Cryptography and Steganography for Information Hiding

Mihir Wagle
B.E. Computer Engineering
V.E.S. Institute of Technology
Mumbai, India

Pavan Chhatpar
B.E. Computer Engineering
V.E.S. Institute of Technology
Mumbai, India

ABSTRACT
Due to rapid advances in consumer computing, most data is stored electronically, without any physical copy. Thus, securing this data has become an important issue to consider. Steganography allows us to hide information within other information or digital media to prevent snooping. Cryptography encrypts the information so that snoopers cannot decipher it with ease. If information hiding employs cryptography, followed by steganography, then even if a third party guesses you have used steganography, it is unable to gain access to the information with ease.

General Terms
steganography, cryptography, security, digital signal processing

Keywords
steganography, cryptography, security, digital signal processing

1. INTRODUCTION
Alongside a rise in storage of data electronically, we are also witnessing a lot of innovation in the mobile communications field. Transmission of data in the open, over largely insecure computer networks has led to security issues which need to be resolved. The field of information hiding is thus in vogue thanks to a need for data integrity and confidentiality.

With a rise in Internet speeds, it is possible to transmit large images or other such media over the internet without any losses. This may lead to unauthorised copying of the media by an unauthorised third party. Information hiding may be used to embed a copyright which can be extracted only in a specific manner, leading to better protection of owner rights.

Information hiding includes applications ranging from copyright conservation or watermarking for machine readable media to steganography. Out of all of these, this paper concentrates on Steganography, which hides a secret message/file within the host data set so that its presence is imperceptible to any snooper.

Further, to reduce the chances of compromising data confidentiality, the message to be hidden is encrypted with cryptographic techniques. In cases where a snooper may realize that steganography is used, it may yet be difficult to obtain the actual data until he manages to crack the cryptographic algorithm.

In Section 2, we talk about steganography in detail in order to avoid confusion between cryptography and steganography. Section 3 discusses several steganographic techniques. Section 4 describes an information hiding system that was developed by using one of the cryptographic and steganography techniques. Finally, section 5 provides a concluding summary of the system and suggests further work that can be carried out to improve on it.

2. STEGANOGRAPHY – OVERVIEW
The word steganography has its roots in the Greek word ‘steganos’ and the Latin word ‘graphia’. ‘Steganos’ means ‘reticent’ and ‘graphia’ means ‘a descriptive science’. It is therefore considered the art of concealing a message or a file within another message or file in such a way that the very existence of that message is concealed.

Steganography tries to pass information in a manner that is undetectable. It does not stop others from looking at the information while it is embedded within the carrier, but instead stops them from thinking that it may be there in the very first place. If a steganography technique allows for a third party guessing the presence of information being transported in a carrier medium, then the technique is largely ineffective.

So far cryptography has enjoyed much more applications when compared to steganography but the outlook is changing since:

- Copyrights need to be securely embedded in carrier media to stop any infringement.
- A rise in the efficacy of cryptanalysis is forcing people to look for alternate means to secure their data.

The basic model of steganography consists of a carrier (relatively large size to hide data), message (data to be hidden), a function (which hides the message in the carrier) and a stego-object (the output of the function). Thus, the carrier and the message are provided to the function which generates a stego-object from the same.

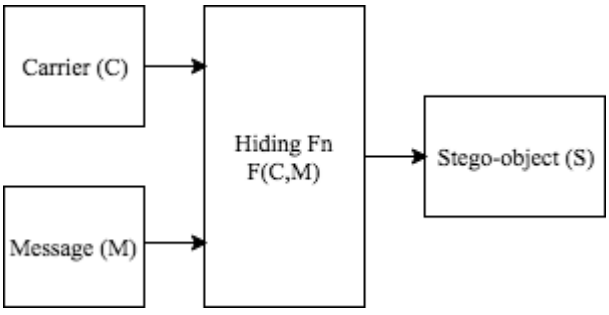

Fig 1: Basic Steganographic Hiding

In order to recover the message, we need only the stego-image, password and the inverse of the function used for hiding the data in most cases. For some techniques, the original carrier may be required for comparison. Such a technique is not preferred since having two similar images on a single system can raise suspicion.

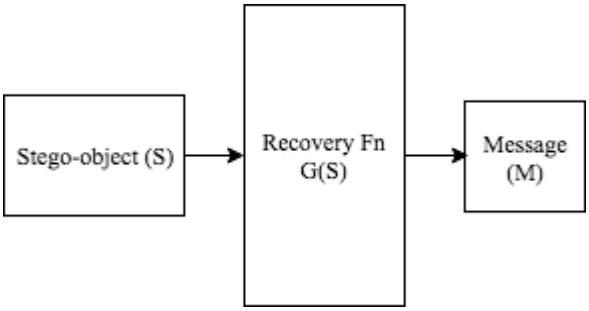

Fig 2: Basic Steganographic Recovery

The carrier mentioned above may be one of two things:

- Uncompressed image - bmp
- Uncompressed audio – wav

Uncompressed media is preferred due to its larger size, allowing us to hide more data.

Information hiding consists of two steps:

- Identification of redundant bits in the carrier.
- Replacement of the redundant bits in the carrier with the message.

Yet, there are limitations with the use of steganography[1]. An image can only hide a limited amount of data until the manipulation of bits create a significant change in the resultant stego-object which is perceptible by human observation. Taking a specific case of image steganography, we may choose only certain number of LSBs to hide information. This puts a restriction on the message size to be hidden which is determined by the image size being used. Another source of limitation arises from a necessity to choose appropriate embedding positions in the carrier image[2]. If the positions are chosen randomly, smooth regions of the carrier image may possibly get affected leading to less security and reduced visual quality. Any reduction in image quality alongside unexpected artifacts in the image are to be avoided in order to stop any third party viewing the image from suspecting that the image has been altered in any way and therefore invalidating the steganography.

2.1 Steganography versus Cryptography

Cryptography encrypts data to stop third parties from accessing it whereas Steganography [3] hides the existence of that data in the first place.

One way to determine the strength of a cryptographic algorithm is through its key domain. In cryptography, the algorithm is known to all but the key being used is unique. Whereas in steganography, there is no such concept of key. So, the strength is entirely dependent on how stealthily is data stored on the carrier and the techniques used for selection of embedding positions. Data can get compromised if someone gets doubtful about a carrier and decides to check it against the various steganography techniques.

The two techniques may be combined by firstly encrypting a message and then hiding it using steganography, thus combining their advantages and removing their disadvantages as shown in Table 1.

Table 1. Steganography versus Cryptography

STEGANOGRAPHY	CRYPTOGRAPHY
Message passing is hidden	Message passing is visible
Multiple carrier formats possible	Techniques used are strong but known to all.
If detected, message is visible to third party.	Resilient to third party attack

3. STEGANOGRAPHY TECHNIQUES

Some of the techniques used in Steganography [4] include:

- Least significant bit (LSB)
- Pixel value differencing
- Edges based data embedding method
- Random pixel embedding method
- Mapping pixel to hidden data method
- Labelling or connectivity method
- Pixel intensity based method
- Texture based method
- Histogram shifting methods

Out of these, the LSB method[1] is going to be primarily used in our system. The LSB method is one in which the bits of the message are embedded in a deterministic fashion, into the least significant bits of the carrier.

4. CRYPTO-STEGANOGRAPHIC SYSTEM (CSS)

In the LSB method, steganalysis can be a comparatively easy task as the embedding positions are deterministic. So steganography can be useful only till someone manages to realize that there might be some information hidden in the seemingly innocuous image. Cryptography can be used as the second line of defence and this gave rise to the system being discussed in this section.

An information hiding system has been developed and named 'Crypto-Steganographic System'. It serves to ensure both data integrity and confidentiality. CSS uses an uncompressed image format like bmp as the carrier and the LSB method mentioned in Section 3 as the function for hiding data.

The system takes the image(carrier), message, and key as input and checks whether the size of the image is large enough to successfully hide the data. For the message, there are two classifications, text file or not text file. In case of a text file, the system extracts the text from the file, encrypts it using a cipher (in this case, a Vigenere cipher) and then hides it in the carrier image file to produce a stego-image. In case of any other file, since extracting text may have unsupported values, the system creates a password protected zip file and then hides it in the carrier image file to produce a stego-image. The password protected zip file acts as the cryptographic layer; the security offered by various zip utilities have been improved through a series of critics it faced which exposed possible attacks on the encrypted information[5]. Today, many of the modern zip utilities offer an acceptable level of security by reducing the number of possible attacks that can be performed. It is therefore a viable choice to avoid reinventing the wheel or using established cryptography providers.

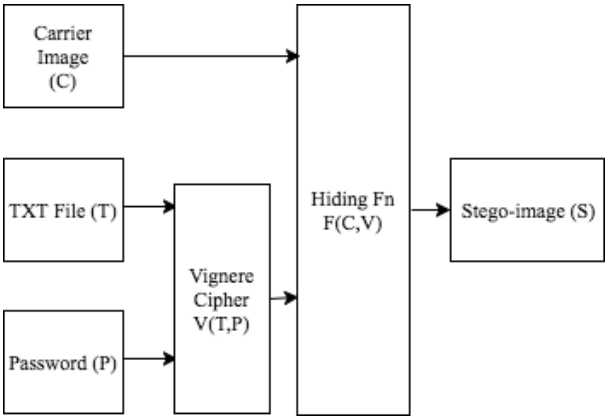

Fig 3: CSS – Text

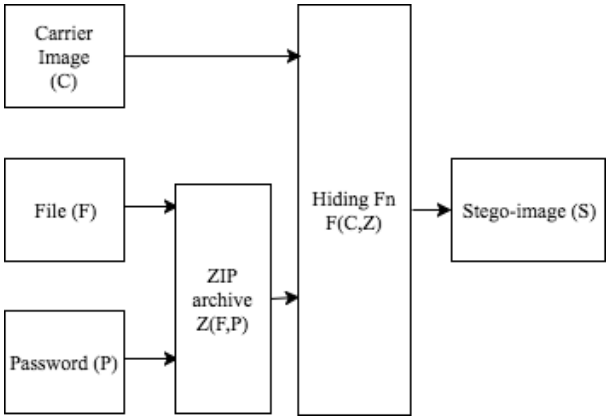

Fig 4: CSS – Files

Carrier Image

Stego-image

Fig 5: Carrier versus stego-image

We find that there is little perceptible difference between the stego-image and the carrier image since the LSB method provides for a change in the LSB which cannot modify an image much.

Recoverability is dependent on an assumption that the stego image remains pristine. We must note that there can be no modifications whatsoever to the stego-image since even an elementary transformation such as rotation can shift bits around which in turn will render the hidden data unrecoverable.

5. CONCLUSION

In this paper, we explain the process of steganography and how it differs from the ubiquitous cryptography. We also designed a system combining the two techniques to ensure data integrity and confidentiality. The system uses a password which can be used with a Vigenere cipher in case of text and as a password for the zip archive in case of any other files. Then LSB method is used to hide the data. While LSB changes are not perceptual, it suffers from the problem that even an elementary transformation can change the overall message and can lose the data being hidden. Thus the further scope would be to design a system that overcomes this limitation by making systems that can endure some amount of contamination. Also, embedding positions for the information are deterministic and cryptography is intended to overcome this loophole. Alternative methods for steganography which will make steganalysis a tough task can be a future scope which will improve security of data from the perspective of steganography.

6. REFERENCES

[1] Chandramouli, Rajarathnam, and Nasir Memon. ”Analysis of LSB based image steganography techniques.” Image Processing, 2001. Proceedings. 2001 International Conference on. Vol. 3. IEEE, 2001.

[2] Weiqi Luo, Member, IEEE, Fangjun Huang, Member, IEEE, and Jiwu Huang, Senior Member, IEEE “Edge Adaptive Image Steganography Based on LSB Matching Revisited” IEEE Transactions on Information Forensics and Security, Vol. 5, No. 2, June 2010.

[3] Desai, Hardikkumar V. ”Steganography, cryptography, watermarking: a comparative study.” Journal of Global Research in Computer Science 3.12 (2012): 33-35.

[4] Hussain, Mehdi, and Mureed Hussain. ”A survey of image steganography techniques.” (2013).

[5] Tadayoshi Kohno∗ “Analysis of the WinZip encryption method” IACR ePrint Archive 2004/07.