ROLE OF DATA COMPRESSION IN CYBER SECURITY

Prof. Swati D. Ghule
Research Scholar TMV
gadeswati@rediffmail.com

Dr. Anup Girdhar
Ph.D. Guide, TMV
anupgirdhar@gmail.com

ABSTRACT

In a digital age due to increase in communication technology there is a huge demand of information security. Data is frequently transmitted via computer networks. Data compression is a technique which makes the communication faster. Transmission of information across untrusted network raises the issue of security of information. Encryption is one way to translate data so that unapproved person cannot recognize its content. For solving network bandwidth and security problem compression and encryption technologies need to be combined. Theoretically it is believed that encryption and compression are two independent processes. But independent compression and encryption of data are slow to meet the demand of many applications. And encrypted files are incompressible due to their randomness and most compression algorithms fail to isolate such redundancies. So compression has to be performed before encryption. The idea is to push encryption operation into compression process and perform them in a single step. Especially by means of dictionary based encoding compression and encryption are addressed simultaneously as a single simplified process. Adjusting encryption into compression process is preferable.

KEYWORDS: *Cryptography, Data Compression, Encoding, Encryption.*

I. INTRODUCTION

Data is frequently transmitted via computer networks. Transmission of information across untrusted network raises the issue of security of information. To protect data from unauthorized access, guarantee their content from change and prevent them from network attacks during transmission. Reliable transmission of data is needed in many applications such as military operations, business transaction and medical systems. The magnitude of data sent has become bigger and thus data compression is becoming a vital tool [1]. And cryptography is the technique for sending message secretly. Independent compression and encryption of data are slow to meet the demand for many applications. One idea is to push encryption operation into compression process and perform them in a single step [11]. Especially by means of dictionary based encoding compression and encryption are addressed simultaneously [4] as a single simplified process.

II. BACKGROUND

The main objective of data compression is to find out the redundancy and eliminate them through different methodology; so that the reduced data can save space: to store the data, time: to transmit the data and cost: to maintain the data. To eliminate redundancy, the original file is represented with some coded notation and this coded file is known

as encoded file. For any efficient compression algorithm this file size must be less than the original file. To get back the original file it needs to decode the encoded file [8].

Theoretically it is believed that encryption and compression are two independent processes. However, because encrypted files are incompressible due to their randomness and most compression algorithms fails to isolate such redundancies compression has to be performed before encryption [8]. Figure 1 depicts the standard way to combine data encryption and compression.

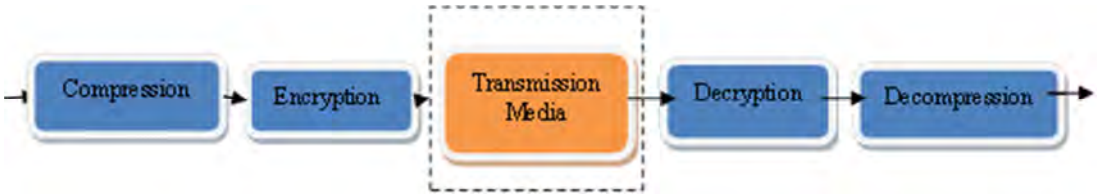

Figure 1: Image Compression and encryption

III. DICTIONARY BASED ENCODING

Mainly there are two types of data compression: Lossy and Lossless. In first type some data is lost at the output which is acceptable and in lateral one, no data is lost and the exact replica of original file can be retrieved by decoding the encoded file. In this type of compression generally the input file is encoded first which is used for storing or transferring data and then at the output the file is decoded to get the original file [8]. Following are the coding techniques which comes under lossless data compression

1. **ENTROPY BASED ENCODING**: It first counts the frequency of occurrence of each unique symbol is given text and then it is replaced by the unique symbol generated by the algorithm. The frequency of symbols varies with respect to the length in given input file [8].

2. **DICTIONARY BASED ENCODING**: Another name is substitution encoding. A data structure known as dictionary is maintained throughout the process. Dictionary based compression uses two kinds of dictionary namely static and adaptive. The static dictionary is built up before compression occurs and it does not change during processing of data. The receiver of message cannot reconstruct the same dictionary by processing the decompressed file. So the dictionary has to be transmitted along with the file that results in certain add to compressed file [6]. Adaptive dictionary scheme helps to avoid this problem by reconstructing the dictionary while data is compressed.

This data structure consists of number of strings. The encoder matches the substring chosen from the original text and finds it in the dictionary. If a successful match is found then the substring is replaced by a reference to the dictionary in the encoded file. The complete LZ family of encoding comes under this technique. LZ string encoding creates a dictionary of encountered strings in a data stream.

The introductory [13] dictionary based algorithm is the beginning of nearly all well-known dictionary compression algorithm such as gZip, Zip RAR and LZMA [5]. The most popular one is standard dictionary based data compression technique LZW.

At first the dictionary only contains each symbol in the ASCII alphabet where the code is the actual ASCII code. Whenever a new string appears in the data stream the string is added to the dictionary and given a

code. When the same word is encountered again the word is replaced with the code in the outgoing data stream. If a compound word is encountered the longest matching dictionary entry is used and over time the dictionary is builds up strings and their respective codes. In some of the LZ algorithms both compressor and decompress or needs to construct a dictionary using the exact same rules to ensure that the two dictionaries match [8]. The method is also found in several image file format such as GIF, TIFF, PDF. By changing the management of dictionary of LZW, a random dictionary is formulated [4]. If file contains repetitive data LZW compression works best.

After the compression and decompression method are fixed then apply encryption on the top of this process. Upon communication to receiver send the encoded file and model parameters encrypted with AES counter mode. The legitimate receiver provided with right key will then recover the encoded file together with model parameter [3].

IV. ENCRYPTION

In a digital age due to increase in communication technology there is a huge demand of information security. Data is extensively used in numerous fields of society and equivalent to its practice, it is essential to keep them safe [9]. Encryption is one way to scramble data so that unapproved person cannot recognize its content. Several text encoding schemes such as DES, IDEA, AES and RSA were used for data.

AES is symmetric block cipher. So information that is to be encrypted will be broken up into 128 bit blocks and the operation will be carried out in blocks. The standard implementation of AES applies 128 bit block size (16 byte). The 16 bytes are treated as 4*4 byte matrix. This matrix is referred to as state array. Padding is added if information is not multiple of 128 bits. The algorithm works for different encryption key lengths [3].

The mode of operation selected determines the security of encryption. Different encryption methods such as Electronic Code Book (ECB), Cipher Block Chaining (CBC), Cipher Feedback (CFB), Output Feedback (OFB) and Counter Mode (CTR) are supported for symmetric block cipher [7]. Most cryptographers recommends the CTR mode for encrypting large files as it is appropriate to function on multi-processor device where blocks are encoded in parallel. The cipher operation takes place for each block using secrete key and unique counter. The counters are started from an initial random value and increment it to process the next block this ensures uniqueness of cipher. In CTR mode the counter bits are encrypted together with secrete key and then XORed with the input plain text blocks. So the input plain text is not directly encrypted. Because each block in CTR mode is independent of each other, CTR mode runs more rapidly than other modes of AES without losing the security requirement of secure cipher. For many purposes AES guarantees the tightest security currently available. A contemporary trend developed due to computational requirement is selective encryption [3].

OpenSSL is one of the most broadly used cryptographic libraries. It implements different cryptographic procedures such as symmetric key encryption, public key encryption, hash function, digital signatures random number generation etc. In general for encrypting large amount of data less computationally demanding algorithms are chosen [3].

V. COMPRESSION AND ENCRYPTION METHODS

In recent years study has been done to combine compression and encryption for secure data communication. Several

procedures have been suggested to combine compression and encryption so as to reduce the whole processing time. However, they are either insecure or computationally rigorous [2].

VI. CRYPTOGRAPHY

Cryptography is concern with the transmission of information in presence of foe, who is working firmly to retrieve contents of data transmitted. The fundamental issue in cryptography is how to make secure communication over insecure network. The answer to this problem is to encrypt the information to be communicated i. e. converting data into a cipher a form that can't easily be understood. An encryption converts plaintext to cipher text and decryption does the reverse. An encryption system must have a property that decrypting a cipher text with appropriate key yields the original message that was encoded.

The key and cryptographic algorithms are two primary components of cryptography. The algorithm is mathematical function and used for applying protection for data, and the key is parameter used by function. The key determines an operation in such a way a person with appropriate key can reproduce the operations. The strength of key is vital since the reliability of entire cryptographic practices depends on them. It is important to utilize a longer key length to resists brute force attack. In computer cryptography integer keys are used. It then requires a generation of key using pseudo-random number generator (PRNG). PRNG is a function that initialized with random value called seed and output sequence of numbers that appears random. If an outsider does not know a value of seed then he cannot differentiate the output of PRNG from that of true random sequence [10]. The process is deterministic as it always produces the same sequence when it initialized with same seed. In cryptography a highly secured number is demanded that resists the attack in which nobody using advanced cryptanalysis methods can predict the output.

VII. CIPHER DESIGN

The cipher is designed to combine encryption and compression of text by attaching encryption as post-processing to LZW compression algorithm. The central idea is to develop a way of handling the dictionary formed in process of compression. The dictionary constructed will have same entries as in LZW but the entries are in some random order. The initial dictionary is filled with all single characters, but unlike to LZW their entry is picked arbitrarily. The dictionary entries are partially altered for each step of algorithm. The dictionary entries are permuted by the pseudo random number generator (PRNG) function G. This means from a set of predefined algorithm G produces a sequence of random integers that is indistinguishable from a true random but on real sense it is a deterministic process. It follows that encryption can be achieved by changing correct indices to some other random index value [12]. To process cipher encryption an initialization vector (IV) is required at the beginning. An IV is an arbitrary number that can be used along with secrete key for data encryption. It avoids repetition in data encryption which makes difficult for cracker to break cipher. One instance could be if there are repeated sequences in encrypted file, an attacker may assume that their corresponding sequences in plain message were also identical. In such cases IV will prevent the appearance of identical duplicate sequence in the cipher text.

VIII. CONCLUSION

Image compression and encryption are growing fields. Currently available methods can further analyzed for different types of data. Compression alone is not sufficient as it can be accessed. But the system where security embedded into compression algorithm provides safety and thus can meet

the demand of fast and secure transmission of data. Adjusting encryption into compression process is preferable.

REFERENCES

[1] Aldossari, M., Alfalou, A., & Brosseau, C. (2014) "*Simultaneous Compression and encryption of closely resembling images: application to video sequence and polarimetric images*". Optics Express, 22(19), 22349-68, doi:10.1364/OE.22.022349

[2] Cheng, H. (2000). "*Partial encryption of compressed images and videos*". IEEE Transaction on signal Processing, 48(8), 2439-2451. doi:10.1109/78.852023.

[3] Getachew Mahabie Mulualem, dissertation on "*Compression and Encryption for Satellite Images: A Compression between Squeeze Cipher and Spatial Simulations*", Enschede the Netherlands, February 2015 pp-10-20.

[4] Kelley, J., & Tamassia, R. (2014). "*Secure compression: Theory and Practice*". Dept. of Computer Science Brown University. Retrieved from http://eprint.iacr.org/2014/113.pdf

[5] Langiu, A. (2013). "*On parsing optimality for dictionary based text compression-the Zip case*". "Journal of Discrete Algorithms", 20, 65-70, doi:10.1016/j.jda.2013.04.001

[6] Nandi, U., & Mandal, J. K. (2013). "*Modified Compression Techniques Based on Optimality of LZW Code (MOLZW)*". Procedia Technology, 10, 949-956. doi:10.1016/j.protcy.2013.12.442

[7] Paar, C., Pelzel, J., & Preneel, B. (2011). "*Understanding Cryptography: A Textbook for students and Practionars"* [E-book] Available through www.Springer.com/in/book /9783642041006 [Accessed 10th Mar 2017].

[8] Samish Kamble, Prof S. B. Patil, "*FPGA based Data Compression using Dictionary based 'LZW' Algorithm*", International Journal of Latest Trends in Engineering and Technology, Volume 5 Issue 4 July 2015, ISSN: 2278-621X.

[9] UR Rehman, A., Liao, X.,| Kulsoom, A., & Abbas, S.A. (2014). "*Selective encryption for gray images based on chaos and DNA complementary rules*". Multimedia Tools and Applications. doi: 10.1007/s11042-013-1828-7

[10] Van Tilborg, H.C.A., & Jajodia, S. (Eds.). (2011). "*Encyclopedia of Cryptography and Security*". Boston, MA:Springer US. DOI:10.1007/978-1-4419-5906-5.

[11] Xiang, T., Qu, J., & Xiao, D. (2014). "*Joint SPIHT Compression and selective encryption*". Applied Soft Computing, 21, 159-170. doi:10.1016/j.asoc.2014.03.009

[12] Xie, D., & Kuo, C.-C. J. (2005). "*Secure Lempel-Ziv Compression with embedded encryption*". In E. J. Delp III & P.W. Wong (Eds.), Electronic Imaging 2005 (pp-318-327). International Society for Optics and Photonics. doi:10.1117/12.590665.

[13] Ziv, J., & Lempel, A. (1978). "*Compression of individual sequences via variable-rate coding*". IEEE Transactions on Information Theory, 24(5), 530-536,doi:10.1109/TIT.1978.105593.