Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 2, Number 5; April-June, 2015 pp. 473-477

http://www.krishisanskriti.org/acsit.html

Improved Techniques for Secure Data Communication System

Jyoti Gupta[1], Asiya Raza[2], Jasbeer Kaur[3] and Garima Verma[4]

[1,2,3,4]CSE Department Galgotias college of engineering and technology, Greater Noida, India
E-mail: [1]Jyoti.suresh.gupta@gmail.com, [2]asiya.raza.rose@gmail.com,
[3]jasbeerkaur.49@outlook.com,[4]garimaverma11@gmail.com

Abstract— *Data compression technique plays a significant role in the area of data transmission for reducing time to transmitting and retrieving data as well consuming a less bandwidth of network and decrease the space required to store it. However, this technique does not make data immune to attack, it may contain secret information from being tampered or hacked by malicious attacks even in the small area network environment. Apart from the reducing data volume, an implementation of encryption is an important concern in order to transmit data over a network. This paper presents feasible approach in ensuring its security and minimizing quantity of data .The data is first encrypted by using an encryption algorithm namely DES or AES followed by a suitable compression algorithm such as Huffman coding Shannon Fano, Lempel-Ziv-Welch (LZW) etc. We propose a tool involving cryptanalysis making the communication system secure my increasing entropy of system along with bringing in anonymity as the best compression algorithm will be picked up by analyzing the best cases and previous statistics by making use of data mining technique. Thus, a lightweight security strategy which includes encryption and compression is an efficient solution to the security issue and to reducing file sizes along with increasing the entropy of the system and bringing in anonymity by introducing encoding scheme achieved by selecting the most appropriate compression technique followed by encryption.*

Keywords: *Compression and encryption, CE, EC, JCE, Huffman Coding, Advanced Encryption Standard*

1. INTRODUCTION

Cryptography is a Greek word, which means “secret writing”. It is the science and art to transform the information to make it secure and resistive to security attacks. It is the technique to provide secure communication in presence of adversaries to maintain information securities such as data integrity, non-repudiation, authentication and data confidentiality. The process to convert the plain text into the cipher text in cryptography is called encryption.

The cipher text is understandable only to intended user, someone who knows how to decrypt it. The message or information (plain text) is encrypted using an encryption algorithm. This is usually done with the use of a key (shared or secret), which specifies how the information is to be encoded. Any malicious user that can see the cipher text should not be able to determine anything about the original message. An authorized user, however, is able to decode the cipher text using a decryption algorithm which usually requires a secret decryption key. Encryption schemes are of two types

Symmetric-key -In this scheme secret key (same key) is used for encryption and decryption it is also known as the secret key encryption

Asymmetric key -In this scheme different keys (public and private) are used for encryption and decryption it is also known as the public-key encryption.

The remainder of this paper is structured into five major parts: Section 2 addresses the motivation, section 3 addresses related work, Section 4 discusses how the proposed methodology is being implemented and Section 5 shows the result. Finally, Section 6 concludes the paper and discussion on the future.

2. MOTIVATION

In particular, this work emphasizes on finding a solution to protect and strengthen communication system which will ultimately enable a secure channel for communication. The ultimate aim is to not allow an eavesdropper or an unauthorized users from accessing the data which is meant to be securely transferred .This is done by cryptanalysis ,more specifically we follow an approach whereby encryption and compression techniques are applied .The compression algorithms themselves in a way encode the encrypted data and if we wish to follow a more advanced approach we can encode each character of compressed data by dictionary encoding to further enhance the security. In addition, it enhances the entropy of the overall system as the tool designed can serve users by itself selecting the most appropriate compression technique based on best cases ,worst cases and average cases of compression algorithms and also on the basis of past results thus the system can be made to learn and itself choose the appropriate technique. Thus, an intruder not only as to break the encoding done but also has to get the key for intruding in the communication system.

3. RELATED WORK

Considering the security issue and limitation of transmission capacity during transmitting information over insecure communication channel, it is necessary to implement a system with security and compression encoding efficiency. In Encryption followed by Compression (EC) size of data do not increase but intruder may have more clues to access data. Furthermore, data are not compressed efficiently due to decrease of sequence size. There are many current compression and encryption techniques of Compression followed by Encryption CE approach that have been presented and analyzed. The author in [1] proposed the use of lightweight symmetric encryption such as AES for securing data and asymmetric encryption using Diffie Hellman Key Exchange method for key exchange protocols in TFTP. In [2], the authors present a preliminary study on transferring variable size of encrypted data using AES through TFTP. This lightweight encryption is compatible to be implemented on TFTP as this protocol itself acts as simple file transfer protocol. However, considering the original TFTP(current versions can support up to 4GB) as a simple file transfer protocol, file size that can be sent is limited to not more than 32 Mb[3][4]. Symmetric encryption of block cipher is considered to be implemented in constrained environment because this algorithm uses less energy as well as the performance of encryption throughput is better in smaller packet structure compared to stream cipher (RC4) [5]. Furthermore, the author in [6] also presented comparative analysis between block cipher and stream cipher. It showed that stream cipher RC4 is faster and consume less power compared to block cipher (AES) but it is not efficient in encrypting small packet size. The author in [7] presented an overview of research work on data compression and encryption technique which is categorized into CE, EC and JCE categories. M. Chaudari and K. Saxena in his paper [8] proposed fast and secure data transmission using symmetric encryption, DES and two types of lossless compression, Huffman coding and LZW. It is concluded that LZW compression is effective for large text files.Besides, the author in [9] proposed on a secure compression algorithm using different compression techniques such as Arithmetic coding, Huffman Coding, Lempel-Ziv coding (LZW), Prediction by Partial Matching and Burrows Wheeler. Another technique which performs both lossless and encryption schemes is presented in [10] based on SCAN patterns generated by the SCAN methodology using binary and gray-scale images. Based on above related works, in order to transfer information fast and securely in a closed system, this paper proposes an implementation of Huffman compression to reduce file size as well as minimize time to retrieve and transmit data, and symmetric encryption for protecting data content. Which. An intelligent system based on automatic selection has also been proposed.

4. PROPOSED METHODOLOGY

Minimization of data size and securing data are some of the important aspects of data transmission. Strategies such as these prevent scoundrel from getting information without permission as well as shorten the transmission time. Even in a local area network such as research laboratory, school or at office and home, it becomes imperative to pay attention to security issue. Data transmission with encryption and compression is one of the best techniques to secure data as well as in a way encoding is done. Compression technique reduces the size of data which also minimize transmission time whereas encryption technique protects data from being read or tampered by an unauthorized user as the original data known as plaintext is changed into unreadable text known as cipher text. Compression can be categorized into lossless and Lossy techniques. If the reconstructed data from the compressed data is identical to the original data, it is a lossless compression otherwise it is a lossy compression. If we want to look into the protocol that will be involved while using a lossless compression techniques ensure that files can be downloaded by clients reliably using TFTP. Huffman coding, LZW method are examples of lossless data compression technique. Moving on to address security requirements for data, the first step would be to implement cryptographic protocols to protect its data against malicious attack during transmission. Encryption process is generally used in computing system which can be classified into symmetric key encryption that uses single key for both sender and recipient to encrypt and decrypt data, and asymmetric key encryption that uses two keys, a public key known by everyone and private key that is only known by the recipient to decrypt data where the value of public key and private key are different. Nevertheless, it is impossible to use common cryptographic encryption with extremely constrained resources available. Therefore, there is a need to utilize an ideal encryption algorithm in a limited local area network which has low supply of essential resources of that protocol. Classification and description of compression and encryption schemes related to the aforementioned scenario are explained as follows:

1) Individual compression and encryption

Compression followed by Encryption (CE): This technique gives an adversary less chance to access data but encryption may increase the size.

- Encryption followed by Compression (EC): Size of data does not increase but adversary may have more clues to access data .Furthermore, data are not compressed efficiently due to decrease of sequence size of data.

2) Joint Compression and Encryption (JCE): This approach is recently used as it is faster compared to the above two techniques but its procedure is complex.

This paper focuses on EC technique in which encryption of the data using symmetric encryption, namely the DES algorithm and this is followed by compression which is accomplished using the lossless compression techniques. Data will be transmitted using TFTP in local network between two computers, data which is encrypted and compressed by server will be sent to client side which will then decompress and decrypt it using same algorithm. Fig. 1 below shows steps involved in transmitting the file between client and server based on proposed EC approach where data is secured by DES encryption algorithm then compression is performed by compression algorithm:

1) Plaintext file is encrypted by sender using DES algorithm.

2) Encrypted file is compressed using suitable compression Algorithms and finally send the encrypted and compressed file to the authorised receiver.

3) Intended receiver only will have the entire tool consisting of all the decompressing techniques so will be able to decompresses the file. It is effective and simple as the data mining statistics after analysing the fed encrypted file will apply the most suitable compression technique and thus encoding of all symbols will be done.

4) Only the authorised receiver will have access to the entire tool which will consist of the decompression techniques. Any unauthorised eavesdropper will neither have the access to the key nor the knowledge of the decompression and most importantly if dictionary based encoding is done then the channel will be triply protected. And finally after decompression, decryption of the file will be done by the authorised or intended receiver.

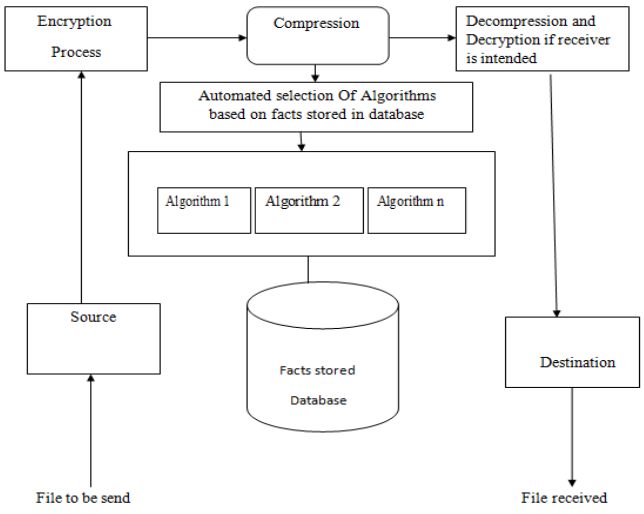

Fig. 1: Schematic diagram on proposed methodology

A. DES Algorithm

The DES (Data Encryption Standard) is a symmetric-key algorithm which uses a block cipher ,published by the National Institute of Standards and Technology (NIST).Data Encryption Standard is based on a cipher known as the Feistel block cipher.. It encrypts the data in a block of 64 bits. It produces 64 bit cipher text, key length is 56 bits. The key is consisting of 64 bits. The bit position 8, 16, 24, 32,40,48,56, 64 discarded from the key length.DES is based on two fundamental attributes of Diffusion (Substitution) and Confusion, (Permutation) consisting of 16 rounds. It consists of a 16 Feistel rounds , each round perform bit-shuffling by using Permutation (P-boxes) ,substitutions (S- boxes) and exclusive OR operations Confusion and Diffusion increases at each round It encrypts the data in block size of 64 bits each. *B. Compression Techniques*

1. Huffman Coding:

Huffman coding used a variable-length code table for encoding a source symbol, variable-length code table has been derived based on the estimated probability of occurrence of the source symbol [11]. Huffman code procedure is based on following observations:

a. Sort or prioritize code word based on number of occurrences of symbol in text.

b. The symbols that occurs more frequently have shorter code words

c. The Huffman code is implemented by merging the lowest probable symbols and this process is continue until only two probabilities of two compound symbols are left and thus a code tree is generated and Huffman codes are obtained from labelling of the code tree.

2. Shannon Fano

According to Shannon's compression, the optimal code length for a symbol is –log b P, b is the number of symbols used to make output codes and P is the input symbol probability. Similar to the Huffman coding, first frequency table is generated and then a particular procedure is followed to produce the code table from frequency.

3. Run Length Encoding (RLE)

Run Length Encoding is the simplest type of the data compression algorithms. It replaces runs of two or more of the same character with a number which represents the frequency of symbol in text, followed by the original symbol. Single symbol is coded as runs of 1. The major task of this algorithm is to identify the run of the each symbol in text file, and to record the symbol and the length of each run. The RLE algorithm uses those runs to compress the original source file while keeping all the non-runs without using for the compression process.

Example -

Input: AAABBCCDDEEEB

Output: 3A2B2C2D3E1B

4. Lempel–Ziv–Welch (LZW)

LZW is created by Abraham Lempel, Jacob Ziv, and Terry Welch. It was published by Welch in 1984 as an improved version of the LZ78 algorithm published by Lempel and Ziv in 1978, algorithm is simple to implement, and has the potential for high throughput in hardware implementations.LZW is based on parsing the sequence of pixels and consider the resulting phrases as events to be encoded. The parsing method **is** based on a tree structure. The number of branches of every tree node depends on the number of states of tree. Therefore, for a binary image, there are only two states, while for **8-**bit images, there are 256 states. This algorithm is adaptive in the sense that it starts with an empty table of symbol strings and builds table during both the compression and decompression processes. This adaptively results in a poor compression in the initial part of the sequence; as a result, the sequence must be long enough to allow gaining enough symbol frequency experience; on the other hand, most finite implementations of an adaptive algorithm lose its ability to adapt after a certain sequence length has been processed.

To enhance the anonymous nature of the tool being proposed and to facilitate enhanced security we also consider the Gzip and Cosmo Compressor and other related compressors as more the number of compression techniques greater is the enhancement in the entropy and encoding gets anonymous.

C. The Actual Methodology Proposed

Based on parameters like compression ratio, time take to compress the encrypted file, the size of the compressed file we prepare a database. We store facts relating to the kind of input fed ,analyze the best case, average case and worst cases of algorithms used apart from the time and size details.

After this we will have the training data set and test data set ,by deciphering the patterns found, as to which parameter gained more impact when a particular type of input type was fed we can make the system learn and thus when a similar kind of situation will emerge the best compression technique will automatically be put to use.

It is interesting to note that only the authentic user will have the knowledge as to which compression technique was selected and finally the authorized user will decompress the file and decrypt it as he will have tool consisting of the decompression module and will know which algorithm will be used to decompress it and finally decrypt it as the authentic user will have the required key.

5. EXPERIMENTAL SETUP AND RESULTS

An experimental setup is established as below:

- Server: Desktop dual-core AMD Opteron(tm) Processor 1218 2.60GHz
- Client: Desktop Intel Xeon® processor CPU E312703.4GHz
- File sizes range: 10Kb to 30Mb.

File is sent through in local area network to decompress and decrypt it. Java language programming is used to code the proposed tool.

Performance evaluation factors on average of file reduction percentage and execution time for file transmission are obtained .Security was enhanced, apart from key, compression also in a way encoded the encrypted symbols.

Based on improved results, there are some points of considerable importance:

i. Original file is the text file without encryption or compression. All files are in txt format.
ii. Encrypted-only file is the file protected with DES encryption algorithm. Key size used is 64bits.
iii. Compressed-only file is the file compressed with one of seven techniques used for compression. For instance, if encoding is also used along with compression. The encoder has 2 pass encoders; the first pass generates Huffman tree and the second pass encodes the data.
iv. Encrypted-compressed file is the file compressed with the best compression technique which was decided base on previous results which were stored as facts in the database and then this file sent to the intended receiver.

6. CONCLUSION AND FUTURE WORK

From the above result, it is concluded that the encryption and compression of data is a significant method in order to secure transmission and reduce file size as well as save more time. Combing both approaches and at the same time making the selection of the most appropriate compression technique which acts as a encoding mechanism (Although separately also we may encode each symbol of the compressed file).The most appropriate algorithm is selected on the basis of various parameters and previous results as recorded previously. The future work will include all types of file to be fed to the tool proposed apart from this the system can be made fully intelligent by taking into consideration all the parameters and analysing all cases of the algorithms used.

REFERENCES

[1] M. A. M. Isa, N. N. Mohamed, H. Hashim, S. F. S. Adnan, J. A. Manan, and R. Mahmod, “A lightweight and secure TFTP protocol for smart environment,” in 2012 International Symposium on Computer Applications and Industrial Electronics (ISCAIE), 2012, pp. 302–306.
[2] N. N. Mohamed, H. Hashim, Y. M. Yussoff, and A. M. Isa, “Securing TFTP packet: A preliminary study,” 2013 IEEE 4th Control Syst. Grad. Res. Colloq., pp. 158–161, Aug. 2013.
[3] “rfc783,” 1981. [Online]. Available: http://tools.ietf.org/pdf/rfc783.pdf.
[4] “rfc2347,” 1998. [Online]. Available: http://tools.ietf.org/pdf/rfc2347.pdf.
[5] P. Prasithsangaree and P. Krishnamurthy, “Analysis of energy consumption of RC4 and AES algorithms in wireless LANs,”

GLOBECOM ’03. IEEE Glob. Telecomm. Conf., pp. 1445–1449, 2003.

[6] H. O. Alanazi, B. B. Zaidan, A. A. Zaidan, H. A. Jalab, and M. Shabbir, “New Comparative Study Between DES , 3DES and AES within Nine Factors,” J. Comput., vol. 2, no. 3, pp. 152–157, 2010.

[7] Razzaque and N. V. Thakur, “Image Compression and Encryption_: An Overview,” Int. J. Eng. Res. Technol., vol. 1, no. 5, pp. 1–7, 2012.

[8] M. Chaudhari and K. Saxena, “Fast and Secure Data Transmission using Symmetric Encryption and Lossless Compression,” Int. J. Comput. Sci. Mob. Comput., pp. 58–63, 2013.

[9] E. Celikel and M. E. Dalkilic, “Experiments on a secure compression algorithm,” in International Conference on Information Technology: Coding and Computing,pp. 150– 152.

[10] S. S. Maniccam and N. G. Bourbakis, “SCAN based lossless image compression and encryption,” Proc. 1999 Int. Conf. Inf. Intell. Syst. (Cat. No.PR00446), pp. 490–499, 1999.