

Steganography in HTML Files Using ROT-N Algorithm

N.K. Gupta^{1*}, B. Paul²

¹Assistant Professor, Department of Computer Science & Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, India

²Teaching Assistant, Department of Computer Science & Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, India

*Corresponding Author: narendra.gupta@shiats.edu.in

Available online at: www.ijcsonline.org

Accepted: 14/Jul/2018, Published: 31/Jul/2018

Abstract — In this modern era, the importance of information security has gained a special importance. Steganography is the art and science of invisible communication. This is accomplished through hiding information in other information, thus hiding the existence of the communicated information. Steganography techniques can be applied to text file, HTML file, XML file, images, a video file or an audio file. This paper intends to give an overview of web steganography techniques. It proposes a new idea to hide information in web pages which includes using ROT-N algorithm and thereby, acquiring the extra security. The experimental results show that the proposed method has high security than other existing methods. The suggested method is implemented by using Java language.

Keywords—Steganography, Cryptography, Security, HTML, ROT-N Algorithm

I. INTRODUCTION

Steganography (pronounced STEHG-uh-NAH-gruhf-ee, from Greek *steganos*, or "covered," and *graphie*, or "writing") is the hiding of a secret message within an ordinary message and the extraction of it at its destination. Steganography takes cryptography a step farther by hiding an encrypted message so that no one suspects it exists. Ideally, anyone scanning your data will fail to know it contains encrypted data. In Steganography unauthorized person will be unaware of the secret data being sent by the end user via any carrier such as web page, video file, audio file or image file.

Whereas, in Cryptography the unauthorized person will be aware of the secret data being sent by the end user but since the message will be encrypted by an unknown key (known to only end users), the information cannot be decrypted by him. The multimedia files are only used as the cover medium but not as a transmission medium. Whereas in web page information hiding process, the web pages can serve as both the cover medium and transmission method to conceal the secret information.

Hyper Text Markup Language (HTML) is used to create pages and make them functional. A Markup Language is a way that computers speak to each other to control how text is processed and presented. To do this HTML uses two things: tags and attributes. Tags are used to mark up the start of an HTML element and they are usually enclosed in angle brackets. Attributes contain additional pieces of information. Attributes take the form of an opening tag and additional info

is placed inside.

In this paper we propose a new approach of information hiding in HTML file by getting the encrypted message by ROT-N encryption algorithm using Value of N as a private key and this obtained message will be used as a name value in HTML name attribute. Unlike existing approaches in web page, information hiding i.e. changing the case of tags which can be done by changing some letters in the tags, storing data in id, this paper is using a name attribute to store the encrypted secret information so that an unintended receiver cannot suspect it as a stego web page. Even if in case unauthorized person suspect, he cannot retrieve the information because of the private key value N which will be shared only between the end user.

Few existing web steganography Techniques are

In the existing models of the web page information concealing methods the researches have been done to conceal the secret information in the tags and attributes of the source code i.e. HTML and XML files [6] as well as the white spaces of the source code [7][8]. The popular techniques used in the existing model are as listed below:

1) Empty tag method

In this process empty tags i.e. either a begin tag quickly taken after by an end tag or an empty tag is used in order to conceal the secret information. Using these types of empty tags in the source code does not affect the content on the webpage[9].

2) Line break approach

In this approach by continuously adding the line break tag at the end of each tag the secret information was concealed in the webpage. Usage of repetitive of line break at the end of each tag does not affect the content of the web page[9].

3) Changing the case of the tag

As the HTML is not a case sensitive language even the changes in the case of the tags may not show any change in the web page while parsing[10][11].

4) Information hiding based on the attribute value string

The strings used in the attribute values are not case sensitive, so attribute values can be taken for information hiding by keeping the uppercase indicates 1 and lower case indicates 0 [2].

5) Colour code or tag id replacement with Hexadecimal data [12]

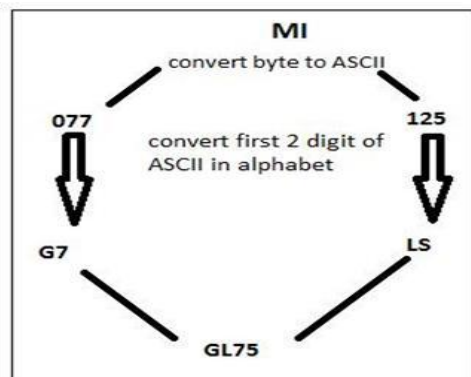


Fig. 1 Tag id replacement

Accordingly, the structure of this paper as follows: Section 2 Methodology Adapted. Section 3 Framework of the proposed Technique Section Finally, conclusion is presented in section 4.

II. METHODOLOGY ADOPTED

One of the attributes that can be used in input tags in HTML is name attribute. The name attribute specifies a name for the element. This name attribute can be used to reference the elements. The name attribute can be used on the following elements-

button,fieldset,form,iframe,input,map,meta,object,output,param,select,textarea Each name will be given with a unique data identity. The value in this tag attribute will be the encrypted data.

ROT-N ALGORITHM

Rotation ciphers have a long history, a famous example being the Caesar Cipher (also Caesar's code or Caesar's shift), and a substitution cipher used to encode messages by substituting letters by other letters a fixed number of positions away in alphabetic location. To cipher and decipher the given text, ROT-N shuffles (rotates) the alphabet by N places, N will be the input number between 1-25. Which means that if we write down the alphabet in 2 rows (each containing 10 characters), then we can transform (encipher, decipher) the text by substituting the characters of the first row by characters of the second row (and vice versa). For example if the value of N is given as 3 the A will be replaced by D while encryption process. And D will be replaced by A in Decryption process. And so on as shown in Fig.2.

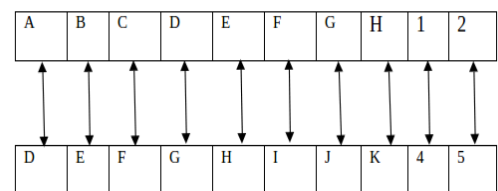


Fig 2 ROT-N – encryption/decryption schema (N=3)

By Encryption process the name value will be generated from each words of the text. It is required to have unique name for each name attribute identity. Hence, a numerical value (1, 2, 3...) will be concatenated with the generated encrypted string as shown in fig.3. And this value will be used as a name for a name attribute in an HTML file. Each name attribute is given a unique name.

In Decryption process, the generated names will be extracted and the last digit i.e. the number, will be removed. After that these words will be given as an input with the value of N in ROT-N Decryption Algorithm. The encrypted secret message will be retrieved as a result as shown Fig 4.

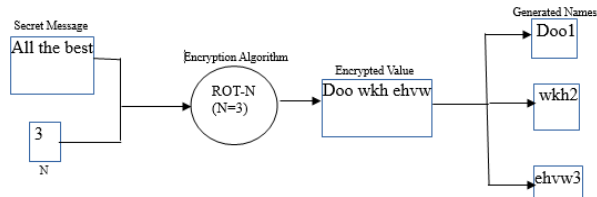


Fig.3 Technique of converting Encrypted value into unique name id

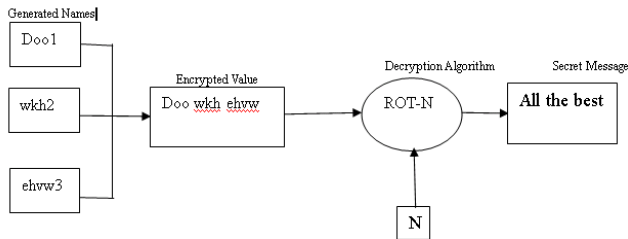


Fig 4 Technique of Retrieving the message from Stego HTML File

Algorithm for encryption & decryption of data

Input: HTML file, Secret Information, Value of N

Output: stego XML file

1. Select the HTML file which contains name as attributes in an input tag.
2. Input the message and value of N as a key.
3. Perform encryption using ROT-N encryption algorithm.
4. Separate the each obtained words and concat it with the number (incremental order) as a last digit.
5. Store all the words one by one into name attribute of an input Tag repository one by one.

Input: stego HTML file, Value of N

Output: Secret Information

1. Select the stego HTML file which contains encrypted secret message in name as attributes in a input tag.
2. Input the Secret key i.e. the value of N.
3. Remove the last digits from all the names.
4. Perform Decryption using ROT-N Decryption algorithm.
5. Retrieve the secret message which is the Decrypted text.

III. FRAMEWORK OF THE PROPOSED TECHNIQUE

The following diagram depicts the fundamental flow of encryption and decryption process of the proposed technique. Here, we use ROT-N Encryption Algorithm for the Encrypting the data. A ROT-N Encryption algorithm uses text message and Value of N as an input. And generates the Encrypted message which includes shuffles (rotation) of the

alphabet by N places. Decryption process requires the Stego HTML file and a Value of N, as shown in Fig. 5, to obtain the secret message as output.

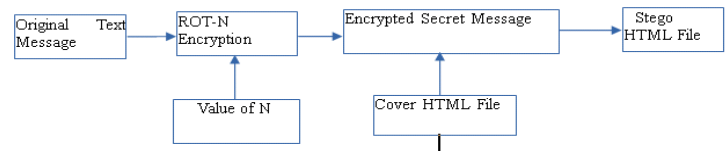


Fig 5 Data Encryption Process

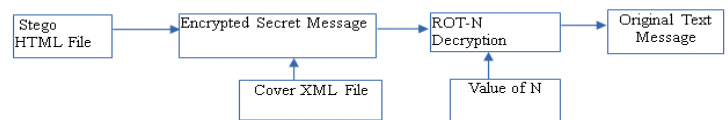


Fig 6 Data Decryption Process

IV. CONCLUSION

Although only some of the main web steganographic techniques were discussed in this paper, one can see that there exists a large selection of approaches to hiding information in web pages. All the different methods of hiding messages in HTML files, with different strong and weak points respectively. Where one technique lacks in security, the other lacks in robustness

Thus, a new method has been proposed in which some sort of encryption is done firstly using ROT-N Encryption Technique. The proposed method offers high security as compared to other available techniques. In addition, this method offers robustness, as the hidden data was inserted inside the name attribute as unique name. Moreover, Using the ROT-N algorithm enhances security by using an encryption mechanism.

REFERENCES

- [1] B.S. Reddy, K.S. Kuppasamy, T. Sivakumar "Towards Web Page Steganography with Attribute Truth Table", 2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS -2016), Jan. 22 – 23, 2016, Coimbatore, INDIA
- [2] X. Yong, L. Juan, and Z. Yilai, "A High Capacity Information Hiding Method for Webpage Based on Tag," 2012 Third Int. Conf. Digit. Manuf. Autom., pp. 62–65, 2012.
- [3] R.K. Singh, B. Alankar, "A Novel Approach For Data Hiding In Web Page Steganography Using Encryption With Compression Based Technique", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. III (May-Jun. 2016)
- [4] C. Dhanani, K. Panchal, "HTML Steganography using Relative links & Multi web-page Embedment" IJEDR | Volume 2, Issue 2 | ISSN: 2321-9939, 2014.
- [5] M. Garg, "A Novel Text steganography Technique Based on HTML Document", International Journal of advanced Science and Technology

- Vol. 35, October 2011.
- [6] Arora, K., & Gandhi, G. "A Review of Approaches for Steganography". International Journal of Computer Sciences and Engineering, 2(5), 118-122, 2014.
- [7] Y. C. Chou and H. C. Liao, "A Webpage Data Hiding Method by Using Tag and CSS Attribute Setting," 2014 Tenth Int. Conf. Intell. Inf. Hiding Multimed. Signal Process., pp. 122–125, 2014.
- [8] I. Lee and W. Tsai, "Secret Communication through Web Pages Using Special Space Codes in HTML Files," Int. J., pp. 141–149, 2008.
- [9] D. V. Dhawase and P. S. Chavan, "WEBPAGE INFORMATION HIDING USING PAGE CONTENTS," vol. 3, no. 1, pp. 182–186, 2014.
- [10] X. Guo, G. Cheng, C. Zhu, A. Zhou, W. Pan, and D. Truong, "Make Your Webpage Carry Abundant Secret Information Unawarely," 2013 IEEE 10th Int. Conf. High Perform. Comput. Commun. 2013 IEEE Int. Conf. Embed. Ubiquitous Comput., pp. 541–548, 2013.
- [11] S. Dey, H. Al-Qaheri, and S. Sanyal, "Embedding Secret Data in Html Web Page," arXiv Prepr. arXiv1004.0459, pp. 1–10, 2010.
- [12] Mohammad Shirali Shahreza, "A New Method for Steganography in HTML Files", Advance in computer, Information and System Science & Engineering , 247-251, 2006 Springer.
- [13] Rao, B. R., Kumar, P. A., Rao, K. R. M., & Nagu, M. "A Novel Information Security Scheme using Cryptic Steganography". Indian Journal of Computer Science and Engineering, 1(4), 327-332., 2010.

Authors Profile:

Dr. N. K. Gupta has completed his Doctoral Degree in database optimization from Sam Higginbottom University of Agriculture, technology and Sciences in 2018. He has teaching experience of more than 18 years and guided around 42 Masters and 60 Graduates. The author has published / Presented around 25 international and 15 national papers in reputed journals/conferences. Presently he is working as Assistant Professor in Dept. of CS&IT at SHUATS, Allahabad.



Er. Binnu Paul has completed his Bachelors in Technology(B.tech) in Computer Science from Karunya University, Coimbatore, Tamil Nadu in 2015 and pursued Masters in Technology (Mtech) in Computer Science from Sam Higginbottom University of Agriculture, technology and Sciences in 2018. He has teaching experience of 2 years and he also attended several seminars and Conferences. Presently he is working as Assistant Professor in Dept. of CS&IT at SHUATS, Allahabad.

