

Data Hiding in Web Pages

Kushagra Kathpalia

Computer Science & Engineering

HMR Institute of Technology & Management

New Delhi, India

Harsh Bhagwani

Computer Science & Engineering

HMR Institute of Technology & Management

New Delhi, India

Puneet Kr. Aggarwal

Computer Science & Engineering

HMR Institute of Technology & Management

New Delhi, India

Abstract—Internet is changing all the time. It continues to be the most democratic of all the mass media. This growth and change of internet has developed the demand for techniques that ensures information security on web pages. There are 3 main methods for securing data on web pages viz. Steganography, Cryptography, Watermarking. Steganography is one of the method to hide message from unapproved audience by embedding data inside various kinds of files like Audio, Video, Images and even text documents. This paper concentrates on various Steganography methods which makes use of web pages to hide the data.

Keywords—Steganography; Data-Hiding Technique; Web Based Steganography; Information Concealing in Web Pages

I. INTRODUCTION

The information plays a vital role in the modern world. The internet is one of the rapidly growing technology in this present era. This growth focuses attention on one of the important aspect of internet viz. information security. Since internet is a public network, securing the information on internet is very important. There are various techniques including Steganography, Cryptography, Watermarking etc. are used to secure information on the internet. Steganography is a Greek word which means concealed writing or hidden writing. It is the art and science of encoding hidden messages in such a way that no one, except, the sender and intended recipient, suspects the existence of the message. Cryptography is the science of converting the messages that are intended to be secret into some other form, such that it cannot be understandable to anyone other than the intended sender and recipients. The advantage of Steganography over Cryptography is that the intended secret message does not attract attention to itself as an object of inspection. The main carrier that are been used in current technology are text, audio, video, image. Using Steganography, we can embed a secret message inside a piece of unsuspicious information and can be sent without realizing the existence of the secret message. Text based Steganography techniques are applied on webpage text and plaintext.

WWW pages have become the main source to provide the information to the users. A webpage text contains HTML, CSS, XML, JavaScript etc. as content. HTML code file can be read, parsed or displayed by browser, e.g. Firefox, Internet Explorer, Chrome. Webpage text Steganography uses tags,

attributes of the tags in varying combinations or in different orders and even use whitespace to hide the data.

A. GENERAL MODEL OF INFORMATION HIDING

The figure below gives the idea about Steganography scheme in which first step is to embed secret message in the carrier using an embedding technique. Then, embedded communication travel through the broadcast media. At the receiver side, receiver decodes the message which is the reverse process of embedding and gets the unique message.

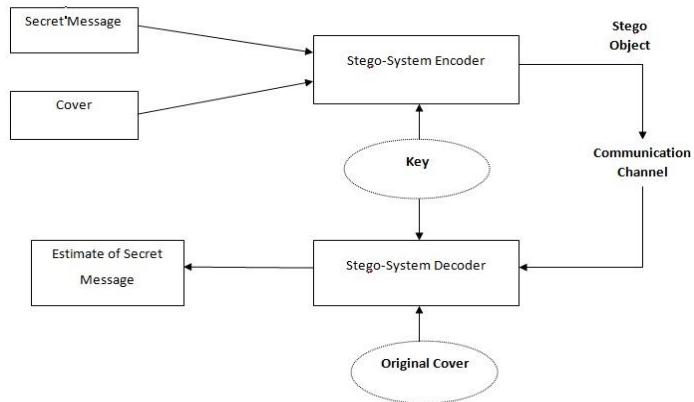


Fig. 1. General model of Information Hiding

In Steganography embedding process, the cover-object or cover- medium is the carrier of the message which can be an image, audio, video, text or some other digital media which embed the message and hides the existence of message. A Stegokey is used to embed the communication in the cover- medium. A Stego-object or Stego-medium is created once the message is successfully hidden in the carrier. And encrypted text is called as Stego data. It consists of Carrier, Message and Password. Carrier is the cover object in which a message is embedded and hides the existence of message. Message is private data which can be of plain text, cipher text, or other image that can be embedded in a bit stream such as copyright mark, covert communication or a serial number. Password is the Stego-key. A Stego-key is used to control the hiding process so as to restrict detection and retrieval of the embedded data to parties who know it.

In extracting process, Recovering message from a Stego-object requires the cover object itself and a corresponding

decoding key if a Stego-key was used during the encoding process. To extract the message, the original image may or may not be required in most applications. The message can be extracted from the cover object only by those receivers who know the corresponding decoding key.

In general, the information hiding process extracts redundant bits from cover-object.

The process consists of two steps:

1. Identification of terminated bits in a cover object. Dismissed bits can be changed without destroying the integrity of the cover-object.

2. The embedding process then selects the subset of the dismissed bits to be replaced with data from a private message. The replacement of the selected dismissed bits with message bits creates the stego object.

B. APPLICATIONS OF STEGANOGRAPHY

Digital steganography of image has many applications, for example protection of copyright, tagging of the constructions and secret communication. Copyright notice or watermark can be embedded inside an image to identify it as intellectual property. If someone attempts to use this image without approval, we can prove by extracting the watermark. In feature tagging the annotations, captions, time prints and descriptive elements can be embedded in an image. Copying the stego-image also copies the embedded features and only bashes which possess the decoding stego-key will be able to extract and view the features. On the other hand, secret communication does not advertise a covert message by using steganography. Therefore, it can avoid Scrutiny of the sender, message and recipient. This is effective only if the hidden communication is not detected by the other people.

II. TECHNIQUES FOR DATA HIDING IN WEB PAGES

There are various techniques to hide the text inside the source code of the xml file, which are already been implemented in the HTML. These techniques are known as a based on the tags.

1) *Representation of empty elements:* The representation of an empty element defines either a start tag immediately followed by an end-tag, or an empty-element tag. By switching these two forms, we can embed the data preserving all meanings of original document. The following example shows that how information can be hidden by altering image element. This method can embed one bit of data per an end-tag of empty elements.

Example

stego key:

<frame></frame> ... 0

<frame/> ... 1

stego data:

<frame="one.html"></frame>
<frame="two.html"/>
<frame="three.html"/>
<frame="four.html"/>
<frame="five.html"></frame>

Embedded data:

01110

2) *White spaces in tags:* Representation of a tag is either including some white spaces before close brackets, or no white spaces. By inserting or deleting spaces, we can embed the data preserving all meaning of original document. The following shows method of information hiding by inserting or deleting a space. This method can embed one bit of data per a tag.

Example

stego key:

<tag>, </tag>, or <tag /> ... 0

<tag >, </tag >, or <tag /> ... 1

stego data:

<real ><name>Alice</name ><roll>01 </roll></real>

<real><name >Bob</name><roll>02</id ></roll >

Embedded data:

101100 010011

3) *Appearing order of the elements:* Secret data can embed XML documents by exchanging of the appearance order of elements. In the example shown, one bit of data can be hidden in the documents per an exchange of two elements.

Example

stego key:

<student><name>NAME</name><roll>ID</roll></student> ... 0

<student><roll>ID</roll><name>NAME</name></student> ... 1

stego data:

<student><name> Alice</name><roll>0 </roll></student>

Embedded data:

01

Conditions to apply this method are,

1. No dependency of the order of the elements should exist in the application.

2. No reorder of the elements is to be done before extracting the secret data.

4) *Appearing order of the attributes:* Secret data can be embedded in XML documents by exchanging of the appearing order of attributes in the element. In example shown, one bit of data can be hidden per an exchange of the order of attributes.

Example

stego key:

<event date="DATE" day="DAY">EVENT</event> ... 0

<event day="DAY" date="DATE">EVENT</event> ... 1

stego data:

<event date="4" day="Tuesday">Tech Fest</event>

<event day="Friday" date="25">Cultural Fest</event>

Embedded data:

01

5) *Elements containing other elements:* This method uses two or more elements such that an element contains other element. Given example shows that exchanging inner-tags and outer-tags are done and per exchange one bit of data can be hidden into it.

Example

stego key:

```

<best><game>SOMETHING</game></best>... 0
<game><best>SOMETHING</best></game>... 1
stego data:
<game><best>Cricket</best></game>
<best><game>Football</game></best>
Embedded data:

```

6) *Change Case of Letter in Tags:* We can embed secret data in HTML codes. In this method that in tags, the character in upper case denoted as 1 bit and lower case denoted as 0 bit. By switching these two forms, we can embed the data preserving all meanings of original document.

Example

Stego key:

Uppercase ...1

Lowercase ...0

Stego data:

```

<html>
<body bColor="gRay">
<table><tr>
<Td width="200" align="center">test</td></tr>
< td wiDth="200" align="center" heigHt="50"> Hiding
information</td></tr>
</table>
</bodY></Html>
Embedded data:
10110110011001

```

7) *Change quotation marks of attributes value in tags:* In this method, secret data hide techniques in the quotations mark of attributes of the HTML documents. Given example shows that exchanging the double quotation mark and single quotation mark of attributes are done by hide bits. In this techniques one or more bits hide in HTML documents.

Example

Stego key:

<td width="500">1

<td width='500'>0

Stego data:

```

<html>
<body bColor='cyan'>
<table><tr>
<td width="500" align='left' height="100"> Hiding
Information</td></tr>
<td width="500" align="right">test <font color='blue'
size="5"></td><tr>
</table>
</body></html>
Embedded data: 01011101

```

8) *Repeat attributes:* Secret data can be embedded in HTML documents by repeating of the appearing attributes in the element. In the following example one bit of data can be hidden per repeating of attributes in the elements.

Example

Stego key:

<td width="200">1

<td width="200" width="200">0

Stego data:

<html>

```

<body bColor="gray">
<table> <tr>
<td width="200" align="center" align="center" height="50"> Hiding information</td></tr>
<td width="200" align="center" align="center"> test <font
color="cyan" color="cyan" size="10" size="10"> </td></tr>
</table>
</body></html>
Embedded data:
11011000

```

9) *Hide data using attribute order:* In this method, data can be hidden by exchanging the order of the attributes exchanging the order of the elements in HTML document.

Example

<abbr id="anId" class="aClass" style="color:blue;"

title="Information Hiding">

and

<abbr class="aClass" id="anId" style="color:blue;"

title="Information Hiding">

This method works on the methodology that attribute order in the tag doesn't affect the output of the HTML web page.

III. PROPOSED WORK

In our proposed techniques, We have used a white space in tags techniques in a attributes for data hiding on the web pages. In this techniques better imperceptibility and largest embedded capacity than other techniques. We have modified a techniques a more bit hiding on HTML tag.

Several tags with several attributes are used for making an HTML page. Our proposed Techniques in based on the tags and the attributes.

A. Algorithm for hiding secret message in HTML file.

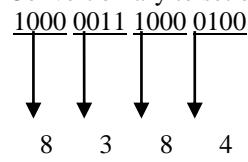
Step 1: Write message that you want to embed. (e.g. CD)

Step 2: Convert it into ASCII code. (e.g. C=67, D=68)

Step 3: Convert it into binary code. (e.g.

10000011 10000100)

Step 4: Convert binary to set of nibble.



Step 5: Prepare mapping table based on html parsing which contains the nibble and its corresponding html tag no based on the table 1.

Step 6: Based on table 1 and message Prepare the sequence number of tag which stores actual message in body part using white space attribute method. (Single space between tag and attribute=0 and double space between tag and attribute =1)

Step 7: End of the embedding procedure.

TABLE I.

	NIBBLE	No. of Tags
0	0000	2 7 12 18
1	0001	3 8 13 19
2	0010	4 9 14 20
3	0011	5 10 15 21

4	0100	6 11 16 21
5	0101	23 28 33 38
6	0110	24 29 34 39
7	0111	25 30 35 40
8	1000	26 31 36 41
9	1001	27 32 37 42
10	1010	28 33 38 43
11	1011	44 49 54 59
12	1100	45 50 55 60
13	1101	46 51 56 61
14	1110	47 52 57 62
15	1111	48 53 58 63

B. Algorithm for extracting secret message from the HTML file.

Step 1: Separate the binary data from webpage embedded using the table #1.

Step 2: Convert binary code to ASCII code.

Step 3: Convert ASCII code to encrypted message M.

Step 4: End of extraction procedure. Get the message.

IV. PERFORMANCE EVALUATION

Table 2 shows the performance parameters of existing method on information hiding on web page which are the Change case of letter in tag, the changing uppercase and lowercase of tag, the order of attributes pair, Representation of empty elements, Change quotation marks of attributes value in tags and the repeat attribute of the tag algorithm. This existing method does not change the display of the content and appearance of web pages after hiding secret information. The hidden secret information was not found by viewing the source code of the webpage. According to the above experiment result, this method has strong anti-testing capability, strong security capability against detection, strong robustness capability, no more change case of the file, good imperceptibility and good larger embedded capacity.

TABLE II. COMPARISON OF EXISTING STEGANOGRAPHY TECHNIQUES ON WEB PAGE

	Imperceptibility	Robustness	Change case of the size	Detection	Security	LEC(Largest Embedded Capacity)
Case of letter in tags	Weak	-	No	Weak	Weak	Yes
Attributes in tags	Good	Strong	No	Strong	Strong	Low
Appearing order of attributes	Good	Strong	No	Strong	Strong	Low
By using white space	Good	-	Yes	Weak	Yes	Yes
Representation of empty elements	Good	-	No	Strong	Yes	Yes
Quotation marks of attribute value in tags	Good	-	No	Weak	Weak	Yes
Repeat attributes	Good	Weak	Yes	Weak	Weak	Yes

CONCLUSION

In this paper, we gave an overview on Steganography and its various techniques to hide the data/information in web pages. We surveyed various html tags and attributes to conceal information. By using these methods, the source code of various web pages can change but not the appearance of the HTML page. We have surveyed that only those people who have a good knowledge about the HTML source code can only identify that there is some information concealed inside the source code.

FUTURE WORK: Using some techniques to be implemented on HTML code we can easily conceal the data/information without changing the source code and not being recognized by any person.

REFERENCES

- [1] Puneet Kumar Aggarwal, Dharmendra, Parita Jain, Teena Verma, "Adaptive Approach for Information Hiding in WWW Pages", IEEE-2014 page 113-118.
- [2] Geetika Dhandh, "Information Hiding Techniques", in proceeding of the national conference: INDIAcom-2008.
- [3] Chintan Dhanani, Krunal Panchal,"Steganography using Web Documents as a carrier": A survey, IJEDR:2013, Page 172-179.
- [4] Mohit Garg: A Novel Text Steganography Technique based on HTML Documents. International Journal of Advanced Science & Technology Vol.35, pp. (2011).
- [5] Chapman, Davida, "Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text," Financial Cryptography, First International Conference , FC'97, pp.335-345, Feb. 1997.
- [6] Matsumoto, Inoue, Kitabayashi, "An information hiding method for Standard MIDI File," Symposium on Cryptography and Information Security, SCIS2000-C03, Jan. 2000(in Japanese).
- [7] Xiajun GUO, Guang CHENG, Chenmgang ZHU, Aiping ZHOU, Wubin PAN, Dinhua TRUONG, "Make your Web Page Carry Abundant Secret Information Unawarely", 2013 IEEE International Conference on High Performance Computing & Communications.
- [8] Dongsheng Shen, Hong Zhao," A Novel Scheme of Web Page Information Hiding Based on Attributes", IEEE 2010.
- [9] Yujun Yang, Yimei Yang, "An Efficient Web Page Information Hiding Method Based on Tag Attributes", IEEE-2010, page 1181-1184.
- [10] Wikipedia, "Steganography Techniques and its Pros and Cons".