Scientific Journal Impact Factor (SJIF) : 3.518
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

IMPROVING COMPRESSION RATIO FOR ENCRYPTED SECRET MESSAGE (IN AES ENCRYPTION) BY USING GZIP COMPRESSION

Dr. Suhad Malalla[1] and Farah R. Shareef[2]
[1]*Assist Prof. Dr. Suhad Malalla, University of Technology, Department of Computer Science*
[2]*PhD student Farah R. Shareef, University of Technology, Department of Computer Science*

Abstract- One of the most popular encryption algorithms in existence today is AES, which provides better security and has less implementation complexity, it is a symmetric block cipher and this require more memory, it works on large chunk of data. So we need a good compression method like GZIP in order to reduce the size of encrypted message. This paper proposed a good combination between AES encryption algorithm and GZIP compression, which is give a good results for compression ratio for large encrypted message.

Keywords- AES encryption, GZIP compression.

I. INTRODUCTION

Cryptography is the mathematical techniques related to information security aspects like as confidentiality, entity authentication, data integrity, and data origin authentication. The cryptography secure the information by convert it into an unknown format. The original plaintext, or text is transformed to an equivalent coded known as “ciphertext” by used algorithm of encryption. Just on who has the secret key can decrypt (decipher) the ciphertext into original text. It can be classified the cryptography systems broadly to tow main types: symmetric-key and public-key systems. The first system (symmetric-key) used single key (password) which most have by both (sender and the receiver), while the second system (public-key) used two keys, a public key have by both (sender and the receiver) and a private key that only have by messages recipient [1].. In this paper, we will used only symmetric-key systems and we used AES algorithm for encryption the secret message.

Compression is an encoding process for data to be fewer bits than the original representation so that it gets minimize storage space and less consuming time when communicating over a network [2]. There are two kinds of compression, lossy and lossless.

In Lossy data compression, some information are losing. When the compression file is decompression, the output is not identical to the original data file. It is compatible for Movies, Sounds and Images. While in lossless, it is used with text files [3].

In compression the data, we can used for processing technique in-network so as to save energy because it decrease the size of data so as to decrease data transmitted and/or reduce transfer time because the amount of data is decreased [4].

II. THEORETICAL BACKGROUND

2.1 CRYPTOGRAPHY

Cryptography provides a very important tool to secure message (especially for messages transmission when it transfer from one location to another). The cryptography disguise the original message by convert it to unreadable form, just intended recipients (who have encryption “key”) can remove the disguise from message and read original message. The secret message may be encrypted using a “code”, or a “cipher” or 'cypher'. In case of “code” each one of characters or a group of characters will be replaced by an alternative one, in case of “cipher” the whole message is converted

instead of individual characters. The systems of Cryptographic commonly classified to three are three methodologies of independent dimensions [5].

1. Text to cipher text transforming Methodology: All algorithms of encryption are based on two common principles: transposition and substitution. In transposition the plaintext elements are rearranged wile in substitution each plaintext element is mapped into another element. The main requirement is not to lose any information be [6].

2. Keys number Methodology: There are many standards methods of cryptography like: hash function, public key, secret key, and digital signature [7]:

A. Symmetric-key cryptography: In symmetric-key encryption, each one of the sender and the recipient should have the code (secret key) that been used to encrypt a packet of information in sender side before it sent to receptor over the network that should have same key to decrypt it. There is two main types of Symmetric-key cryptography AES (Advanced Encryption Standard) and DES (Data Encryption Standard).

1. AES: this encryption algorithm is based on substitution-permutation network (SPN) which is a linked mathematical operations series that have been used in block cipher algorithms as in AES. This algorithm is fast in both software and hardware. AES differs from DES (its predecessor) it's not use a Feistel network. AES is a Rijndael variant that has a 128bits fixed block size, and (128, 192, or 256 bits) key size. By contrast, Rijndael the key and block sizes is any 32 bits multiple, both with a minimum of 128bits and a maximum of 256bits [8].

2. DES: is the an archetypal block cipher algorithm which is take a fixed-length string of plaintext bits then converts it to another cipher-text bit string with same length by a series of complicated operations. In DES the block size is 64bits. The DES uses a particular key for transformation process, so only persons that know the decryption key can view the original continent. Ostensibly the key is consists of 64 bits, but actually the algorithm are used only 56 of these bits, the remained eight bits have been used just to checking parity then it discarded. Thus, the length of effective key is 56bits [9].

B. Public-key cryptography: It also well known as "asymmetric cryptography", which is an algorithm of cryptographic that need two separated keys (public key and secret (or private) key). The "asymmetric" term come from the use of two different keys to achieve these opposite functions, as contrasted with symmetric cryptography that use a same key to achieve `both. In spite of differences, this two separated keys are mathematically linked. The public key is used to verify encrypted plaintext or the digital signature, while the private key has been used to decrypt cipher-text or generate a digital signature [10].

1. RSA: a crypto-system which is one of the first workable public-key crypt- systems which is vastly used to secure data transmission. For this type of crypto-system, the key of encryption is public and different from the key of decryption that is kept secret. The first publicly described RSA algorithm are Leonard Adleman, Adi Shamir and Ron Rivest in 1977. In 1973, an English mathematician (Clifford Cocks), had been developed system equivalent to RSA, but it was not declassified until 1997 [10].

C. Digital Signature: Digital signature has been used due to needs the ensuring of the authentication. The digital signature is like sender signature or stamp that embedded with data together and use private key to encrypt it then send it to other side. Additionally, the signature assures that receiver will be detect any change may be made to the data that has been signed [11].

D. Hash Function: It is a one way encryption, which is a mathematical formula or well-defined procedure which is represent a small size of bits that created from a file of big sized, the function result can be called hashes or hash code. The hash code generating is faster from other methods so it much

desired for integrity and authentication. Hash functions are more used for digital signature and it is highly desirable because of cheap constructions. Recently, the use of hash functions become a standard approach for message authentication in different applications, especially for internet security protocols. The integrity and the authentication considered an important issues in secure the information. It can be attached the hash code to the original file, then the users at any time be able to check the integrity and authentication after sending the secure data through put same hash function again to the received message then comparing the hash result to the sender hash code, if it's similar, it means that the received message are came from the original sender with no change in its content, because any changed in original data will changed the receiver side hash code [11].

1. Methodology for processing plain text.

The method of processed the plaintext. A two type of processes: stream and block cipher. In block cipher a one input block of elements are processed at a time, generating an output block for each input block, while the in stream cipher the input elements are processed continuously, generating one output element at a time, as it goes along [6].

2. 2 GZIP

It is a general-purpose compression benefit, widespread and is used in many commercial implementations. The advantages of using such an instrument would be [12]:

- ✓ It is ready in both open-source and commercial implementations
- ✓ Has a better compression rate (40-50%) and freedom from patented algorithms
- ✓ It is built into http and web-servers as a standard feature.

So the purpose of GZIP Compression is to compressed pages before being transmitted from server to browser, in order to get small transmitted data, then to be quick in delivery. This is useful for servers running Windows operating system [13].

So, gzip is a format of file and a software application used for compression and decompression the data in the file. Jean-Loup Gailly and Mark Adler are created the program as an alternative free software for the compress program utilized in system like Unix, and prepared for use by the GNU Project (the "g" is from "GNU")[14].

III. THE COMBINATION BETWEEN AES ENCRYPTION AND GZIP COMPRESSION

The suggestion in this paper is to improve the compression ratio for the encrypted secret message by combining AES algorithm with GZIP compression.

As for us, we first encrypted secret message with AES algorithm. In general, a certain information has statistical characteristics. The number of 0s and 1s are different. This is the evidence of statistical attacks against security methods. In this reason, we aimed to reduce the differences of the number of 0s and 1s. The compression has been used to reduce the bits by eliminating and identifying statistical redundancy. Algorithms of data compression usually utilize statistical redundancy to represent data more concisely. So, this provides us 2 benefits; reducing the length of secret data and revealing statistical redundancy.

3.1 Algorithm for the proposed system

The architecture of this system is organized with two portions: sender side which consists of encrypted secret message, compressed secret message and receiver side which consists of decompression and decryption for the secret message as shown in figure 1.

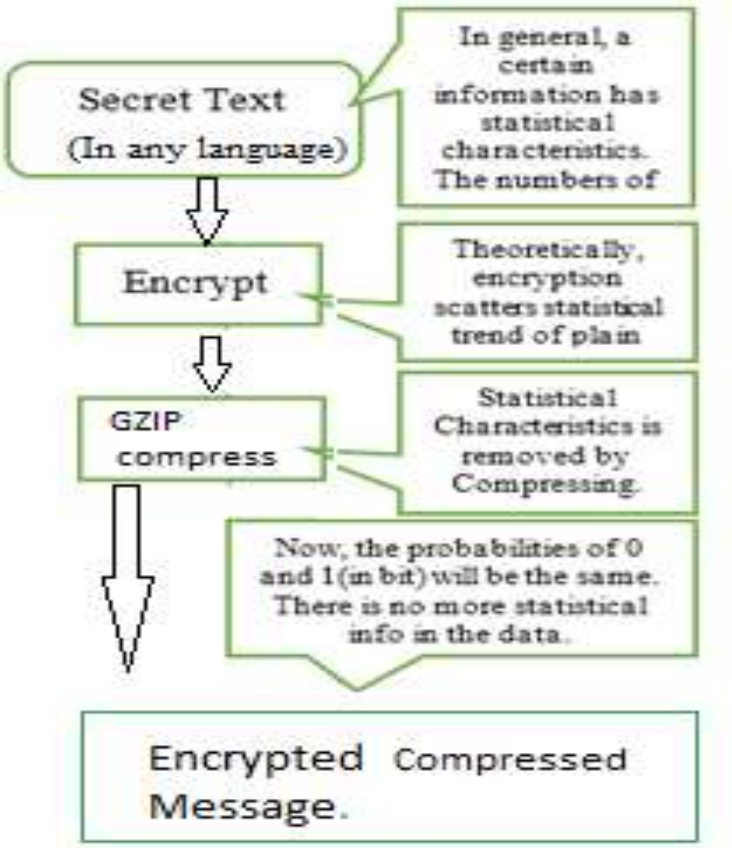

Figure (1)**:** The combination architecture of AES-GZIP

3.2 Encrypted secret message

An advanced encryption standard (AES) has been used to encrypt secret message that based on Rijndael algorithm. The symmetric block cipher can be processing (128 bits) data blocks by using cipher keys of (128, 192, and 256) bits lengths. The input and output sequences length of Rijndael can be any of the three allowed values (128, 192, and 256) bits, but for the (AES) the only length allowed is 128.
The common best practice for symmetric encryption is to use AEAD (Authenticated Encryption with Associated Data). In this paper, we use AES then HMAC (keyed-hash message authentication code method). A HMAC is a special structure used to calculating the MAC that including a hash function combination with a secret cryptographic key. We uses AES256 and then HMAC SHA256, a two-step Encrypt then MAC that needs more keys and more overhead. The method function takes key(s), secret message string, and an optional non-secret payload then return then authenticated encrypted string optionally prepended with the non-secret data with a 256bit key(s) randomly generated. In addition, it have a helper methods which used a string password for keys generation.

3.3 Compress Encrypted Secret Message

The gzip algorithm has been used to compress the encrypted secret message. The gzip give a good compression ratio to secret message that encrypted with AES Algorithm. For Compressed Module basically we compressed the encrypted secret message by using gzip (Figure 2). (gzip depend on deflate algorithm that contains LZ77, Huffman encoding[static or dynamic], RLE[for dynamic Huffman tree] and Huffman encoding for RLE compressed tree).

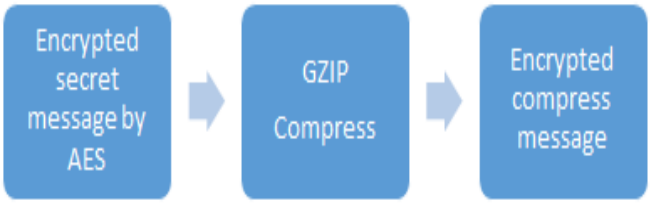

Figure (2)- Encrypted compress message Model.

❖ The AES then HMAC algorithm with GZIP compression described in follows:

Input: Secret message, cryptography keyword. Output: Encrypt compressed secret message (ECSM).

Step 1: Start.
Step 2: Insert text for encryption.
Step 3: Use Random Salt to block pre-generated weak password attacks. The salt bit size is 64(at first salt1 is created then derived and used for crypto key in AES and also slat2 is created then derived and used for authentication key for HMAC).
Step 4: Apply AES encryption algorithm by used a 128 Block bit Size and 256 key bit size.
- Convert secret message text to UTF8.
- Convert cipher text to Hexadecimal format, based 64 string and ASCII code respectively.

Step 5: Encryption (AES) then Authentication (HMAC) of a UTF8 message
- Prepend non-secret payload
- Prepend IV
- Write Cipher text.
- Authenticate all data
- Gather encrypted message and using HMAC SHA256 to add authentication.

Step 6: - Call GZIP compression to compressed data.
- Generate encrypt compressed secret message (ECSM).

Step 7: End.

❖ The Decompression of GZIP with decryption of AES -HMAC algorithm as :

Input:	Encrypted Compressed Secret Message (ECSM).
Output:	Secret Message.

Step 1: Start.
Step 2: Decompressed encrypted secret message by used GZIP algorithm.
Step 3: Convert the message (which is decoded from stego-cover) to UTF-8.
Step 4: Grab *salt1* and *salt2* (8-byte for each one) from Encrypted Message.
Step 5: Derive Crypto key from salt1 and Authentication key from salt2.
Step 6: Use Authentication key for HMAC SHA256:
- Grab (*Sent-Tag* 32byte) from Encrypted Message.
- Calculate Tag (HMAC SHA256) by using Authentication key and Encrypted Message.
- Compare between *Sent-Tag* and *Calculation Tag* to check the integrity of the message

Step 7: Grab *IV* (16 byte) from Encrypted Message:
- Use Crypto key and IV for AES decryption in order to get bytes of secret message
- Convert bytes to string

Fixed Secret Message (length)	Secret language	No. of 1's Before (Encrypt + Compression)	Length (bits) Before (Encrypt + Compression)	No. of 1's After Encrypt	A-Length (bits) After Encrypt	No. of 1's After Encrypt + compression	B-Length (bits) After Encrypt + compression	Compression Ratio C=B/A *100
5	English	22	56	326	656	317	656	100%
1340	Arabic	8871	19152	9897	19736	4234	8472	42.9%
3212	English	11395	25720	13122	26264	6759	13592	51.7%
8938	Arabic	62655	130352	65463	130968	2696	5400	4.12%

Step 8: End.

IV. RESULTS

This part shows the experiments results that leads to measure the performance of the proposed system. The system has been designed by used c# language and includes. The tested has been run by used a workstation laptop (Dell) with following specifications:

- CPU 1.8 GHz core i3
- RAM 4GB DDR3
- OS Windows 8 64bit
- Visual studio 2013

In our test we used some secret messages to calculate their bit size, bit size after encryption, bit size after compression.

4.1 Capacity

For our test we calculated the secret message bit size before encryption and after encryption and with and without compression for four secret messages with different sizes to calculate the compression ratio.

Table (1)- The secret messages capacity (with & without) encrypted and compression.

From table 1, the results shows that compression with (gzip) is very useful with large secret message that is be practically efficient.

V. CONCLUSION

This paper emphasis on improving the compression ratio of the encrypted secret message, the algorithm of AES is used to encrypt the secret message which is given a good security.
So that usage a compression model which be efficient to compress large amount of data will be useful to reduce the encrypted message size and gzip is a good choice for this purpose.

REFERENCES

[1] Siddharth T.,Prashant K.K., "A Novel Information Security Scheme by Creptic Video Stegnography", International Journal of Computer Technology and Electronics Engineering (IJCTEE), Volume 2, Issue 1, ISSN: 2249-6343, pp.65-69, (2012).
[2] Sidhu ,A., S., Garg ,Er. M.," Research Paper on Text Data Compression Algorithm using Hybrid Approach",2014.
[3] Forouzan.B,M.,"Foundations of computer science", Second Editon,2008.
[4] Capo-Chichi, E.P.,Guyennet,H., Friedt,J.," K-RLE : A new Data Compression Algorithm forWireless Sensor Network",2009.
[5] Prof. S.D.Joshi, Anil G., and Sunita B. "Information security using encrypted Steganography", National Conference on Advanced Computing and Communication network, (9- 10 March 2007).
[6] Gurtaptish K., "An Efficient Text Storage Security Algorithm Research fellow" International Journal of Application or Innovation in Engineering & Management (IJAIEM), Vol.2, Issue 6, ISSN 2319 - 4847 (June 2013).
[7] A.Joseph Raphael,Dr.V.Sundaram ., Head & Director, "Cryptography and Steganography-A Survey",(2010).

[8] Himanshu G., "Twin Key Implementation in AES", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, Vol.16, Issue 5, pp.01-05, (2014).
[9] Vibha V., Avinash D., "Analysis of comparison between Single Encryption(Advance Encryption Scheme (AES)) and Multicrypt Encryption Scheme", International Journal of Scientific and Research Publications, Vol.2, Issue 4, ISSN: 2250-3153, (April 2012) .
[10] Sadkhan Al Maliky, Sattar B., "Multidisciplinary Perspectives in Cryptology and Information Security: Advances in Information Security, Privacy, and Ethics", Book, IGI Global, ISBN : 9781466658097, (2014).
[11] Shailendra M. P., Sandip R. S., Vipul D. P., Puja S., "A Survey on compound use of Cryptography and Steganoghaphy for Secure Data Hiding", International Journal of Emerging Technology and Advanced Engineering (IJETAE), Vol.3, Issue 10, ISSN: 2250-2459, (October 2013).
[12] Smitha S. Nair, " XML Compression Techniques: A Survey", University of Iowa, USA. Available at: http://www. Zen cart support, what is GZIP, and why should I use it?
[13] Available at: gzip - Wikipedia, the free encyclopedia.html.

