden data [27].

Recently, there have been powerful statistical techniques reported in literature. Westfeld and Pfitzmann presented a method for analysis of LSB embedding based on a definition of Pairs of Values. This principled method provides reliable results for known placement of the message embedding and states, if the message is scattered randomly within the cover its length is comparable to that of the cover's LSB array.

A more accurate technique was reported by Jessica Fridrich and R Du which they published in October 2001. It is known as RS attack or "Reliable Detection of LSB Steganography in Grayscale and color images". The algorithm they use is very precise for the detection of pseudo-aleatory LSB steganography. Its precision varies with the image but, its referential value is a 0.005 bits by pixel or .5% of the space the the information occupies. In other words if the hidden information exceeds the 0.005 bits per pixel, the algorithm will detect it and give an estimate of the size.

## **Attack Testing**

### **Overview**

It is necessary to test the attacks such that we can be confident that they are working as expected. It can be a little subjective when testing steganalytical functions and therefore a little harder to test. However, through the research carried out we can run sufficient tests to see that they are behaving as expected. This part discusses the testing methodology that was carried out for each of the functions that comprise the system.

### **Chi-Squared Attack**

Andreas Pfitzmann and Andreas Westfeld [24] created a method based on statistical attack called Chi-Squared which analysis the Pair of Values (PoV) that are exchanged during sequential embedding. They claim that the LSB in images are not completely random. Rather the frequencies of each of the two pixel values in each PoV tend to lie far from the mean of the PoV. In other words, it is rare that for the frequency of pixel value

$2_k$  and  $2_{k+1}$  become equal or near equal Pair of Values in images and bases the probability of embedding on how close to equal the level pixel are and their corresponding odd pixel values are in the test image [25]. This attack works on any sequential embedding. This type of embedding technique makes PoVs in the values we embed in. This will affect the histogram  $Y$  of the image's pixel value  $k$ , while the sum of  $Y_{2i} + Y_{2i+1}$  will remain unchanged. Thus the expected distribution of the sum of adjacent values is 2 and the  $\chi^2$  value for the difference between distributions with  $v-1$  degrees of freedom is 3, from 2 and 3 we get the  $\chi^2$  statistic for our PoVs as in 4.

$$E(Y_{2i}) = \frac{1}{2}(Y_{2i} + Y_{2i+1}) \quad (2)$$

$$\chi^2 = \sum_{i=1}^v \frac{(F - E(F))^2}{E(F)} \quad (3)$$

$$\chi_{PoV}^2 = \sum_{i=1}^{127} \frac{\left(Y_{2i} - \frac{1}{2}(Y_{2i} + Y_{2i+1})\right)^2}{\frac{1}{2}(Y_{2i} + Y_{2i+1})} \quad (4)$$

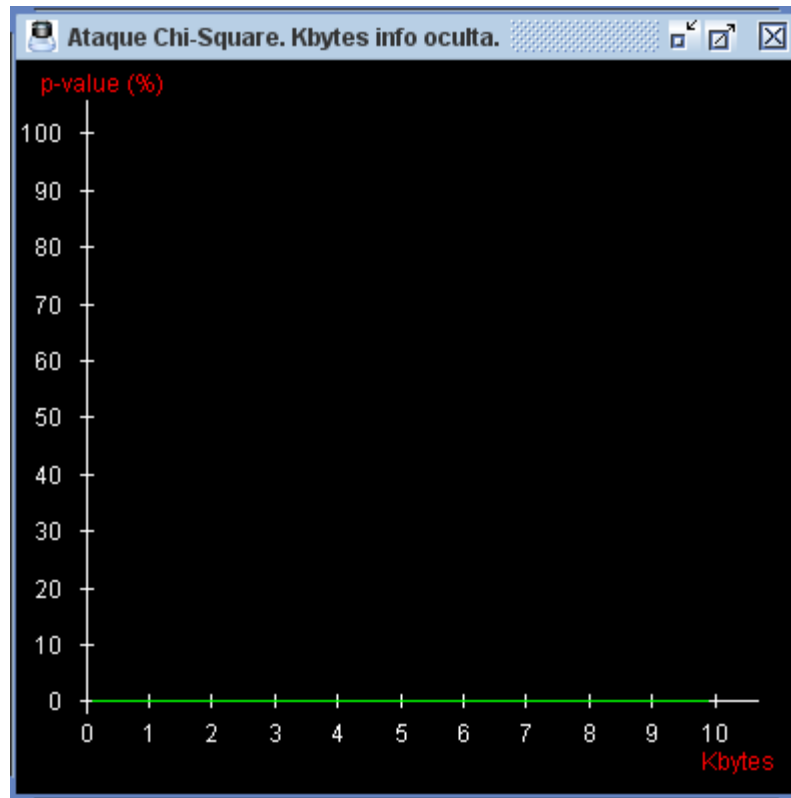
The probability of embedding is determined by calculation the p-value 5 for a sample from the values examined, which starts at the beginning of the image and gets increased for each measurement. This attack though, does not work for pseudo-random type of embedding.

$$p = \Pr(\chi_{k-1}^2 \geq \chi_{PoV}^2) = 1 - \frac{1}{2^{\frac{k-1}{2}} \Gamma\left(\frac{k-1}{2}\right)} \int_0^{\chi_{PoV}^2} e^{-\frac{x}{2}} x^{\frac{k-1}{2}} dx, \quad (5)$$

### Chi-Squared Attack Results

I embedded an image with data and ran some tests using several statistical tools we will first look at the results using the attack known as Chi-Squared Attack. Using the tool known as StegSecret and a specialized tool for Chi-Squared attacks I was able to determine that an image had hidden data when the technique used to embed the data was

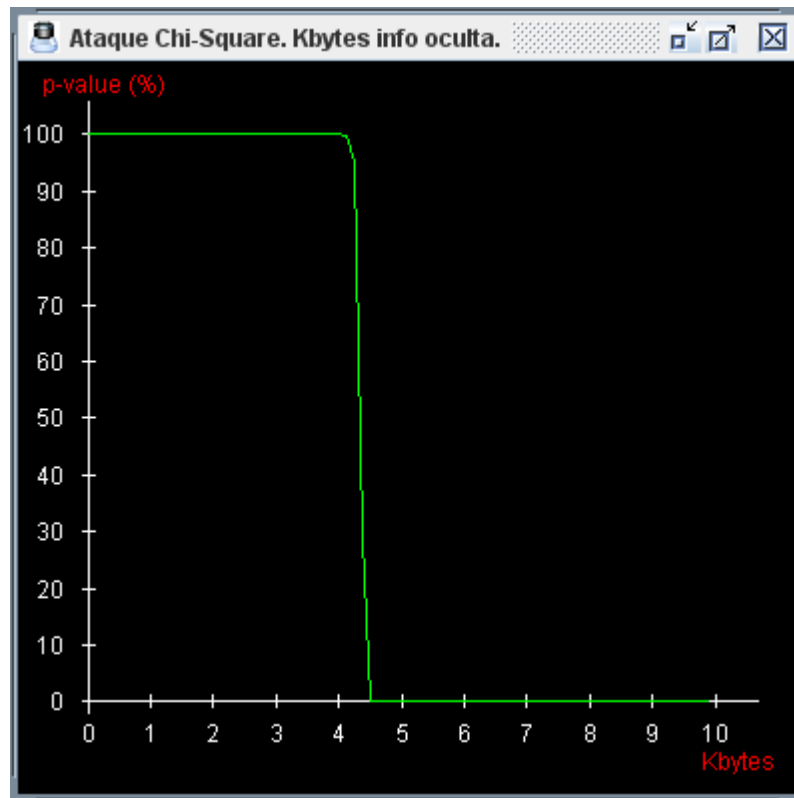
sequential. The downside to using this type of attack is that it is really hard to determine if an image has data embedded if the technique used was aleatory or not sequential. This occurs because the nature of this particular algorithm is to find information that is organized on an unorganized universe.



**Figure 1. Results of an image without any hidden information using StegSecret to implement the Chi-Squared attack.**

On the figure 1 and 2 we have the results of the original image after I applied the steganalysis algorithm. The tool used to run the statistical attack on this picture was the StegSecret tool which detected that everything was normal. This could mean two different things either the image had no data embedded in it or the steganographic method used to hide the data was not sequential. In which case the algorithm saw the picture as it should, which is a lot of pixels without any Pair of Values formed. This is where this type of attack fails; there are methods that can randomly choose the LSB on which the data is going to be hidden. On the other hand having the original image is not necessary to find that it has hidden or embedded data.

On figure 2 below a Sequential method was used to embed data onto de digital image. This goes against the nature of digital image LSB which is not completely random. This means that in this picture the  $2k$  and  $2k + 1$  pixels where found to be equal or nearly equal Pair of values. We can see that the percentage of the PoV's is high. This means there is hidden data embedded.


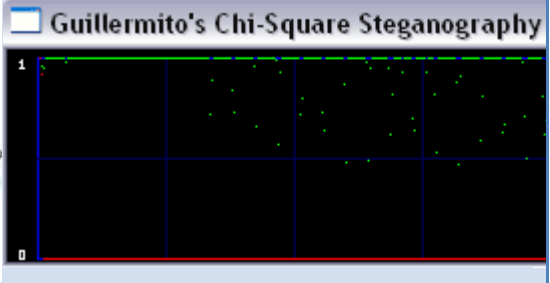

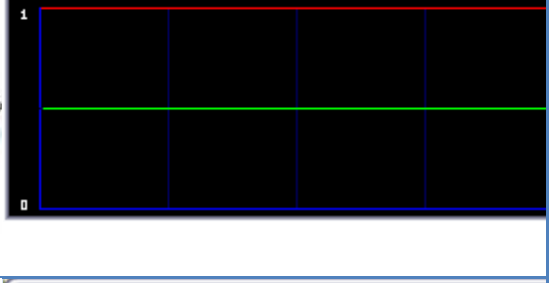

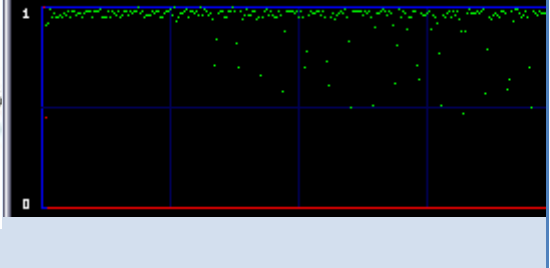


**Figure 2. Results of the same image with hidden information embedded using a sequential technique.**

I performed a statistical attack on figure 3 using a Chi-Squared tool. On this tool instead of only considering the LSB, it also takes into account more bits. Then it checks for the Pair of Values distribution is close to 50/50. So it actually considers 8 bits of each byte, having 256 possible values, and 128 Pairs of Values.

The first step on the process is to calculate the 128 PoV's that we extract from the image by measuring frequency of the Pair of Values. After this is done a table is created by the program. This table is then compared to a second table created by calculating the theoretical frequency which is 50% each of the PoV that were already calculated.

The program will output a graph with two curves. The first one in red is the result of the Chi-Squared test. If it is close to one, then the probability for a random embedded message is high. The second output is a simple idea, to add a second layer of verification: it's the average value of the LSBs on the current block of 128 bytes. So, if there is a random message embedded, this green curve will stay at around 0.5. On the graph network, every blue line represents 1 kilobyte of embedded data.

Fig	Hidden	Image	Graph
3	Nothing		
4	4kb of Sequential data		
5	4kb of Non-Sequential data		

We can see that the red line is practically on top of the 0 which means there are no messages hidden on the image. Also the green line is at 1 which concurs with the first layer of the detection and there is no blue line which means that there is 0 kilobyte of embedded data.



In figure 4 we have approximately 4 kilobytes of hidden information. The results came out positive the tool detected that there was information embedded, this is represented by the red line which is on top of 1. Also the test detected that there were in fact 4kb of information in the digital image. Lastly the green line is at 0.5 which means there is a random message hidden within the image. On the other hand I embedded the same information of figure 5 using an non-sequential method. As expected the results were negative there was no data found using the Chi-Squared attack. This proves that although this method is excellent for detecting information inserted using a sequential LSB technique, it has some flaws. As represented by the red line which is in 0, there is no data embedded. There is no blue vertical line covered and this means that there are 0 kilobytes of hidden information.

### RS Attacks

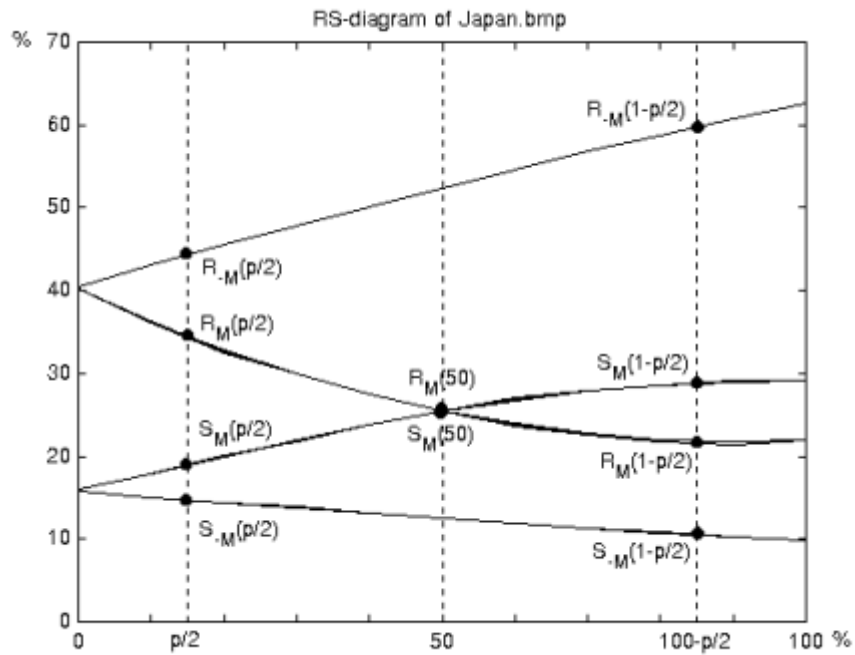


Figure 6. RS-diagram

The principle of the RS method is to estimate the four curves of the RS diagram and calculate their intersection using extrapolation. The general shape of the four curves in the diagram varies with the cover-image from almost perfectly linear to curved. Fredrich collected experimental evidence that  $R_M$  and  $S_M$  curves are well modeled with straight

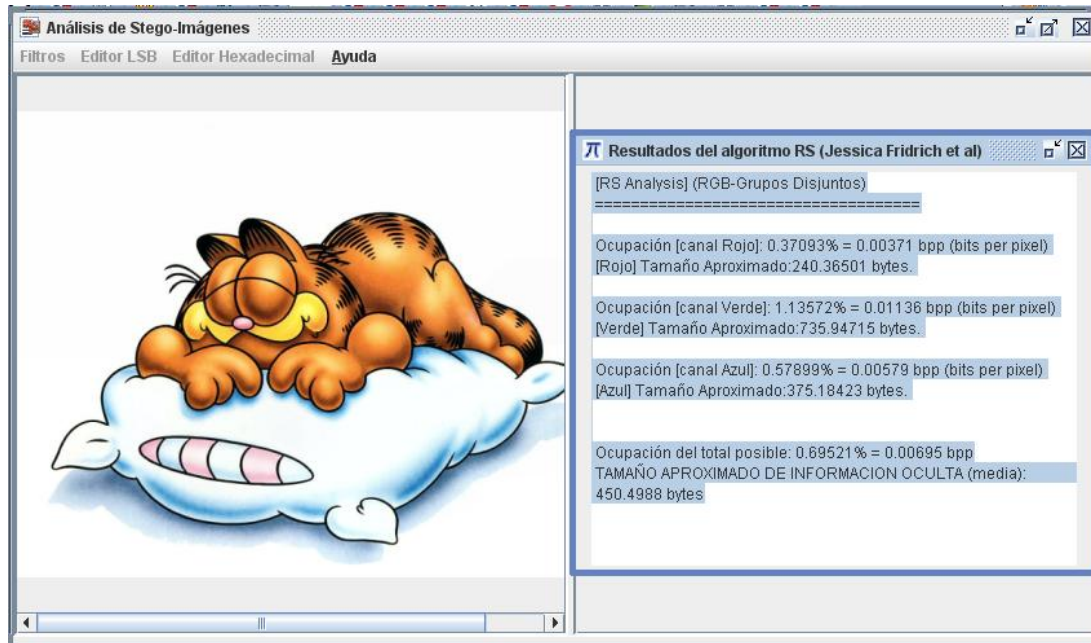
lines, while the “inner” curves  $R_M$  and  $S_M$  can be reasonably well approximated with second degree polynomials. The parameter of the curves can be determined from the points marked in Figure 6. RS works by partitioning images into groups consisting of  $n$  adjacent pixels, and computes the “smoothness” of each group using a discrimination function [3].

$$f(G) = f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

Each group is classified as “regular” or “singular” depending on whether the pixel noise within the group is increased or decreased after flipping the LSBs of a fixed set of pixels within each group. The classification is repeated for a dual type of flipping. Some theoretical analysis and some experimentation show that the proportion of regular and singular groups form curves quadratic in the amount of message embedded by the LSB method. Under these assumptions the proportions of regular and singular groups with respect to the standard and dual flipping, sufficient information can be gained to estimate the proportion of an image which data is hidden. The estimates in this method can be extremely accurate, but fails when the assumption does not hold.

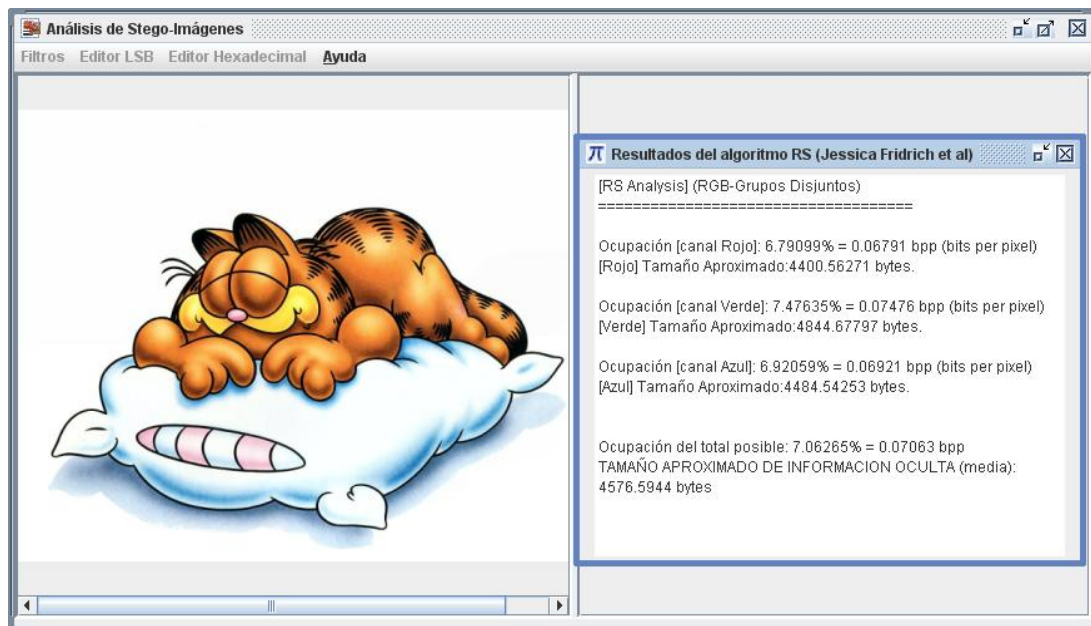
### **Accuracy**

Like any other steganalysis attack there are several different factors that can influence the accuracy of their results. First it is important to mention that even original images may indicate a small non-zero message length due to random variations. This could be both positive and negative and it puts a limit on the theoretical accuracy of the RS method. Also very noisy images have a difference between the number of regular and singular pixels in the image are small. This causes the lines in the RS diagram to intersect at a small angle which causes the accuracy to decrease. Lastly the message placement has a direct impact on the accuracy. This method is more accurate for messages that are randomly scattered in the stego-image than for messages concentrated in localized areas.



**Figure 7. RS Attack applied to an image that does not contain hidden data.**

On Figure 7 I used a RS attack on an image that did not contained any hidden information. The percentage was 0.69521 % which means it has given us a false positive which is natural for any kind of test, no test is foolproof. This occurred because the sample reflected the limit for detection on this image. An image free of hidden information is suppose to show a really low percentage which indicates



**Figure 8. Contains hidden data using a non-sequential technique.**

On figures 8 and 9 the images had information hidden using sequential and non-sequential techniques. The tests ran on the images can back positive with a high percentage of hidden information on each. This proves that RS method can be used to attack different steganography techniques and can even calculate

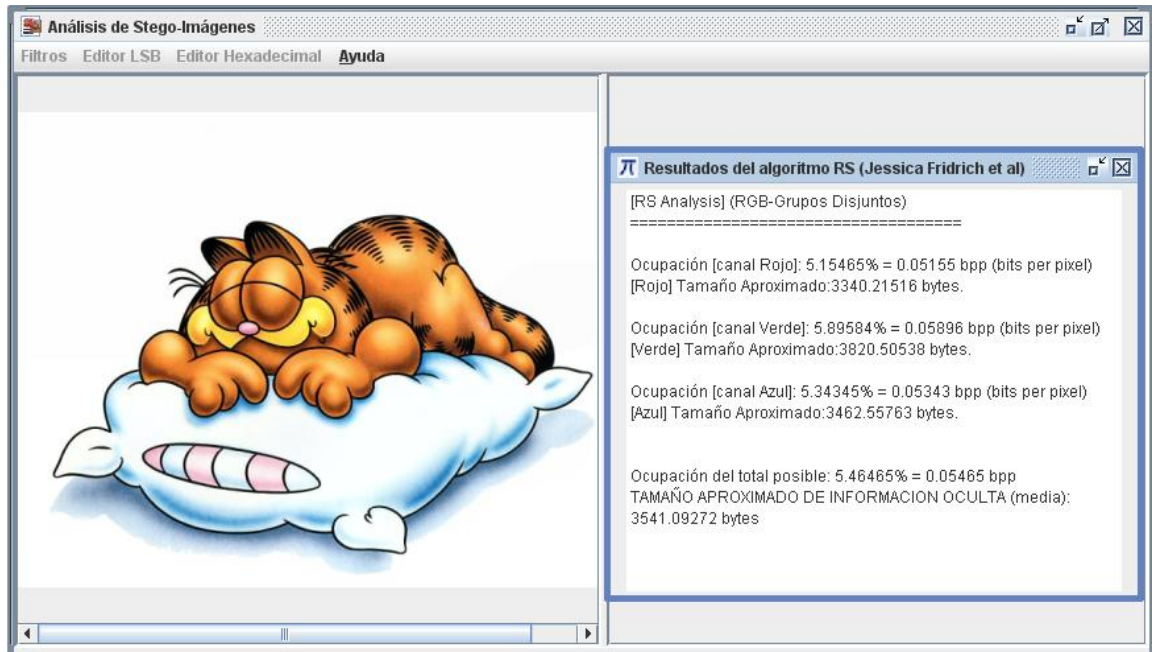


Figure 9. Contains hidden data using a sequential embedding technique.

## Entropy

The entropy is a measure of uncertainty which is associated with a random variable. This is usually referred to as Shannon entropy, which qualifies the expected value of the information contained in a message, usually bits. Shannon is a measure of the average information content one is missing when one does not know the values of the random variable. This method represents an absolute limit on the best possible lossless compression, under certain constraints, on images. This treats messages to be encoded as a sequence of independent and identically distributed random variables.

Digital images are composed of different brightness pixels which occupy different regions of an image. The different pixels show different shapes, and different shapes include different information in the image. Image entropy reflects the information amount of the two-dimensional digital image pixel gray values, and they can be used to describe the pixel gray value information distribution. The difference among the image

entropy corresponds to the visual differences among the images. Steganography will change the image pixel distribution probability and color information. The more information is hidden, the greater change the entropy will have.

## Entropy Attack

Experimental results showed that entropy is a suitable method for discriminating stego-images from suspicious images. It can differentiate stego-images from cover images with a certain rate of success. The problem is that the original image is required for comparison when using the MaxEntropy method in R image which is an extension or package in R Language.



**Figure 10.** Comparison between two images. The left image has hidden information using a sequential technique and the image on the right is the original image.

The image was converted into 8 bit grayscale which is required by the method and then the MaxEntropy plugin was used. There are very notable changes and characteristics on the images. Figure 11 is the same test used on a different image. I show the differences in a small part of the image.



**Figure 11**

## Conclusion

In general, steganographic algorithms have determined thresholds of detection, in other words if the information is embedded having these algorithms in mind they will not work correctly. In the same way their results can vary depending from image to image. This works this way because even if the image has no hidden information, in some cases, it will return a false positive in the form of a few hundred bytes. On the other hand if we embed very little information the algorithms might not detect them and it would not notice a difference between the original image and the altered one. This is not applicable to RS attack.

The Chi-Squared algorithm is excellent for detecting embedded information using LSB sequential techniques. It is even better when you could use different filters to determine where information might be hidden and then apply the Chi-Squared attack to the suspected area. On the contrary the results might vary and it effectively might be lower to what it could be when we have pre-selected areas to work on.

The entropy methods plugins in R language required us to have the original image to compare the change in noise of the images. This puts this method behind the other two because the original image is not always going to be available and even if we acquire the image there is no way to tell if it has been altered which would also change the noise in it.

There is no perfect or magical way to determine if images have hidden information. This is like a cat and mouse game where you can only react to the circumstances. Every day a new method for embedding information is created. This is a continually evolving field as new steganography methods emerge new attacks will have to be developed to counter them. The best way to determine whether an image has hidden data is to use different attacks and techniques on them and working the images out. Also part of the key to a successful analysis is to have a experienced analyst, that is consistent with the results uses the right tools and the right interpretation to the results.

## References

- [1] Kelley, J., & TODAY, U. (n.d.). Terror groups hide behind Web encryption. *News, Travel, Weather, Entertainment, Sports, Technology, U.S. & World - USATODAY.com*. Retrieved April 18, 2012, from <http://www.usatoday.com/life/cyber/tech/2001-02-05-binladen.htm>
- [2] Mohammed Ishaque, Syed Abdul Sattar, "Quality Based Jpeg Steganography Using Balanced Embedding Technique," *icetet*, pp.215-221, 2009 Second International Conference on Emerging Trends in Engineering & Technology, 2009
- [3] Jessica Fridrich, Miroslav Goljan, Rui Du, "Detecting LSB Steganography in Color and Gray-Scale Images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22-28, Oct.-Dec. 2001, doi:10.1109/93.959097
- [4] Bin Xia, Xingming Sun, Jiaohua Qin, "Steganalysis Based on Neighbourhood Node Degree Histogram for LSB Matching Steganography," *mines*, vol. 2, pp.79-82, 2009 International Conference on Multimedia Information Networking and Security, 2009
- [5] Adel Almohammad, Gheorghita Ghinea, Robert M. Hierons, "JPEG Steganography: A Performance Evaluation of Quantization Tables," *aina*, pp.471-478, 2009 International Conference on Advanced Information Networking and Applications, 2009
- [6] Adel Almohammad, Robert M. Hierons, Gheorghita Ghinea, "High Capacity Steganographic Method Based Upon JPEG," *ares*, pp.544-549, 2008 Third International Conference on Availability, Reliability and Security, 2008
- [7] Qingzhong Liu, Andrew H. Sung, "Detect Information-Hiding Type and Length in JPEG Images by Using Neuro-fuzzy Inference Systems," *cisp*, vol. 5, pp.692-696, 2008 Congress on Image and Signal Processing, Vol. 5, 2008
- [8] Abbas Cheddad, Joan Condell, Kevin Curran, Paul Mc Kevitt, "Biometric Inspired Digital Image Steganography," *ecbs*, pp.159-168, 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ecbs 2008), 2008

- [9] Mankun Xu, Tianyun Li, Xijian Ping, "Estimation of MB steganography based on least square method," *icassp*, pp.1509-1512, 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009
- [10] Xiaoyi Yu, Noboru Babaguchi, "An Improved Steganalysis Method of LSB Matching," *iih-msp*, pp.557-560, 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008
- [11] Xing Li, Tao Zhang, Kaida Li, Xijian Ping, "A Blind Detection Method for Additive Noise Steganography in JPEG Decompressed Images," *mines*, pp.489-493, 2011 Third International Conference on Multimedia Information Networking and Security, 2011
- [12] Jifeng Huang, "The Algorithm of Estimating Location of the Embedded Secret Message in Stego Image," *itcs*, vol. 2, pp.205-208, 2009 International Conference on Information Technology and Computer Science, 2009
- [13] Qingzhong Liu, Andrew H. Sung, "Detect Information-Hiding Type and Length in JPEG Images by Using Neuro-fuzzy Inference Systems," *cisp*, vol. 5, pp.692-696, 2008 Congress on Image and Signal Processing, Vol. 5, 2008
- [14] Neil F. Johnson, Sushil Jajodia, "Exploring Steganography: Seeing the Unseen," *Computer*, vol. 31, no. 2, pp. 26-34, Feb. 1998, doi:10.1109/MC.1998.10029
- [15] Rajni Goel, Moses Garuba, Chunmei Liu, Trang Nguyen, "The Security Threat Posed by Steganographic Content on the Internet," *itng*, pp.794-798, International Conference on Information Technology (ITNG'07), 2007
- [16] Rajni Goel, Moses Garuba, Chunmei Liu, Trang Nguyen, "The Security Threat Posed by Steganographic Content on the Internet," *itng*, pp.794-798, International Conference on Information Technology (ITNG'07), 2007
- [17] Greg Goth, "Steganalysis Gets Past the Hype," *IEEE Distributed Systems Online*, vol. 6, no. 4, pp. 2, April 2005, doi:10.1109/MDSO.2005.22
- [18] George Abboud, Jeffrey Marean, Roman V. Yampolskiy, "Steganography and Visual Cryptography in Computer Forensics," *sadfe*, pp.25-32, 2010 Fifth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering, 2010



[19] Hsien-Wen Tseng, Chin-Chen Chang, "Steganography Using JPEG-Compressed Images," cit, pp.12-17, Fourth International Conference on Computer and Information Technology (CIT'04), 2004

[20] McCullagh, D. (2001, February 7). Bin Laden: Steganography Master?. *WIRED*. Retrieved April 18, 2012, from [www.wired.com/politics/law/news/2001/02/41658?currentPage=all](http://www.wired.com/politics/law/news/2001/02/41658?currentPage=all)

[21] Schneier, B. (2001, July 24). Terrorists and steganography - ZDNet. *Technology News, Analysis, Comments and Product Reviews for IT Professionals | ZDNet*. Retrieved April 18, 2012, from <http://www.zdnet.com/news/terrorists-and-steganography/116733>

[22] KOLATA, G. (2001, October 30). Veiled Messages of Terror May Lurk in Cyberspace - NYTimes.com. *NY Times Advertisement*. Retrieved April 18, 2012, from <http://www.nytimes.com/2001/10/30/science/veiled-messages-of-terror-may-lurk-in-cyberspace.html?pagewanted=all>

[23] SITE. (n.d), "The Search for International Terrorist Entities", <http://www.siteinstitute.org/>

[24] Westfeld, Andreas and Andreas Pfitzmann. "Attacks on Steganographic Systems: Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S-Tools—and Some Lessons Learned." 3rd International Workshop on Information Hiding (2000).

[25] J. Fridrich. "Feature-Based Steganalysis for JPEG Images and Its Implications for Future Design of Steganographic Schemes", *Lecture Notes in Computer Science*, vol. 3200, pp. 67-81, 2004.

[26] Saeed R. Khosravirad, Taraneh Eghlidis, Shahrokh Ghaemmaghami, "Higher-order statistical steganalysis of random LSB steganography," *aiccsa*, pp.629-632, 2009 IEEE/ACS International Conference on Computer Systems and Applications, 2009