 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

Text Steganography Basedon Inter-Word Space Statistics

[1]K.Kavitha, [2]B. Kishore Kumar
[1,2]Dept. of IT, AITAM, Tekkali, Andhra Pradesh, India

Abstract
Steganography can be used for hiding the secret message within a larger one in such a way that others can't discern the presence or contents of hidden message. The Steganography algorithms employ multimedia as the medium to ensure hidden exchange of information between multiple contenders and to protect the data from unauthorized access. Text documents can be watermarked by patterning the inter word spaces.
Steganography with watermarking adds additional security to the document and it provides high levels of exploitation of novel concepts of word classification and inter-word space statistics. The text steganography algorithms are based on modification of word style, spaces etc. Here, the nonlinear word positions of data in the document are targeted with insignificant modifications and by grouping into segments and their classification using word class information where the amount of information is dependent on requirement. This information is encoded by modifying some statistics of inter-word spaces of segments belonging to same class.

Keywords
Encryption, Steganography

I. Introduction

Steganography 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. Steganography sometimes used when encryption is not permitted or more commonly, Steganography is used to supplement encryption. An encrypted file may still hide information using Steganography, so even if the encrypted file is deciphered, the hidden message is not seen. Text documents can be watermarked by patterning the inter word spaces. Steganography provides additional security to the document and it provides high levels of exploitation of novel concepts of word classification and inter-word space statistics. The text Steganography algorithms are based on modification of word style, spaces etc. Here, the nonlinear word positions of data in the document are targeted with insignificant modifications and by grouping into segments and their classification using word class information where the amount of information is dependent on requirement. This information is encoded by modifying some statistics of inter-word spaces of segments belonging to same class.

A. Steganography

Steganography is the art or practice of concealing a message, image, or file within another message, image, or file. Although Steganography is an ancient subject the modern formulation of it is often given by where two inmates wish to communicate in secret to hatch an escape plan. There are four types of steganography as shown in Fig. 1. All of their communication passes through a text, image or file.

1. Types of Steganography

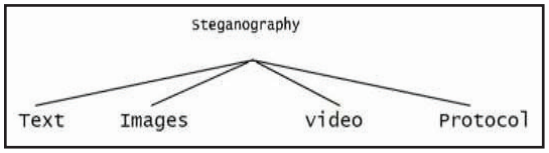

Fig. 1: Types of Steganography

Almost all digital file format scan be used for Steganography, but the form that are more suitable are those with a high degree of redundancy. The redundant bits of an object are those bits that can be altered without the alteration being detected easily. Image and audio files especially comply with this requirement, while research has also uncovered other file formats that can be used for information hiding. Hiding information in text is historically the most important method Steganography. An obvious method was to hide a secret message in every letter of every word of a text message. It is only since the beginning of the Internet and all the different digital file formats that is has decreased in importance. Given the proliferation of digital images, especially on the Internet, and given the large amount of bits present in the digital representation of images are the most popular cover objects for Steganography. Steganography will focus on hiding information in other information.

2. Text Steganography

Text Steganography is a method used for hiding secret information inside some cover text. Here the text Steganography algorithms based on modification of word style etcetera, has advantages of great capacity, good imperceptibility and wide application range.

3. Image Steganography

As stated earlier, images are the most popular cover objects used for Steganography. In the domain of digital images many different image file formats exist, most of them for specific applications. For these different image file formats, different algorithms exist.

4. Audio Steganography

Audio Steganography is a young branch of this discipline. An encoding mechanism is used for embedding the message into the audio file. The quality of the audio file after encoding remains unaffected. A public key cryptographic algorithm can also used to ensure greater security.

B. Advantegs of Steganography

1. Robustness

Robustness is the ability of a computer system to cope with errors during execution or the ability of an algorithm to continue to operate despite abnormalities in input, calculations, etc. The harder it is to create an error of any type or form that the computer cannot handle safely the more robust the software is. Formal techniques, such as fuzzy testing, are essential to showing robustness since this type of testing involves invalid or unexpected inputs. In addition, fault injection can be used to test robustness.

2. Invisibility
Making things invisible is the stuff of science fiction. Our fascination is with the idea of not being able to see things that are before our eyes, but invisibility in a different sense may be useful too.

3. Signal to Noise Ratio Capacity
Signal-to-noise ratio (often abbreviated SNR or S/N) is a measure used in science and engineering that compares the level of a desired signal to the level of background disturbance.

4. Tamper Resistance
Tamper resistance is resistance to tampering (intentional malfunction) by the normal users of either a product, or system or others with physical access to it. There are many reasons for employing tamper resistance.

II. Literature Survey
Steganography is derived from the Greek word stegano, which literally means "Covered" and graphy means "Writing", i.e. covered writing. Steganography refers to the science of "invisible" communication. Digital form of media as a cover-object being use in Steganography are images, video clips, music or sounds. Text Steganography also have been used since 2000 BC as a cover media. Text Steganography is the most difficult kind of Steganography, due largely to the relative lack of redundant information in a text file as compared to image or sound.

The existing method is used to hide information in any word instead of pointed ones only. Here at encryption side consider a large text document with a number of pages. Let, all the lines contains almost same number of words in every line what we done to align the text both the left and right margins except the last line of any paragraph. This creates a clean look along the left to right side of the page. After that, the message is taken. Initially, the string length is calculated. The corresponding 8-bit data for length is positioned into array. The characters are converted into its 8-bits data using ASCII-8. Here consider 2-bit at a time to hide information- changing style on selected words of cover text of selected page. So, calculate the number of word positions to strike.

The cover text is taken as normal text with Times New Roman font and size of 10. Here there are four different styles to hide the data within text. We have used the styles like Bold, Italic and Underline. For two data bits may occur at four different orders starting from 00 to 11 i.e. four different combination of style of cover text can represent 8-bit embedding. The presence of zero in LSB (Least Significant Bit) is as word starting with vowel and one as word starting with consonant. The presence of zero in MSB (Most Significant Bit) is as word with odd number of letters and one as word with even number of letters.

First two MSB columns for array data bits and remaining two LSB columns for targeted word style. Depending on the data bits from array and selected word, the font styles are included with Bold, Italic and Underline. For Example, if any two-array data bits are of 11 and targeted word starting with vowel and even numbers of letters, the corresponding word will be altered to Arial Narrow and Bold. After applying styles to the existing text then provide different spaces between the words in the text.

At the decryption side receiver may decrypt the encrypted text to get the original text document by applying the reverse procedure of the encryption procedure.

In the proposed system, initially consider the original text document and encrypt the document by applying two encryption procedures. Initially we have to replace the words in the text document in the first encryption phase. After the first encryption we have to apply the font styles to the given text document in the second encryption phase. Here we can get the encrypted text document. The encrypted text document can be send to the receiver.

At the receiver side, the encrypted text document can be decrypted by applying the reverse procedure of the encryption. Initially we decrypt the fonts from the encrypted text document and we can replace the text document with the original words. Now receiver can get the original text document.

III. Problem Statement
Existing method is used to hide information in any word instead of pointed ones only. We have pointed words all through the text in a number of pages nonlinearly. First, we have taken a large text document with a number of pages. Let, all the lines contains almost same number of words in every line what we done to align the text both the left and right margins except the last line of any paragraph. This creates a clean look along the left to right side of the page. After that, the message is taken. Initially, the string length is calculated. The corresponding 8-bit data for length is positioned into array. The characters are converted into its 8-bits data using ASCII-8. Here we have taken 2- bit at a time to hide information-changing style on selected words of cover text of selected page. So, calculate the number of word positions we have to strike. The cover text is taken as normal text with Times New Roman font.

The proposed system deals with the encryption and decryption of text document using some functions in Mat lab. The text file can be encrypted by replacing so many words with the other secret words in the source file. For this encrypted text file depending on the number of characters in the text some font styles are applied for providing more security. After encryption, the encrypted file will be send to receiver. Here receiver can decrypt the encrypted file by using the same key words in the encryption and applying the font styles to get back the original text file.

In the proposed system, initially consider the original text document and encrypt the document by applying two encryption procedures. Initially we have to replace the words in the text document in the first encryption phase. After the first encryption we have to apply the font styles to the given text document in the second encryption phase. Here we can get the encrypted text document. The encrypted text document can be send to the receiver.

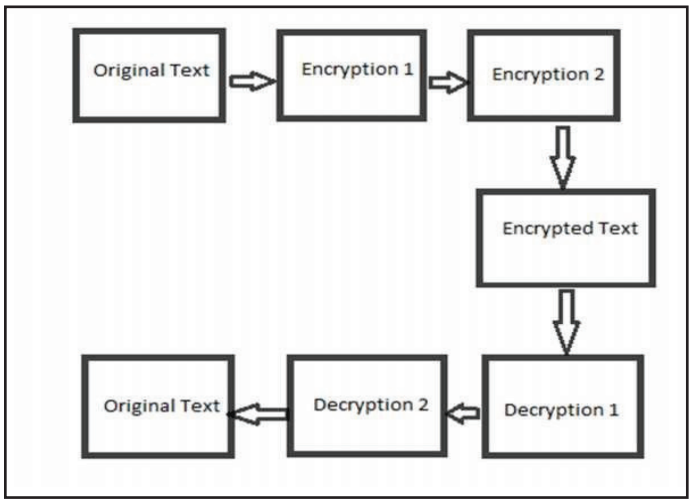

Fig. 2:

At the receiver side, the encrypted text document can be decrypted by applying the reverse procedure of the encryption. Initially we decrypt the fonts from the encrypted text document and we can replace the text document with the original words. Now receiver can get the original text document.

IV. Algorithms

A. Encryption

For text encryption and decryption we are going to use the basic functions in MATLAB software. The GUI consists of browse, upload, encrypt, encrypt1, clear, save and close options. First of all we are going to browse the file which is going to be encrypted and upload it. Now on clicking the encrypt button the following algorithm will be performed.

Step 1: Retrieve the text from the browsed file.
Step 2: Now apply the encryption on the text by replacing the key with the specified word in the text file.
Step 3: Finally save the replaced text in another file.

Now click on the encrypt1 button to perform the following algorithm

Step 1: Retrieve the modified text from the browsed file.
Step 2: Now apply different font styles to the text based on the number of characters.
Step 3: Finally save the encrypted text.

B. Decryption

In this section, we will decrypt the encrypted text to get the original text which actually needs to be received by the member or user. Here the decryption is done by using the key which is used at the encryption side.

Now on clicking the decrypt button the following algorithm is performed:

Step 1: Retrieve the text send by the sender.
Step 2: Now decrypt back the modified font into basic font style.
Step 3: Finally save the modified text.

Now on clicking the decryptt1 button the following algorithm will be performed

Step 1: Browse the text from the last saved file.
Step 2: Now apply the decryption on the text by replacing the specified word with the key in the text file.
Step 3: Finally save the decrypted text in another file.

V. Conclusion and Future Scope

In our system, we have performed double encryption to the text by replacing the word and applying font styles to it. So, it provides an additional security to the text. Even though if hacker get the file cannot get the original text data and that hacker may think this is the original message but this is not an original message the receiver only know the existence of the original message.
Future Scope

This can be applied to large data and we can exchange for more number of pages. Here we applied one font style to the whole text but we can extend to apply different fonts on single line or single word in the text document and we can apply different font sizes to the text file.

Refernces

[1] Sabyasachi Samanta, Saurabh Dutta, Goutam Sanyal, "Text Steganography: Using Nonlinear Word Positions (NWP)", IJAIR, Vol. 2, Issue 8, 2013.
[2] "A Research paper: An ASCII value based data encryption algorithm and its comparison with other symmetric data encryption algorithms", Akanksha Mathur Computer Sc. & Engg. JIET Group of Institutions Jodhpur, India, International Journal on Computer Science and Engineering (IJCSE), Vol. 4, No. 09, 2012.
[3] M. Grace Vennice, Prof. Tv. Rao, M. Swapna, Prof. J. Sasi kiran,"Hiding the Text Information using Steganography" International Journal of Engineering, Vol. 2, Issue 1, pp. 126-131, 2012.
[4] Debnath Bhattacharyya, Poulami Das, Samir Kumar Bandyopadhyay, Tai-hoon Kim,"Text Steganography: A Novel Approach", International Journal of Advanced Science and Technology, Vol. 3, 2009.

Smt. K. Kavitha M.Tech (CSE), MISTE, Senior Assistant Professor in Department of Information Technology, AITAM, Tekkali. She received B.Tech in Information Technology in Aditya Institute of Technology and Management, TEKKALI, INDIA and received M.Tech from JNTU Kakinada. She is having 11 years teaching experience.

Mr. Kishor Kumar Bhupati M.Tech (CSE), MISTE, Senior Assistant Professor in the department of Information Technology. He completed M.Tech in Computer Science and Engineering from AITAM, Tekkali, India and B.Tech in Information Technology from GMR Institute of Technology, Rajam, India. He is having 10 years of experience.