

WEBPAGE INFORMATION HIDING USING PAGE CONTENTS

Dhammjyoti V. Dhawase , Prof. Sachin Chavan

Abstract- “Webpage Information hiding”, is the one of the Hot-Spot area of research; many techniques, methods and algorithms worked on it. Webpage information security is a way to forefend information against its confidentiality, reliability and availability. Considering webpage as covered medium to transmit the information which is to be hide and retrieve. Traditional algorithms used the tags, attributes to hide information. We have integrated the Tags, Tag Attributes and Attribute values to form Advanced Tag dictionary. Here in this paper we are encrypting webpage information using webpage contents only. A technique applying on web tools like modified HTML tag dictionary have been proposed. Advanced modified technique consists of change of states of elements or methods used on created dictionary. It can increase the more chaotic sequence and Randomness to hide secretes information and achieves better security and privacy.

This paper discusses proposed methods, implementations of obnubilating data and additionally a comparative analysis is made predicated upon some security variables.

Keywords-webpage information hiding; secure web; webpage; encrypted contents; tag; tag dictionary; case-insensitive ,content randomness, etc.

INTRODUCTION

Everyday activity now covered by anyone of internet usage. Data transmission through web based information everyday needs security.

For security reasoning, many different methods have been implemented and new methods are evolving in everyday life. Cryptography, Steganography and Watermarking are well known ways of securing information but they all work under different mechanisms. Cryptography makes data unreadable by writing into secret code and it ensures authentication, confidentiality and integrity [7]. Steganography hides the existence of data and it ensures transparency, robustness and capacity. Whereas, watermarking technique provides evidence

for the intellectual property rights over certain content some information in it.

Web based communication has a great amount of bandwidth and hence can be used for secret communication. HTML and XML are two basic but important and universal tools for web development. Scripting languages are used for dynamic web development but all browsers at the end have to translate the scripting code into HTML format.

Text Steganography is an extension of Steganography. A cover medium is required to hide information in the same way. In this paper some Text Steganographic methods are applied on HTML file and is further combined with the cryptography to add another layer of security.

HTML data is used as a cover medium and modified advance encryption algorithm is used to secure it further. This algorithm uses ShiftRowPhase as a key point of encryption. Bytes [ROWS] and [COLOUMNS] values decides which and how much shifting is done.

Combining Cryptography with Steganography has been applied in different ways but for multimedia based information [2] but problem of less research has so far been noticed in the field of Text Steganography. HTML and XML is considered as a medium because of its flexibility by creating our own tags which could be exploited to hide information in a better way, since there is less chance of vulnerability of a document.

LITURATURE SUREY

Work has to be done by respective Researchers and scientists related to webpage information hiding problem. They have put and proposed their some

meaningful and result-oriented conclusions and several schemes proposed several different schemes to reduce the problem of imperceptibility or to increase the hiding capacity of webpage.

In paper [1], proposed the method for information hiding on the basis of Tag. This paper contains the three methods. Amount of information hidden expanded to three from one bit. Therefore information hiding is greatly increased. In reference paper [2], proposed the method for information hiding on the basis of Tag Attributes. This paper overcomes the problem of ability of imperceptibility and ability of contradict with machine filtration. Based on Tag Attributes; letter upper and lower cases in tags and on the changing of attribute pair order in tags. This method does not change the display of the content and appearance of web pages after hiding secret information. In reference paper [3], proposed the method for information hiding on the basis of Tag Attributes. This paper contains the HTML standard does not define a preferred order of attributes, what means that any order can be used without affecting a web page appearance. This method is probably the most often mentioned one in the context of hiding data in the mark-up language documents. Its main advantage lies in its security, but in practice it allows only to send a small amount of data. In reference paper [4], proposed a Technique based on Attributes; To overcome the previous problems like imperceptibility and robustness. Based on the HTML attributes the different binary conventions is. In this scheme map between set of all (n-1) bit binary string is established Algorithms used 5 different definitions. In reference paper [5], demonstrated the Invisible characters, such as spaces and tabs, can be hidden at the end of each line of webpage, and covert communication happens. To determine whether there is hidden information in webpage, the statistical analysis of stego-webpage is anatomized, and a detection method is presented. In [6], paper presents a procedure to transform the binary code string into octal string to capture the randomness, from which some statistical features have been discovered. Though there still exist some bugs in the algorithm, we expect that, as a common steganalysis, the steganalytic method should be improved continuously. Shingo and Kyoko in [9] have proposed some techniques for hiding information using XML file. Techniques discussed were empty elements, white spaces in tags, attribute and element ordering. All these techniques have been proposed as a communication model; none of these

were implemented to show the results of these techniques but proposed as a future recommendation. Experiments and analysis prove that different ways can be adopted to use each method hence varieties are noticed amongst different authors. Some more techniques have also been proposed in [10] and [11] but only for HTML files. Mohammed and Sun in [10] have proposed some watermarking techniques for HTML based pages. They have focused on exploiting white space, line breaks, attributes ordering, string delimiter and color values. All above mentioned techniques were not implemented however, some of these have been tested to show sample results. In [11] by Ala'a and Mazin have also used HTML files to achieve secret communication. Their idea was to hide a secret data in an HTML file by using white space inside the webpage text and then to encrypt only colored data by using DES algorithm. Capacity of hidden information is not enough in this case as only white spaces were considered for hiding textual information. Aasma, Sumbul and Asadullah in [12] have also studied information hiding methods using XML files. They have tested random character, reverse character, tags shuffling and attribute shuffling techniques and have shown results with respect to the security and bandwidth.

PRAPOSED METHOD

HTML files/pages are used comprehensively in this structure for achieving a secret communication for textual information only. On the sender side, HTML file is used as a carrier with cover data which is further encrypted using MAES Modified Advanced Encryption Standard) algorithm, where text information is converted into a string message. This string message is embedded into HTML file using nine different embedding techniques which are described below accordingly. All embedding techniques have therefore been implemented to see the actual working scenario. A reverse procedure is applied on the receiver side to get back the original file by first decrypting and then the original HTML where embedded data is extracted.

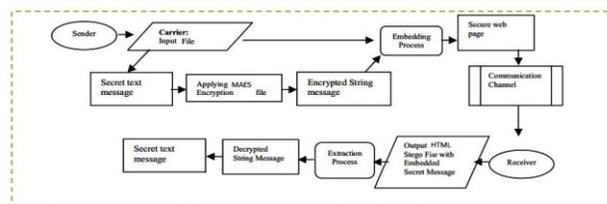


Figure 1: Flow Chart of System

Algorithm:

Html file is taken as an input on sender side which is embedded using one of the nine implemented embedding methods and is further, encrypted using a cryptographic algorithm using MAES (Modified Advanced Encryption Standard).

A reverse procedure is applied on stegoed-encrypted XML file which will first be decrypted using MAES and then extracted to get back the original file.

Other Embedding Techniques

i. White Space Method: Data is embedded by inserting a white space in html tags. In this method an extra space is inserted in alignment with tags. A space is inserted after reading "<" and before reading ">" character. White space represents a one-bit data in html file. A reverse procedure is applied in extraction phase when all extra spaces in tags are removed.

ii. White Space Replacement Method: Data is embedded by replacing all white or extra spaces in XML file with " " character value. A reverse method of replacing the " " character value with empty space is applied to extract the original message.

iii. Empty Tags Method: Data is embedded by using empty tags. Representation of an empty tag is either a start-tag immediately followed by an end tag or an empty tag. Usually this technique can be exploited or implemented using the tag. In this method the first image tag is taken and a closing character "/" is inserted before reading the end character ">". We need to have both <tag/> and <tag> </tag> to carry out the extraction process. So that when "/" character is erased from the first tag <tag/>, it will have another closing tag </tag> to avoid any error. In reverse process "/" is deleted before reading the end character ">".

iv. Random Characters Method: Data is embedded by inserting random characters in tags. A character is inserted after reading the first character of first tag. Similarly, after each word one random character is inserted. In case of a special character, full stop (.),

wild character or white space the process is repeated again. Embedding process is recursively applied to all tags. At the reverse process, all inserted characters are deleted from the file.

v. Color Replacement Method: Data is embedded by replacing color name with its hexadecimal value. In this method it first needs to find the color attribute followed by character "=" and then the color name is replaced with its hexadecimal value. The process is recursively applied on all color tags in XML file. At reverse, the hexadecimal value is replaced back to its color name.

vi. Div tag placement: Data is embedded by adding parameter inside div element.

vii. Table tag method: Data will be added by adding hidden row in table

viii. Line Break Method: Data is embedded by exploiting Line Break. In this method, first tag is taken and a line break is inserted after reading the closing character ">". The process is recursively applied to all tags. A reverse process is applied to erase all line breaks to get back the original format.

ix. Word Space Method: Data is embedded using <p> tag. In the method Text is first divided into blocks of word. A data bit is then embedded by adjusting the width of spaces between the characters within a Block, according to some predefined rule. A block size of 18-20 characters is predefined and within this block size 3-bits data (secret message) is embedded

Encryption Algorithm

We will modify the tag dictionary using MAES [Modified Advanced Encryption Standards] to be more efficient and secure way by adjusting the ShiftRow Transformation.

A. ShiftRow Transformation

a- Examine the value in the first row and first column, (state [0] [0]) is even or odd?

b- If it is odd, The ShiftRows step operates on the rows of the state; it cyclically shifts the bytes in each row by a

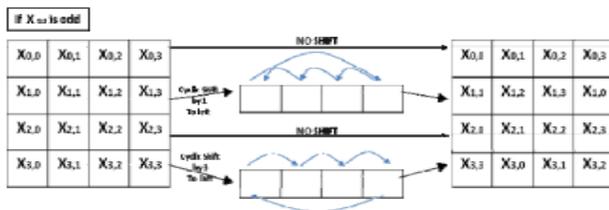
certain offset. For MAES, the first and third rows are unchanged and each byte of the second row is shifted one to the left. Similarly, the fourth row is shifted by three to the left respectively.

The pseudo code for ShiftRows is as follows.

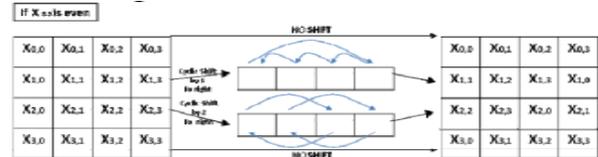
```

ShiftRows (byte state [4, Nb])
Begin byte t [Nb]
If state [0] [0] odd numbers
  For r = 1 step 1, 3
    x = r mod 4
    If x = 0 step 0 to x + 1
      For c = 0 step 1 to Nb - 1
        T[c] = state[r, (c + x) mod Nb]
      End for
      For c = 0 step 1 to Nb - 1
        State[r, c] = t[c]
      End for
    End for
  Else
    For r = 2 step 2, 4
      k = 0
      x = r mod 4
      If x = 0 step 0 to 3
        For c = Nb - 1, c >= 0, c - 1
          T[c] = state[x, (c + x) mod Nb, k + 1]
        End for
        For c = 0, c < Nb, c + 1
          State[x, c] = t[c]
        End for
      End for
    End for
  End
End.
    
```

Fig: Steps of Algorithm Flow



c- If it is even, The ShiftRows step operates on the rows of the state; it cyclically shifts the bytes in each row by a certain offset. The first and fourth rows are unchanged and each byte of the second row is shifted three to the right. Similarly, the third row is shifted by two respectively on to the right.



CONCLUSION

In this paper, Secret information is hidden using webpage contents. The traditional methods of webpage information hiding is modified by constructing Tag Dictionary of HTML elements. The modified encryption method applied on this dictionary elements or methods of HTML/XML and obtain more security and privacy of data transmitted over web.

We have Predicted result that it is method which gives better encryption in terms of security, privacy and capacity of webpage,

REFERENCES

- [1] Xie Yong; Li Juan; Zhang Yilai , “A High Capacity Information Hiding Method for Webpage Based on Tag “, Third International Conference on Digital Manufacturing and Automation (ICDMA) 2012
- [2] Yujun Yang; Yimei Yang ,” An efficient webpage information hiding method based on tag attributes ”, Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2010.
- [3] Xingming sun - et al “An algorithm of webpage information hiding based on attributes permutation”, IEEE computer society 2008.
- [4] Dongsheng Shen; Hong Zhao “A novel scheme of webpage information hiding based on attributes ”, IEEE International Conference on Information Theory and Information Security (ICITIS), 2010
- [5] Huajun Huang, Xingming Sun, Zishuai Li, Guang Sun,” Detection of Hidden Information in Webpage”, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) 2007.
- [6] Junwei Huang; Xingming Sun; Huajun Huang; Gang Luo,” Detection of Hidden Information in Webpages Based on Randomness”, Third

International Symposium on Information Assurance and Security, IEEE computer society 2007.

[7] S. Al-Riyami, K Paterson, "Advances in Cryptology-ASIACRYPT" 2003, Springer.

[8] Secure web-based communication Nighat Mir a, Sayed Afaq Hussain b, WCIT 2010.

[9] Herman Kabetta¹, B. Yudi Dwiandiyanta², Suyoto "Information hiding in css: A secure scheme text steganography using public key cryptosystem", IJCIS, December 2011-13.

[10] Mohammad Laheen, Sun XingMing, "Techniques with Statistics for Web page Watermarking" 2005.

[11] Ala' a H., Mazin S., Mohammad A. Al Hamami, "A Proposed Method to Hide inside HTML Web Page File".

[12]. Aasma, Sumbul, Asadullah, "Steganography: A New Horizon for Safe Communication through XML", 2005-2008 JATIT.

AUTHORS

<p>Ms.Dhammjyoti Dhawase Perusing M.E. Computer Engineering from MGM's College of Engineering and Technology, Navi mumbai, India.</p>	
<p>Prof. Sachin Chavan Working as Assistant Professor in Computer Engineering at MGM's College of Engineering and Technology, Navi mumbai, India. Having 9 years of Teaching Experience. Published paper is two various National, International conferences and Journals. His basic area of interest is system security, wireless networking etc.</p>	