Novel Approach for the Text Authentication with Tamper Detection

Prerna Gupta[1], Amit Kumar[2]
[1]Research Scholar, [2]HOD (Head of the Department) of CSE Branch
[12]Kashi Institute of Technology, Varanasi

Abstract- The various methods are given by the researchers in the area of text documents authentication. This describes different studies related to the field of text documents authentication. The proposed approach is intended to design a security algorithm against the previously available algorithm for text cryptography. The previously available method having some limitations such as less cipher strength, limited for detection of text and the size of cipher text therefore a new fragile watermarking approach is required to design for finding the more appropriate and efficient solution. When the tampered portion in image text file, this was very difficult to spot in text file because text can easily be modified. After ensuring that image text file is altered somewhere, when it convert those image text file in to text file then it will not be similar like original one. Hence by proposed algorithm then one can easily decide that integrity of text file is compromised.

Keywords- Text Authentication, Tamper detection, encryption

I. INTRODUCTION

Cryptography is a method of storing and transmitting data in a particular form so that only those for whom it is intended can read and process it. Cryptography is about constructing and analyzing protocols that prevents third parties or the public from reading private messages [1]. In modern Cryptography, constructing and analyzing protocols means that used to the intersection of the disciplines of mathematics, computer science, and engineering for prevent the data in various aspects, such as confidentiality, data, authentication, and non-repudiation. In recent years the text document venerability attacks increasing easily by the unintended users. In same manners the cases of security issues and frauds are also increasing. The main reason is unsecure data transfer in the network [2]. To prevent user data over the un-trusted network most of the time cryptographic approaches are utilized, but traditionally available algorithms are not much suitable in the new generation computing technology. Symmetric-key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. Symmetric-key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of text. Symmetric encryption is the oldest and best-known technique. A secret key, which can be a number, a word, or just a string of random letters, is applied to the text of a message to change the content in a particular way. This might be as simple as shifting each letter by a number of places in the alphabet [3]. As long as both sender and recipient know the secret key, they can encrypt and decrypt all messages that use this key. Symmetric-key encryption can use either stream ciphers or block ciphers. Asymmetric encryption is the approach in which there are two related keys--a key pair. A public key is made freely available to anyone who might want to send you a message. A second, private key is kept secret, so that only you know it. Any message (text, binary files, or documents) that are encrypted by using the public key can only be decrypted by applying the same algorithm, but by using the matching private key [4]. Any message that is encrypted by using the private key can only be decrypted by using the matching public key. This means that you do not have to worry about passing public keys over the Interne. In a public key encryption system, any person can encrypt a message using the public key of the receiver, but such a message can be decrypted only with the receiver's private key. A watermark is a more or less transparent image or text that has been applied to a piece of paper, another image to either protect the original image or text. Digital Image Watermarking is a technique for inserting information (the watermark) into an image, which can be later extracted or detected for variety of purposes including identification and authentication purposes [5]. Usually watermarking is divided into visible and invisible, fragile and robust, spatial and frequency.

II. TEXT CRYPTOGRAPHY

In text cryptography, encryption is the process of transforming information using algorithm, which means that original text transform into cipher text, cipher text known as encrypted or encoded information that is unreadable by a human or computer without the proper cipher to decrypt it. Decryption is the inverse of encryption; Decryption is the process of turning cipher text into readable plaintext. In the earliest time, the classical type of cipher methods is using such as transposition cipher, substitution ciphers. In the modern time, modern encryption methods can be divided into two categories such as symmetric, asymmetric. In bit plane slicing techniques the image are sliced at different planes [6]. It ranges from bit level 0 to bit level 7, where bit level 0 is the least significant bit (LSB) and 7 is the most significant bit (MSB). The advantage of doing these techniques is to urge the vital knowledge of every little bit of image. During in this techniques, solely in last four higher order bits planes important knowledge in envisioned. The lower level bit plane doesn't provide a lot of detail because they are made up of lower contrast. Bit plane shows the least significant bit (LSB) and the most significant bit (MSB). Digitally, an image is represented in terms of pixels. These pixels can be expressed further in terms of bits [7]. It uses a 'binary' form of numbers (ones, twos, fours, eights, etc.) made up entirely of the digits '0' and '1'. With decimal figures it only requires three digits to write the highest value of '255', but using binary figures of just '0' and '1' it takes a total of eight digits. This means that each gray scale pixel has a number from '00000000' for black, to '11111111' for white, as the binary value of '11111111' is equal to the decimal value of '255'.

III. LITERATURE SURVEY

Keerthi K, et.al (2017) proposed a novel mapping approach using which the message can be encoded on the elliptic curve in the form of finite number of points. The plain text is transformed into ASCII values using mapping technique [8]. Further, HEXADECIMAL is generated by converting these ASCII values. The x and y coordinates are generated by integrating the Hex values. For avoiding any kinds of security attacks to enter the network, the encryption of converted values is done in reverse order. The overhead generated for a common look up table that is shared by sender and receiver is minimized by this approach. Thus, the speed at which encryption method works is higher. Further, in case when group count is odd, extra padding is avoided.

Ertan Atar, et.al (2016) presented that there are very less numbers of linear measurements as compared to the genuine signal vectors' lengths when compressed sensing is applied [9]. A measurement matrix is used here through which a very short signal is generated from original signal. The asymmetric cryptography is applied to transmit the asymmetric keys to the receiver such that the security can be increased. The orthogonal matching pursuit (OMP) approach is used here such that the compressive sensing can sense the inputs. Similar to the 4f optical cryptography systems, the hybrid cryptography is used to encrypt the input. Data compression and encryption are achieved by the overall system as per the simulation results achieved.

Raghunandan K R, et.al (2017) proposed a novel mechanism through which the issues faced by RSA are eliminated. Unlike the previous RSA method, the cubic power of Pell's equation follows a completely different key generation method [10]. The alpha, prime numbers and variables of Pell's equation are the three different factors that affect the K public key. Using public key exponent "K" to achieve private key is difficult. By making comparisons with RSA, the efficiency, security and reliability of mechanism can be achieved. An analysis of several attacks and histogram is performed in this paper to achieve results. In terms of mean value, a good variation or deviation is achieved in the proposed mechanism.

M. Saritha, et.al (2016) proposed the integration of cryptography and steganography in order to develop a highly secure model [11]. In order to perform steganography, sequential algorithm is applied and cryptography is performed for Symmetric XOR algorithm within this proposed work. The effort and time can be minimized manually by the software within this proposed technique. Any application can easily apply this proposed mechanism. Also, the users can operate and easily understand the operations of this system. Both, video as well as audio files apply this concept. There are varieties of files and images to which this approach can be applicable efficiently.

Deepali Bhat, et.al (2017) proposed a secure data transmission mechanism in order to hide the information in this paper [12]. This approach included both, text steganography and cryptography algorithms in it. In order to perform cryptography, a symmetric key algorithm called Data Encryption Standard (DES) was used as a base for the proposed mechanism. Text steganography which is a mechanism through which secret message can be concealed into other text is utilized for data hiding. The dynamic construction of text is done for hiding the sensitive information. The presence of any secret data is hidden safely here. Thus, the information can be hidden and data can be transmitted across secured channels efficiently by integrating the two cryptographic mechanisms.

Sudipta Singha Roy, et.al (2017) proposed a novel approach with the help of time varying delayed Hopfield neural network to encrypt the text messages [13]. Further, this paper proposed a posterior DNA cryptographic model in this paper. A binary sequence is created with the help of implementing chaotic neural network. Further, the permutation function is applied such that first level encryption can be performed by creating a key. The ASCI value is converted by converting plain text to a relevant binary sequence. Further, the chaotic neural network maps and a permutation function that completely relies on the binary sequence are switched to perform encryption. Thus, by integrating DNA cryptography and delayed chaotic neural network, the performance of proposed mechanism is known to be better in terms of security.

IV. RESEARCH METHODOLOGY

As we know that text file is nothing but the combination of various alphanumeric characters which includes number, characters and special symbols. All given four steps are hierarchal. It means output of previous step will be the input of next steps. Proposed approach consists of four steps as shown in figure 1 and figure 2.

1. Text files compression using run length encoding.

2. Authentication bit generation and text file to image conversion
3. Tamper detection
4. Text file decompression

A. Text file compression

Since text files may be very huge in size and may contain redundant information hence before proceeding we need to compress it. There are following steps are provided to compress the text file.

Step1-Take a text files as an input and read its all characters row wise.

Step2-Convert all characters into its ASCII form, which will be decimal number.

Step3-All decimal numbers are converted into eight bit binary form.

Step4-Apply run length encoding on binary bit stream.

Step5-Output of step 4 will be represented as compressed text file which will be no more readable.

B. Authentication bit generation and text file to image conversion

Once we get the compressed text file it will be embedded with authentication bit in order to make self-authenticating text files. These authentication bits are nothing but fragile watermark, which will be destroyed if any alteration is done to text file, hence tamper can be easily localized. Block diagram is shown in figure. The detail steps are shown below:

Step1- Take the compressed text file as an input.

Step2- Convert the compressed text file into binary vector.

Step3-Create clusters of five bits for all bits of text file binary vector.

Step4-Generatethree authentication bits for each cluster of five bits by following way.

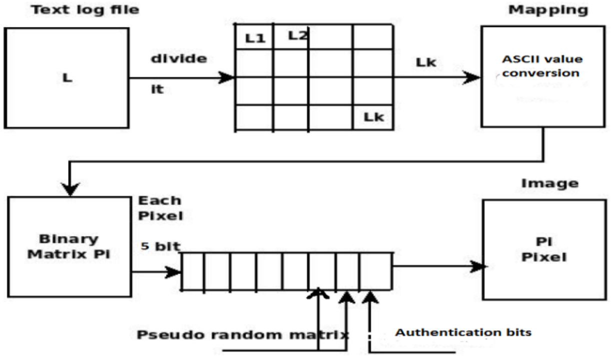


Fig.1: Authentication bit generation for image text file

For 1st bit generation-

Step 5-Take bit wise XOR of five bits of each cluster and calculate the modulo2 of the sum of them.

$$1stbit = \left(\sum_{i=1}^{4} b_i \oplus b_{i+1}\right) mod\ 2 \qquad (1)$$

Where b_i represents the i^{th} bit of vector.

For 2nd bit generation-

Step 6- Using a secret key, generate a random matrix *Rm* of same size r x c as image whose values ranges from 0 to 31. Do bit wise XOR between corresponding cluster's bits and bits of *Rm*.

$$2ndbit = \left(\sum_{i=1}^{5} b_i(Im) \oplus b_i(Rm)\right) mod\ 2 \qquad (2)$$

For 3rd bit generation-

Step7- Calculate the decimal value of five bit cluster.

Step 8- Calculate the complement of the each decimal value of cluster and take the bit wise XOR with its original value.

$$3rdbit = \left(\sum_{i=1}^{5} b_i(Im) \oplus b_i(31 - Im)\right) mod\ 2 \qquad (3)$$

Step 9- Append all generated three bits to five bits of its corresponding cluster as shown in figure3.1.

Step10- Now takes the decimal value of each eight bit binary cluster.

Step11-The resultant matrix will be represented as image.

A. TAMPER DETECTION

It may be possible that text file is tampered intentionally or unintentionally. But proposed scheme creates fragile watermark for text file which itself is sufficient enough to detect the alteration as shown in figure. The detailed procedure is as follows:

Step1- Take the altered text file image as an input.

Step2- Take a tamper localization matrix with all initially assigned values as 0.

Step3- Extract the last three bits of all pixels of text file image.

Step4- Recalculate the three bits for all clusters using eq. 1, 2 and 3.

Step5- Compare the extracted and recalculated three bits for all corresponding pixels.

Step6- If any mismatch is found then marks the corresponding location in tamper localization matrix.

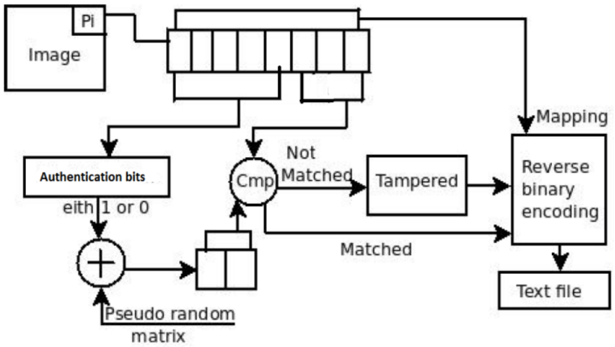


Fig.2: Tamper localization procedure.

B. TEXT FILE DECOMPRESSION

Once image text file is checked for tamper, we need to decompress it in readable format. Following steps explain the decompression approach.

Step1- Extract five MSBs of each pixels of image text file and make a vector.

Step2-Now creates clusters of eight bits from given vector.

Step3-All eight bit clusters are converted into decimal format.

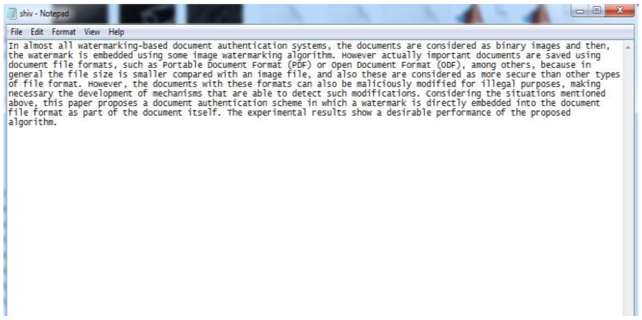
Step4-Now again creates clusters of eight bits from given vector.

Step5-Using run length decoding just expands each intensity by its frequency count.

Step6-Now replace all decimal value with its ASCII values and the resultant text in copied in output file which is nothing but the extracted decompressed text file.

V. RESULT AND ANALYSIS

Proposed algorithm for text file security is simulated in Mat lab 2010 and we have taken the text files of various sizes from our colleges' web server texts. According to our scheme we first take text files as and input to the algorithm as shown in figure 3.



In almost all watermarking-based document authentication systems, the documents are considered as binary images and then, the watermark is embedded using some image watermarking algorithm. However actually important documents are saved using document file formats, such as Portable Document Format (PDF) or Open Document Format (ODF), among others, because in general the file size is smaller compared with an image file, and also these are considered as more secure than other types of file format. However, the documents with these formats can also be maliciously modified for illegal purposes, making necessary the development of mechanisms that are able to detect such modifications. Considering the situations mentioned above, this paper proposes a document authentication scheme in which a watermark is directly embedded into the document file format as part of the document itself. The experimental results show a desirable performance of the proposed algorithm.

Fig.3: Example of text files

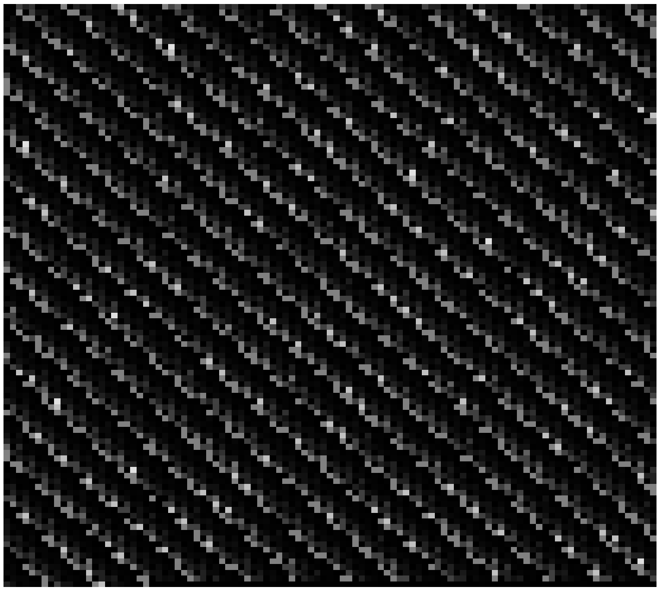


Fig.4: Encrypted image text file

First of all text file is compressed by proposed compression technique. After compression it will be no more readable in nature. We can see that image text file consist few gray colors .Where non-black color is occupied by the valuable information of text file whereas black portion shows the absence of information or remaining space in image text file as shown in figure. Text file size and black color present in image text file is inversely proportional means when we increases the size of text file then the area in image text file which contains black portion will decrease. Now this compressed text file again embedded with some authentication bits in order to make its self-authenticable text files as shown in figure 4.

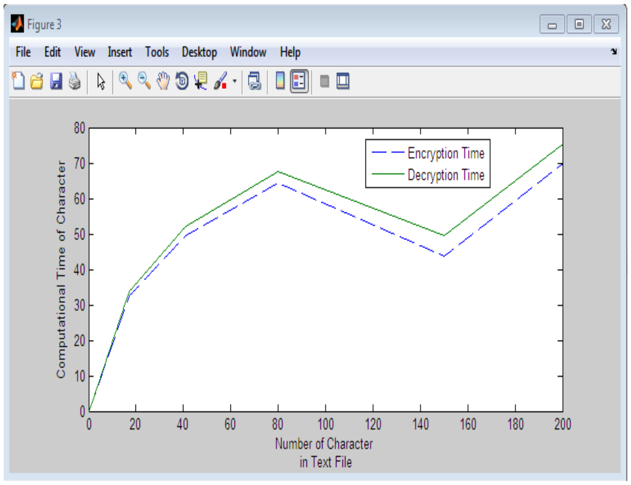


Fig.5: Computational time of character

Figure 5 shows the computational time of character with encryption time and decryption time. It shows the how much take time at the side of encryption, whenever text file is encrypted and also shows the decryption time whenever text file is decrypted at the side of decryption. Each time vary the time of encryption time and decryption time, while it take same file of text each time.

VI. CONCLUSION

The implementation of the proposed methodology is provided using Mat Lab environment. In this a data hiding method by improved LSB substitution process is proposed. The text documents authentication quality of the crypto-image can be greatly improved with low extra computational complexity. The performance of the system is evaluated in terms of security and detection complexity. The run length encoding technique is implemented successfully and the performance is better than the previously used encryption method in terms of Memory, Compression Time, Encryption Time, decryption time, Detection Time and Data Security.

VII. REFERENCES

[1]. S .M . C Vigila and Munseeswaran,"Implementation of Text based cryptosystem using elliptic curve cryptography", 2009 1st International conference on advances computing ,ICAC 2009,pp.82-85,2009.

[2]. P.Bh,D.Chandeavathi , and P .P. Roja,"Encoding and Decoding of a message in the Implementation of Elliptic Curve Cryptography using Koblitz's Method", International Journal on computer science ,vol.02,no .05,pp-1904-1907,2010.

[3]. F.Amounas and E. H. E. Kinani,"Fast Mapping Method based on Matrix Approach For Elliptic Curve Cryptography", vol. 1, no. 2, pp. 54- 59,2012.

[4]. L. D. Singh and K. M. Singh,"Implementation of Text Encryption using Elliptic Curve Cryptography", Procedia Computer Science,vol.54, no.1, pp. 73-82,2015

[5]. C. T. M and V. Paul, "Secure Method for embedding plaintext on an elliptic curve using TDMRC code and Koblitz method", Journal of Theoretical and Applied Information Technology,vol. 84,no. 2,pp. 298- 304,2016.

[6]. J. Muthukuru, B. Sathyanarayana, "Fixed and Variable Size Text Based Message Mapping Techniques using ECC", vol. 12,no. 3, 2012.

[7]. Candès, E., Romberg, J. and Tao, T., ''Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information'', IEEE Trans. on Information Theory, 52(2) pp. 489-509, February, 2006.

[8]. Keerthi K, Dr.B.Surendiran, "Elliptic Curve Cryptography for Secured Text Encryption", 2017 International Conference on circuits Power and Computing Technologies [ICCPCT]

[9]. Ertan Atar, Okan K. Ersoy, Lale Özyılmaz, "Character/Text Data Compression and Encryption by Compressive Sensing and Hybrid Cryptography", 2016, IEEE

[10]. Raghunandan K R, Rashmitha Shetty, Ganesh Aithal, "Key Generation and Security Analysis of Text Cryptography using Cubic Power of Pell's Equation", 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)

[11]. M. Saritha, Sushravya. M, Vishwanath. M. Khadabadi, "Image and Text Steganography with Cryptography using MATLAB", International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016

[12]. Deepali Bhat, Krithi V, Manjunath KN, Srikanth Prabhu, Renuka A., "Information Hiding through Dynamic Text Steganography and Cryptography", 2017, IEEE

[13]. Sudipta Singha Roy, Shaikh Akib Shahriyar, Md. Asaf-Uddowla, Kazi Md. Rokibul Alam, Yasuhiko Morimoto, "A Novel Encryption Model for Text Messages using Delayed Chaotic Neural Network and DNA Cryptography", 2017 20th International Conference of Computer and Information Technology (ICCIT)