Text Steganography Techniques: A Review

Neha Rani[#1], Jyoti Chaudhary[#2]

[#1]*M.Tech Scholar, CSE Department, M.D.U. Rohtak, Haryana, India*
[#2]*Assistant Professor, CSE Department, M.D.U. Rohtak, Haryana, India*

Abstract— **Steganography has been in use since the presence of secret messages and this use has gained popularity as the internet became well-liked. Text steganography is a sub part of steganography that hides the message behind other cover text file. Moreover, hiding the text behind HTML coding of web pages makes the detection of steganography impractical as web pages are a fundamental building blocks of the internet. This paper reviews some of most common text steganography techniques that mainly focus on protecting text behind HTML and CSS (Cascading Style sheets) coding of web pages. This paper also proposes a new technique based on hiding text behing HTML and CSS portion of web page coding.**

Keywords—**Steganography, Text, HTML, CSS.**

I. INTRODUCTION

There has always been a need of protecting the messages that are sensitive in nature. Such messages if exposed to some intruder may pose a threat to nation's security or company's critical decisions. Thus, such information must be secured at any cost and to serve the purpose there has been a trend to encrypt or hide the secret information. Cryptography (derived from Greek work 'kryptos' meaning hidden and 'graphein' meaning to write) is used to encode the text to make it understandable. Steganography (composed of Greek word 'steganos', meaning covered and 'graphein' meaning to write) on the other hand, is used to hide the text behind some other media. Cryptography may draw the suspicion of the intruder towards the text that is in encoded format. Steganography donot lures the eavesdropper as it hides the message. Steganography can be classified based on the type of media it uses to hide the text[1]
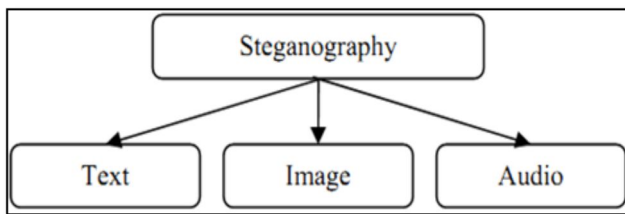
These are as follows:
1) *Text Steganography:* It hides the text behind some other text file. It is the a difficult form of steganography as the redundant amount of text to hide the secret message is scarce in text files.

2) *Image Steganography*: It is one of the most commonly used technique because of the limitation of the Human visual System(HVS). Human eye cannot detect the vast range of colors and an insignificant change in the quality of an image that results from steganography.

3) *Audio steganography*: It is also a difficult form of steganography as humans are able to detect a minute change in the quality of audio.

A. Text Steganography Techniques

1) *Selective hiding:* This hides the characters in the first (or any specific location) characters of the words. Concatenating those characters help extracting the text. But this technique requires huge amount of plain text.

2) *HTML web pages*: This may hide text using the fact that attributes of HTML tags are case insensitive. Those characters can then be used to retrieve the original text.

3) *Hiding using Whitespace*: Fewer number of whitespaces may specify a 0 and more number of whitespaces between words may determine a 1.

4) *Semantic Hiding*: Uses synonyms to hide the message.

Text steganography can use HTML pages to hide the text behind them because of two factors:

1) Web pages are present in a vast amount and detecting which one is containing hidden information is next to impossible.
2) The order of tags used for formatting the appearance of a web page does not matter and this can help to hide one bit of text behind the tags.

Thus this paper reviews some of the common text steganography techniques. The rest of the paper is organized as follows. Section II to IV reviews some text steganography techniques. Section V concludes the paper.
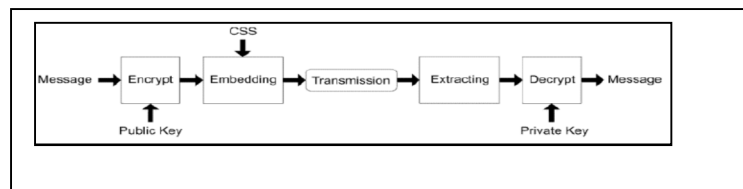
II. A NOVEL TEXT STEGANOGRAPHY BASED ON HTML DOCUMENTS

Text steganography is most challenging due to the presence of very less redundant information in text documents as compared to the images and audio[3].HTML Tags and associated members are case insensitive e.g., <html>, <HTML> or <hTmL> will have the same impact on the document"s outlook. Bits are hidden in TAGS by changing the caseof the alphabets based on the bit as either 0 or 1 [5][6]. This paper hides the text in HTML documents. HTML tags have a property that ordering of attributes does not affect the appearance and look of web page. Initially the text is encrypted by playfair cipher. The resulting text is converted to binary form before hiding it behind HTML tags. Then, this technique encompasses three key steps:

1) Key file generation: This step generates the key file that is composed of key combinations. These combinations are arranged in the form of rows and columns. These combinations consist of prime and secondary attributes. Ordering of prime and secondary attributes can hide one bit of data depending upon their ordering. The pair of attributes that is frequently used is selected as the key combination. Any attribute of the combination can be prime provided that it follows the uniqueness constraint. Prime followed by secondary means bit 1 else 0[2].

First Attribute (Primary)	Second Attribute (Secondary)

2) Hiding the message: Scan the HTML document to determine key combinations. If an attribute matches the prime attribute and corresponding secondary attribute is available in the tag, then only this can hide a bit. To hide a 1, prime attribute must precede the secondary one else the reverse must happen[2].

3) Extracting the message: It follows just the reverse process. For every combination of prime and secondary attributes determine the order. If prime follows secondary then 0 has been hidden behind them else 1[2].

III. INFORMATION HIDING IN CSS: A SECURE SCHEME TEXT-STEGANOGRAPHY USING PUBLIC KEY CRYPTOSYSTEM

This approach hides the text in the CSS portion of a web page using End of Line(EOL) spacing. Two parties are involved in the exchange . Receiver has the job of generating the pair of public and private keys using RSA encryption scheme[1].The publicly available key is transmitted to the other end, i.e, sender. Sender encrypts the text using the public key. The encoded copy is converted to the binary form to be masked in the web page. The text steganography scheme finds the semicolon for hiding the encoded text.

Whitespaces are used to hide the text. After the semicolon, a whitespace denotes a 0 bit and a tab after the semicolon denotes the 1 bit.

## IV. A NOVEL APPROACH OF TEXT STEGANOGRAPHY BASED ON NULL SPACES

A. *Block Diagram of the Model*

A new model is proposed using white/null space technique for hiding the secret message in text steganography.
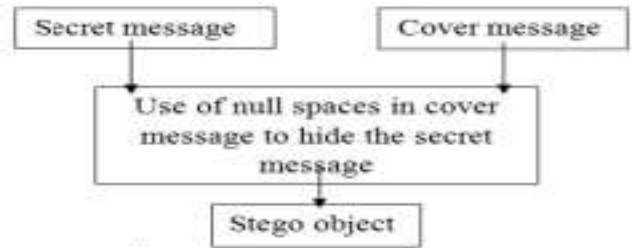
B. *Working of the model*

Information hiding within spaces seems to be possible as people scarcely can know about the existence of the hidden bits.This method was applied in the space which appears between the words. In this approach single space is interpreted as "0" whereas two spaces are interpreted as "1". The disadvantage of this technique is that it requires a great deal of space to encode few bits. For example, a character is equivalent of 8 bits, and it requires approximately 8 inter-spaces to encode one character[4].
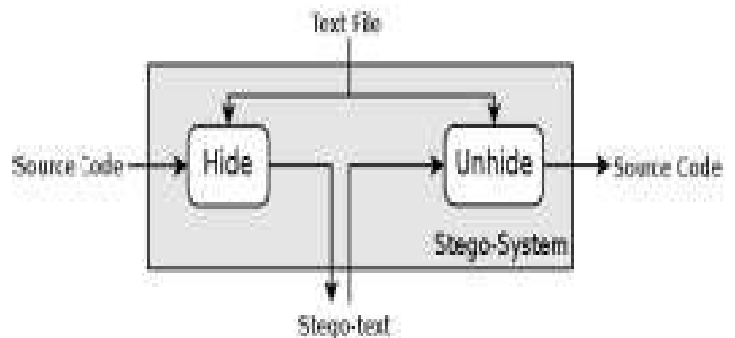

For hiding the secret message in the cover message we present a new approach on text steganography through the white/null space technique.

C. *STEPS IN THIS TECHNIQUE*

This technique consist of two programs mainly. One of them is hiding program which is the program of converting text that is hiding data in cover message. Other one is Extractor program which is the program of extracting data from the stego text (text containing hidden data).

1) *Hiding program:* First we have to enter the secret message that we want to hide. Now, Enter the name of the file as a cover text in which we want to hide our secret message.

Then, generate a covered text file as a stego-text in which our secret message is hidden[1].
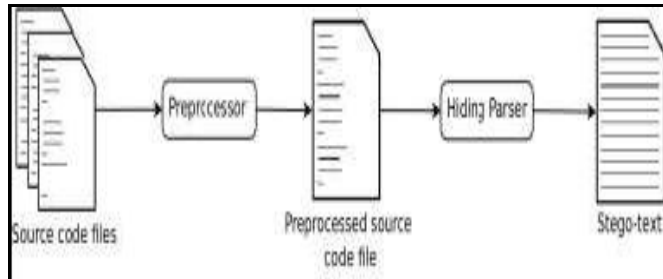
2) *Unhiding program:* First we have to enter the name of the text file (stego-text) that generate after the hiding of the secret message. Now, Extract the secret message.
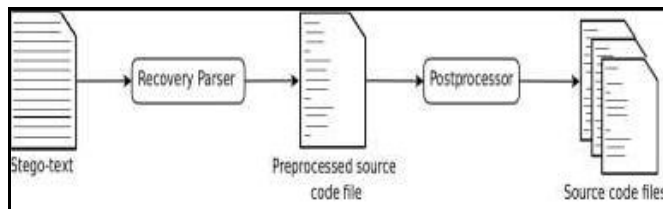
V. CONCLUSIONS

This paper reviews some text steganography methods that hide text behind HTML coding of web pages. This technique is very effective as web pages are building blocks of internet and they are present in vast amount. So, suspecting the presence of hidden text in web pages is next to impossible. Security can further be enhanced by hiding text in CSS coding of pages as Cascading Style sheets are used extensively for formatting and controlling the appearance of web pages. Furthermore, CSS coding cannot be seen at the client side thus adding to the security.

REFERENCES

[1] H.Kabetta, B.Y. Dwiandiyanta, Suyoto, "Information hiding in CSS: A secure scheme text-stegangraphy using public key Cryptosystem",IJCIS,pp. 13-22 ,Vol.1, No.1,December 2011.
[2] Mohit Garg," A Novel Text Steganography Technique Based on Html Documents",International Journal of Advanced Science and Technology,pp.132-138,Vol. 35, October, 2011.
[3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", pp. 313-336, IBM Systems Journal, vol. 35, Issues 3&4, 1996.
[4] P.Singh, R.chaudhary and A.Agarwal," A Novel Approach of Text Steganography based on null spaces", IOSRJCE, PP 11-17,Volume 3, Issue 4 (July-Aug. 2012).
[5] M.S. Shahreza," A New Method for Steganography in HTML Files", Computer, Information, and Systems Sciences, and Engineering, ProceedingsIETA 2005, TeNe 2005, EIAE 2005, 247-251, Springer.
[6] K. Bennett, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", Purdue University, CERIAS Tech. Report 2004-13.