Chapter 1

Introduction

One of the reasons why data breaches are successful is that most of the information acquired from a computer system are in forms that are comprehendible. Intruders may reveal the information to others, modify it to misrepresent an individual or organization, or use it to launch an attack (Identity Theft). One solution to this problem is, through the use of steganography. Steganography is a technique of hiding information in digital media. In contrast to cryptography, it is not to keep others from knowing the hidden information but it is to keep others from thinking that the information even exists.

Steganography become more important as millions of people join the internet revolution. Steganography is the art of concealing information in ways that prevents the detection of hidden messages. Steganography include an array of secret communication methods that hide the message from being seen or discovered.

* 1. Background of Study
  2. Theorical Background

In order to understand how steganography is achieved in computer science, a little knowledge about computer system is necessary. All information in a computer system is represented as a bunch of bits, each with a value of 0 or 1. The only thing that distinguishes different data objects is the context in which we view them. The same sequence of bits might represent a number, a sequence of text, or a machine instruction in different contexts. For example, the bit sequence “0110 1000 0110 0101 0110 1100 0110 1100 0110 1111” means “hello” if translated using the ASCII standard. It will be most likely meaningless if translated using some other standards. Image, audio, and video files are no exception. They are all sequences of bits stored in the hard disk.

During the 15th and 16th centuries, many writers including Johannes Trithemius (author of Steganographia) and Gaspari Schotti (author or Steganographica) wrote on Steganographic techniques such as coding techniques for text, invisible inks, and incorporating hidden messages in music.

Between 1883 and 1907, further development can be attributed to the publications of Auguste Kerckhoff (author of Cryptographic Militaire) and Charles Briquet (author of Les Filigranes). These books were mostly about Cryptography, but both can be attributed to the foundation of some steganographic systems and more significantly to watermarking techniques.

Concepts such as null ciphers (taking the 3rd letter from each word in a harmless message to create a hidden message, etc.), image substitution and microdot (taking data such as pictures and reducing it to the size of a large period on a piece of paper) were introduced and embraced as great steganographic techniques.

In the recent digital world of today, namely 1992 to present, Steganography is being used all over the world on computer systems. Many tools and technologies have been created that take advantage of old steganographic techniques such as null ciphers, coding in images, audio, video and microdot. With the research this topic is now getting we will see a lot of great applications for Steganography in the near future.

Steganography replaces unneeded or unused bits in regular computer files (Graphics, sound, text) with bits of different and invisible information. Hidden information can be any other regular computer file or encrypted data.

Steganography differs from cryptography in a way that it masks the existence of the message where cryptography works to mask the content of the message.

Steganography sometimes used in conjunction with encryption. An encrypted file may still hide information using steganography, so even if the encrypted file is deciphered, the hidden information is not seen.

1.2.1 TYPES OF STEGANOGRAPHY

There are different ways to hide the message in another, well known are Least Significant bytes and Injection.

When a file or an image is created there are few bytes in the file or image which are not necessary or least important. These types of bytes can be replaced with a message without damaging or replacing the original message, by which the secrete message is hidden in the file or image.

Another way is a message can be directly injected into a file or image. But in this way the size of the file would be increasing accordingly depending on the secrete message

**STEGANOGRAPHY IN IMAGE**

Digital images are the most widely used cover objects for steganography. Due to the availability of various file formats for various applications the algorithm used for these formats differs accordingly.

An image is collection of bytes (known as pixels for images) containing different light intensities in different areas of the image. When dealing with digital images for use with Steganography, 8-bit and 24-bit per pixel image files are typical. Both have advantages and disadvantages 8-bit images are a great format to use because of their relatively small size. The drawback is that only 256 possible colors can be used which can be a potential problem during encoding. Usually a gray scale color palette is used when dealing with 8-bit images such as (.GIF) because its gradual change in color would be harder to detect after the image has been encoded with the secret message. 24-bit images offer much more flexibility when used for Steganography. The large numbers of colors (over 16 million) that can be used go well beyond the human visual system (HVS), which makes it very hard to detect once a secret message, has been encoded.

Large amount of data can be encoded in to 24-bit images as it is compared to 8-bit images. The drawback of 24-bit digital images is their size which is very high and this makes them suspicious our internet due to their heavy size when compared to 8-bit images. Depending on the type of message and type of the image different algorithms are used.

Few types in Steganography in Images:

* Least significant bit insertion
* Masking and filtering
* Redundant Pattern Encoding
* Encrypt and Scatter
* Algorithms and transformations

**Least significant bit insertion**

Least Significant Bit (LSB) insertion is most widely known algorithm for image steganography, it involves the modification of LSB layer of image. In this technique, the message is stored in the LSB of the pixels which could be considered as random noise. Thus, altering them does not have any obvious effect to the image.

**Masking and filtering**

Masking and filtering techniques work better with 24 bit and grey scale images. They hide info in a way similar to watermarks on actual paper and are sometimes used as digital watermarks. Masking the images changes the images. To ensure that changes cannot be detected make the changes in multiple small proportions. Compared to LSB masking is more robust and masked images passes cropping, compression and some image processing. Masking techniques embed information in significant areas so that the hidden message is more integral to the cover image than just hiding it in the "noise" level. This makes it more suitable than LSB with, for instance, lossy JPEG images.

**Redundant Pattern Encoding**

Redundant pattern encoding is to some extent similar to spread spectrum technique. In this technique, the message is scattered throughout the image based on algorithm. This technique makes the image ineffective for cropping and rotation. Multiple smaller images with redundancy increase the chance of recovering even when the stegano-image is manipulated.

**Encrypt and Scatter**

Encrypt and Scatter techniques hides the message as white noise and White Noise Storm is an example which uses employs spread spectrum and frequency hopping. Previous window size and data channel are used to generate a random number. And within this random number, on all the eight channels message is scattered throughout the message. Each channel rotates, swaps and interlaces with every other channel. Single channel represents one bit and as a result there are many unaffected bits in each channel. In this technique it is very complex to draw out the actual message from stegano-image. This technique is more secure compared to LSB as it needs both algorithm and key to decode the bit message from stegano-image. Some users prefer these methods for its security as it needs both algorithm and key despite the stegano image. This method like LSB lets image degradation in terms of image processing, and compression.

**Algorithms and transformations**

LSB modification technique for images does hold good if any kind of compression is done on the resultant stego-image e.g. JPEG, GIF. JPEG images use the discrete cosine transform to achieve compression. DCT is a lossy compression transform because the cosine values cannot be calculated exactly, and repeated calculations using limited precision numbers introduce rounding errors into the final result. Variances between original data values and restored data values depend on the method used to calculate DCT

**STEGANOGRAPHY IN AUDIO**

Implanting secrete message into an audio is the most challenging technique in Steganography. This is because the human auditory system (HAS) has such a vibrant range that it can listen over. To put this in perspective, the (HAS) recognize over a range of power greater than one million to one and a range of frequencies greater than one thousand to one making it extremely hard to add or remove data from the original data structure. The only weakness in the (HAS) comes at trying to differentiate sounds (loud sounds drown out quiet sounds) and this is what must be exploited to encode secret messages in audio without being detected.

Below are the lists of methods which are commonly used for audio Steganography.

* LSB coding
* Parity coding
* Phase coding
* Spread spectrum
* Echo hiding

**LSB coding**

Using the least-significant bit is possible for audio, as modifications usually would not create recognizable changes to the sounds. Another method takes advantage of human limitations. It is possible to encode messages using frequencies that are indistinct to the human ear. Using frequencies above 20.000Hz, messages can be hidden inside sound files and can not be detected by human checks.

**Parity coding**

Instead of breaking a signal down into individual samples, the parity coding method breaks a signal down into separate regions of samples and encodes each bit from the secret message in a sample region's parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process flips the LSB of one of the samples in the region. Thus, the sender has more of a choice in encoding the secret bit, and the signal can be changed in a more unobtrusive fashion.

**Phase coding**

Phase coding attends to the disadvantages of the noise inducing methods of audio Steganography. Phase coding uses the fact that the phase components of sound are not as audible to the human ear as noise is. Rather than introducing perturbations, this technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, attaining an indistinct encoding in terms of signal-to-perceived noise ratio.

**Spread spectrum**

In the context of audio Steganography, the basic spread spectrum (SS) method attempts to spread secret information across the audio signal's frequency spectrum as much as possible. This is comparable to a system using an implementation of the LSB coding that randomly spreads the message bits over the entire audio file. However, unlike LSB coding, the SS method spreads the secret message over the sound file's frequency spectrum, using a code that is independent of the actual signal. As a result, the final signal occupies a bandwidth in excess of what is actually required for broadcast.

**Echo hiding**

In echo hiding, information is implanted in a sound file by introducing an echo into the separate signal. Like the spread spectrum method, it too provides advantages in that it allows for a high data transmission rate and provides superior strength when compared to the noise inducing methods. If only one echo was produced from the original signal, only one bit of information could be encoded. Therefore, the original signal is broken down into blocks before the encoding process begins. Once the encoding process is completed, the blocks are concatenated back together to create the final signal.

**STEGANOGRAPHY IN VIDEO**

In video steganography, a video file would be embedded with supplementary data to hide secret messages. In the process, an intermediate signal which is a function of hidden message data and data of content signal would be generated. Content data (video file) is then combined with this intermediate signal to result encoding. The supplementary data can include copy control data which can be brains by consumer electronic device and used to disable copying.

The intermediate signal may also contain a pseudo arbitrary key data so as to hide encoding and decode needs corresponding key to extract hidden information from encoded content. In some implementations regulation data is embedded in the content signal with auxiliary data. This regulation data consists of known properties enabling its identification in the embedded content signal. This encoding is robust against scaling, resampling and other forms of content degradation, so that the supplementary data can be detected from the content which might have been degraded.

There are different approaches for video steganography apart from the above mentioned. Most widely known are listed and discussed below.

**Least Significant Bit Insertion**

This is the most simple and popular approach for all types of steganography. In this method the digital video file is considered as separate frames and changes the displayed image of each video frame. LSB of 1 byte in the image is used to store the secret information. Effecting changes are too small to be recognized by human eye. This method enhances the capacity of the hidden message but compromises the security requirements such as data integrity.

**Real time video steganography**

This kind of steganography involves hiding information on the output image on the device. This method considers each frame shown at any moment irrespective of whether it is image; text. The image is then divided into blocks. If pixel colors of the blocks are similar then changes color characteristics of number of these pixels to some extent. By labeling each frame with a sequence number it would even be easy to identify missing parts of information. To extract the information, the displayed image should be recorded first and relevant program is used then.

**STEGANOGRAPHY IN DOCUMENT**

Steganography in documents just focuses on altering some of its characteristics. They can either be characteristics of text or even text formatting. Below are few ways listed and discussed to implement the same.

Since everyone can read, encoding text in neutral sentences is doubtfully effective. But taking the first letter of each word of the previous sentence, one can see that it is possible and not very difficult. Hiding information in plain text can be done in many different ways. One way is by simple adding white space and tabs to the ends of the lines of the document. The last technique was successfully used in practice and even after a text has been printed and copied on paper for ten times, the secret message could still be retrieved.

Another possible way of storing a secret inside a text is using a publicly available cover source, a book or a newspaper, and using a code which consists for example of a combination of a page number, a line number and a character number. This way, no information stored inside the cover source leads to the hidden message. Discovering it depends exclusively on gaining knowledge of the secret key.

Setting background color and font color is one of the mainly used steganographic approach. This method is focused for Microsoft word documents. Choose predefined colors and set font and background colors of invisible characters such as space, tab or the carriage return characters. R,G,B values are 8 bits means we have allowed range of 0 to 255.Most of the viewers would not feel interested about color values of these invisible characters hence 3 bytes of information is easily hidden in each occurrence of space, tab or carriage return. This approach needs no extra information to hide required bits.

* 1. Problem statement

There is a need to communicate sensitive information or personal messages secretly to a particular target destination without the fear of the message being breached or intercepted by a malicious third-party, considering the security challenges posed to users over the internet and possible identity thefts, we are forced to take matters into our own hands and protect ourselves from malicious people. These means of protection needs to be fast, very accessible and handy.

* 1. Aim and Objectives

The aim of this project is to develop an android based image steganography system as security tool based on steganography system.

The objectives are as follows:

* determine the best and most efficient image steganography algorithm.
* implement the chosen algorithm in the most suitable language.
* design the workflow and interfaces following best practices mobile design.
* build an android application around the algorithm and interface designs.
  1. Methodology
  2. Scope of Work

The scope of the project is implement steganography on an android device as a tool for hiding information in any image file, information to be hidden may include any form of files, i.e. text, media or data files, and also upon retrieval of the concealed file from the image, the user can determine where the retrieved file will be stored. It also tries to make information hiding more simple and user-friendly

* 1. Justification of Study

Steganography offers to help us hide information we hope to transfer over a medium, using various algorithm, depending on how robust our host machine or how we design our system. Users information can be well hidden and protected from malicious party.

* 1. Definition of terms

1. **Data:** The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media. Raw facts or unprocessed information.
2. **Security:** In information technology, security is the protection of information assets through the use of technology, processes, and training
3. **Cryptography:** Cryptography is a method of storing and transmitting data in a particular form so that only those for whom it is intended can read and process it
4. **Steganography:** Steganography is the practice of concealing a file, message, image, or video within another file, message, image, or video.
5. **Information:** Information (shortened as info) is that which informs. In other words, it is the answer to a question of some kind
6. **Techniques:** a way of carrying out a particular task, especially the execution or performance of an artistic work or a scientific procedure.