Ashwini Palimkar, Dr. Suhas. H. Patil / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 2, March -April 2013, pp.379-384

Using Stegnography Technique for Data Leakage Problems Detect

Ashwini Palimkar (student), Prof. Dr. Suhas. H. Patil
Bharati Vidyapeeth Deemed University College of engineering, Pune, Maharashtra, India

Abstract:
Our goal is to detect when the distributor's sensitive data have been leaked by agents, and if possible to identify the agent that leaked the data. In this project, we are giving the methodology for adding fake object into data. Fake object will be added through stegnography concept.
Stegnnography is an ancient technique of data hiding or the stegnography is the art and science of hiding the existence of information, we contrast it with the related disciplines of cryptography
And traffic security.
Many techniques are used to hide data in various formats in steganography. The most widely used mechanism on account of its simplicity is the use of the Least Significant Bit. Least Significant Bit or its variants are normally used to hide data in a digital image. This paper discusses technique to hide data in a colorful image using least significant bit.

1. Introduction

Sometimes data is leaked and found in unauthorized Places Nowadays; more and more data are sold and transmitted on the internet. Databases are being used widely in many important fields, such as, banking and so on. With the fast growth of database business on the net, the data may be unsafe after passing through the unsecure network.
The recent surge in the growth of the Internet results in offering of a wide range of web-based services, such as database as a service, digital repositories and libraries, e-commerce, online decision support system etc. These applications make the digital assets, such as digital images, video, audio, database content etc, easily accessible by ordinary people around the world for sharing, purchasing, distributing, or many other purposes. In computer based stegnographhy, several forms of digital media may Be used as cover for hidden information, photos, documents, web pages and even MP3 music files may all serve as innocuous looking hosts for secret messages
The data leakage application will be shown through News channel.
In this project we are finding out the data is leaked or not. The agent will receive the information to broadcast the data on the news media from server. We will check whether the authorized user leaked the data to another news media channel.
Assumption1:
Each and every agent has its unique data.
Assumption2:
Another news channel will received the data only from agent.
Description of this paper in short:
The data leakage application will be shown through BNN News channel.
In this paper, we are finding out the data is leaked or not. The agent will receive the information to broadcast the data on the news media from server. We will check whether the authorized user leaked the data to another news media channel.
In this, for the agent has its unique username will be considered as the fake object. All over information of the agent is stored into the database (MySQL)

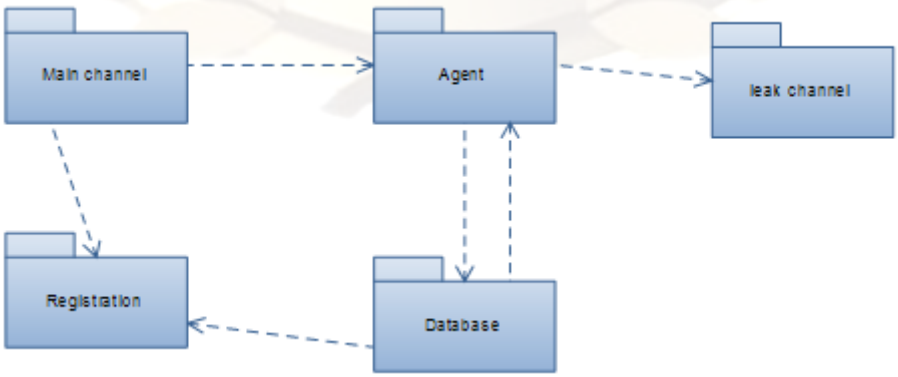

Figure 1: BNN News Channel

2. Modules Information

Module1:

In this module consist of the design for the website as well as database design for the application. The authorized agent will log in. for the new agent has to fill the register form, after successful submission of data he can logged into our system. The agent will request for the data (jpeg image), he will receive the data by adding the fake object into data.

Module2:

If the agent leaks that particular data then other news channel will be updated by those images. At server side the Admin is user for data security purpose. It will check whether the leaked data is same or not.

Module3:

The agent request for the data by explicit request and the data is mp3 file. The user will receive the data by adding the fake object which is unique.

Module4:

This module checks whether our system data as well as another system data is same or not. If he found the same data then the agent will be called guilty agent.

3. ADDING OF FAKE OBJECT

The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. However, fake objects may impact the correctness of what agents do, so they may not always be allowable. The idea of perturbing data to detect leakage is not new. However, in most cases, individual objects are perturbed, e.g., by adding random noise to sensitive salaries, or adding a watermark to an image. In this case, perturbing the set of distributor objects by adding fake elements is done. In some applications, fake objects may cause fewer problems that perturbing real objects. For example, say the distributed data objects are medical records and the agents are hospitals. In this case, even small modifications to the records of actual patients may be undesirable. However, the addition of some fake medical records may be acceptable, since no patient matches these records, and hence no one will ever be treated based on fake records. A trace file is maintained to identify the guilty agent. Trace file are a type of fake objects that help to identify improper use of data.

4. STEGANOGRAPHY

Steganography, as defined above is a technique to hide a data in an image in such a manner that it is unperceivable.

Many a times Steganography is related to Cryptography. It may be a misleading statement with respect to a Steganography approach. Steganography is related to Cryptography by meaning that both are used for security purposes but with different approach or implementation. Steganography is, along with Cryptography, a very ancient concept but its application varies according to emerging technologies.

The steps in steganography,

1. Include the writing the text messages, encryption of the text message is one of the options available.
2. Later, text is hidden in the selected media and transmitted to recipient.
3. At receiver end, reverse process is implemented to recover the original text message. Various techniques used in the art of steganography is the arrangements of various bits of the characters of the text in an image or other media. Keeping in mind the above, two files are needed; the image file and the text file that contains the data.

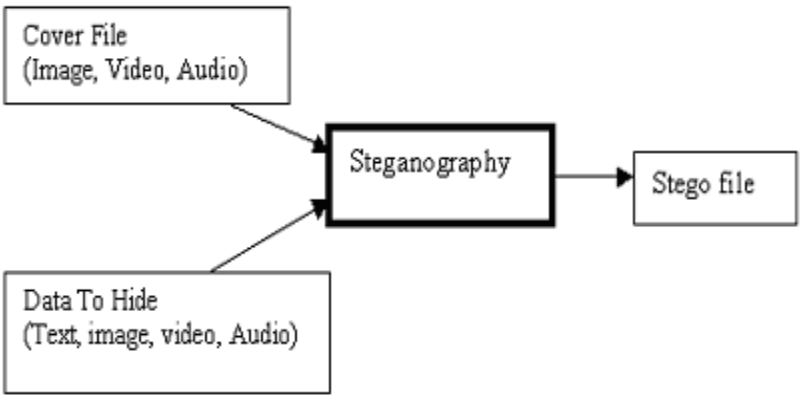

Figure 2: The Process of hiding data

LSB affects the smallest changes of the 8 bits therefore it alters the image to minimum. The most common method used is called LSB (Least Significant Bit) Mechanism that is hiding if the data in the least significant Bit (LSB) of the message.

However, one of its major limitations is small size of data which can be embedded in such type of images using only LSB. LSB is extremely vulnerable to attacks. LSB techniques implemented to 24 bit formats are difficult to detect contrary to 8 bit format. The other techniques include Masking and Filtering. It is normally associated with JPEG. In this technique image data is extended by masking secret data over it. Therefore, experts do not include this as a form of Steganography.

All algorithms employed for any type of format have pros and cons and depend upon the environments used. It also depends upon the information to be embedded. Various techniques developed were compared.

5. IMPLEMENTATION

Steganography, as defined above is a technique to hide a data in an image in such a manner that it is unperceivable. To achieve such outcome one may think of chopping the raw data that is the data to be hidden in equal number of block and hide it in specific areas with in an image. Such a thought interprets that the concept is not vivid. That is, one is not able to extract the true beauty of Steganography.

Presently the technology being mostly used is Digital Images [6]. By digital Images we presume to deal with bits that is 0's and 1's.

Digital Images we have selected are 24-bit depth color images using RGB color model. 24-bit refers to 8-bit for each RGB color channel, i.e. 8-bits for red.8-bits for green and 8-bits for blue and 24-bit depth with width and height of 800 x 600 pixels. It must be remembered that image resolution is highly dependent on the monitor screen resolution.

The idea is to hide text in image with the conditions that the image quality is retained along with the size of the image. Here, a thought may arise that why we need to hide text in an image if we can easily encrypt is using several ways. This is the point where Cryptography and steganography differs. Applying, cryptography result in an output of an unreadable text which when send over an internet is easily detectable that some important information is being conveyed. Hiding message in an image, along with the conditions, may seem just an exchange of picture between two ends.

At this point mind wanders that how an image can contain a message with no change in its quality and size. From this end onwards we will give a brief vision on the image design. To begin with, one must have a clear concept of an Image layout which includes the Image Header information, Image block information, Image extensible information (if any).Most of this information is clearly specified in standard technical documents.

We are using 24-bit BMP over 24-bit JPEG because it is lossless compression. JPEG is lossy i.e. to save space on disk it just eliminates parts of any image. If you compare the bit depth and pixel value of a BMP and JPEG both have the same except the file size on disk. The reason lies in the compression scheme used by JPEG and BMP. To make this compression efficient the color model much be changed from RGB to YUV color model, why YUV model, reason lies in the way this color model represents color. Simply, RGB stores image color in combination but YUV separates these color in Y for Luminance (color Brightness), U for chrominance(color difference) and V for color information. This approach is used because human eye is more sensitive to Luminance then chrominance. Fewer bits can be used for representing chrominance information. Reducing bits in chrominance is achieved by applying chrominance down sampling. Thus, reducing the overall size of an image.

Our algorithm is simple and flexible using LSB technique. We have selected the formats that commonly use lossless compression that is BMP, PNG, TIFF and GIF. We can make use of any of these formats or convert BMP into any of the above said formats. When data is streamed, it is captured after the header and chopped into 8 bits.

Figure 3: Simple conversion of a BMP to GIF

It has been analyzed that the conversions do not distort the images to a level where the degradation can be felt with the naked eye.

We recommend this algorithm to be used to hide small amount of data although large amount of data may also be incorporated in an image. The concept is similar to somewhat one time pad. A large number of images from the said formats were chosen to hide the data. Results confirm the use of these formats for successful Steganography techniques.

In 24-bit BMP, using RGB color model, with header size of 54 bytes.54 bytes onwards starts the pixel value of an image. Each pixel value contains the value of the color and is represented in bits (0 & 1).Similarly, text to hide is also represented in bits (0&1).Therefore comparing bit values byte by byte result in hiding the bit values of Text in bit value of an Image. The technique we are using is Least Significant Bit (LSB) i.e. storing in LSB of a byte (pixel). As mentioned above, the RGB model is used, we first stream an Image file and read the file in bits and then seek the position ahead the header bits.

Figure 5 Resultant image after hiding data in GIF format

6. IMAGE ANALYSIS

A. LSB in BMP

The BMP file format also called bitmap or DIB file format (for *device-independent bitmap*), is an image file format used to store bitmap digital images , if it is transmitted with an LSB stego. When image are used as the carrier in Steganography they are generally manipulated by changing one or more of the bits of the byte or bytes that make up the pixels of an image. The message can be stored in the LSB of one colour of the RGB value or in the parity bit of the entire RGB value.A BMP is capable of hiding big message. LSB in BMP is most suitable for applications, where the focus is on the amount of information to be transmitted and not on the secrecy of that information. If more number of bits is altered, it may result in a larger possibility that the altered bits can be seen with the human eye. But with the LSB the main objective of Steganography is to pass a message to a receiver without an intruder even knowing that a message is being passed is being achieved.

B. LSB in PNG

Portable Network Graphics (PNG) is a bitmapped image format that employs lossless data compression. PNG was created to improve upon and replace GIF. When images are used as the carrier in Steganography they are generally manipulated by changing one or more of the bits of the byte or bytes that make up the pixels of an image. The message can be stored in the LSB of one colour of the RGB value or in the parity bit of the entire RGB value .A PNG is capable of hiding quite a large message. LSB in PNG is most suitable for applications where the focus is on the amount of information to be transmitted and not on the secrecy of that information. If more number of bits is altered it may result in a larger possibility that the altered bits can be seen with the human eye. But with the LSB the main objective of steganography is s to pass a message to a receiver without an intruder even knowing that a message is being passed is being achieved.

C. LSB in GIF

Graphics interchange format also known as GIF is one of the machine independent compressed formats for storing images. Since GIF images only have a bit depth of 8, amount of information that can be hidden is less than with BMP. Embedding information in GIF images using LSB results in almost the same results as those of using LSB with BMP. LSB in GIF is a very efficient algorithm to use when embedding a reasonable amount of data in a grayscale image. GIF images are indexed images where the colours used in the image are stored in a palette. The colours of the palette are typically ordered from the most used colour to the least used colours to reduce lookup time. Some extra care is to be taken if the GIF images are to be used for Steganography. This is because of the problem with the palette approach. If the LSB of a GIF image is changed using the palette approach, it may result in a completely different colour. This is because the index to the colour palette is changed. The change in the resulting image is noticeable if the adjacent palette entries are not similar. But the change is not noticeable if the adjacent palette entries are similar. Most applications that use LSB methods on GIF images have low security because it is possible to detect even moderate change in the image.

Solutions to these problems could be ,

1. Sort the palette so that the colour difference between consecutive colours is minimized
2. Add new colours, which are visually similar to the existing colours in the palette.
3. Use Gray scale images. In a 8 bit Gray scale GIF image, there are 256 shades of gray. This results in gradual changes in the colours and it is hard to detect.

CONCLUSION

In some cases “realistic but fake” data records are injected to improve the chances of detecting leakage and identifying the guilty party.

Using containers (cover messages) to embed secret messages into is by far the most popular use of Steganography today. This method of Steganography is very useful when a party must send a top secret, private or highly sensitive document over an open systems environment such as the Internet. By embedding the hidden data into the cover message and sending it, you can gain a sense of security by the fact that no one knows you have sent more than a harmless message other than the intended recipients.

Steganography is in the nascent stage of development. This allows users to hide files of larger sizes while at the same time preserve the general appearance of any cover image used.

Furthermore, the implementation of various security measures provides a high level of protection for the hidden data. It is concluded that the, it is seen that the original image and the final embedded image appear to be identical to the human eye.

LSB makes use of BMP image, to be able to hide a secret message inside a BMP file, one would require a very large cover image. BMP images of 800×600 pixels found to have less web applications. Moreover such uses are not accepted as valid. For this reason, LSB Steganography has also been developed for use with other image file formats. All the major image file formats have different methods of hiding messages, with different strong and weak points respectively. LSB in GIF images has the potential of hiding a large

message, but only when the most suitable cover image has been chosen.

REFERENCES:

[1] Panagiotis Papadimitriou, Hector Garcia-Molina (2010)'Data Leakage Detection', IEEE Transactions on knowledge and data engineering, Vol.22, No.3.

[2] P. Papadimitriou and H. Garcia-Molina, "Data Leakage Detec- tion," technical report, Stanford Univ., 2008.

[3] "A Stegnography Implementation", Beenish Mehboob and Rashid Aziz Faruqui.

[4] Moerland, T., "Steganography and Steganalysis", *Leiden Institute of Advanced Computing Science.*

[5] Silman, J., "Steganography and Steganalysis: An Overview", *SANS Institute*, 2001 Jamil, T.,

[6] Bryan. "Steganography: How to Send a Secret Message." 8 Nov. 2001 www.strangehorizons.com/2001/20011008/steganography.shtml.

[7] Johnson, Neil F., and Sushil Jajodia, "Steganalysis of Images created using Current Steganographic Software", Proceedings of the Second Information Hiding Workshop, April 1998.

[8] Krenn, R., "Steganography and Steganalysis", http://www.Krenn.nl/univ/cry/steg/article.pdf Kafa Rabah. Steganography - The Art of Hiding Data. Information technology Journal 3 (3) - 2004

[9] Sellars, D., "An Introduction to Steganography",