ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

A Novel Encryption Technique For Securing Text Files

K.Berlin
Ph.D Research Scholar
Computer Science and Engineering
Alagappa University, Karaikudi
berlinjenson@gmail.com

Dr.S.S.Dhenakaran
Professor
Computer Science and Engineering
Alagappa University, Karaikudi
ssdarvind@yahoo.com

Abstract—**All over the world, the transaction of data via networking systems or internet is in inestimable mode. The entire transactions that are taking place via internet depend on the security issues. For improvising the security system, huge amount of cryptographic mechanisms are designed and implemented in various sites. Even though the prevailing cryptographic techniques remain good, better secured mechanisms are needed. A lot of crypto techniques exist with some limitations such as fixed key generation techniques, same keys used for encrypting data that include both public and private. In the situation of same key usage, there is a chance for hackers to break keys easily. So it turns to be a great issue for every authorized user to maintain the privacy of their respective message. To overcome this kind of hurdles, the present work endeavors a new mechanism called "Dynamic Usage of Crypto Algorithm". The proposed method is designed for text files. At the outset, the input text file is compressed and segmented. Then the segmented files are encrypted using three cryptographic algorithms. Selection of these algorithms is in dynamic mode, based on the size of the compressed text file. Final encrypted message is the combination of the outcome of three algorithm.**

Keywords—Compression, Cryptography, Fixed Key Generation, Networking, Security Mechanisms.

I. INTRODUCTION

Cryptography is the process of transmitting data from applications into unreadable format to protect data from the unauthorized access [10]. It consists of several mathematical techniques to solve the complexity of problems. It provides the aspects of secured data as confidentiality, integrity and Compression authentication [9]. Cryptography algorithms play a crucial role in the information society. When we use our credit card, make call to someone through mobile phone, get access to health care services or get something on the web, cryptographic algorithms used to protect the original data from the unauthorized access [5]. These algorithms guarantee that nobody can whip money from our account. While cryptography is a necessary component, the importance of cryptography should be put in the correct perspective.

A. Compression

The process of converting an input data stream or the original raw data into another output stream (compressed data) is called as data compression. There are many known methods for data compression. They are having different kinds ofideas, are suitable for different types of data, and provide various results with expected outcome, but based on the same principle, like compress data by removing redundancy in the source file. The compression or encoder is the program that compresses the raw data in the input stream and creates an output stream with compressed data. The decompression or decoder converts in the opposite direction. There are two types of compression techniques: 1. Lossy Compression Technique 2. Lossless Compression Technique. In the kinds of lossy, there is a chance to loss original data during the process of decompression. No loss can be taken placed while using lossless Compression techniques.

B. Encryption

The combined process of encryption and decryption is called as Cryptography, it is the science of using mathematics to encrypt and decrypt information. The combination of an encryption algorithm and an encryption key can be used to

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

encrypt data from plain text into cipher text. The purpose of using encryption algorithm is to protect data from the third party access and gives more confidentiality of data stored on computer systems and transmitted over the internet or any other computer networks. Encryption algorithms are categorized into two types: Symmetric and Asymmetric. Symmetric encryption ciphers use same key for encrypting and decrypting a massage. Symmetric key encryption is much speed than the Asymmetric encryption, but the sender needs to share key with receiver to decrypt original message. In public-key cryptography, there are two keys needed to process the message: 1. Public 2. Private. The public key can be shared with everyone, but the private key should be maintained as secret.

II. RELATED WORKS

Debashis Chakraborty et.al., discussed on the Dictionary based Text Compression technique using replacement strategy [1]. The authors of the algorithm maintain a dictionary, in which for every string of length six, it is compressed by assigning a single character to it. Finally, the single character is used to decompress the encoded file. This gives a good compression ratio irrespective of the content of the text file. Character Replacement is the main feature that is highlighted in the algorithm.

Strahil Ristov and Damir Korenci, explained the concept a suffix array and the derived Longest Common Prefix interval tree [2]. The results of the author's algorithm reveal that it is possible to obtain equal time complexity when changes in the tree structure modifies for a dynamic application. The tree index does not include the insertions and substring orders for the substitutions. Moreover, the algorithm works for larger files in index substitution method which in turn causes searching complexities.

Ahmad Steef, M. N. Shamma and A. Alkhatib, designed an algorithm based on public-key cryptography[3]. ASCII code is used to encrypt and decrypt a message, the message text is converting into binary representation by RSA algorithm. Finally, dividing this binary representation of the message to bytes (8s of 0s and 1s) and applying a bijective function between the group of those bytes and the group of characters of ASCII and then using this mechanism to be compatible with using RSA algorithm.

Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh, explained a new encryption method for text files using Elliptic Curve Cryptography. The classic technique used for mapping the characters to affine points in the elliptic curve has been removed [4]. The corresponding ASCII values of each plain text are paired, the paired value is works as input of elliptic curve cryptography. It avoids costly operation of mapping and the need of sharing common lookup table between sender and receiver. Finally, the algorithm is used to encrypt any type of script message.

Arafat Awajan & Enas Abu Jrai, jointly takes the credit for proposed a hybrid technique for Arabic text compression [7]. A hybrid technique that comprises of the linguistic features of Arabic language to improve the compression ratio of Arabic texts is performed by the authors. This technique works in phases. In the first phase, the text file is split into four different files using a multilayer model-based approach followed by; each one of these four files is compressed by the Burrows-Wheeler compression algorithm. This approach gives a better compression ratio apart from traditional model.

Pooja Jain et.al discuss on the improvement of data compression ratio by using the optimization of Lempel –Ziv - Welch and adaptive Huffman algorithm [8]. A comparative study is made based on the compression ratios between various existing algorithms. The author's algorithm proves to have better compression ratio for different files as well as different file size than other available data compression techniques.

III. PROPOSED ALGORITHM

The proposed algorithm is designed based on the combination of both compression and encryption. Process of the entire mechanism is explained below in detail.

A. A new TextComp16 Compression Technique for Text files using Matrix Mathematical Model

The proposed compression algorithm is designed to compress text files under lossless technique. The text input file has alphabets, numbers, special characters etc. ASCII (American Standard Code for Information and Interchange) values are used to fulfill the process of text compression. Particularly ASCII values ranging between 32 to 127 printable characters are used for further processing. Generally every character is having position to represent a numeric value. The proposed compression method is initiated with alphabet positions for conversion process.

The given text file is divided as groups each having six characters. Each character is allotted a position and partitioned. The concept of multiplicative array is processed between the character positions. The outcome of multiplicative array is stored on variable A. The process of additive array also applied on these same character positions. The outcome of the additive array is stored on variable B. Now the difference between A and B is calculated and the difference value is stored on variable D. Based on the difference value the character conversion process is taken place.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

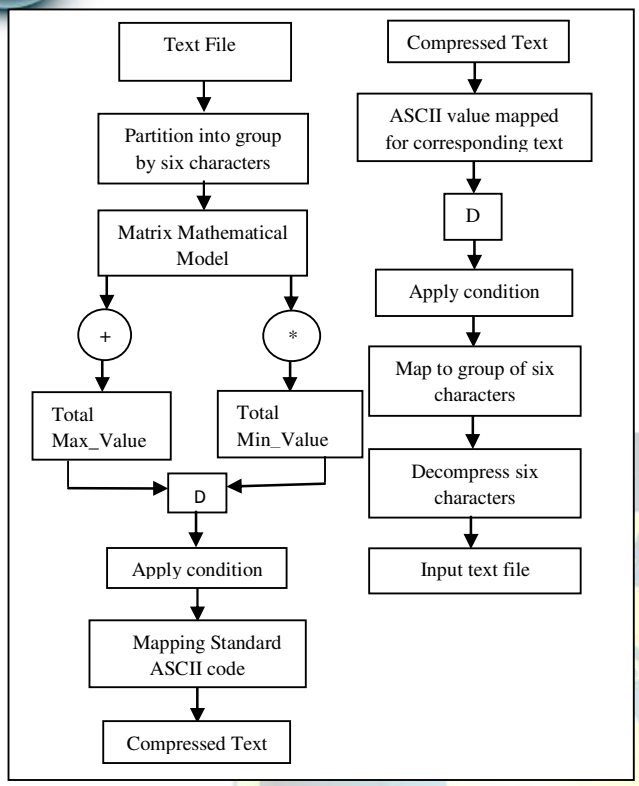

Fig. 1. Compression and Decompression Process

The final phase of the compression method depends fully on the ASCII values. Based on the difference (D) value, the compression mechanism chooses ASCII values to provide final compressed text data. If difference value is less than 32, it is needed to add 32 to maintain D always as greater than or equal to 32.

If D value is greater than 32, just write corresponding ASCII values instead of numbers (D). If the difference value is greater than 32 and less than 127, just print equivalent ASCII value as output. Subtract 32 from D, if the value of D is greater than 127 and less than or equal to 160. If D is greater than 160 and less than 256, subtract 127 from D. Output of the above mentioned procedure, finally, is printed as prefixes of ASCII values.

Table 1. Results of Compression Algorithm

Input text file (size in byte)	Compressed file(size in byte)	Compression Ratio (%)	Saving Space (%)	Speed (s)
132	18	13.63	86.37	0.053
117	15	12.82	87.18	0.054
176	23	13.06	86.94	0.064
352	51	14.48	85.52	0.106

B. RSA Cryptosystem

Rivest, Shamir and Adelman published one public key cryptography named RSA algorithm on 1978. It uses two different keys, public key is known to everyone while the private is kept as a secret. The authorized users only know how to open the message. The encryption ratio of RSA algorithm is high and processing speed is also fast. Key length of this algorithm is >1024 bits. The method of timing attack provides security from the cryptanalytic. Block size of RSA algorithm is 446 bytes and 1 round for encryption. RSA is implemented using stream cipher. Loss will arise while decrypt the data. For key generation choose two distinct prime numbers p and q. compute n=p*q. compute $\varphi(n)=\varphi(p)\ \varphi(q)$, where φ is Euler's function and this value kept as secret. Choose an integer e such that $1 < e < \varphi(n)$ and $\gcd(e, \varphi(n))=1$ and $\varphi(n)$ are coprime. Determine d as $d = e^{-1} \pmod{(\varphi/n)}$.

Encryption: $c = m^e \bmod n$

Decryption: $m = c^d \bmod n$

C. ElGamel Cryptosystem

The ElGamel algorithm is one of the asymmetric key cryptographic techniques to encrypt and decrypt the message. ElGamal is the predecessor of DSA. Generally the ElGamel cryptosystem is used in a hybrid cryptosystem[5]. For example, the message itself encrypts using symmetric cryptosystem and ElGamal is then used to encrypt the key that is used for the symmetric cryptosystem. This is because asymmetric cryptosystems like Elgamal are usually slower than symmetric ones for the same level of security. So it is faster to encrypt the symmetric key (which most of the time is quite small if compared to the size of the message) with

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

Elgamal and the message (which can be arbitrarily large) with a symmetric cipher.

For key generation of ElGamal cryptosystem, Alice generates an efficient description of a cyclic group G of order q with generator g. Alice choose the value x randomly from { 1,……, q-1} and computes $h = g^x$. For encryption Bob chooses a random y from { 1,……, q-1}, and calculates $c_1 = g^y$. shared secret $s = h^y$ calculates by Bob and maps the secret message m with an element $m^!$ of G. finally Bob calculates $c_2 = m^! . s$Bob sends the ciphertext ,

$$(c_1, c_2) = (g^y, m^! . h^y)$$

To decrypt cipher text (c_1, c_2) with private key x, Alice calculates shared secret $s = c_1^x$. Then Alice computes $m^! = c_2 . s^{-1}$.

$$C_2 . s^{-1} = m^! . h^y . (g^{xy})^{-1} = m^! . g^{xy} . g^{-xy} = m^!$$

D. Elliptic Curve Cryptography

Elliptic Curve cryptography is a public key cryptography in the recent researches. ECC depends on the hardness of the discrete logarithm problem[6]. Let P and Q be two points on an elliptic curve such that kP = Q, where k is a scalar. Given P and Q, it is hard to compute k; k is the discrete logarithm of Q to the base P. The main operation is point multiplication. The important features in Elliptic Curve Cryptography are: smaller key size, working speed is faster than RSA and Good for handhelds and cell phones. For key generation Alice and Bob chooses a finite field fp over an Elliptic curve E and choose base point B. Alice choose a random secret integer e and calculate eB € E. Bob choose a random integer d and calculate dB € E. now eB and dB are public and e,d are secret.

Finally Alice and compute edB = s = (s_1, s_2). Then compute k = (s_1 * s_2) mod n and compute c = (k * m), send it to Bob. Alice receive c and decrypt it as follows: Compute k = (s_1 * s_2)mod N and also compute k^{-1} mod N, where N is a highest prime number. Finally got original message as,

$$k^{-1} * c = k^{-1} * k * m = m$$

Table 2. Compare key sizes of three Cryptographic algorithms

ECC Key Size(bit)	RSA key size(bits)	Key size Ratio	AES key size(bits)
256	3,072	1:12	128
384	7,680	1:20	192
512	15,360	1:30	256

E. Compound Compression and Encryption Technique

Enormous crypto techniques are designed to secure data while transmitting them among more than two end users yet it needs more and more techniques to maintain data secrecy. One new security mechanism called dynamic usage of crypto algorithm with compression is designed here to save text data. The procedure of this algorithm is as follows: read input text file from the user. The new TextComp16 technique is applied on the Input text file. The process of this compression technique is explained above in detail. Now the compressed text file is split here. Since a maximum number of three segmented files are created from the compressed input file, a maximum of three asymmetric encryption algorithms are utilized in the encryption process. These three algorithms are :RSA, ElGamal and ECC. The selection process of the encryption algorithm for segmented file is also designed newly. After the process of encryption, the maximum of three encrypted text file is created. Finally these three encrypted files are combined and converted into the form of graphical representation.

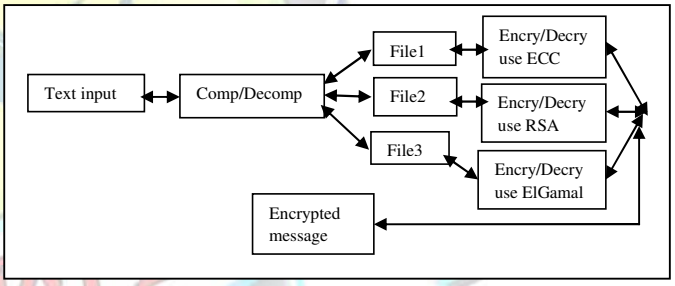

Fig. 2. Compound Compression and Encryption Technique

IV. CONCLUSION

The process of compression in encryption algorithm is newly designed here to secure the text files. Lot of encryption algorithms based on both symmetric and asymmetric techniques that are designed for text files. Generally most of the encryption algorithms using more than two keys to encrypt data for provide security. But the proposed encryption algorithm working differently. Before enter into the encryption part, the input text file is encrypted and segmented. The segmented files dynamically choosing an asymmetric algorithm (RSA, ElGamal, ECC) to encrypt it. So the proposed encryption algorithm on segmented files is changing dynamically. So this novel encryption technique provides highest level of security with high speed and the process of encryption also fast.

REFERENCES

[1] Debashis Chakraborty, Debajyoti Ghosh and Piyali Ganguly, "A Dictionary based Efficient Text Compression Technique using Replacement Strategy", in International Journal of Computer Applications, Volume 116 – No. 16, April 2015, pp: 19-23.

[2] Strahil Ristov and, Damir Korenci, "Using static suffix array in dynamic application: Case of text compression by longest first substitution", in Information Processing Letters, Vol: 115, 2015, pp: 175–181.

[3] Ahmad Steef, M. N. Shamma and A. Alkhatib, "RSA Algorithm with a New Approach Encryption and Decryption Message Text by ASCII" in International Journal on Cryptography and Information Security (IJCIS), Vol. 5, No. 3/4, December 2015.

[4] Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh, "Implementation of Text Encryption using Elliptic Curve Cryptography" in Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 73 – 82

[5] https://en.wikipedia.org/wiki/ElGamal_encryption

[6] http://www.ccs.neu.edu/home/riccardo/courses/cs6750-fa09/talks/Ellis-elliptic-curve-crypto.pdf

[7] Arafat Awajan & Enas Abu Jrai, "Hybrid Technique for Arabic Text Compression", in Global Journal of Computer Science and Technology: C Software & Data Engineering Volume: 15, Issue: 1, 2015, pp: 1-7.

[8] Pooja Jain, Anurag jain and Chetan Agrawal, "Improving Data Compression Ratio By The Use Of Optimality Of Lzw & Adaptive Huffman Algorithm (OLZWH)", in International Journal on Information Theory,Vol.4, No.1, January 2015.

[9] A. Biryukov, A. Shamir, D.Wagner, "Real time cryptanalysis of A5/1 on a PC," fast software Encryption, LNCS 1978, B. Schneier, Ed., Springer-verlag,2000,pp.1-18.

[10] Branovic, R.Giorgi and E.Martineli "Memory performance of public-key cryptography methods in mobile environments".