Data hiding on web using combination of Steganography and Cryptography

Lipi Kothari
Computer Engineering
Silver Oak College of Engineering & Technology
Ahmedabad, India
lipi.kothari123@gmail.com

Rikin Thakkar
Assistant Professor
Computer Engineering
Silver Oak College of Engineering & Technology
Ahmedabad, India
rikinthakkar.it@socet.edu.in

Satvik Khara
Head of Department
Computer Engineering
Silver Oak College of Engineering & Technology
Ahmedabad,India
satvik@socet.edu.in

***Abstract*—Today's world is digital era, everyone looks for information on Web. The web is not only space for information, but most importantly, it is a tool to connect people. People used to share information and transfer confidential data on the Web. Since Internet is publicly available securing data on Web is much important, some techniques are needed to hide this data. There are different techniques available to hide the data, for example Steganography, cryptography and so forth. The benefit of steganography over cryptography is that no one except the sender and receiver can see the message. This paper concentrates on different steganography techniques to hide the data on Web. One more advantage of using steganography to hide data on Web is that it doesn't look suspicious. This research paper gives another skyline to safe correspondence through information hiding on Internet. The experimental results show that the proposed method has high security, larger embedding capacity and best Imperceptibility than others.**

Keywords*—Steganography, Web based Steganography, data hiding on web, Steganography + Cryptography

I. Introduction

Information hiding is to hide some secret information in cover objects, such as an image, audios, videos, texts, etc. [1]. Cryptography and Steganography are basic methods which will help us to secure data from unauthorized access. Steganography is one of the best techniques to hide messages from unauthorized audience. In cryptography the unofficial user can decrypt the encrypted message if he has key while in steganography unofficial person can't view the message because the message is hidden cover media and user will not have any idea about algorithm or method.

Steganography is derived from the Greek words. "stegos" is termed as "roof or covered" and "graphy" as "writing or drawing"[1,5]. Thus Steganography is the hidden writing or secret writing. With the help of this, a secret message can be set inside a piece of trustful information and can be sent without anyone being aware of the secret message.

In Steganography unauthorized person can't able to view the message because the message is hidden in a carrier and travels through the carrier. The carrier of the message can be plaintext , audio, images, video, web etc.

II. STEGANOGRAPHY MODEL

Steganography model is explained in the fig. 1[4]. In the first step original message embed in the carrier using any embedding technique. After that, embedded message travels through the transmission media. The receiver decodes the message at the receiver end, which is the reverse process of embedding and gets the original message.

Carrier is one type of cover object which hides message in it. The types of information carriers used in steganography are audio, video, text, image and web.

Message is nothing but confidential and private data.

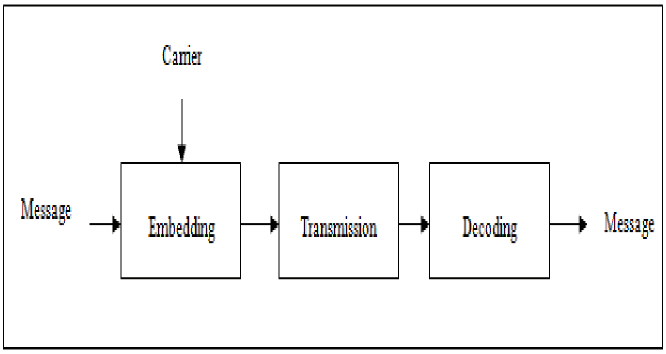


Fig. 1. Model of steganography[4]

A. *Steganography vs Cryptography*

Difference between technologies is shown in table I.

Steganography	Cryptography
Message passing is unknown	Message passing is known
Less known Technology	Known Technology
Technology is being developed in certain areas	Most algorithms are known to all
Once it is detected, the message is known	Strong algorithm is currently resistant to attack

It covers the existence of communication	Encryption prevents an unofficial party from discovering the contents of the communication
The secret message is indistinguishable to anyone	A person can easily detect and modify the encrypted message

Table I. Comparison of Steganography and Cryptography [1, 4,5]

B. *Literature Survey*

In paper[1], Authors have used one data hiding method named “change order of elements” to hide and extract data. They have used only one method to hide data and they have not encrypted plain text so this method is not that much secure.
In

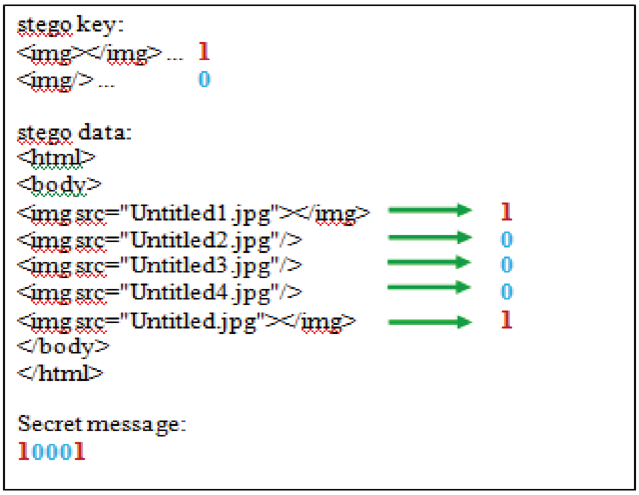
III. Data Hiding On Web Using Steganography Techniques

There are various methods to hide data inside the source code on Web pages. These techniques are based on tags.

A. *Representation of empty elements*

An empty element can be represented either by or an empty-element tag or by open- tag immediately followed by close-tag [1]. By swapping these tags, data can be embedded maintaining the originality of the document. In the following illustration the method of message hiding is shown by transformation of image tag. One bit of data is set per close-tag of empty elements.

Illustration:

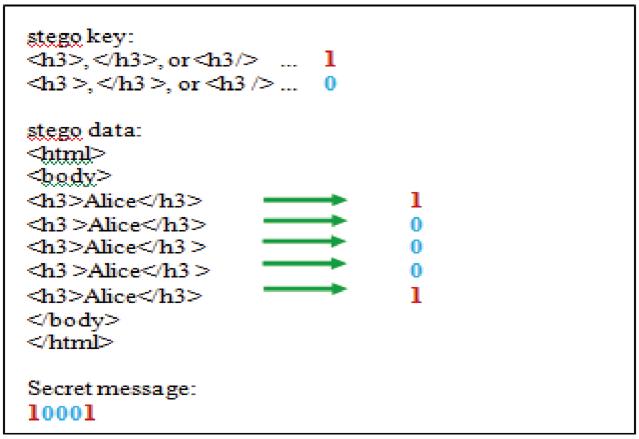


```
stego key:
<img></img> ...  1
<img/> ...       0

stego data:
<html>
<body>
<img src="Untitled1.jpg"></img>  →  1
<img src="Untitled2.jpg"/>       →  0
<img src="Untitled3.jpg"/>       →  0
<img src="Untitled4.jpg"/>       →  0
<img src="Untitled.jpg"></img>   →  1
</body>
</html>

Secret message:
10001
```

B. *White spaces in tags*

A tag can be represented either by placing white spaces before brackets are closed, or no white spaces. By placing or removing spaces, data can be set conserving originality of the document.[1,4].
Here a message is hidden by placing or removing a space. one bit of data embeds per tag.

Illustration:

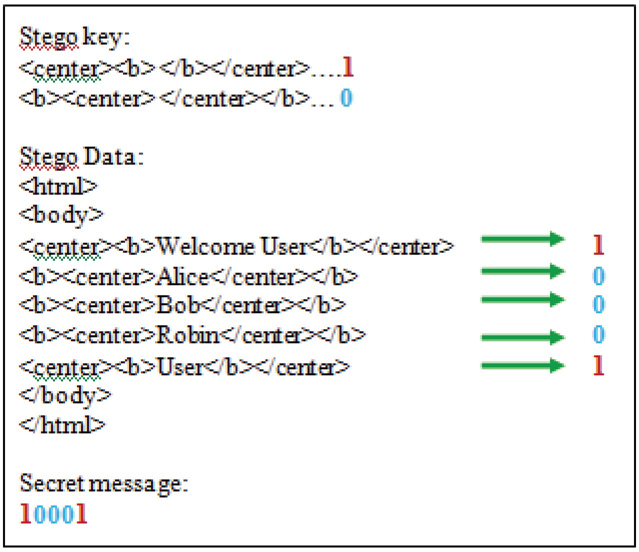


```
stego key:
<h3>, </h3>, or <h3/>  ...     1
<h3 >, </h3 >, or <h3 /> ...   0

stego data:
<html>
<body>
<h3>Alice</h3>     →   1
<h3 >Alice</h3>    →   0
<h3>Alice</h3 >    →   0
<h3 >Alice</h3 >   →   0
<h3>Alice</h3>     →   1
</body>
</html>

Secret message:
10001
```

C. *Appearing order of the elements*

A secret message can also be embedded by interchanging order of elements. Here one bit of data is hided in the page by interchanging two elements [1].

Illustration:

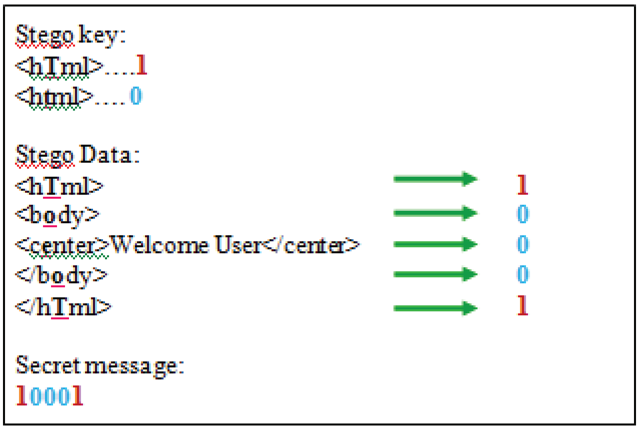


```
Stego key:
<center><b> </b></center>....1
<b><center> </center></b>... 0

Stego Data:
<html>
<body>
<center><b>Welcome User</b></center>   →   1
<b><center>Alice</center></b>          →   0
<b><center>Bob</center></b>            →   0
<b><center>Robin</center></b>          →   0
<center><b>User</b></center>           →   1
</body>
</html>

Secret message:
10001
```

D. *Change case of Letters in tags*

HTML tags are case insensitive, hence we can take its advantage to hide a message within a document by changing the case of specific letters in a tag’s name. For example, <BR>, <br>, <Br> and <bR> means exactly the same.Big capacity is the main advantage of this method. But the pitfall of this method is it very easy to discover the secret message since it is very uncommon to use small and capital letters alternative.

Illustration:

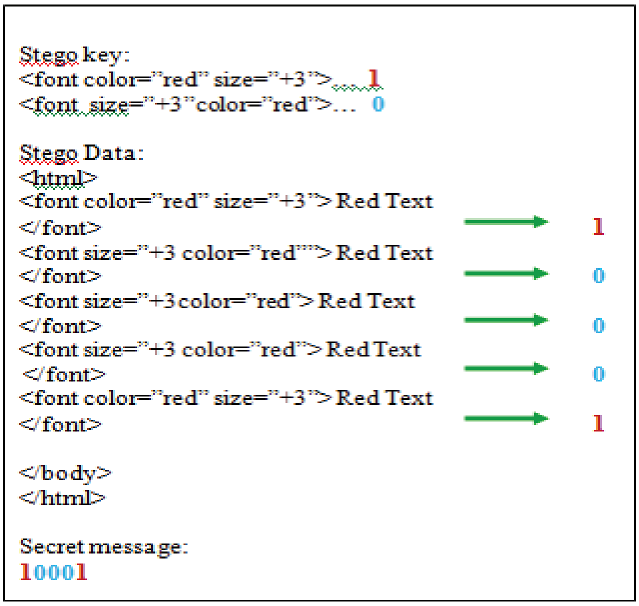


```
Stego key:
<hTml>....1
<html>.... 0

Stego Data:
<hTml>                              →    1
<body>                              →    0
<center>Welcome User</center>       →    0
</body>                             →    0
</hTml>                             →    1

Secret message:
10001
```

E. *Appearing order of the attributes*

A secret data can also be hidden by swapping the order of attributes in the element. In this illustration, one bit of data is covered per by interchanging order of attributes [1,6].

Illustration:

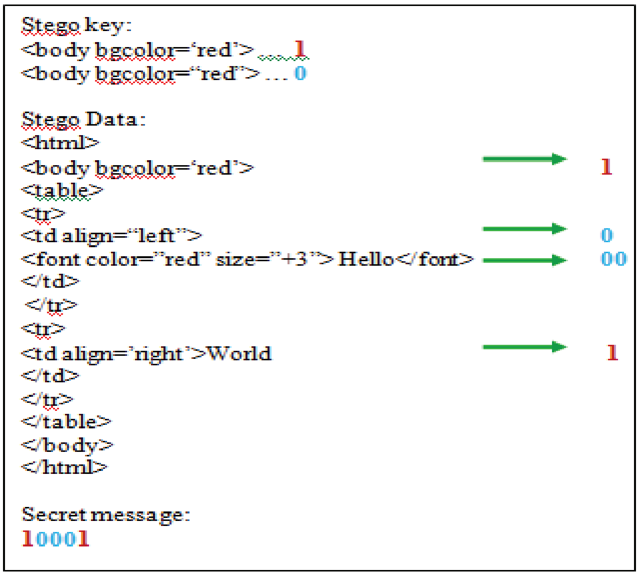


```
Stego key:
<font color="red" size="+3">... 1
<font size="+3"color="red">... 0

Stego Data:
<html>
<font color="red" size="+3"> Red Text
</font>                                  →    1
<font size="+3 color="red""> Red Text
</font>                                  →    0
<font size="+3color="red"> Red Text
</font>                                  →    0
<font size="+3 color="red"> Red Text
 </font>                                 →    0
<font color="red" size="+3"> Red Text
</font>                                  →    1

</body>
</html>

Secret message:
10001
```

F. *Change quotation marks of attribute values in tags*

Attribute values can be enclosed with single inverted comma, double inverted comma or without commas. It does not affect the output of the HTML page [5].

Illustration:

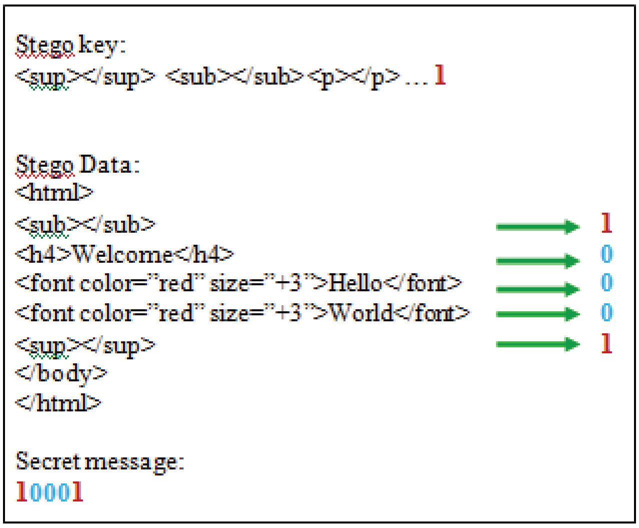


```
Stego key:
<body bgcolor='red'> ... 1
<body bgcolor="red"> ... 0

Stego Data:
<html>
<body bgcolor='red'>                                   →    1
<table>
<tr>
<td align="left">                                      →    0
<font color="red" size="+3"> Hello</font>             →    00
</td>
 </tr>
<tr>
<td align='right'>World                                →    1
</td>
</tr>
</table>
</body>
</html>

Secret message:
10001
```

G. *Add useless tags*

Secret messages can be embedded further by inserting useless tags intermediary in the Html documents. Here in this illustration, one bit of data is covered per number of useless tags in the document.

Illustration:



```
Stego key:
<sup></sup> <sub></sub> <p></p> ... 1


Stego Data:
<html>
<sub></sub>                                   →    1
<h4>Welcome</h4>                              →    0
<font color="red" size="+3">Hello</font>      →    0
<font color="red" size="+3">World</font>      →    0
<sup></sup>                                   →    1
</body>
</html>

Secret message:
10001
```

IV. PROPOSED WORK

A. Flow Diagram Of Proposed Work

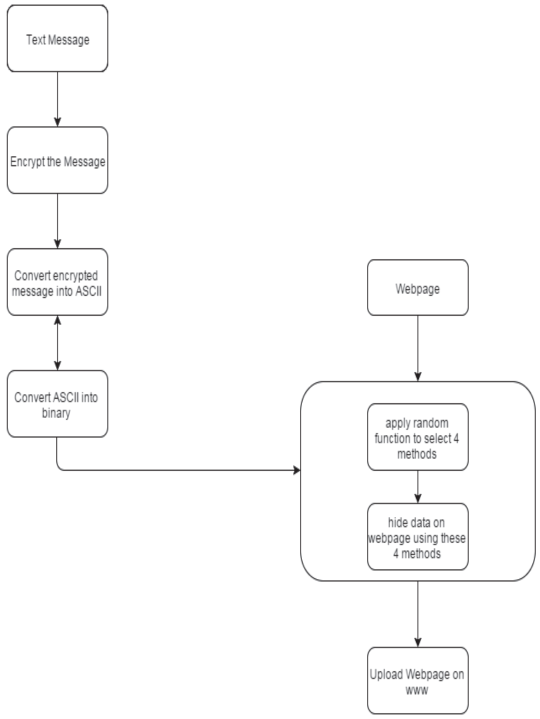

Text Message
Encrypt the Message
Convert encrypted message into ASCII
Convert ASCII into binary
Webpage
apply random function to select 4 methods
hide data on webpage using these 4 methods
Upload Webpage on www

Fig. 2. Flowchart For Embedding Process

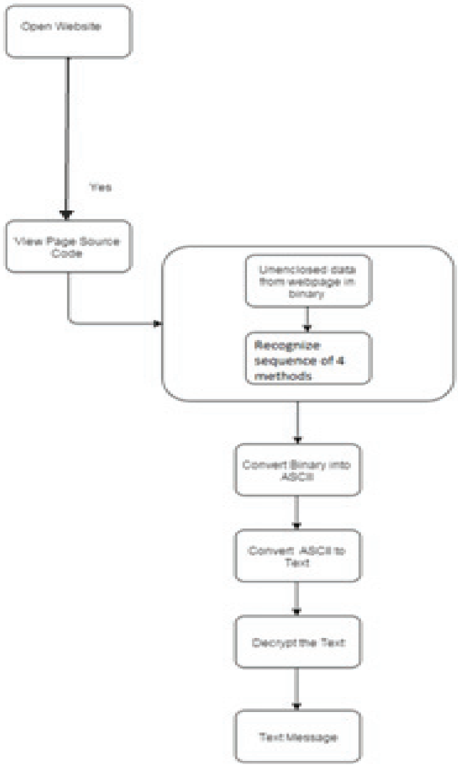

Open Website
Yes
View Page Source Code
Unenclosed data from webpage in binary
Recognize sequence of 4 methods
Convert Binary into ASCII
Convert ASCII to Text
Decrypt the Text
Text Message

Fig. 3. Flowchart For Embedding Process

B. Algorithm of Proposed Work

Figure 2,3 shows our proposed algorithm. Our proposed algorithm consists of 2 steps. 1st step is embedding process in which plain text is embedded in webpage. 2nd step is reverse process of 1st step in which plain text is extracted from webpage. Details of these 2 steps is explain below from the figure 2 and 3.

a) Algorithm for embedding process

Step 1. Take original message (called plain text) that you want to hide and encrypt the original message using encryption algorithm using key k.

Step 2. Convert the encrypted message into ASCII code.

Step 3. After that this ASCII code is converted into binary i.e. in 0's and 1's form.

Step 4. Take a web page on which yu want to hide data. Select 4 random methods of hiding data and hide that converted binary data in a webpage.

Step 5. Now your webpage is ready in which you have hided your data. Now make this webpage online so receiver can extract data from this webpage.

b) Algorithm for extracting process

Step 1. Open the website in a browser and check the url.

Step 2. View its source code.

Step 3. Fetch the binary data from the webpage.

Step 4. Recognize the random methods of steganography techniques being applied to the data on webpage.

Step 5. Convert the binary data into ASCII and this ASCII is converted into text.

Step 6. Original message will be enhanced by decrypting the text using key k.

V. EXPERIMENTAL RESULT

Table II. Experimental Result of Steganography

Techniques	Imperceptibility	Change in file size	Security	LEC
Change case of letters in tags	Weak	No	Weak	100 %
By using white space	Good	Yes (minor)	Yes	100 %
Appearing order of the attributes	Good	No	Strong	5%
Change of quotation marks in attribute values of tags	Medium	Yes (minor)	medium	80%
Proposed Method	Very Good	Yes (minor)	Very Strong	285 %

Experimental result table shows the comparision of 4 tradidional data hiding methods on web with our proposed algorithm. Results shows that imperceptibility of our algorithm is better than other traditional algorithms. Security is also better than other traditional algorithms. As we are combining 4 methods in our proposed algorithm, LEC is also very good. Only one disadvantage of this algorithm is that, it increases the page size more than other algorithms. But overall performance of algorithm is very good.

VI. CONCLUSION

HTML steganography is new period of concealing information and it gives more attainability to shroud information on the grounds that there is tremendous number of pages accessible on the web and information taken cover behind HTML pages is less suspicious. We reviewed different steganography methods that utilization the html labels and ascribes to conceal the mystery message. These strategies connected on line of source code in HTML website page archives such a path, to the point that it won't influence the first substance of source code. With the assistance of quality based proposed method, our point is to enhance the Largest embedded Capacity (LEC) of spread page with keeping great intangibility.

References

[1] Puneet Kumar Aggarwal, Dharmendra, Parita Jain, Teena Verma, "Adaptive approach for Information Hiding in WWW Pages",IEEE-2014.

[2] Xiaojun GUO, Guang CHENG, Chengang ZHU, Aiping ZHOU, Wubin PAN, Dinhtu TRUONG "Make Your Webpage Carry Abundant Secret Informatin Unawarely",IEEE-2013

[3] Chintan Dhanani, Krunal Panchal "Steganography using web documents as a carrier:A Survey",IJEDR-2013

[4] Chintan Dhanani, Krunal Panchal, "HTML Steganography using Relative links &Multi web-page Embedment",IJEDR-2014

[5] Yung-Chen Chou, Hsin-Chi Liao "A Webpage Data Hiding Method by Using Tag and CSS Attribute Setting",IEEE-2014

[6] L.Polak,Z.Kotulski "Sending Hidden Data Through Www Pages:Detection And Prevention",IFTR-2010

[7] Kapil Kumar Kaswan, Dr. Roshan Lal " Use of Stegnography in Hiding Text Using CSS in Markup Language",IJARCSSE-2013

[8] Mengmeng Wang, Guiliang Zhu, Xiaoqiang Zhang "General Survey on Massive Data Encryption"

[9] Haythem Hayouni1, Mohamed Hamdi1, Tai-Hoon Kim "A Survey on Encryption Schemes in Wireless Sensor Networks",IEEE-2014

[10] M. Thangavellll, P. Varalakshmi "A Survey On Security Over Data Outsourcing"l,IEEE-2014

[11] Babita Ahuja,Anuradha, Dimple Juneja "Dynamic Query Processing for Hidden Web Data Extraction",IEEE-2015