iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4776

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Steganography in XML Files Using RC4 Stream Encryption Algorithm

Binnu Paul[1], NK Gupta[2], Anchit S Dhar[3]
[1]*Student of M-tech CSE 4th SEM, Department of Computer Science & Engineering, Sam Higginbottom University of Agricuture, Technology and Sciences,*
[2]*Assistant Professor, Department of Computer Science & Engineering, Sam Higginbottom University of Agricuture, Technology and Sciences,*
[3]*Assistant Professor, Department of Computer Science & Engineering, Sam Higginbottom University of Agricuture, Technology and*

Abstract: In this modern era, the importance of information security has gained a special importance. Steganography techniques can be applied to text file, images, a video file or an audio file. This paper gives a new insight that how Private Key Steganography can be used to give an extra-secure method. This paper gives another dimension of safe correspondence through information hiding on Internet. The experimental results show that the proposed method has high security than others. The suggested method is implemented by using Java language.
Keywords: Steganography, Cryptography, Security, XML, RC4 Algorithm

I. INTRODUCTION

Steganography is derived from the Greek words. “stegos” is termed as “roof or covered” and “graphy” as “writing or drawing”. Thus Steganography is the hidden writing or secret writing. With the help of this, a secret message can be set inside a piece of information and can be sent without being aware of the secret message. 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. Extensible Mark-up Language (*XML*) defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. XML is used to store or transport data, while HTML is used to format and display the same data. Whereas these both languages use tags with certain attributes.
In this paper we propose a new approach of information hiding in XML or HTML file by getting the encrypted message by RC4 encryption algorithm using a private key and this obtained message will be used as an id value in XML or HTML id 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, this paper is using an id 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 encryption. 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:

A. 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].

B. 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].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

C. 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].

D. 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].

E. Colour Code or tag id Replacement with Hexadecimal Data [12]

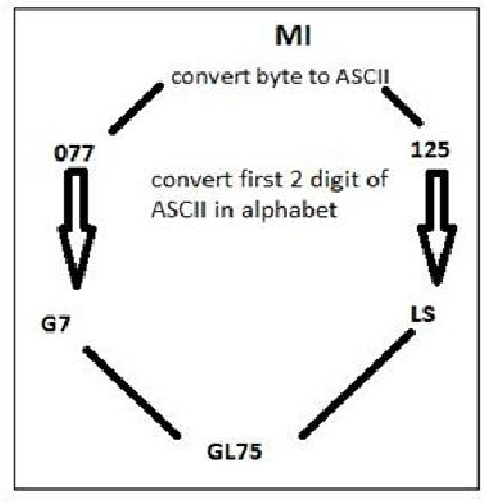


Fig. 1 Tag id replacement

II. METHODOLOGY ADOPTED

One of the attributes that can be used in all the tags in XML as well as in HTML is id attribute. Each id will be given with a unique identity. The value in this tag attribute will be the encrypted data.

A. Rc4 Algorithm To Encrypt Message Using Private Key[13]

RC4 is an encryption algorithm that was created by Ronald Rivest of RSA Security. It is used in WEP and WPA, which are encryption protocols commonly used on wireless routers. The workings of RC4 used to be a secret, but its code was leaked onto the internet in 1994. RC4 was originally very widely used due to its simplicity and speed. Typically, 16 byte keys are used for strong encryption, but shorter key lengths are also widely used due to export restrictions. Over time this code was shown to produce biased outputs towards certain sequences, mostly in first few bytes of the keystream generated. This led to a future version of the RC4 code that is more widely used today, called RC4-drop[n], in which the first n bytes of the keystream are dropped in order to get rid of this biased output. Some notable uses of RC4 are implemented in Microsoft

Excel, Adobe's Acrobat 2.0 (1994), and Bit Torrent clients. To begin the process of RC4 encryption, you need a key, which is often user-defined and between 40-bits and 256-bits. A 40-bit key represents a five character ASCII code that gets translated into its 40-character binary equivalent (for example, the ASCII key "pwd12" is equivalent to 0111000001110111011001000011000100110010 in binary). The next part of RC4 is the key-scheduling algorithm (KSA), listed below.

```
for i from 0 to 255
S[i] := i
endfor
j := 0
for i from 0 to 255
j := (j + S[i] + key[i mod keylength]) mod 256
swap(S[i],S[j])
endfor
```

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

KSA creates an array S that contains 256 entries with the digits 0 through 255, as in the table below.

0	1	2	.....	i	i+1	.....	253	254	255

Each of the 256 entries in S are then swapped with the j-th entry in S, which is computed to be
j = [(j + S(i) + key[i mod keylength]) mod 256], where j is the previous j value (which is initially zero). S[i] is the value of the current entry in S. key[i mod keylength] is either a zero or a one. F or example, if we are at the 52th entry in S and the keylength was 40-bit, then 52 mod 40 = 12. The 13th element (because numbering for arrays begins at zero) in the binary version of "pwd12" is 0. For example, consider the first iteration of KSA with key "pwd12". Then, since i = 0, i mod 256 = 0. So, theNelement at the index 0 of the key is p, and its ascii value is 112. So, the new j is computed as
j = [(0 + 0 + 112) mod 256] = 112.
So, swapping the i-th and the j-th elements, we obtain the following array after the first iteration:

112	1	2	...	111	0	113	114	...	255

The next part of RC4 is the pseudo-random generation algorithm (PRGA). The PRGA is below:

```
i := 0
j := 0
while Generating Output:
i := (i + 1) mod 256
j := (j + S[i]) mod 256
swap(S[i],S[j])
output S[(S[i] + S[j]) mod 256]
end while
```

In PRGA, we begin with the array S that was swapped in the KSA. In PRGA, an element in S (at index i) is swapped with another element in S (at index j). Then, the next element in the encrypted text is the element of S at the index calculated by (S[i] + S[j] mod 256). At each iteration, i is recalculated as (i + 1) mod 256, and j is recalculated as (j + S[i]) mod 256. The number of iterations performed is the length of the key, and every value of S is swapped at least once beyond 256 iterations (due to the fact that i and j are calculated by some number n mod 256). The result of this is the code. Following up with the previous example, let us examine the first iteration of PRGA. Since i, j =0, i becomes 1 and j becomes (0 + S[1]) mod 256. Since S[1] = 124 (see the resulting S from KSA), j becomes 124. Then, the elements of S at 1 and 124 are swapped.
Then, we add the two numbers that we have just swapped: 232+124 = 356, mod by 256 to get 100. then output the 100th element of S, which is 33. We then look at the i-th index of the input string, "Math 310 Proves!", which gives us 'M'. We take the Unicode code of that character, which is 77, and perform a bitwise AND with 33 to get 108. Finally, get the character value for the Unicode code 108, which is 'l'.
By this process the random value for each character of the text will be generated which will be below 256.And This value will be used as a id for an id attribute in an XML or HTML file. Each tag is given a unique ID.Our goal is to hide information in this tag attribte.
After the encryption process the random number between 0-256 will be produced for each character of a message. This random generated number will be further converted to a combination of letters and digits so that they can be used in the ID attribute. First, we save the last digit of the byte code exactly as it is. The remaining two digits are between 0 and 25, which exactly corresponds to the number of English letters. For example the encrypted value for ‘x’ is 186.18 corresponds to R so the converted value will be R6 so on. As shown in fig. Four byte will be combined together and it will be used in an XML as a unique ID for ID attribute.

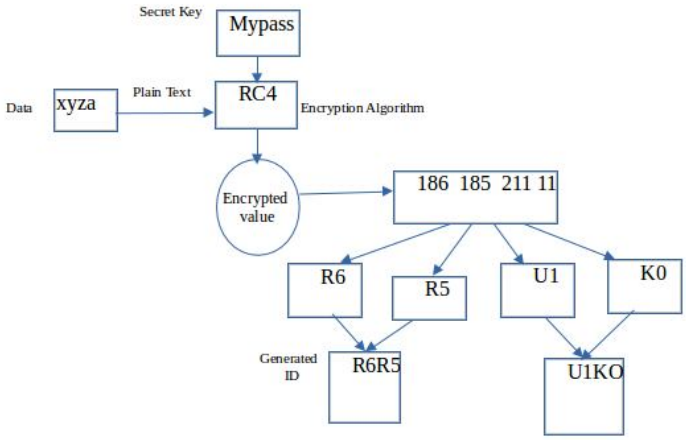


Fig.2 Technique of converting 4 bytes into eight characters

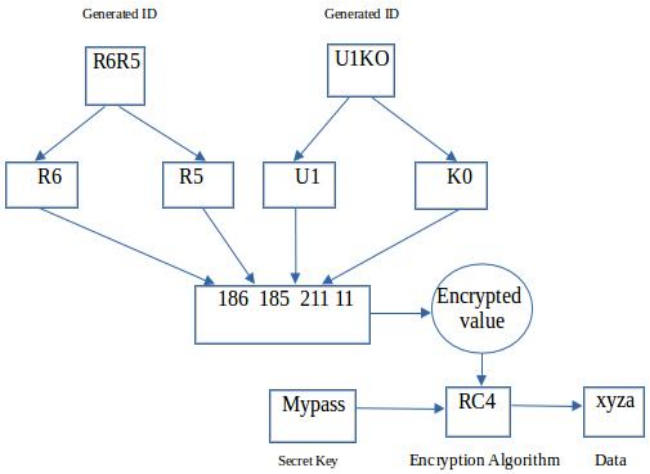


Fig.3 Technique of Retrieving the message from Stego XML File

Algorithm for encryption & decryption of data

1) Select the XML file which contains id as attributes in a tag
2) Input the message and Secret key
3) Perform encryption using RC4 encryption algorithm
4) Separate the first two digits to its corresponding alphabet and concat it with the third digit.
5) Store all the encrypted code into id attribute of a Tag repository one by one.

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 RC4 Encryption Algorithm for the Encrypting the data. A RC4 Encryption algorithm uses text message and secret key as an input. And generates the Encrypted message which includes random number between 0-256 for each characters of a text message. Decryption process requires the key file and a secret key, as shown in Fig. 6, to obtain the secret message as output.

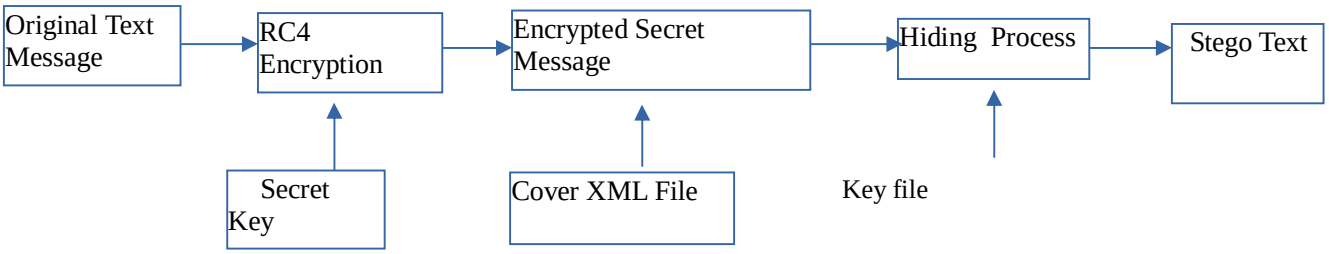


Fig. 4 Data Encryption Process

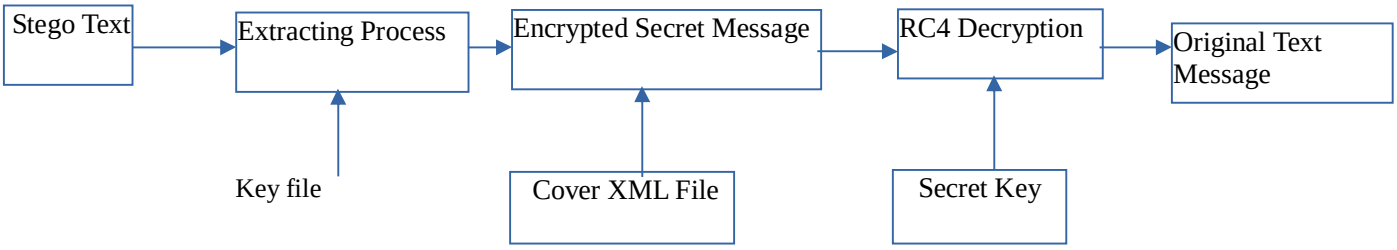


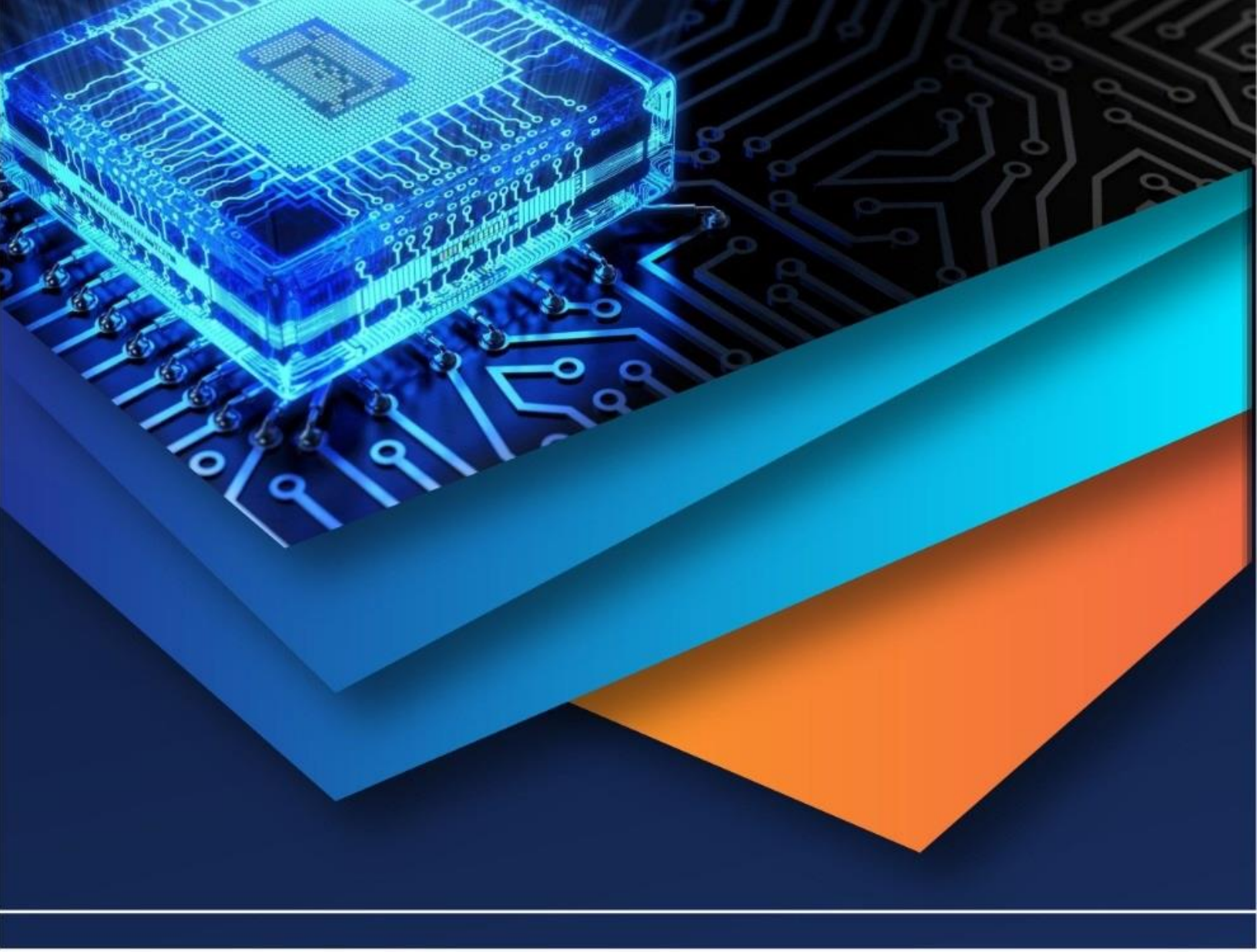
Fig. 5 Data Decryption Process

IV. CONCLUSION

Different algorithms have been presented to hide data inside files. Some of these methods were designed to be applied in specific languages, while others can be applied regardless of the language. In this paper, we presented a promising algorithm that can be applied to XML or HTML files. The proposed method offers high security as compared to other. In addition, this method offers robustness, as the hidden data was inserted inside the in tag attribute as id, and the Internet browser does not show it. Moreover, Using the RC4 algorithm enhances security by using an encryption mechanism.

REFERENCES

[1] Sreekanth Reddy,K.S. Kuppusamy,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] Ram Krishna Singh , Bhavya 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] Chintan Dhanani, Krunal Panchal,”HTML Steganography using Relative links & Multi web-page Embedment” 2014 IJEDR | Volume 2, Issue 2 | ISSN: 2321-993
[5] Mohit Garg, “A Novel Text steganogrpahy Technique Based on HTML Document”, International Journal of advanced Science and Technology Vol. 35, October 2011
[6] “Steganography.” [Online]. Available: https://en.wikipedia.org/wiki/Steganography
[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 Steganogrpahy in HTML Files”, Advance in computer, Information and System Science & Engineering , 247-251, 2006 Springer.
[13] RC4 Encrytion Available:sites.math.washington.edu./~nichifor/310_2008_Spring/Pres_RC4%20Encryption.pd
[14] William Stallings, The RC4 Stream Encryption Algorithm, https://people.cs.clemson.edu/~jmarty/courses/Spring-2017/CPSC424/papers/RC4ALGORITHM-Stallings.pdf
[15] Shingo Inoue, Ichiro Murase, Osamu Takizawa, Tsutomu Matsumoto, Hiroshi Nakagawa, “A Proposal on information Hiding Mehods Using XML”
[16] Xin-Guang Sui, Hui Luo, “A new Steganography method based on Hypertext”, IEEE-2004
[17] Yujun Yang, Yimei Yang, “An Efficient webpage Information Hiding Method Based on tag Attributes”, IEEE-2010.
[18] 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
[19] S. Dey, H. Al-Qaheri, and S. Sanyal, “Embedding Secret Data in Html Web Page,” arXiv Prepr. arXiv1004.0459, pp. 1–10, 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY