A Novel Method to Analyze Steganographic Content in Internet

V. Manikandan
Vellore Institute technology
Chennai

Abstract

Internet is a medium for data transformation used by all set of people. The dawn of data communication over Internet works, has become crucial to ensure security of information. The medium of internet, which is been used by terrorists who uses Steganography to communicate in secret with their fellow mates. *In this paper, we have proposed a novel method to analyze the Steganography content, which comes from unauthenticated sources. The proposed method uses a framework, which will store the Steganography content and analysis further to trace the hidden content. The result of analyze would be Steganography content and other will be the hidden content stored on Steganography. In order to ascertain the hidden content, we have involved dictionary attack rather than brute force attack since the hidden content will not been poorly organized. As a step forward towards the systematic application, this article proposes an extension to the JCA (Java Cryptography Architecture) framework to rivet steganalysis. We* believe that such a novel method would help to ensure the hidden content from Steganography medium to increase the level of security in Internet.

Keywords: steganography, steganalysis, stegframework, cryptography in steganography, Dictionary attack.

1. INTRODUCTION:

Steganography refers to the science of "invisible" communication. Unlike cryptography, where the goal is to secure communications from an eavesdropper, steganographic techniques strive to hide the very presence of the message itself from an observer. For example, with images, it would make sense that we are allowed to make changes as long as she does not alter significantly the subjective visual quality of a suspected stego-image.

It should be noted that the main goal of steganography is to communicate securely in a completely undetectable manner. That is, we should not be able to distinguish in any sense between cover-objects (objects not containing any secret message) and stego-objects (objects containing a secret message). In this context, "*steganalysis*" refers to the body of techniques that are designed to distinguish between cover-objects and stego-objects. It should be noted that nothing might be cleaned about the contents of the secret message. When the existence of hidden message is known, revealing its content is not always necessary. Just disabling and rendering it useless will defeat the very purpose of Steganography. In this paper, we present a steganalysis technique for detecting *stego-images*, i.e., still images containing hidden messages, using LSB technique.

Among all the image information hiding methods, LSB embedding is widely used for its high hiding capacity, and simplicity to realize. Many public steganographical software, such as S-Tools, EZStego and Steganos apply this technique. Therefore, it’s with great significance to detect the images with hidden messages produced by LSB embedding effectively, accurately and reliably. And many experts made efforts on the LSB Steganography and steganalysis research over the years.

2. STEGANOGRAPHY:

Steganography [1] works by replacing bits of useless or unused data in regular computer files (such as graphics, sound, text, HTML, or even floppy disks) with bits of different, invisible information. This hidden information can be plain text, cipher text, or even images and sound wave. In the field of Steganography, some terminology has developed. The adjectives cover, embedded and stego were defined at the Information Hiding Workshop held in Cambridge, England. The term “cover” is used to describe the original, innocent message, data, audio, still, video and so on.

When referring to audio signal Steganography, the cover signal, which is sometimes, called the “host” signal. The information to be hidden in the cover data is known as the “embedded” data. The “stego” data is the data containing both the cover signal and the “embedded” information. Logically, the processing of putting the hidden or embedded data, into the cover data, is sometimes known as embedding. Occasionally, especially when referring to image Steganography [2], the cover image is known as the container.

2.1 Methods in Steganography:

- Least Significant Bit (LSB) Encoding,
- Low Frequency Encoding,
- Mid Frequency Encoding,
- High Frequency Domain Encoding,
- Watermarking

There are many methods available in Steganography. However, each method has its own drawbacks. In Least Significant Bit, encoding message is hard to recover if image is subject to attack such as translation and rotation. In Low Frequency Encoding significant damage to picture appearance, message difficult to recover. In mid Frequency Encoding, it is easy to detect the information. In High Frequency Domain Encoding Image is distorted. Message easily lost if picture subject to compression such as JPEG.

A disadvantage of digital watermarking is that a subscriber cannot significantly alter some files without sacrificing the quality or utility of the data. This can be true of various files including image data, audio data, and computer code.

3. STEGANALYSIS

Steganalysis [3] is the discovery of the existence of hidden information; therefore, like cryptography and cryptanalysis, the goal of steganalysis is to discover hidden information and to break the security of its carriers.

The goal of steganalysis[4] is to identify suspected packages, determine whether or not they have a payload encoded into them, and, if possible, recover that payload.Unlike cryptanalysis, where it is obvious that intercepted data contains a message (though that message is encrypted), steganalysis generally starts with a pile of suspect data files, but little information about which of the files, if any, contain a payload. The steganalyst is usually something of a forensic statistician, and must start by reducing this set of data files (which is often quite large; in many cases, it may be the entire set of files on a computer) to the subset most likely to have been altered.

3.1 Methods In Steganalysis :

There are several type of steganalysis method available in trend some of these

- **Stego-only attack:** Only the stego-object is available for analysis. For example, only the stego-carrier and hidden information are available.

- **Known cover attack:** The original cover-object is compared with the stego object and pattern differences are detected. For example, the original image and the image containing the hidden information are available and can be compared.

- **Known message attack:** A known message attack is the analysis of known patterns that correspond to hidden information, which may help against attacks in the future. Even with the message, this may be very difficult and may be considered the same as a stego-only attack.
- **Chosen stego attack:** The steganography tool (algorithm) and stego-object are known. For example, the software and the stego-carrier and hidden information are known.
- **Chosen message attack:** The steganalyst generates a stego-object from some steganography tool or algorithm of a chosen message. The goal in this attack is to determine corresponding patterns in the stego-object that may point to the use of specific steganography tools or algorithms.
- **Known stego attack:** The steganography tool (algorithm) is known and both the original and stego-object are available.

3.2 Steganography signatures

Unusual patterns in the stego-image are obvious and create suspicion. For example, unused areas on a disk can be used to hide information. A number of disk analysis utilities such as EnCase and ILook Investigator are available, which can report on and filter hidden information in unused clusters or partitions in storage devices. Filters can also be applied to capture TCP/IP packets that contain hidden or invalid information in the packet headers. TCP/IP packets have unused space in the packet headers. The TCP packet header has six reserved or unused bits, and the IP packet header has two reserved bits. Information can also be hidden in the unused bits found in the Type of Service (TOS) Field and Flags of IP headers. Other methods to hide information under TCP/IP are exploiting the optional fields in IP headers, Timestamp, and Time to Live (TTL). These techniques can also be applied to other protocols such as Novell NetWare .Thousands of packets are transmitted with each communication channel, which provide an excellent way to communicate secretly. This technique of hiding information is unsafe because TCP/IP headers might be overwritten in the routing process, and reserved bits could be overwritten, thus rendering the hidden information useless. The technology of firewalls is also greatly improving. For example, you can set filters to determine if packets are coming from within the firewall’s domain. In addition, with the validity of the SYN and ACK bits, the filters can be configured to catch packets that have information in presumed unused or reserved space, just as if you can set certain firewalls to exclude such packets with spoofed addresses.

3.3 Visual detection

By looking at repetitive patterns, you can detect hidden information in stego images. These repetitive patterns might reveal the identification or signature of a steganography tool or hidden information.

Even small distortions can reveal the existence of hidden information. You can analyze these patterns by comparing the original cover images with the stego images and try to see differences. This is called a known-cover attack. By comparing numerous images, patterns become possible signatures to a steganography tool. A few of these signatures might identify the existence of hidden information and the tools used to embed the messages. With this information, if the cover images are not available for comparison, the derived known signatures are enough to imply the existence of a message and identify the tool used to embed the message.

There are lots of methods to analyze the stegnography content from the stegoimage. In our proposed method detect the LSB Steganography content from the stegoimage in the internet environment. For that we have proposed two method.

4. Proposed Method:

The proposed technique is to detect LSB Steganography [7] in continuous-tone natural images and JPEG images. It is based on an important statistical identity related to some sets of pixels. But this identity is very sensitive to LSB embedding, and the change in the identity can quantify the length of the embedded message. Consider the partition of an image into pairs of horizontally adjacent pixels. Let *P* be the set of all these pixel pairs.

4.1. Set Analysis (Method1):

Define the subsets *X* and *Y* of *P* as follows:

- *X* is the set of pairs $(u,v) \in p$ such that *v* is even and *u* < *v*, or *v* is odd and *u* > *v*.
- *Y* is the set of pairs $(u,v) \in p$ such that *v* is even and *u* > *v*, or *v* is odd and *u* < *v*.

The significance of *X* and *Y* in steganalysis of LSB embedding comes from the following assumption.
Statistically we have

$$|x| = |y| \qquad (1)$$

The assumption is true for natural images. This is because natural images are isotropic in terms of the gradient of intensity function. In other words, the gradient in any direction has equal probability to be positive and negative.
Now let *Z* be the subset of pairs $(u,v) \in p$ such that *u* = *v*. Furthermore, partition set *Y* into two subsets *W* and *V*, with *W* being the set of pairs in *P* of the
form (2*k*, 2*k*+1) or (2*k*+1, 2*k*), and *V* = *Y* −*W*. Then $p = x \cup w \cup v \cup z$
. In the sequel we call the sets *X*, *V* , *W* and *Z* primary sets.

Let us analyze now what happens when a message is embedded in the last bit plane of pixel values. The embedding modifies the values of some pixels by switching the LSB. Hence the relative ranking of some affected pixel pairs in *P* will be changed. Given a pixel pair (*u*, *v*), there are four situations:

- both values *u* and *v* remain unmodified;
- only *u* is modified;
- only *v* is modified;
- both *u* and *v* are modified.

If we are in situation ”a” (b, c, d) we say that the modification pattern due to LSB embedding is 00 (10,01, 11, respectively). The embedding process leads to change of membership of some pixel pairs between the primary sets, as illustrated by Figure 1. In Figure 1, each arrow drawn from a set *A* to a set *B*, labelled by a modification pattern, means that any pixel pair of *A* becomes a pixel pair of the set *B*, if modified by the specified pattern.

For each modification pattern $\pi \in \{00,10,01,11\}$ and any subset $A \subseteq p$, denote by $\rho(\pi,A)$ the probability that the pixel pairs of *A* are modified with pattern π. The following assumption about a statistical property of LSB steganography is crucial for the success of the steganalytic method to be proposed.

4.2. Noise Analysis(Method2):

In some cases only a single image is available that time more complicated analysis techniques may be required. In general, steganography attempts to make distortion to the carrier indistinguishable from the carrier's noise floor. In practice this is often improperly simplified to deciding to make the modifications to the carrier resemble white noise as closely as possible, rather than analyzing, modeling, and then consistently emulating the actual noise characteristics of the carrier. In particular, many simple steganographic systems simply modify the least-significant bit (LSB) of a sample; this causes the modified samples to have not only different noise profiles than unmodified samples, but also for their LSBs to have different noise profiles than could be expected from analysis of their higher-order bits, which will still show some amount of noise. Such LSB-only modification can be detected with appropriate algorithms, in some cases detecting encoding densities as low as 1% with reasonable reliability.

4.3. JPEG Detection(Method3):

In the next section, we present "Stegdetect," an automated utility to analyze JPEG images for steganographic content.

4.3.1 Stegdetect:

Stegdetect detects images that have content hidden with JSteg, JPHide and OutGuess 0.13b. For each system that we want to detect, we select the DCT coefficients in the order that they are modified and apply x^2 test.

The output from Stegdetect lists the steganographic systems found in each image or "negative" if no steganographic content could be detected. Stegdetect expresses the level of confidence of the detection with one to three stars. Figure 1 shows some sample

output.

```
misc/0003-wonder-2.jpg : jphide(*)

misc/dscf0001.jpg : outguess(old)(***)

misc/dscf0002.jpg : negative

misc/dscf0003.jpg : jsteg(***)
```

Figure1: The output from Stegdetect contains an estimate of the detection confidence.

4.3.2 JSteg Detection:

Detection of content hidden with JSteg is similar to the approach outlined by Westfeld and Pfitzmann.

JSteg does not modify the DCT coefficients zero and one. For that reason, they are ignored in the x^2-test. We sample the DCT coefficients starting from the beginning of the image and compute the probability of embedding. This process is repeated with increasing sample size until all DCT coefficients are contained in the sample. As a performance optimization, we stop computing the probability of embedding once it falls below a certain threshold.

4.3.3 JPHide Detection

Because JPHide modifies the DCT, coefficients in a fixed order determined by a table, we rearrange the coefficients in that order before computing the probability of embedding. However, two exceptions influence the detection.

JPHide modifies the DCT coefficients −1, 0 and 1 in a special way. As a result, the modifications to these coefficients can not be detected by the x^2-test. However,simply ignoring these coefficients still allows us to detect content embedded with JPHide. We also ignore modifications to the second-least-significant bits, which are not as frequent as modifications to the least-significant bits.

Similar to JSteg, we stop computing the probability of embedding once it falls below a certain threshold.

4.3.4 OutGuess Detection

Detecting content embedded with OutGuess 0.13b is complicated by the fact that the coefficients are selected pseudo-randomly, there is no fixed order in which to apply the x^2-test. However, Provos has shown that the x^2-test can be extended to detect content hidden with OutGuess 0.13b .

Instead of increasing the sample size and applying the test at a constant position, we use a constant sample size but slide the position where the samples are taken over the entire range of the image.

The test starts at the beginning of the image, and the position is incremented by one percent for every application of the x^2-test. The extended test does not react to an unmodified image, but detects the embedding in some areas of the stego image.

To find an appropriate sample size, we choose an expected distribution for the extended x^2-test that should cause a negative test result. Instead of calculating the arithmetic mean of coefficients and their adjacent ones, we take the arithmetic mean of two unrelated coefficients,

$$y_i = \frac{n_{2i-1} + n_{2i}}{2} \qquad (2)$$

A binary search on the sample size is used to find a value for which the extended x^2-test does not show a correlation to the expected distribution derived from unrelated coefficients.

5. Future Work:

In this paper we provide detection mechanism for LSB Steganography and JPEG images. If any picture comes with different format along with different Steganography method, our propsed method won’t give good solution. Now we are working to resolve this problems.

6. Conclusion:

Providing the security mechanism quite easy compare with analysis method like cryptanalysis,steganalysis. Eventhough we provided efficient mechanism to detect the Steganography content in internet. It will highly helpfull to prevent our system from the terrorist and also from malicious intruder attack in the internet environment.

sdaf

7. References:

[1] Information Hiding-A Survey-Fabien A.P.Petitcolas, Ross J.Anderson and Markus G.Kuhn.
[2] Steganography In Gray Images Using Wavelet -Bo Yang and Beixing Deng.
[3] Steganalysis Using Image Quality Metrics-Ismail Avcıbas¸, Nasir Memon and Bülent Sankur.
[4] New features for speci_c JPEG Steganalysis -Johann Barbier,Eric Filiol, and Kichenakoumar Mayoura.
[5] Benchmarking steganographic and steganalysis techniques-Mehdi Kharrazia, Husrev T, Sencarb Nasir Memonb.
[6] StegoWall: Blind statistical detection of hidden data-Sviatoslav Voloshynovskiy, Alexander Herrigel, Yuriy Rytsar and Thierry Puna.
[7] Analysis of LSB based Image Steganography Techniques-R.chandramouli,Nasir Memon.