Web Based Steganography Using Combination of Three Methods

Karan H. Parmar (Research Scholar), Avaniba Parmar(Assistant Professor),

Hasmukh Goswami College of Engineering, Vahelal, Ahemdabad, Gujarat, India

Abstract- **As about today's generation everywhere there is a problem of threats, hacking which is always towards how to steal the information or private data. So it must be required to protect your information or data. There are many techniques for protection of data like cryptography, Steganography, etc. This paper is mainly focus on how to protect your data or information and how to do safe communication using steganography. Steganography is used to hide the particular message or data. It is used through html, www pages, php coding.**

Index Terms- **Steganography, Steganography algorithm, HTML web pages. HTML tags, HTML attributes.**

I. INTRODUCTION

STEGANOGRAPHY

The word steganography comes from the Greek Steganos, which mean covered or secret and graphy mean writing or drawing. Thus steganography is the hidden writing. This is the technique of hiding a fact that communication is taking place. Using steganography, a secret message can be embedded without any doubt of stolen information and can be sent without anyone knowing of the existence of the secret message.

All net user frequently used to send , receive, communicating or sharing their data. Therefore they want to find good and safe way for keep their message or data secret. So Steganography is the one of the good method for it.

Early steganography was messy. Before phones, before mail, before horses, messages were sent on foot. If you wanted to hide a message, you had two choices: have the messenger memorize it, or hide it on the messenger.

Steganography is the art and science of hiding information such that its presence cannot be detected [4] and a communication is happening [5]. Secret information is encoded in a manner such that the very existence of the information is concealed. Steganography includes secret communication methods that hide the message from being seen or discovered.

The main objective of Steganography is to hide the information under a cover media so that the outsiders may not discover the information contained in the said media. Most Steganography works have been performed on images, video clips, text, music and sounds. However text Steganography is the most difficult kind of Steganography [7].

Steganography uses audio, video, text, etc as carrier of information. HTML steganography is one part of text steganography which uses HTML web document as a carrier. To use HTML as a carrier has some benefits like large amount of cover documents available to hide data & by that decoding of that data by any unauthorized user is very difficult. Simply we can say that "steganography" means hiding one piece of data. Steganography is not new, it is from thousands of years ago, but it is changing time to time.

I. HIDE DATA INSIDE THE TEXT:-

As a information on interenet are interpreted by any user, encoded text in nueral languages is doubtfully effective. The last technique is so powerful and effective that if a text is printed and copied on paper for several times, the secret message cannot be retrieved [8]. Recently this technique has been used successfully.
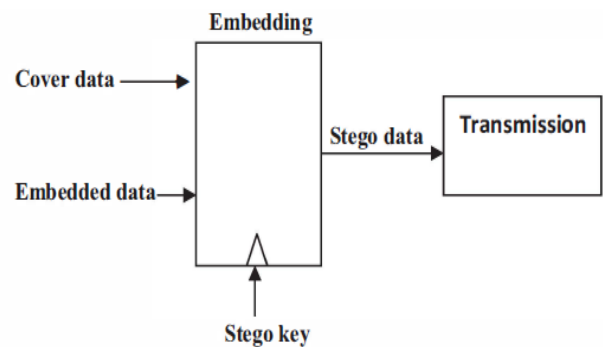

Fig 1. Model for information Hiding[5]

In steganography, the cover-object is the carrier of the message which may image, audio, video, text or some other digital media. The embedded message is the message to be hidden in the cover-data. A stego-key is used to embed the

message in the cover-data. A stego-object is created once the message is successfully hidden in the cover-data.

I. MODEL FOR STEGANOGRAPHY:-
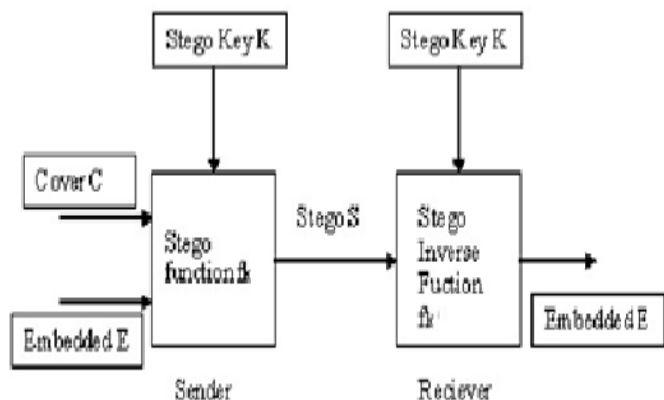
Fig 2. Model of Steganography

The basic model of steganography is shown in fig.2. It consists of Carrier, Message and Password.

Carrier is the cover object in which a message is embedded and hides the existence of message.

II. APPLICATION OF STEGNOGRAPHY

The area differs in what feature of the steganography is utilized in each system.

1. **Confidential communication and secret data storing:-**

   The privacy of the embedded data is shown in this area. Historically Steganography provides us with:
   (A) capability to hide the existence of confidential data
   (B) Hardness of detecting the hidden data
   (C) Strengthening of the secrecy of the encrypted data.

2. **Protection of data alteration:-**

   In this application it is assured that the data which is altered it must have a protect from unsupicious thing. A "Digital Certificate Document System" is a technique to alteration of data in the given scenario.

3. **Access control system for digital content distribution:-**

   In this application the data is hidden but it is explaind to publicly view.

   Today digital content is more popular rather than internet.it is a good way to reach over the nations. The system is mostly used to develop the message over the public without any unauthorized occurrens.

   We have developed a prototype of an "Access Control System" for digital content distribution through Internet. The following steps explain the scheme.

   A content owner classify his/her digital contents in a folder-by-folder manner, and embed the whole folders in some large vessel according to a steganographic method using folder access keys, and upload the embedded vessel (stego data) on his/her own Webpage.

4. Media Database systems

In this steganography secrecy is not important but it defined two data.

Media data. A photo picture, for instance, may have the following.
(1) The title of the picture
(2) The date and the time when the picture was taken
(3) The camera and the photographer's information

Recently, almost all cameras are digitalized. They are cheap in price, easy to use, quick to shoot. They are very helpful to people with good result and on time attractive result. Using this Digital camera you can easily connect with the pc and securely transfer your data in small no of time.

HTML

HTML stands for **H**yper**T**ext **M**ark-up **L**anguage .

HTML is the standard markup language used to create web pages.

HTML is written in the form of HTML elements consisting of *tags* enclosed in angle brackets like <html>. HTML tags most commonly come in pairs like <h1> and </h1>. although some tags represent *empty elements* and so are unpaired, for example<img>. The first tag in a pair is the *start tag*, and the second tag is the *end tag* (they are also called *opening tags* and *closing tags*).

```
HTML Example. <!DOCTYPE html>
<html>
<head>
<title>This is a title</title>
</head>
<body>
<p>Hello world!</p>
</body>
</html>
```

I.   HTML ELEMENTS:-

HTML documents are composed of HTML elements.

These are indicated in the document by HTML *tags*, enclosed in angle brackets thus: <p>. These are defined by three components ,

- Tags… "Start tags" and "End tags".
- Attributes within tags.
- Any textual representation between tags.

General form of an HTML element is:

<tag attribute I ="value I" attribute2="value2">content</tag>

Form of Empty elements:

<tag attributeI="valueI" attribute2="value2">.

Example:

<**head**>
<**title**>The Title</**title**>
</**head**>

This defines header of document.

II.   HTML ATTRIBUTES:-

HTML elements have attributes which provide additional information about an element and are specified in start tag. Most of the attributes of an element are name-value pairs, separated by "=" and written within the start tag of an element after the element's name. The value may be enclosed in single or double quotes.

Following attribute are commonly used in the elements of HTML:

a.   Id attributes:-

provides a document-wide unique identifier for an element. This is used to identify the element so that stylesheets can alter its presentational properties, and scripts may alter, animate or delete its contents or presentation

b.   Class Attributes:-

The class attribute provides a way of classifying similar elements. This can be used for semantic or presentation purposes. For example, an HTML document might semantically use the designation class="notation" to indicate that all elements with this class value are subordinate to the main text of the document.

c.   Title Attributes:-

The title attribute is used to attach subtextual explanation to an element.

d.   Style Attributes:-

The use of the style attribute to assign presentational properties to a particular element.

e.   Lang Attributes:-

The lang attribute identifies the natural language of the element's contents, which may be different from that of the rest of the document.

METHODOLOGY

I. REPRESENTATION OF AN EMPTY ELEMENT:-

In this kind of techniques it defines that either a start tag followed by an end-tag or an empty element tag. Means we can apply the code like this,

Example:

stego key:

<img></img> ... 0
<img/> ... 1

stego data:

<img src="img l.jpg"></img>

<img src="img5.jpg"></img>

Embedded data:

01110

II. APPEARING ORDER OF AN ELEMENTS:-

We can embed secret data in XML documents by exchanging of the appearing order of elements [5].In given example one bit of data can be hidden in the documents per an exchange of two elements.

Example:

stego key:

<user><name>NAME</name><id>ID</id> </user> ... 0
<user><id>ID</id><name>NAME</name> </user> ... 1

stego data:

<user><name> Karan</name><id>E0001 </id></user>
<user><id>E0002</id><name>Krunal</name></user>

Embedded data:

01

III. APPEARING ORDER OF AN ATTRIBUTES:-

Secret data can be embedded in XML documents by exchanging of the appearing order of attributes in the element [5]. In the following example one bit of data can be hidden per an exchange of the order of attributes.

Example:

stego key:

<event month="MONTH" date="DATE">EVENT</event> ... 0
<event date="DATE" month="MONTH">EVENT</event> ... 1

stego data:

<event month="DEC" date="3">My Birthday</event>
<event date="31" month="DEC">New Year Day</event>

Embedded data:

01

IV. WHITE SPACES IN TAGS:-

Representation of a tag is either including some white spaces before close brackets, or no white spaces [5]. The following is an example method of information hiding by inserting or deleting a space.

Example 2.

stego key:

<tag>, </tag>, or <tag/> ... 0
<tag >, </tag >, or <tag /> ... 1

stego data:

<user ><name>Karan</name ><id >E0001 </id></user>
<user><name >Krunal</name><id>E0002</id ></user >

Embedded data:

101100 010011

V. CHANGE OF SPELLING:-

In his research paper [6] presented a method to exploit same words which case are differently in Uppercase and Lowercase for hiding secret message bits.

Lowercase`	Uppercase
Fulfill	FULFILL
Criticize	CRITICIZE
Favorite	FAVORITE
Center	CENTER

Table 1 : Change case of Spelling[6]

VI. ADD USELESS TAGS:-

This method is use to add some unless tags to hide the data or message.

Example:-

```
<html><body bcolor="gray">
<table><tr>
<td width="200" width="200" align="center" align="center">
<font color="red" size="5" size="5">Test</font></td>
</tr><tr>
<br></br> <rt></rt> <dt></dt> <sub></sub>
<td width="260" align="right" height="50" height="50">
Hiding Information
</td></tr></table>
</body></html>                    secret message: 1001011001011011
```

VII. HIDE DATA USING ATTRIBUTE ORDER:-

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

Example:-

<abbr id="anld" class="aClass" style="color:blue;" Title="Information Hiding">

and

<abbr class="aClass" id="anld" style="color:blue;" Title="Information Hiding">

VIII. HIDE DATA BASED ON ATTRIBUTE PERMUTATION:-

Example:-

The largest embedded capacity (LEC) is the largest Integer which a tag or a webpage can be embedded.

Theorem 1. Let M_T be the LEC of a tag

$$T (a_1, a_2, ... , a_n). \text{ Then}$$

$$M_T = n! \qquad (1)$$

IX. MISCELLANEOUS TECHNIQUES:-

A large number of idiosyncrasies methods have been given by the authors in [6] that may be used for hiding secret message bits inside a cover text e.g., by introducing modification or injecting deliberate grammatical word/sentence errors with in text. Some of the suggested techniques / procedures given in this context include:

(i) *Typographical errors - "tehre" rather than "there".*
(ii) *Using abbreviations / acronyms - "yr" for "your" / "TC" in place of "Take Care".*
(iii) *Transliterations – "gr8" rather than "great".*

PROPOSED ENHANCEMENT

In some previous research papers it is found that the data can be hidden with the help of specific tag. Here Combination of Best two methods can hide large amount of data and increase in LEC(Largest Embedded Capacity) and because of this the content of the source code remain unchanged.
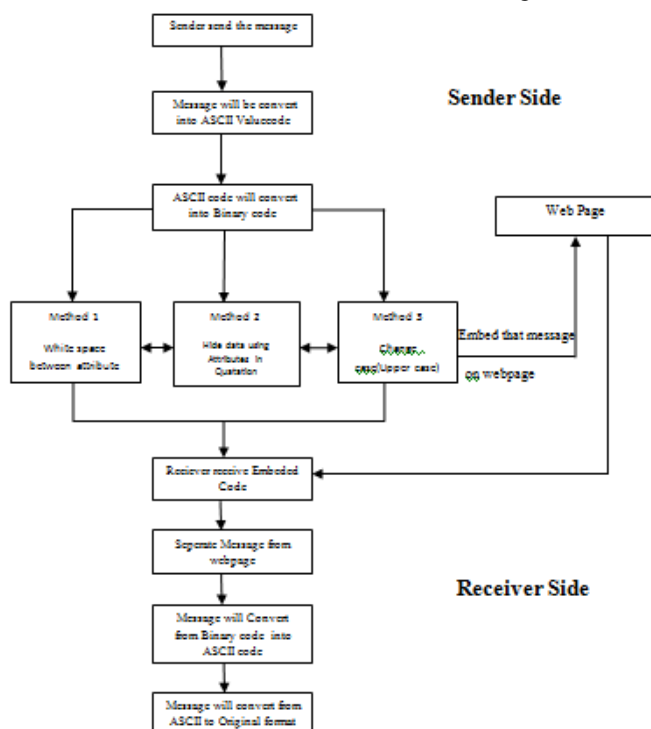

Fig 3. Proposed methods Flowchart

Proposed Algorithm:

Algorithm for embedding the data:

Step1: Input the tag document files that to be used as cover for the secret message.

Step2: Input the text message need to be hidden.

Step3: Read the tag document and analyze the structure of the document.

Step4: Define the key element or method.

Step5: If Input is valid then go to **Step6,** otherwise display the message "Input is wrong" and exit

Step6: Embed Message

 6.1 Get a Collection of tag from cover file

 6.2 For each tag in collection of tags do

  6.2.1 Convert Message into ASCII code Then

  6.2.2 Convet ASCII Code into Binary code

Step7: Save the Cover document as a Stego File and Put the message on the webpage code.

Step 8: End of Embedding Procedure.

Algorithm for Extracting the Data:

Step 1: Input the tag stego file.

Step2: Read the tag document .

Step3: Extract Message

 3.1 Get the message on webpage.

 3.2 Separate the message like cover and data.

 3.3 For each element ,

  3.3.1 Read the message in binary code.

  3.3.2 Convert that message from binary to ASCII value code.

  3.3.3 Convert same ASCII code into Original alphabetic value.

Step4: Display the secret message.

Step5: End of extraction procedure.

Here are the steps to generate algorithm which have to follow and do the process for the Steganography.

We propose a method that will transfer the hidden data or information through web pages and that will be secure way over steganography. This will give good results. For that I have refer all the method which are used before this and how it works properly.

We will compare these methods and find the result. The final methods can be selected by this result, and will be apply on best two method accordingly and improve result.

COMPARISION TABLE

Method	LEC	Increase in Space	Security
White space between attributes	25%	1-2%	less
Hide data using attribute in quotation	25%	0%	less
Change case(upper case) of tag	12.5%	0%	less
Proposed Method	62.5%	1-2%	

Table 2 :Result Analysis

CONCLUSION

In this paper, we have given an overview of web based steganography. It can enhance confidentiality of information and provide a means of communicating privately. All techniques of information hiding using HTML have been shown in this paper. These all techniques will not affect the visual contents of web page. We have proposed new hybrid technique of web based steganography and results show that, largest embedded capacity and security is increasing using our proposed algorithm.

REFERENCES

[1] Xiaojun GUO, Guang CHENG, Chengang ZHU, Aiping ZHOU, Wubin PAN, Dinhtu TRUONG, "Make Your Webpage Carry Abundant Secret Informatin Unawarely" Southeast University, Nanjing 210096, China, vol. A247, pp. 542–548, IEEE 2013.

[2] Chintan Dhanani, Krunal Panchal, "Steganography using web documents as a carrier: A Survey", ISSN 2321-9939,INTERNATIONAL JOURNAL OF ENGINEERING AND RESEARCH, pp.172–179.

[3] Neha Rani, Jyoti Chaudhary, "Text Steganography techniques: A Review", International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 7- july 2013, pp. 3013-3015.

[4] Bachar Alrouh, Adel Almohammad, Gheorghita Ghinea, "Information hiding in SOAP messages: A Steganographic method for web services", International Journal for Information Security Research (IJISR), Volume 1, Issues 1/2, March/June 2011, pp.61-70.

[5] Puneet Kumar Aggarwal, Dharmendra, Parita Jain, Teena Verma, "Adaptive approach for information hiding in www pages", 2014 Internationai Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), ©2014 IEEE, pp.113-118.

[6] Prem Singh, Rajat Chaudhary and Ambika Agarwal, "A Novel Approach of Text Steganography based on null spaces", IOSR Journal of Computer Engineering (IOSRJCE), pp.11-17.

[7] Chintan Dhanani, Krunal Panchal. "HTML Steganography using realative links and multi web page embedment", International journal Of Engineering Development and Research,volume 2 Issue 2. ISSN: 2321-9939, pp.1960-1965.

[8] L.Polak, Z.Kotulski, "SENDING HIDDEN DATA THROUGH WWW PAGES: DETECTION AND PREVENTION", Engineering transactions,Polish academy of sciences, Institute of fundamental technological research, 58 1-2 75-89 2010.

[9] Kapil Kumar Kaswan, Dr. Roshan lal, "Use of steganography in hiding text using css in markup language", volume 3, Issue 11, November 2013, pp 823-827.

[10] Shingo INOUE, Kyoko MAKINO, Ichiro MURASE, Osamu TAKIZAWA, Tsutomu MATSUMOTO, Hiroshi NAKAGAWA, "A Proposal on Information Hiding Methods using XML".

[11] Kapil Kumar Kaswan, Dr. Roshan lal, "Use of steganography in hiding text using css in markup language", volume 3, Issue 11, November 2013, pp 823-827.

[12] Mohit Garg, "A Novel Text steganography technique based on html documents", International Journal of Advanced Science and Technology Vol. 35, October, 2011, pp.129-138.

[13] Aasma Ghani Memon, Sumbul Khawaja and Asadullah Shah, "STEGNOGRAPHY: A NEW HORIZON FOR SAFE COMMUNICATION THROUGH XML", Journal of Theoretical and Applied Information Technology, 2005-2008, pp.187-202.

Sheetal Mehta, Kaveri Dighe , Meera Jagtap, Anju Ekre, "Web based BPCS Steganography", Internationa Journal of Computer Technology and Electronics Engineering(IJCTEE), pp. 126-130.