ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

Cryptography and Steganography Algorithm for Hiding Text in HTML Document

Dr. Swapna Borde, Yatisha Suresh Bhoir, Nikita Krishnarao Deshmukh, Utkarsha Bhushan Kamble

***Abstract*— As we know, our internet is growing fast and rapidly. It is very essential to secure the information. In this paper, we use text steganography technique which uses HTML document as the cover medium to hide secret messages. We are using C# .net technology for implementing the technique. The technique of cryptography with steganography provides security that ensures the safe and secure delivery of message to the receiver.**

***Index Terms*— Steganography, Cryptography, HTML Document, Attributes (Primary & Secondary)**

I. INTRODUCTION

The internet is a huge collection of networks. It is a super-highway that connects places all over the world. Internet is one of the rapidly growing technologies in the present era. This growth has focused attention on one of the most important aspect of internet viz. information security. Since internet is a public network, securing the information on internet is very important. Various techniques including cryptography, steganography etc. are used to secure information on the internet. Cryptography is the science of converting the messages that are intended to be secret into some other form, such that it is not understandable to anyone other than the intended sender and recipients.

Steganography is a technique for securing information by hiding it in some other medium, such that the existence of information is concealed to everyone except for the intended sender and receiver. Steganography refers to the art and science of hiding secret information in some other media. The information to be hided is called the secret message and the medium in which the information is hided is called the cover document. The cover document containing hidden message is called stego-document. The algorithms employed for hiding the message in the cover medium at the sender end and extracting the hidden message from the stego-document at the receiver end is called stego system.[3][6].

II. EXISTING SOLUTIONS

A. Digital messages

Concealing messages within the lowest bits of noisy images or sound files. Concealing data within encrypted data or within random data. The data to be concealed are first encrypted before being used to overwrite part of a much larger block of encrypted data or a block of random data (an unbreakable cipher like the one-time pad generates cipher-texts that look perfectly random if one does not have the private key). Mimic functions convert one fit have the statistical profile of another. This can thwart statistical methods that help brute-force attacks identify the right solution in a cipher text only attack. Concealed messages in tampered executable files, exploiting redundancy in the targeted instruction set. Pictures embedded in Video material (optionally played at slower or faster speed). Injecting imperceptible delays to packets sent over the network from the keyboard. Delays in key presses in some applications (telnet or remote desktop software) can mean a delay. [1][6][4]

B. Social Steganography

In communities with social or government taboos or censorship, people use cultural steganography: hiding messages in idiom, pop culture references, and other messages that are shared publicly and assumed to be monitored. This relies on social context to make the underlying messages visible only to certain readers. [3][6]

C. Network

All information hiding techniques that may be used to exchange steganograms in telecommunication networks can be classified under the general term of network steganography. This nomenclature was originally introduced by Krzysztof in 2003. Contrary to the typical stenographic methods which utilize digital media (images, audio and

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

video files) as a cover for hidden data, network steganography utilizes communication protocols control elements and their basic intrinsic functionality. As a result, such methods are harder to detect and eliminate. Typical network steganography methods involve modification of the properties of a single network protocol. Such modification can be applied to the PDU (Protocol Data Unit to the time relations between the exchanged PDUs, or both (hybrid methods). Moreover, it is feasible to utilize the relation between two or more different network protocols to enable secret communication. These applications fall under the term inter-protocol steganography. [3][5][6]

D. Printed

Digital steganography output may be in the form of printed documents. A message, the plaintext, may be first encrypted by traditional means, producing a cipher text. Then, an innocuous cover text is modified in some way so as to contain the cipher text, resulting in the stegotext. For example, the letter size, spacing, typeface, or other characteristics of a cover text can be manipulated to carry the hidden message. Only a recipient who knows the technique used can recover the message and then decrypt it. Francis Bacon developed Bacon's cipher as such a technique. The cipher text produced by most digital steganography methods, however, is not printable. Traditional digital methods rely on perturbing noise in the channel file to hide the message, as such; the channel file must be transmitted to the recipient with no additional noise from the transmission. Printing introduces much noise in the cipher text, generally rendering the message unrecoverable. There are techniques that address this limitation, one notable example is ASCII Art Steganography. [2][4][6]

E. Using Puzzles

This is the art of concealing data in a puzzle can take advantage of the degrees of freedom in stating the puzzle, using the starting information to encode a key within the puzzle image. For instance, steganography using Sudoku puzzles has as many keys as there are possible solutions of a Sudoku puzzle, which is 6.71×1021. This is equivalent to around 70 bits, making it much stronger than the DES method which uses a 56 bit key. [3][6][5]

III. WORKING OF NOVEL TEXT STEGANOGRAPHY SYSTEM USING HTML DOCUMENT

A. Key file generation

This is one of the main techniques of this project. It is very essential to generate the key file. The key file initially contains two main attributes first is Primary attribute and the second is secondary attribute. This attributes consists of bit 1 or 0.

B. Hiding the message

To hide a message, we convert the plain text into binary number. Then scan the html document to find the attribute combinations that can be used to hide a bit.

Procedure for hiding the message

1. Encrypt the secret message using play fair cipher encryption mechanism and convert the message in binary format.

2. Scan the html document. Analyse each attribute of each tag of the html document.

3. If this attribute is found in the primary attribute field of the key file:

i) Check if its corresponding secondary attribute is present in the currently being processed tag. If yes, then this pair of attribute can hide a bit.

ii) To hide a bit, read one bit of secret message if it is 1, and then compare the actual order of this pair of attributes in the tag with the desired order according to key file.

iii) If primary attribute lies before the secondary in the tag, then order is retained, else it is reversed. Mark both attributes as processed.

iv) If it is 0, then retain the order if secondary attribute lies before the primary, else reverse the order. Mark both attributes as processed.

v) If the attribute is not found in the primary attribute field of the key file or if it is marked as processed then skip this attribute and move to another attribute.[3][6]

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

C. Analysis

In Analysis, It shows the remaining capacity of html document to hide more bits. To know the capacity of html document, we have to click on the analysis button provided in form. For example, it shows the notification as “The file can hide 4177 bits (522 characters). Add more attributes to your key to raise the capacity"

D. Extracting the message

The extraction component is quite simple. First, it scans the document to find the attribute pairs that hides a bit, using a similar procedure as used for hiding. Once it finds the attribute pair, it compares the positions of the attributes according to the key file. If primary attribute lies before (as determined by the positions) the secondary attribute, then a bit 1 is recorded else a bit 0 is recorded.

Procedure for extracting the message

1. Scan the html document. Analyze each attribute of each tag of the html document.
2. If this attribute is found in primary attribute field of the key file:
   i) Check if its corresponding secondary attribute is present in the currently being processed tag. If yes, then this pair of attribute hides a bit. To retrieve the hidden bit, check the ordering of attribute.
   ii) If the primary attribute is followed by secondary attribute, record a bit 1, else record a bit 0. Mark the attributes as processed after retrieving the bit.
   iii) If the attribute is not found in the primary attribute field of the key file or if it is marked as processed, then skip this attribute and move to another attribute.
3. Convert the bit stream obtained after the completion of step2 into stream of characters. This is the extracted secret message in its encrypted form.
4. Decrypt the encrypted secret message using play fair cipher decryption mechanism to recover the original secret message. [5][6]

IV. RESULT

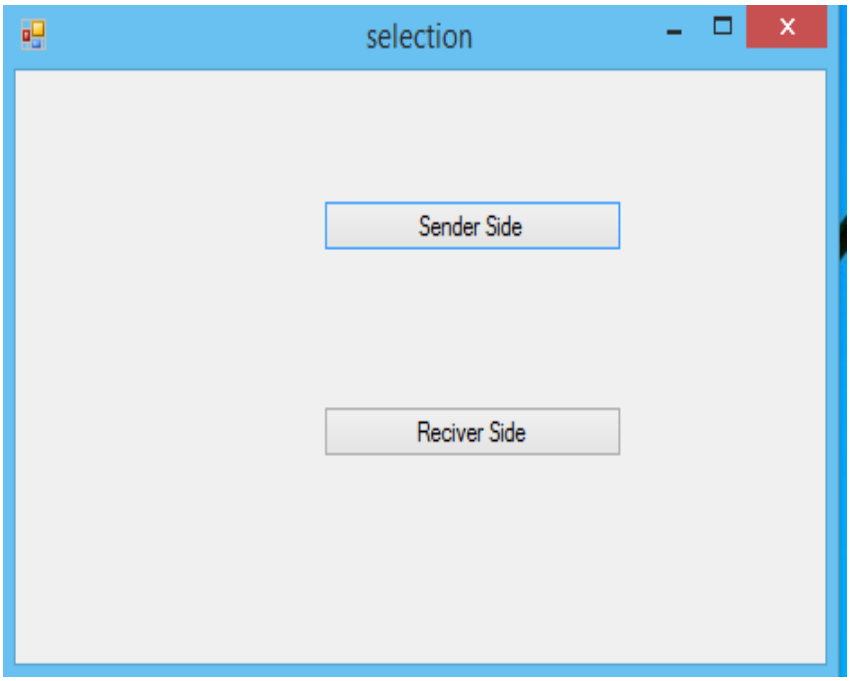

Fig.(a) Selection of Sender’s or Reciever’s Side

Fig.(a) shows that create a form of sender side and reciever side.This is our first form which consists of process called “Selection”.The user can select the sender side for the further process of the cryptography.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

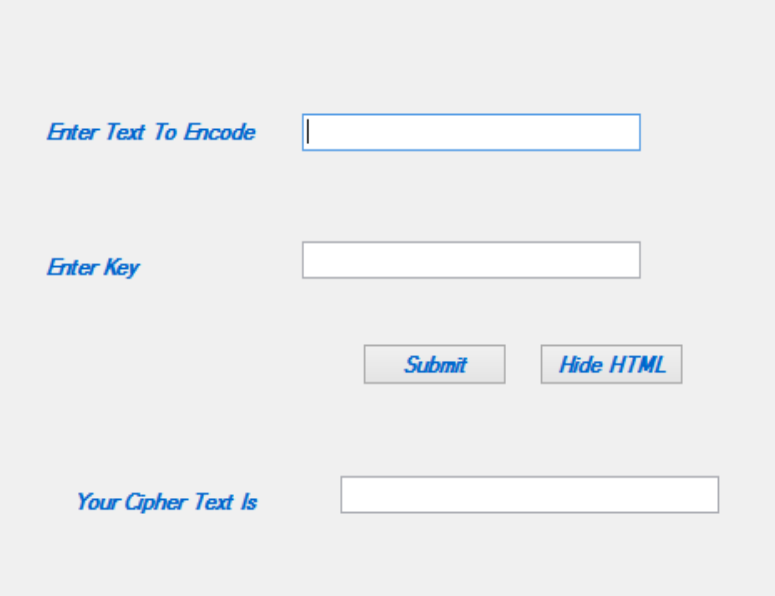

Fig.(b) Encoding Of Plain Text into Cipher Text On The Sender's Side

Fig.(b) shows that when the user clicks on sender side the next form appears. The process of converting the plain text into cipher text is done here. First will be entering text to encode,then the key has to be entered in order to hide the text. As soon as we click on submit button we receive a cipher text for our given plain text.

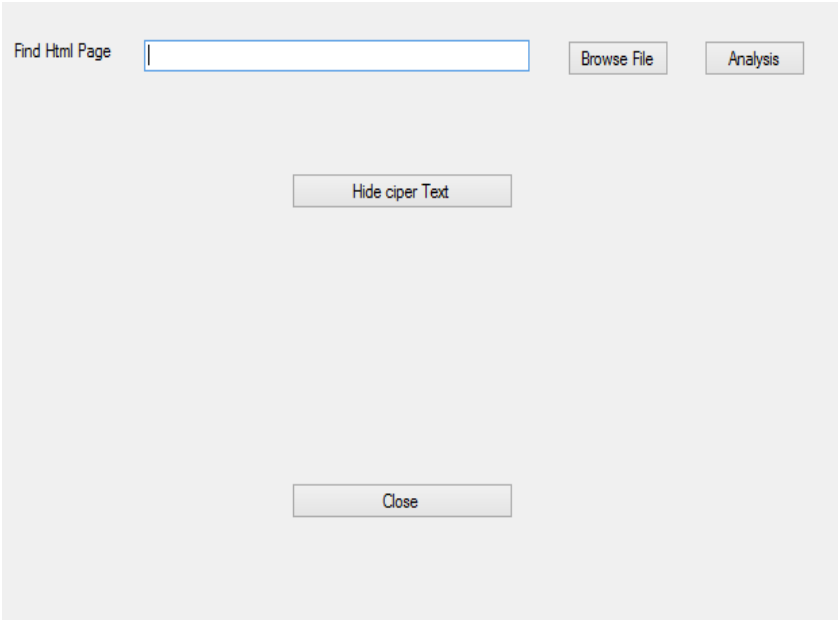

Fig.(c) Hide Cipher Text into Html Document

Fig.(c) shows that after we get cipher text the next form we get is for html document. We need to browse any file into our html page in order to hide the cipher text.In analysis, it shows the remaining capacity of html document to hide more bit to know the capacity ofhtml document,we have to click on Analysis button provided in form.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

```
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
<meta content="msapplication-tooltip" name="IMDb Web App">
<meta name="IMDb Web App" content="msapplication-tooltip">
```

Fig.(d) Hided Cipher Text in HTMLFile

Fig.(d) shows that it is the hided cipher text in html file. It contains two types of attributes corresponding to columns; Primary Attribute and Secondary Attribute. The primary attribute is something that represents a bit 1 or 0,depending on its order relative to its secondary attribute.

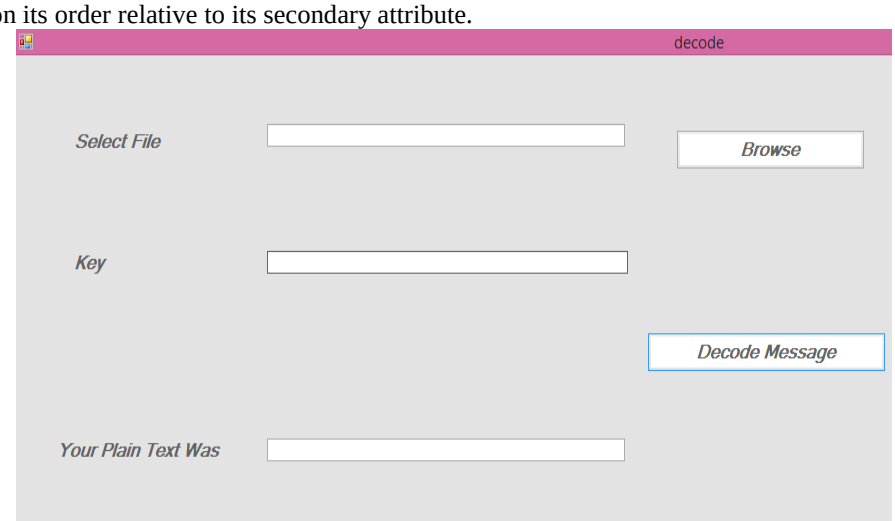

Fig.(e) Decoding Of Cipher Text into Plain Text On The Reciever's Side

Fig.(e) shows that when the user clicks on reciever side the next form appears. The process of converting the cipher text into plain text is done here. First will be entering text to decode,then the key has to be entered in order to decode the text. When we click on decode message button we received plain text for our given cipher text.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

V. CONCLUSION

We propose a novel approach of text steganography that uses the html tags and attributes to hide the secret messages. The basic idea of the proposed technique is to hide the messages by changing the order of attributes as the ordering of attributes does not affect the appearance of the html documents. The html documents are fundamental elements of the web.

ACKNOWLEDGMENT

We are grateful to Dr. Swapna Borde from Vidyavardhini's college of engineering and technology, (Mumbai University) for guiding and providing knowledge for the same.

REFERENCES

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", IBM Systems Journal, vol. 35, Issues 3&4, 1996, pp. 313-336.

[2] T. Moerland, "Steganography and Steganalysis", www.liacs.nl/home/tmoerlan/privtech.pdf, May 15, 2003.

[3] S.H. Low, N.F. Maxemchuk, J.T. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting", Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95), vol.2, 2-6 April 1995, pp. 853 - 860.

[4] A.M. Alattar, and O.M. Alattar, "Watermarking electronic text documents containing justified paragraphs and irregular line spacing", Proceedings of SPIE -- Volume5306, Security, Steganography, and Watermarking of Multimedia Contents VI, June 2004, pp. 685-695.

[5] Y. Kim, K. Moon, and I. Oh, "A Text Watermarking Algorithm based on Word Classification and Inter word Space Statistics", Proceedings of the Seventh International Conference on Document Analysis and Recognition(ICDAR‟03), 2003, pp. 775–779.

[6] Mohit Garg,"A Novel Text Steganography Technique Based On Html Document"- International Journal Of Advanced Science and Technology.

AUTHOR BIOGRAPHY

Dr. Swapna Borde
Ph.D, Head of Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, India

Yatisha Suresh bhoir
Bachelor of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, India

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

Nikita Krishnarao Deshmukh
Bachelor of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, India

Utkarsha Bhushan Kamble
Bachelor of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, India