# A Novel Approach For Data Hiding In Web Page Steganography Using Encryption With Compression Based Technique

Ram Krishna Singh[1], Bhavya Alankar[2]

[1](Research Scholar, Computer Science, Jamia Hamdard, India)
[2](Assistant Professor, Computer Science, Jamia Hamdard, India)

***Abstract:*** *Cryptography, Steganography and Watermarking are three rudimentary techniques which will avail us to secure data from unauthorized access. Steganography is one of the best techniques to obnubilate message from unauthorized users. It is utilized to obnubilate data through sundry medium like audio, video, Text and image. In our research work, we fixate on text Steganography through html web pages to obnubilate data behind it. Because html pages are facilely transferable that contains secret data behind it. There are sundry methods of web pages predicated Steganography. Some of the mundane methods are white spaces, order of attributes, change case letter, change quotation marks, reiterate attributes to obnubilate data etc. But these all methods have some inhibition like iota of data can be obnubilated behind html page & do not have most sizably voluminous embedding capacity etc. To surmount or enhance these circumscriptions we optate to apply any compression technique afore embedding process so that it can obnubilate substantial amount of data and secondly security is much more consequential thus we have to apply any encryption techniques so that data cannot transferable in its pristine form. Thus in our research work we fixates on embedding techniques, compression and encryption techniques. So that system will be more power full as well as secure.*

***Keywords:*** *Compression Ratio, Embedding Technique, Encryption, Huffman Algorithm, Steganography.*

## I. Introduction

With the rapid development of internet era, the information plays a consequential role. The information is sent and received from one place to another. So the main aspect is to secure that information. Thus sundry techniques including Steganography, cryptography etc. is utilized to secure information over the cyber world. Steganography is a technique for securing information by obnubilating it in some other medium, cryptography is the science of converting the messages that are intended to be secret into some other form, such that it is not understandable to anyone other than the intended sender and recipients. Steganography is a Greek word which betokens concealed inditing or obnubilated inditing. Steganography is an art and science of encoding obnubilated messages in such a way that no one accepts the sender and recipient ken about the subsistence of the data. In Steganography an unauthorized person cannot able to optically discern the message because message is obnubilated in carrier and peregrinate through the carrier. The carrier of the message may be plain text, image, audio, video etc. Text predicated Steganography techniques can be applied on webpage text and plaintext. WWW pages have become the main way to provide the information for users. A webpage text contains HTML, CSS, XML, and JavaScript as content.

The bearer record postulates an imperative part in outlining Steganography calculations. Picture, sound, video, and content are distinctive media utilized as often as possible over the Internet [13]. Each of these conveyor documents sorts has certain attributes that empower the client to embed the information inside. Picture documents are the most broadly utilized records as bearer documents which contain high proportion of information recurrence [14]. Then again, it is arduous to utilize the same picture to stow away distinctive messages, since contrasting comparative pictures may sanction assailants with unearth the covered information. Sound documents are verbalized with as sine or cosine waves. A few methods propose to move the stage to conceal zeros and ones [15]. Content documents verbalize with the most onerous conveyor records, since content documents contain minimal perpetual information contrasted with different bearers [16].

Content Steganography is grouped into sundry classes. A standout amongst the most prominent content Steganography strategies is semantic Steganography [17]. This procedure makes utilization of equipollent words in the same dialect or comparative dialects, for example, American English and British English. This is culminated by making a lexicon of equipollent words and trading words to pass zero or one. Different relegations conceal information relying upon the dialect sentence structure. This is kenned as the most punctual procedures that utilize the physical configuration of content to obnubilate data. Different situations utilize phonetic properties to shroud information and rely on upon the document era to pass on the data [18].

## II. Types Of Steganography

There are many types of Steganography methods. Fig. 1.2 below shows four main categories that can be utilized for Steganography techniques viz text, audio, video and image Steganography.

1) Text Steganography: It obnubilates the text abaft some other text file or HTML Page. It is an arduous form of Steganography as the redundant amount of text to obnubilate the secret message is scarce in text files.

2) Image Steganography: It is one of the most commonly used techniques because of the constraint of      the Human visual System (HVS). Human ocular perceiver cannot detect the astronomical range of colors and a nugatory vicissitude in the quality of an image that results from Steganography.

3) Audio Steganography: It is additionally an arduous form of Steganography as humans are able to detect a minute transmutation in the quality of audio.
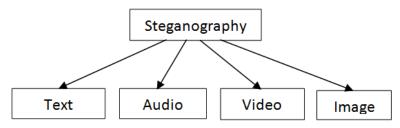


**Fig 1:** Types of Steganography

## III. Motivation For Web Steganography

As the information is transmitted over the cyber world so security of information is most paramount. The web page is most consequential carrier for travelling the information because web pages are facilely transferring from one place to another place no one can visually perceive the information. Albeit there are many techniques for transfer the information like image, audio, and video etc. But html predicated Steganography is facile and secure technique

## IV. Framework Of The Proposed Technique

The proposed technique essentially has the following schema. It depicts the fundamental flow of the both the obnubilating and extracting process of the proposed technique. Here we used adaptive Huffman coding for the compression. An adaptive Huffman algorithm withal called dynamic Huffman coding. A variation called adaptive Huffman coding involves calculating the probabilities dynamically predicated on recent genuine frequencies in the sequence of source symbols, and transmuting the coding tree structure to match the updated probability estimates. It is utilized infrequently in practice, since the cost of updating the tree makes it more gradual than optimized. Adaptive arithmetic coding that is more flexible and has a better compression.
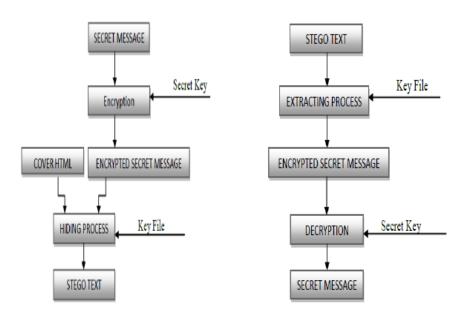


**Fig 2:** proposed framework.

# V. Experimental Results

The simulated work is performed on Matlab on Intel core i3 processor.

reading data for encryption...

data to be hide:

*************************************************************************

 PREETI

*************************************************************************

*************************************************************************

Reading HTML page for encryption...

Data to be hide after huffman compression:

*************************************************************************

huff_data =00001111010001

Original message length: 48bits

Compressed message: length: 14bits

*************************************************************************

<------------------------------------------------------------------------------------------>

Encrypting data in html....

*************************************************************************

hiding "0" in string

orignal string

<img src="./RKGIT-College, Ghaziabad_files/logo.jpg" width="498" height="65px" border="0" alt="RKGIT Logo"> encpted string

<img src="./RKGIT-College, Ghaziabad_files/logo.jpg" width="498" height="65px" border="0" alt="RKGIT Logo">

*************************************************************************

hiding "0" in string

orignal string

<img src="./RKGIT-College, Ghaziabad_files/sp.png" alt="" width="12" height="30" border="0">

encpted string

<img src="./RKGIT-College, Ghaziabad_files/sp.png" alt="" width="12" height="30" border="0">

*************************************************************************

hiding "0" in string

orignal string

<img src="./RKGIT-College, Ghaziabad_files/sp.png" alt="" width="12" height="30" border="0">

encpted string

<img src="./RKGIT-College, Ghaziabad_files/sp.png" alt="" width="12" height="30" border="0">

*************************************************************************

hiding "0" in string

orignal string

<img src="./RKGIT-College, Ghaziabad_files/sp.png" alt="" width="12" height="30" border="0">

encpted string

<img src="./RKGIT-College, Ghaziabad_files/sp.png" alt="" width="12" height="30" border="0">

*************************************************************************

hiding "1" in string

orignal string

<img src="./RKGIT-College, Ghaziabad_files/CCT_LOGO.png" width="180" height="120" border="0" alt="Center for Cloud Technologies (CCT)" title="Center for Cloud Technologies (CCT)">

encpted string

<img src="./RKGIT-College, Ghaziabad_files/CCT_LOGO.png" height="120" width="180" border="0" alt="Center for Cloud Technologies (CCT)" title="Center for Cloud Technologies (CCT)">

*************************************************************************

hiding "1" in string

orignal string

<img src="./RKGIT-College, Ghaziabad_files/journal1.png" width="210" height="56" border="0">

encpted string

<img src="./RKGIT-College, Ghaziabad_files/journal1.png" height="56" width="210" border="0">

Decompression:

reading HTML page for encryption...

*******************************************

decrypting data from html....
00001111010000
************************************
decompression of decypted data...

Msg decoded:

PREETI

Performance Parameter:
- encypt coputational time:1.4895sec
- CR (compression ratio):29.1667%
- decypt coputational time:0.17026sec
- BER (bit error rate):0.0. (10)

## VI. Conclusion

Web page Steganography is new era of hiding data and it gives more feasibility to hide data because there is large number of pages available on the internet & data hidden behind web pages is less suspicious. In our work, we proposed an enhanced cryptosystem which may be more secure as well as having the largest embedding capacity. Because information security is most important so to do this we use one compression techniques along with one encryption techniques. So that large amount of data can be handling easily and transferable behind web pages easily and securely. But now in era of internet web pages are more transferable or sharable easily between two users, no one can access these secret data. So web page steganography might be a good carrier of transmission of secure data. For the text Steganography various methods have been proposed. In this paper, we propose a novel approach of text Steganography that uses the html tags and attributes to hide the secret messages. The basic idea of the proposed technique is to hide the messages by changing the order of attributes as the ordering of attributes does not affect the appearance of the html documents. The html documents are fundamental elements of the web.

These documents are used very commonly on the internet and hence are less prone to arouse suspicion in the intruder of the existence of the secret message. Moreover, any html document has a considerable number of tags and attributes. Thus the capacity of the hiding process to hide secret messages is also high in the proposed technique.

## VII. Future Scope

Future work will include investigating the application and efficiency of these methods, and the methods utilizing HTML document. Here we can apply arithmetic coding for compression as well as DES or RSA encryption technique to amend the result. This is the future enhancements of the work.

## References

**Journal Papers:**
[1] Petitcolas, Fabien AP, Ross J. Anderson, and Markus G. Kuhn. "Information hiding-a survey." *Proceedings of the IEEE* 87.7 (1999): 1062-1078.
[2] Dhanani, Chintan. "HTML Steganography using Relative links & Multi web-page Embedment." *International Journal of Engineering Development and Research*. Vol. 2. No. 2 (June 2014). IJEDR, 2014.
[3] Dhawase, Dhammjyoti V., and Sachin Chavan. "WEBPAGE INFORMATION HIDING USING PAGE CONTENTS."
[4] Yang, Yujun, and Yimei Yang. "An efficient webpage information hiding method based on tag attributes." *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*. Vol. 3. IEEE, 2010.
[5] Huang, Huajun, Shaohong Zhong, and Xingming Sun. "Steganalysis of information hidden in webpage based on higher-order statistics." *Electronic Commerce and Security, 2008 International Symposium on*. IEEE, 2008.
[6] Odeh, Ammar, et al. "Novel Steganography over HTML Code." *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*. Springer International Publishing, 2015. 607-611.
[7] Ren, Junling, and Li Zhang. "Webpage Information Hiding Algorithm Based on Integration of Tags and Data." *ICICA (LNCS)*. 2012
[8] Shen, Dongsheng, and Hong Zhao. "A novel scheme of webpage information hiding based on attributes." *Information Theory and Information Security (ICITIS), 2010 IEEE International Conference on*. IEEE, 2010.
[9] Yong, Xie, Li Juan, and Zhang Yilai. "A High Capacity Information Hiding Method for Webpage Based on Tag." *Digital Manufacturing and Automation (ICDMA), 2012 Third International Conference on*. IEEE, 2012.
[10] Al-Hamami, A. H., M. S. Al-Hakeem, and M. A. Al-Hamami. "A proposed method to hide text inside HTML web page file." (2009).
[11] Mir, Nighat, and Sayed Afaq Hussain. "Secure web-based communication."*Procedia Computer Science* 3 (2011): 556-562.
[12] Inoue, Shingo, et al. "A proposal on information hiding methods using XML."*The 1st Workshop on NLP and XML*. 2001.
[13] P. Niels and H. Peter, "Hide and seek: An introduction to steganography," Security & Privacy, IEEE, vol. 1, pp. 32-44, 2003.

[14]    C. Abbas, C. Joan, C. Kevin, and M. K. Paul, "Digital image steganography: Survey and analysis of current methods," Signal Processing, vol. 90, pp. 727-752, 2010.
[15]    D. Poulami, B. Debnath, and K. Tai-hoon, "Data Hiding in Audio Signal: A Review," International journal of database theory and application, vol. 2, pp. 1-8, 2009.
[16]     A. Odeh, A. Alzubi, Q. Bani, and K. Elleithy, "Steganography by multipoint Arabic letters," in Systems, Applications and Technology Conference (LISAT), 2012 IEEE Long Island, 2012, pp. 1-7.
[17]    A. Odeh, K. Elleithy, and M. Faezipour, "Text Steganography Using Language Remarks," presented at the The American Society of Engineering Education, 2013.
[18]    A. Odeh and K. Elleithy, "Steganography in Arabic Text Using Zero Width and Kashidha Letters," International Journal of Computer Science & Information Technology (IJCSIT),, vol. 4, pp. 1-11, 2012.
[19]    W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", IBM Systems Journal, vol. 35, Issues 3&4, 1996, pp. 313-336.
[20]    T. Moerland, "Steganography and Steganalysis", www.liacs.nl/home/tmoerlan/privtech.pdf, May 15, 2003.pp. 243-254). IEEE.
[21]    R.E. Moore, *Interval analysis* (Englewood Cliffs, NJ: Prentice-Hall, 1966).