STEGANOGRAPHY:

TEXT FILE HIDING IN IMAGE

YAW CHOON KIT

CA10022

FACULTY OF COMPUTER SYSTEM AND SOFTWARE

ENGINEERING

2012/2013

ABSTRACT

Steganography is the art or science in hiding. It is origin from the Greek work where stegano(hiding) + graphy(writing). It is hiding a message rather than encoding it. Basically it can be understand as hiding something into another something else. The something else which embedded the hidden object look no differ from original, thus it is hard to arise suspicious from others.

Through the concept of steganography, this research wishing to hide the text file in the image. The purpose to doing so is to create a steganographic message when only sender and receiver know the trick behind the steganographic message.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity. The word steganography is of Greek origin and means "concealed writing" from the Greek words steganos (στεγανός) meaning "covered or protected", and graphei (γραφή) meaning "writing". The first recorded use of the term was in 1499 by Johannes Trithemius in his Steganographia, a treatise on cryptography and steganography disguised as a book on magic. Generally, messages will appear to be something else: images, articles, shopping lists, or some other converted text and, classically, the hidden message may be in invisible ink between the visible lines of a private letter.

The advantage of steganography over cryptography alone is that messages do not attract attention to themselves. Plainly visible encrypted messages, no matter how unbreakable will arouse suspicion, and may in themselves be incriminating in countries where encryption is illegal. Therefore, whereas cryptography protects the contents of a message, steganography can be said to protect both messages and communicating parties.

Using the theory of steganography, this research is wishing to hide a text file in the image. The encrypted image will looked like not differ to the original image. This can eliminate the suspicion of third party toward the encrypted image. The text file embedded in the image should be able to retrieve it back to the original state.

1.2 Problem Statement

As stated earlier in the introduction, visible encrypted messages will draw the suspicion from others. This will make them attempt to decrypt the cipher text. Indirectly it arise the chance of message being viewed by third party in the exception of sender and viewer.

However by hiding or embedding the text file into image, it is hard to arise the suspicion of other whether it is contain any secret message or not as the processed image had not look differ from the original image in naked eye. This will make the message more secure. Indirectly, only the sender and receiver know the treat behind the image and hence ensure that the message can only view by sender and receiver.

1.3 Objective

The objectives of the research are to:

1. To store a text file (.txt) in image.
2. To protect the text file (.txt) from being viewed by third party using encryption.
3. To restore the stored text file in image.

1.4 Scope

The scopes involved in the project are steganography, cryptography, security issues and a slightly watermarking technique.

Steganography, as the introduction state, basically steganography was hiding something inside into other medium. This was what exactly will be done by me in this project as I will ongoing to attempt to hide the text file in image.

Cryptography involved the encryption and decryption process. I will encrypt by hiding the text file in image. The receiver shall able to decrypt the steganographic message in order to read the original message which wish to be sent by the sender.

The security issues involved here is what secures will the steganographic message be. It is really able to protect it from being view by third party?

Watermarking technique shared some similar concept with steganography as both are wish to embedded one thing going into other medium. Hence, I should use the help of watermarking technique to embedded my text file(.txt) into the image.

Basically the overall process of this project can be viewed as below

1. Choose the ideal image and the text file that want to send to others.
2. Embedded the text file into the image
3. Send the steganographic message to the receiver.
4. Receiver decrypt the steganographic message and restore back the text file hid in image.

1.5 Thesis Organization

This thesis consists of five (5) chapters. Chapter 1 will discuss on introduction to the system which will explain the introduction, problem statement, objectives, and scope. For chapter 2, it will discuss on the literature review on the existing research and system to figure out the existing problem or solution done by other parties. For chapter 3, it will discuss on the method during design and implementation phase, hardware and software used, and Gantt chart planned for the whole PSM period. For chapter 4, it will discuss on the process and data gathering and the work flow will also be shown. In the same chapter, the algorithm on how to implement the model will also be explained. Lastly is chapter 5, which will be the result and discussion. In this chapter, result analysis and research constraints will be discussed in this chapter.

CHAPTER 2
LITERARURE REVIEW

2.1 Introduction

In this chapter, it is will going to focus on discussing the result or finding based on the article, journals or any other related reference material. Some original word from the reference materials may be citied in order to enhance the review. Purpose of this chapter is to explain about the selected project.

Basically, it is dividing into few sub-sections as well. Those sub-section including some little explanation of basic concept of selected project, research of some already existed similar problem or solution done by others and the hardware, technique or method which will going to apply or used in the selected project.

This review will to do research and describe about the existing problem or solution done by other parties. It is will also study about other systems which related to selected project.

This chapter explain in detail regarding techniques/method/hardware or technologies which are suitable to be adapt into the project. This chapter contains information about the study of the project in general.

2.2 What is Steganography?

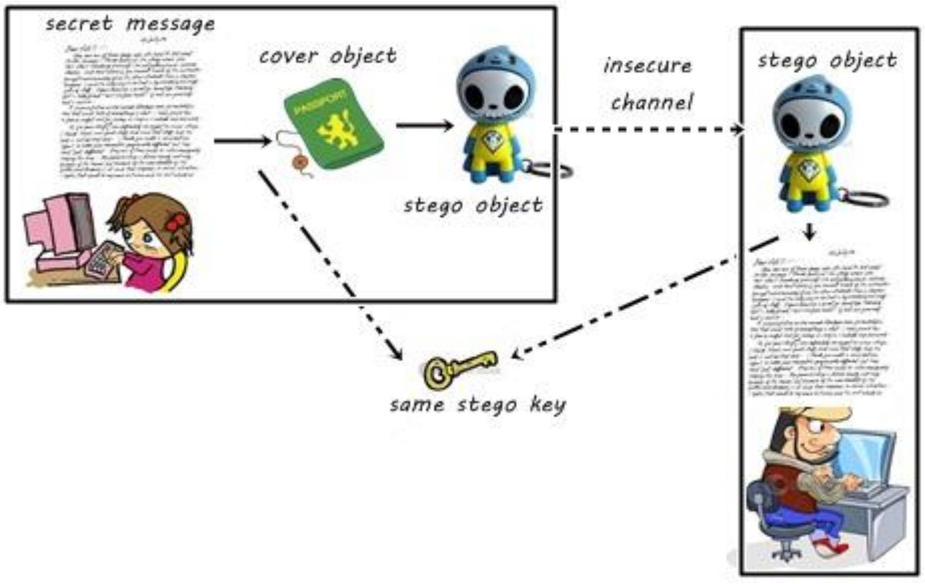
Steganography is the art or science of writing hidden messages in such a way that only both sender and intended recipient, knowing the presence of the message. It is a kind of security via obscuration. The word steganography is come from the Greek origin and means "concealed writing". Steganos (στεγανός) means "covered or protected", and graphei (γραφή) means "writing" in the Greek. Combination of these two Greek words made the existence of the word of steganography (stego [covered] + graphy [writing]).

For the first time, this term was used in years 1499 when Johannes Trithemius is using it in his “Steganographia”, a dissertation on cryptography and steganography. In general, messages will be shown to be other thing else such as articles, images or some other covert text and so on. Classically, this hidden message can be even in invisible ink between the visible lines of a personal letter.

Steganography is hiding a message, rather than encoding it. If a message is not being suspected then it is quite hard to start to decode or decrypt it. It includes a vast array of skill and techniques for hiding messages in a variety of media.

The benefit of steganography is that it do not arise the attention to itself. A plainly visible encrypted message will raised doubt no matter how unbreakable it was. Sometimes, the encrypted message using cryptography may be incriminating in some countries where encryption is considered illegal. Hence, whereas cryptography protects the contents of a message, steganography can be said to protect both communicating parties and messages

Using the theory of steganography, I was wishing to hide a text file in the image. The encrypted image will looked like not differ to the original image. This can eliminate the suspicion of third party toward the encrypted image. The text file embedded in the image should be able to retrieve it back to the original state.



The Concept of Steganography

2.3 Information Hiding in Text Using Typesetting Tools with Stego-Encoding

TeX is a famous typesetting tool invented by Donald E. Knuth. It is a very useful in producing technical or scientific documents with professional page layout. No similar to others type word processing software such as MS Word, TeX does not have the instant formatting results on the screen, but it is using the control sequences commands to manipulate the scheme and page layout

ASCII file is the source the TeX. It can be compiled to produce an independent DVI file. This DVI file later can be change to other portable formats like pdf or ps for the purpose of easier distribution of file.

TeX is designed to work with auxiliary packages that contain higher-level properties. Leslie Lamport wrote out the LeTeX in the TeX format later. He believes that the author should more focus on logical design than visual design in their work, thus it is much easier in use.

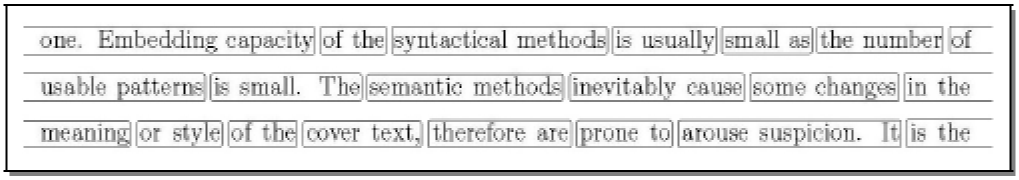
Data hiding without altering the output text can be realized, for instance, by making the control sequence\begin{equation}\label{eq14} represent 0 whereas \begin{equation} \label{eq14}represent 1. The source file is then transferred to convey the secret information that can be extracted directly by checking presence of space.

Inter-word space coding

As arbitrarily editing inter-word spaces may influence the line-feeding places, caution must be taken in the coding so that the original pattern of word grouping is preserved in order to make sure the right synchronization in data extraction. A rule of thumb is to make the total of space widths in each line as small as possible. Hence the below algorithm is applied, as shown in Figure 2.

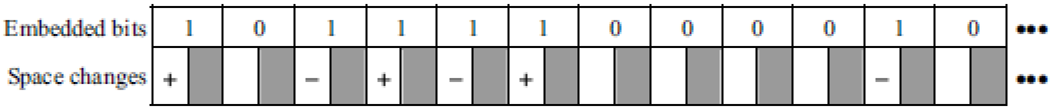
- Spaces between groups must be keep remain unchanged.
- The inter-word space in the group is remained constant if a 0 is to be embedded.

- The inter-word space in the group will be widened for the first encounter, if a 1 is to be embedded. Next it is alternately widened and shrunk for the following subsequent 1s.



one. Embedding capacity of the syntactical methods is usually small as the number of usable patterns is small. The semantic methods inevitably cause some changes in the meaning or style of the cover text, therefore are prone to arouse suspicion. It is the

Figure 1. Grouping of words. Every group has just one usable inter-word space.

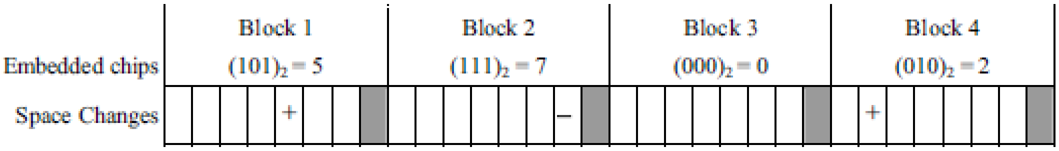


Embedded bits	1		0		1		1		1		1		0		0		0		0		1		0		•••
Space changes	+	▨		▨	−	▨	+	▨	−	▨	+	▨		▨		▨		▨		▨	−	▨		▨	•••

Figure 2. Embedding scheme

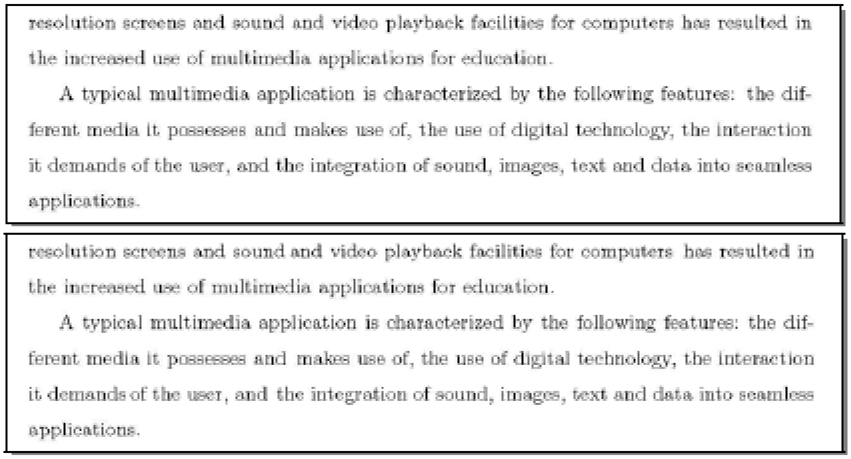
Stego Encoding for Better Stealthiness

The thought of stego encoding is to deplete more cover data. Let’s say n symbols existed per block, than the number of bits, l, to be embedded, but to practically edit only a small portion of these symbols. Each l bits of secret data we call it a chip. We are now use inter-word spaces as symbols for embedding instead of using word pairs. Next, we divide the text into a set of blocks, every set block containing n+1 inter-word spaces, where n =$2^{i} + 1$. One secret chip is being carried by each block.



	Block 1								Block 2								Block 3								Block 4							
Embedded chips	$(101)_2 = 5$								$(111)_2 = 7$								$(000)_2 = 0$								$(010)_2 = 2$							
Space Changes					+			▨							–	▨								▨		+						▨

Figure 3. Data embedding using stego-encoding



resolution screens and sound and video playback facilities for computers has resulted in the increased use of multimedia applications for education.

A typical multimedia application is characterized by the following features: the different media it possesses and makes use of, the use of digital technology, the interaction it demands of the user, and the integration of sound, images, text and data into seamless applications.

resolution screens and sound and video playback facilities for computers has resulted in the increased use of multimedia applications for education.

A typical multimedia application is characterized by the following features: the different media it possesses and makes use of, the use of digital technology, the interaction it demands of the user, and the integration of sound, images, text and data into seamless applications.

Figure 4. Data hiding in text. Top: original text; Bottom: stego-text with hidden characters

2.4 Authentication of Secret Information in Image Steganography

In this section, a steganography skill that hides secret message or information into the spatial domain of the cover image will be presented. This method can affirm the integrity of the secret message by using the Discrete Wavelet Transform (DWT). The integrity can check whether the data had been modified by others when it is being swapped.

This suggests technique is able to affirm the integrity of secret message from the stego image. In order to do so, we need to produce a Verification code *"V "*by using two alternate coefficients in the DWT domain that are diagonally located. This Verification code is switched with the secret message *"X"* and is then embedded in the spatial domain of the cover image.

The overall flow of the process of embedding is shown in figure 1. The embedding process is as follows: After change the current row in the cover image into block form, DWT is used to the blocks. Two special coefficients that are present diagonally are selected to produce the Verification code *"V"* in this block. The gained Verification code *"V"* is then switched with the secret message *"X"* that is to be embedded in the cover image. The switching is achieved by using secret key to obtained secret information. The secret key determines the method on how we switch the verification code with the secret message to produce the secret information *"I"*.

Thus we can say that the gained secret information is the combine of both secret message and the verification code. The embedded image is now the intermediate stego image. The intermediate stego-image is now undergone the similar procedure of generating the verification code. This code is later again attached to the intermediate stego-image to produce the actual stego image to be sent to the receiver.

Figure 2 depict the flowchart of the extracting process. The extraction process is as follows: Alter the current row in the stego image in the block form and then employ DWT. Create the Verification code *"Z"* by selecting two special coefficients of the gained DWT block. Extract the secret information from the stego image. The secret information *"I"* then will be reversed from the stego image. Separate and divide extracted verification code *"V"* and extracted message *"X"* .Now confirm the integrity of secret message *"X"* by comparing *"Z"* with *"V"*. Repeat all the above steps for all the rows of the stego image for the purpose to verify the integrity of the secret message. If 5% or more of the total rows in the stego image fail to examine its integrity, then it is means that someone else has attempted to make modification on it.

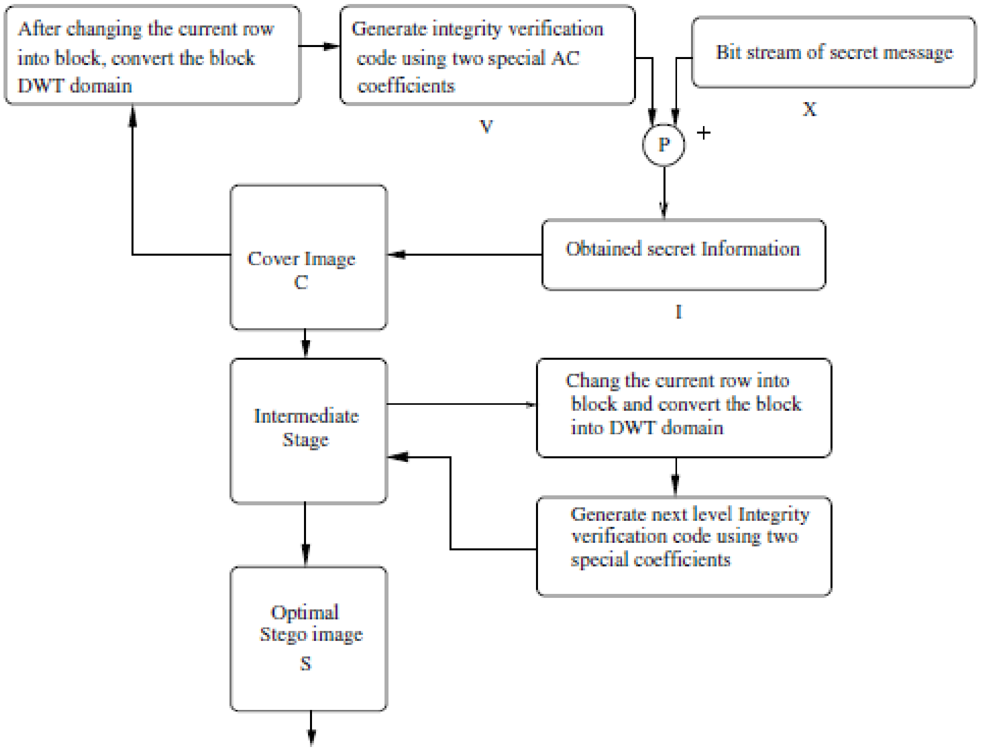

After changing the current row into block, convert the block DWT domain
Generate integrity verification code using two special AC coefficients
V
Bit stream of secret message
X
P
+
Obtained secret Information
I
Cover Image
C
Intermediate Stage
Chang the current row into block and convert the block into DWT domain
Generate next level Integrity verification code using two special coefficients
Optimal Stego image
S

Fig. 2. Flow chart for Embedding process

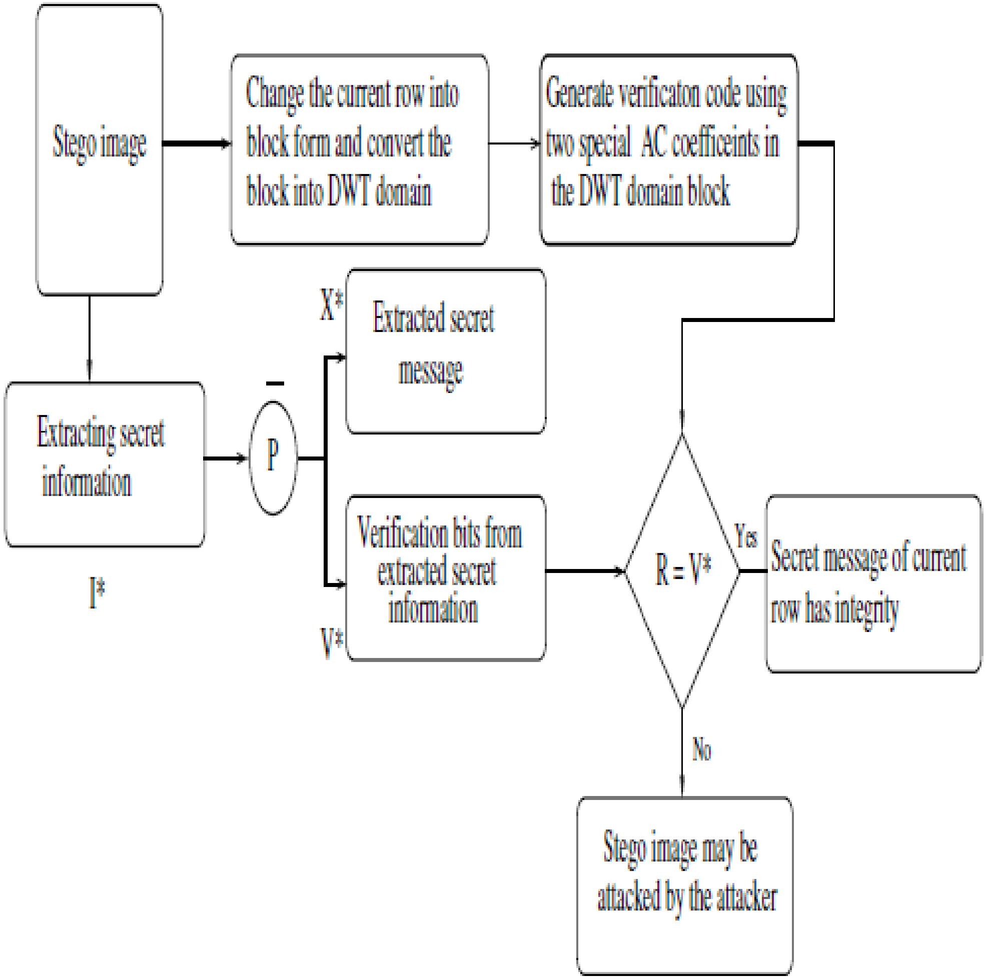

Stego image
Change the current row into block form and convert the block into DWT domain
Generate verificaton code using two special AC coefficeints in the DWT domain block
X*
Extracted secret message
Extracting secret information
P
Verification bits from extracted secret information
V*
I*
R = V*
Yes
Secret message of current row has integrity
No
Stego image may be attacked by the attacker

Fig. 3. Flowchart for Extracting process

2.5 Visual Cryptographic Steganography in Images

In the multimedia steganocryptic system, public key encryption algorithm will be first in used to encrypt the message. Then this ciphered data will be hidden into an image file thus completing both data ciphering and hiding. The multimedia data will be used to give the cover for the information. Each color in the multimedia data when deliberated as an element in an arrangement of 3D matrix with R, G and B as axis can be used to write a cipher (encoded message) on a 3D space. Block or a grid cipher is the technique that use in data mapping. This cipher will consists the data which will be mapped in a 3-D matrix form where the x-axis equal to R (red), y-axis equal to G (green) and z-axis equal to B (blue).

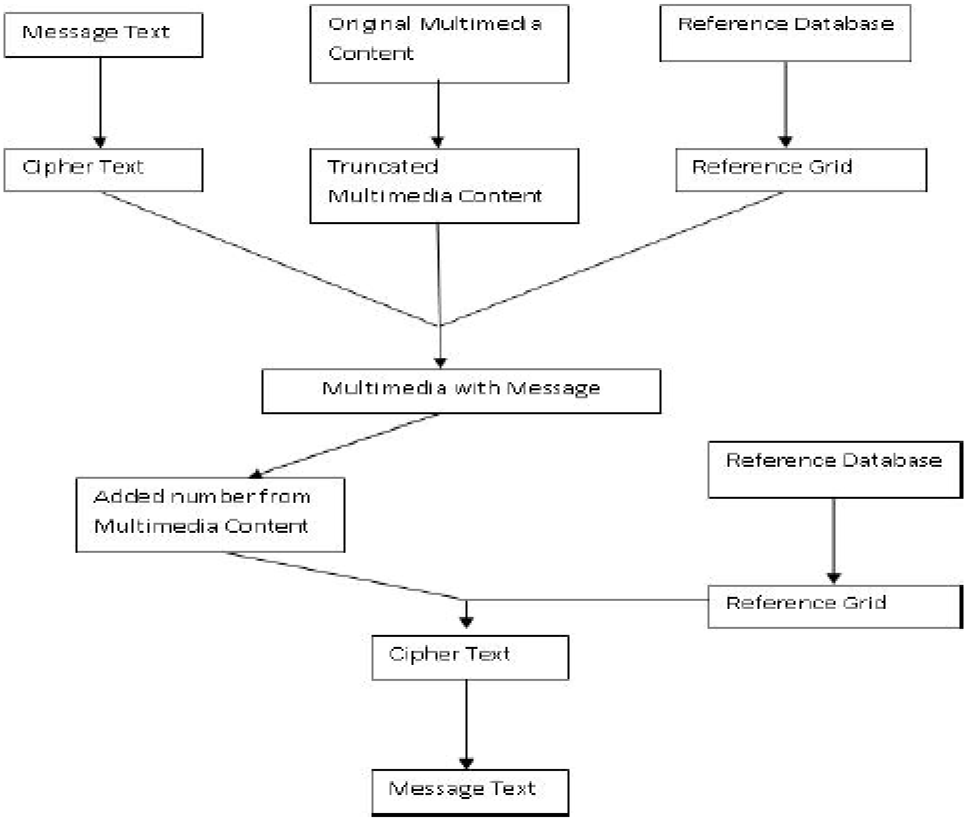
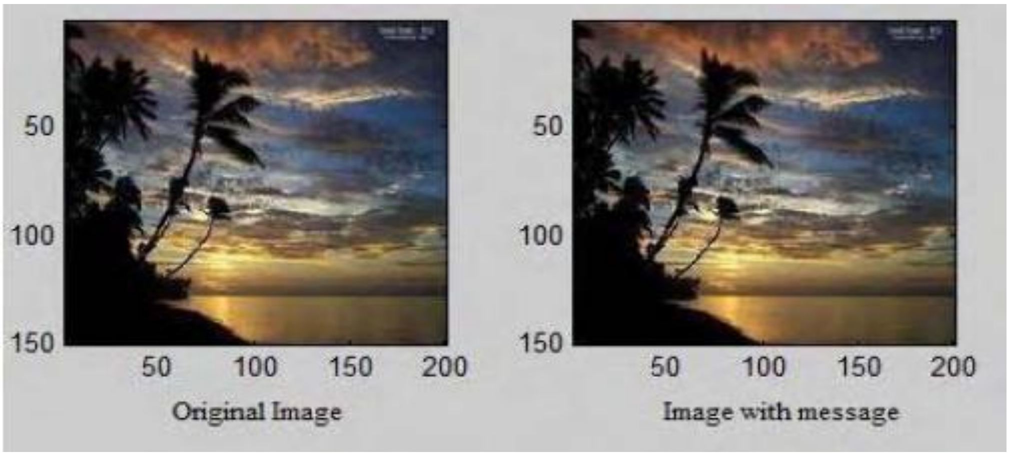


Fig. 1 System Flow Chart

Encryption Algorithm

- Asymmetric Key Cryptography technique will be first in used to encrypt the message. Basic DES algorithm is the chosen algorithm to encrypt the data. Now, this encrypted data can be hidden into the multimedia file.
- Using a modified bit encoding technique, the encrypted data will be saved in the picture by truncating the pixel values to the nearest zero digit (or a predefined digit). A specific number which defines the 3-D representation of the character in the cipher code sequence can be added to this number. For every character in the message a particular change will be made in the RGB values of a pixel. (This change should be less than 5 for each of R,G and B values) This deviation from the original value will be unique for each character of the message. This deviation is depending on the particular data block (grid) chosen from the reference database too. For every byte in the data one pixel will be modified, one byte of data will be saved per pixel in the picture.
- The cipher sequence can be deciphered without the original picture and the receiver will only received the modified picture. The attributes of the picture will be encrypted in the first few lines of its properties, and saved so as to give us the information if the image is modified or the picture extension has been altered such as from bmp to jpg or gif. These properties can be applied in the decrypting So in short, only the right coded image in the right scheme will create the sent out message.



2.6 Technique, Method, Software Going to Use

Well, my selected project full title is “Steganography, Text File Hiding in Image”. As the title sound said, I am going to attempt hiding text file in image. This will be my first ever steganography project. Naturally I will try to make one small text file (approximately may be 5 to 10 KB size) succeed to be hidden in a small image (may be will start with .bmp file).

Upon the success of embedding, I will try to enhance the security further using cryptography technique. Which is mean I am going to play around with encryption and decryption as well.

In the end of my work, I shall able to encrypt wanted text file into chosen image. The text-file embedded image will be the stego message. This stego message should look no different from the original chosen image in naked eye.

After success to create the stego message, I shall able to decrypt the stegno message as well. This is meaning that I shall capable to retrieve back the embedded text file from the stego message. After the decryption, I should have the text file and image back.

CHAPTER 3

METHODOLOGY

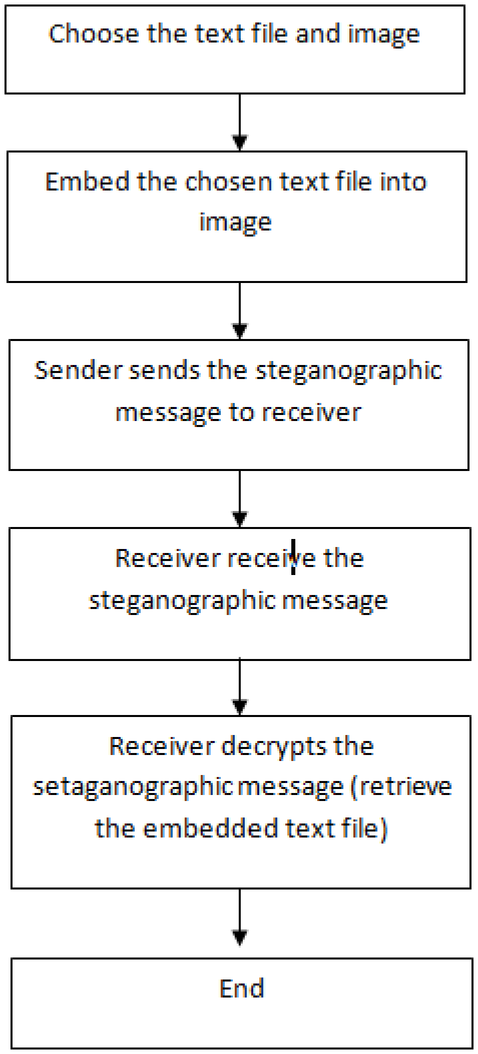
This chapter is the main issue of the overall research because it contains the technique, method, research and related software and hardware for the text file hiding in image.

3.1 INTRODUCTION

This chapter contains of section 3.2 depict the overall storyboard of the research which is the step by step of overall formation of steganograpic message. Section3.3 explained the research methodology apply in this chapter. Section 3.4 show the Graphic User Interface (GUI) of the application created for the aid of formation of steganographic message. Section 3.5 list out all the hardware and software used in doing this research. Section 3.6 attached the Grant Chart of the work planning in doing the research. Last but not least section 3.7 had a small and simple conclusion of the chapter.

3.2 OVERVIEW

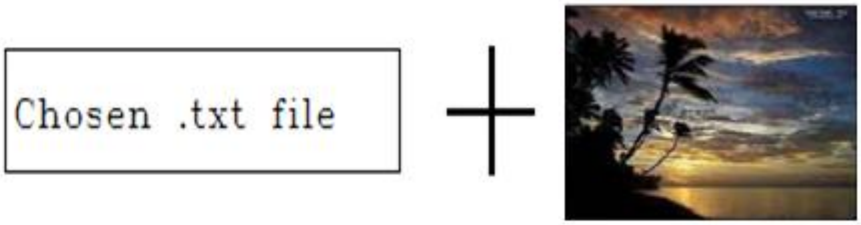
This chapter contains the flow chart or the story board of formation of the steganographic message. This chapter will explain every step involved in the formation of steganographic message. Nevertheless, this chapter will reveal also the sending and receiving process of the steganographic message.


Choose the text file and image
Embed the chosen text file into image
Sender sends the steganographic message to receiver
Receiver receive the steganographic message
Receiver decrypts the setaganographic message (retrieve the embedded text file)
End

Overall Flow Chart of the steganographic message

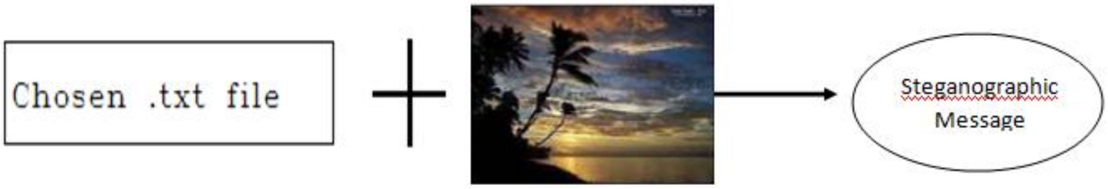
3.2.1 Choose the Text File and Image

It is the first step of the formation of steganographic message. Sender selects the text file which he or she wanted to be hidden as well as an image. After chose the wanted text file and image, sender will further proceed to embed the text file into image to produce the steaganographic message.



3.2.2 Embed the Chosen Text File into Image

After confirm the wanted text file and image, sender then will proceed to the formation of steganographic message. This will start by embedding the chosen text file into the wanted image.

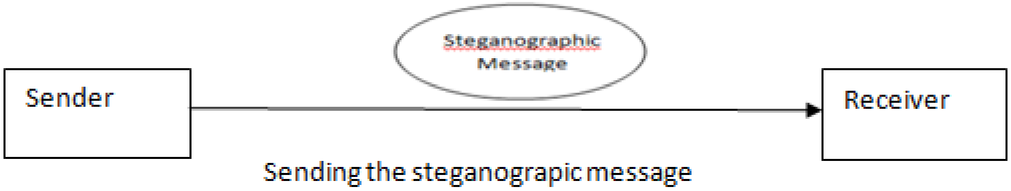


The formation of Steganographic Message

In the naked eye, the steganographic message will just look alike as the normal image. No one will know that this image had been modified and hid a text file. It is almost impossible to discover the secret of the image just by look on it through the naked eye.

3.2.3 Sender Sends the Steganographic Message to Receiver

After the sender succeed to produce the steganographic message (ie, embedded the text file into image), he or she may start to send the steganographic message to the receiver.



3.2.4 Receiver Receive the Steganographic Message

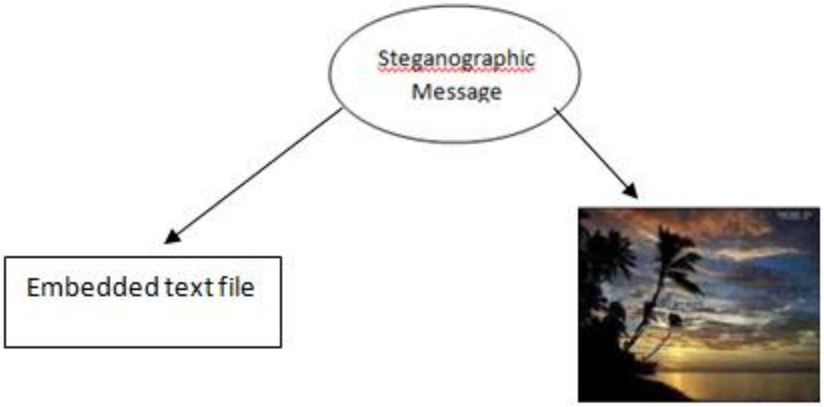
After the sender sends the steganographic message to the receiver, he or she should inform it to the receiver. Receiver now will check his or her mailbox and then acknowledge the sender that the steganographic message had been received successfully.

Again, the steganographic message received by the receiver is just a image (which is look alike to the chosen image of the sender).

This is what wills receiver received. It is just look alike with the original image

3.2.5 Receiver Decrypts the Setaganographic Message (Retrieve the Embedded Text File)

In order to know the content of the text file which is equivalently same to the message that the sender wish to tell receiver, receiver has to decrypt the steganographic message. Receiver has to retrieve back the text file so that he or she can read the content of the message.



Receiver Decrypt the Steganographic Message

3.2.6 End

After the receiver decrypts the steganographic message, the receiver will gain two objects now which are the text file and the image. Now, the receiver can read the text file ready. Therefore, we can say that by this point, the massage which wishes to conveyed by the sender are successfully conveyed.