International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471
Vol. 2 Issue 4, April-2013, pp: (33-39), Available online at: www.erpublications.com

Simplified Compression Crypto System using PGP

Saumya Jindal[1], Mohd Mohiuddin Ansari[2]
Department of Computer Science and Engineering, Sharda University, Gr. Noida, India

Abstract: Traditional Arithmetic coding provides very little or no security but has higher compression efficiency. For the Security purpose, we need to introduce cryptographic techniques in compression process. Several encryption algorithms, such as block cipher, have been introduced. Block ciphers like AES and DES accept all types of input type like text, audio, video as a transparent binary input and perform complex operation to generate the cipher text but the efficiency varies with the input types. This paper focuses on text data type. For text data PGP works more efficiently. This paper presents new technique of compression crypto system using PGP. In the proposed scheme chaos is also introduced to increase the security of the system. In the purposed scheme chaos is used for the key generation in one time pad.

Keywords: Arithmetic coding, one time pad, Chaotic function, RSA, Pretty Good Privacy.

Introduction

A Compression mechanism can provide security to some extent, if the compression algorithm is kept secret itself. In case of conventional compression algorithm which is known to everyone cannot expect secrecy. Hence if the security needed, compression mechanism combined with encryption algorithm which provides higher security with reduced size of original data. Joint Compression and Encryption has gained a higher attention from past five years to reduce the computational complexity & to provide encryption of text and multimedia content. This technique is important because of the increased use of compressed text files in many applications such as in emailing, mobile SMS and wireless sensor nodes.

Encryption process converts the message from comprehensible to incomprehensible structure, which makes typical to compress the encrypted data using any of the compression algorithms. Hence encryption always follows the compression and this scheme more probable in most of the situations. In some situation where the user and network is not trusted there a scheme is used in which compression follows the encryption process [1]. Using the secret key chaotic map model for arithmetic coding is determined and keeps changing. Moreover, the compressed sequence is masked by a pseudorandom key stream generated by another chaotic map. This scheme has two level protection with the use of secret key and better because both the position and the direction of the line segments in the piecewise linear chaotic map are controlled using the secret key [2]. Encryption generates a stream of random data which will not be compressed. Hence compression cryptosystem is proposed in which first compress and then encrypt, using the arithmetic coding.

A. Arithmetic coding

In lossless data compression the idea of arithmetic coding (AC) was suggested by Rissanen. To improve the security of arithmetic coding, AC have need some modification. Two main modifications in arithmetic coding are Randomized AC and key-based interval splitting AC.

Randomized Arithmetic Code (RAC) to protect JPEG2000 images in [3], [4]. The idea of RAC is to selectively swap the encoding intervals of A and B for each input symbol based on random bits. Kim et al. presented Secure Arithmetic Code (SAC) in 2006 which is based on Interval Splitting Arithmetic Code (ISAC) in 2005 [5]. ISAC separates a continuous symbol interval into two portions and reassign the values of these intervals.

In both modified AC same key is used to encrypt different messages which introduce Chosen-plaintext attacks. If random key is used to compress the different message even then RAC is insecure [4]. Both modified AC does not meet the diffusion property then the result is it can be easily broken [6]. In RAC another problem occur at decryption side to reconstruct the random bits. The security problem of Secure Arithmetic Coding (SAC) under an adaptive chosen-cipher text attack. This indicates that the SAC is not suitable for those applications where the attacker can have access to the decoder and use Randomized Matrix Arithmetic Coding (RMAC), in which the security depends on a randomized matrix formed by the random key based on the user profile in [7].

In arithmetic coding a modified integer arithmetic code and Pseudo-Random Bit Generator (PRBG) and the Secure Hash Algorithm (SHA-256) is used to construct the key vector [8]. A new method of lossless image compression based on combining AC with the Run Length Encoding (RLE) [9]. AC is a form of variable-length entropy encoding. AC is one of the well known powerful textbook and multimedia compression algorithm and has higher compression efficiency than other entropy coders [10]. AC is a technique which converts a given probability distribution into an optimal code and is commonly used in compression scheme. Despite its computation complexity, it has a very good compression ability, high speed, low storage requirement, effectiveness of compression [11].

- Low = 0
- High = 1

Loop For all the symbols.
Range = high – low
High = low + range * high_range of the symbol being coded
Low = low + range * low_range of the symbol being coded
Where

- Range, keeps track of where the next range should be.
- High and low, specify the output number.

Arithmetic coders can generate near-optimal output for any given set of symbols and probabilities but Huffman coding is optimal when the probability of each input symbol is a negative power of two [12].

Arithmetic coding does not have this restriction. It works by representing the file by an interval of real numbers between 0 and 1. When the file size increases, the interval needed to represent it becomes smaller, and the number of bits needed to specify that interval increases. Successive symbols in the message reduce this interval in accordance with the probability of that symbol. The more likely symbols reduce the range by less, and thus add fewer bits to the message.

Arithmetic coding differs from other forms of entropy encoding such as Huffman coding in that rather than separating the input into component symbols and replacing each with a code, arithmetic coding encodes the entire message into a single number, a fraction n where ($0.0 \leq n < 1.0$). The arithmetic encoding does not influence the key stream generation of the chaotic system.

B. Pretty Good Privacy

Pretty Good Privacy (PGP) was created by Phil Zimmermann in 1991. Pretty Good Privacy (PGP) is a popular program used to encrypt and decrypt e-mail over the Internet and digitally sign files and emails.
It is a high security cryptographic program which is used to exchange messages and files with privacy, integrity and convenience.

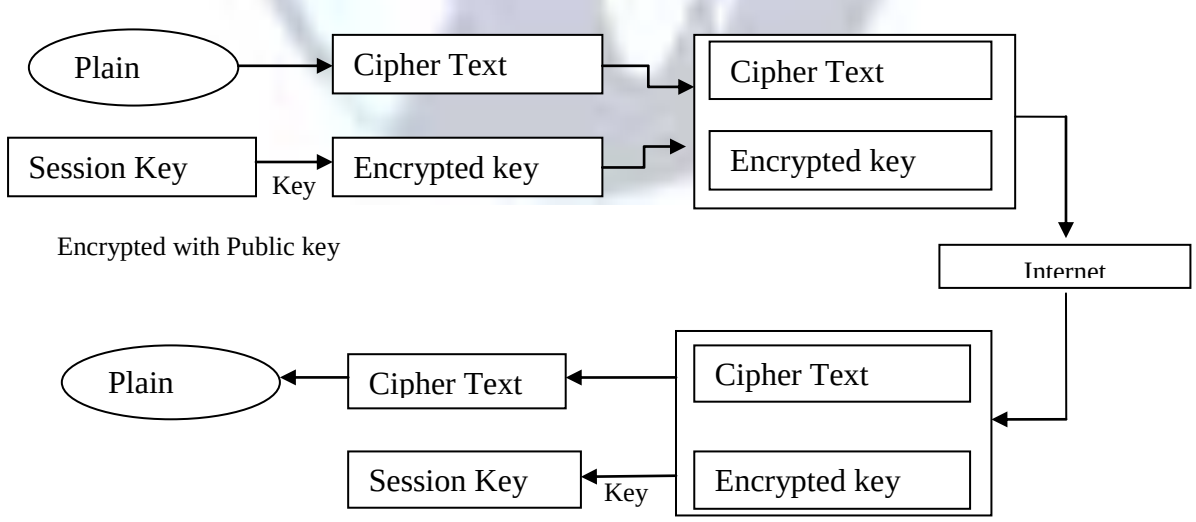

Figure 1 : Pretty Good Privacy

Generally PGP provides:

- **Privacy** - Store and transmit your data so that only select people may view their contents.
- **Integrity** - Ensure your files, data, and applications have not been modified without your permission.
- **Authentication** - A way to verify that people actually are who they claim to be.

When someone wants to using PGP, they generate a key-pair.
- Private key
- Public key

A Private key that a specific user uses to decrypt messages intended for them & to sign messages to authenticate themselves as the sender. A public key which you provide to other people who need to send you confidential messages. To provide a greater degree of assurance that a public key really belongs to a specific person, public keys can be "signed" by other users who attest to your identity.
In public key cryptography, if we use bigger the key, then more secure the cipher text. Public key size and conventional cryptography's secret key size are totally unrelated. A conventional 80-bit key has the equivalent strength of a 1024-bit public key. A conventional 128-bit key is equivalent to a 3000-bit public key.
PGP then creates a session key which is a one-time secret key and generate random collection of letters. It make impossible to recover the message without knowing the key. Generated random key used only one time and then sender and receiver destroy that key after the use.
One time pad starts in 1882. One-time secret key is a perfect cipher and also called a vernam-cipher.

C. Need of Compression Cryptosystem

When the data is transmitted over an insecure bandwidth limited channel, then efficient and secured transmissions are necessary to ensure proper data protection and transmission speed. Hence demand of data compression and encryption increased and are used to meet these requirements.
For the Fast development of data application, demand of compressed text files is growing day by day to meet the high data rate requirements of bandwidth transmission systems. Data compression involves transformation of source message into the smaller size format called code words. Data compression is basically used to meet the high data rate requirements of bandwidth transmission systems and reducing storage and communication cost. Compression not only reduces the amount of disk space occupied by the data in the huge document collections, it also decreases the overall processing time and transmission time. Encryption involves transformation of original message called plaint text into the scrambled form called cipher text using the encryption key. It basically used to protect the data from opponent. Encryption processes generates the cipher-text is at least as large as the plain-text. Therefore, some lossless compression schemes such as Huffman coding and arithmetic coding are usually employed to compress the source message, and then encryption process as applied on them .
In the traditional approaches compression and encryption processes are both performed separately, and take more time hence our scheme performs both jointly in one pass.

Proposed Scheme

In this paper, we propose a communication system with the use of compression and cryptography to reduce and to make secure the data, prior to sending the data over an insecure communication channel. In this section, an arithmetic coding and PGP for joint compression and encryption is described. Joint compression and encryption scheme, variable string is the input of arithmetic coding and based on frequency and probability it generate a tag value which is a fraction n where $(0.0 \leq n < 1.0)$. A secure algorithm using the arithmetic coding based on integer implementation. They use integer implementation of arithmetic coding instead of floating point implementation. This modification reduces the computational cost of AC without affecting significant compression loss. But this technique does not compromise the coding efficiency and in worst case the modified AC has compression ratio similar to that of the traditional one.
In cryptography, to generate a key stream using the chaotic system is a good choice due to its good properties such as ergodicity, sensitivity to initial values and sensitivity to control. In the encryption process firstly stream cipher such as one time pad, is used which provide more security using a key which is different for every new message. We generate a key using the chaotic function which generates random value at each step, which also provide more security.

Proposed Algorithm of Compression Cryptosystem

At Sender End
- Input a string
- Generate tag value a fraction n where $(0.0 \leq n < 1.0)$ or compressed the string.

- Randomly generated a key using the chaotic function (OTP)
- Key (OTP) encrypt the compressed string
- Key (OTP) is encrypted with public key of the receiver
- Encrypted message and key send to the receiver at Receiver End
- Input is encrypted message and key
- Key (OTP) is decrypted with private key of the receiver
- Decrypted key decrypt the compressed string
- Decompress the compressed string

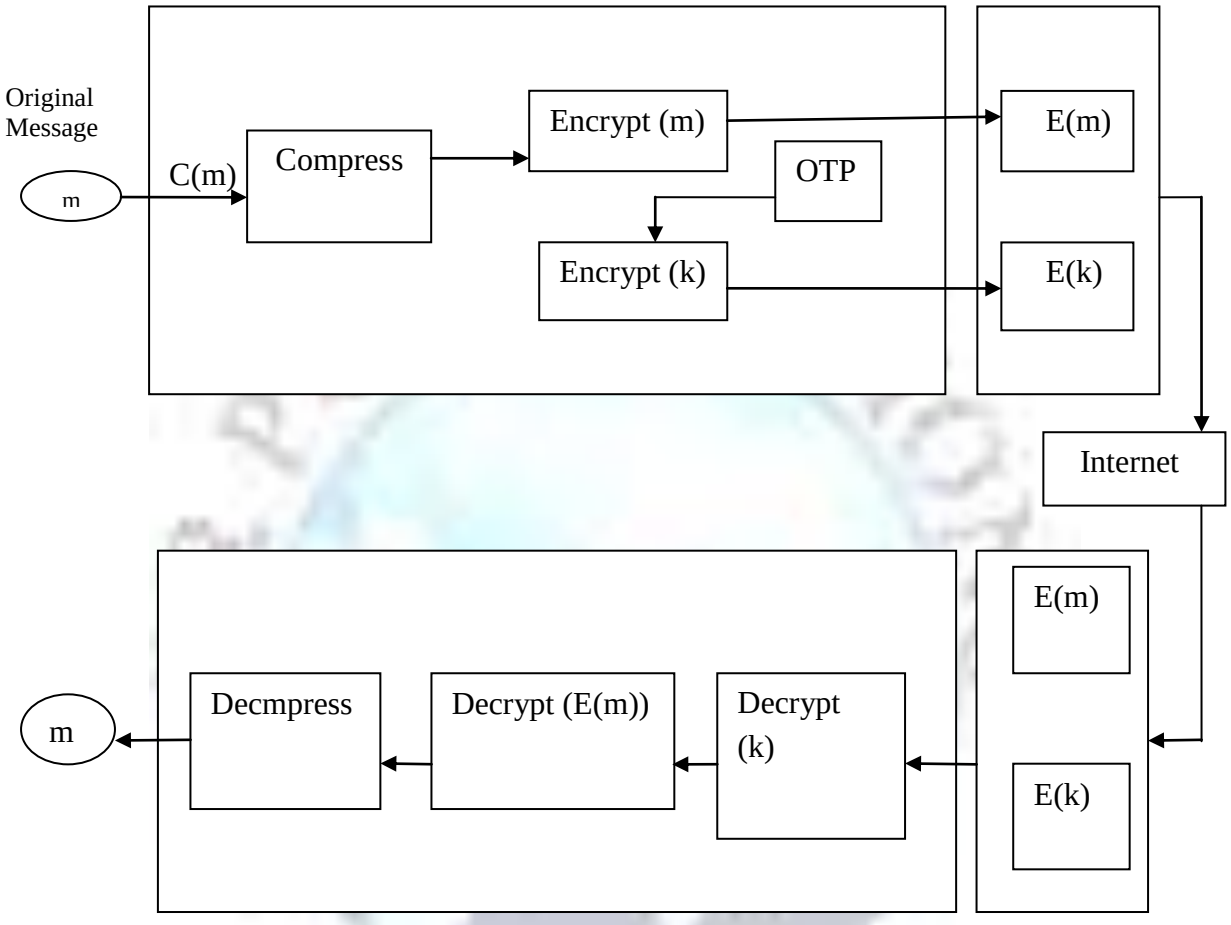

Figure 2 : Compression Cryptosystem

Result and Analysis

Huffman coding is one of the best-known compression techniques that has an optimal expected length H<= $L_{Huffman}$<Hþ1, where H represents the entropy of the Source. Huffman coding also used for encryption has been considered which shown that, if the cryptanalyst does not know the probability mass function (PMF) of the source, then a Huffman coded file is difficult to cryptanalyze. However, this assumption of the cryptanalyst not knowing the PMF is often invalid, especially for text data. If the cryptanalyst also knows the construction rule, then Huffman codes can be easily decoded and essentially provide no security because the probabilities of source symbols must be ordered no increasingly before encoding, resulting in a unique possible codeword set. Adaptive Huffman tree is mutated by a key-stream generated by two chaotic maps, and the probabilistic model is not changed after encryption. In this scheme more time is required.

Universal codes such as Lempel-Ziv coding work without knowledge of the source PMF. However, they need a longer block length to obtain the same performance as a code designed specifically for a known PMF. Consequently, this leads to the increased complexity of the encoder and decoder. Run-Length compression algorithm provides simplicity and efficiency and is suitable for storing highly repetitive collections of texts.
For the sake of best compression, the length of the outputs generated by the algorithm should be minimized. Arithmetic coding encodes the entire message into a single number, a fraction n where ($0.0 \leq n < 1.0$).

Table 1: Coding Efficiency

		File Size(bytes)	
File Name	Original File	Traditional AC	% of compressed file
alice29.txt	152089	86852	57.10604
asyoulik.txt	125179	75242	60.10753
cp.html	24603	16087	65.38633
Fields.c	11150	6984	62.63677
grammar.lsp	3721	2159	58.02204
kennedy.xls	1029744	460040	44.67518
lcet10.txt	426754	249097	58.37016
Out	777	299	38.48134
plrabn12.txt	481861	273002	56.65576
ptt5	513216	77965	15.19146
Sum	38240	25481	66.63441
xargs.1	4227	2593	61.34374

Public key encryption algorithms are slower than Private key algorithms, hence private key encryption algorithms are fast giving the speed advantage to private key encryption. Due to the fact that public key schemes can use longer keys, resulting in longer encryption times, larger encrypted messages and longer transmission times. Public key encryption requires a lot of computer resources when compared to Single-key encryption. Private key encryption can achieve the same level of protection as public key encryption. But private key algorithms also have some disadvantages first the keys must be shared before they can be used. This can cause a delay in communications that need to be encrypted by the keys before transmission. The sharing process has to be secure, since a person who intercepts the private key undetected can decipher messages that he subsequently intercepts.

Another is that if the key becomes known by unauthorized individuals, the key is compromised and must be regenerated and redistributed. Of course, the redistribution must be done in a secure manner, so complex logistics can be involved. One of the uses of public key cryptography is to send the private key in a public key "digital envelope" to guarantee its secure delivery.

Channel must kept secure for secret key exchange: In symmetric key encryption, sharing of secret key is a problem in the beginning. It has to be ensures that exchanged medium should remain secure.

Too many keys: Creates a problem with managing and ensuring the security of all new shared keys which has to be generated for communication with every different party.

A public key encryption algorithm solves the problem of distributing the key for encryption. Everyone publishes their public keys and private keys are kept secret. It allows the use of digital signatures which enables the recipient of a message to verify that the message is truly from a particular sender. Symmetric block cipher algorithm like DES is the old “data encryption standard” has too short key size 56 bits for proper security. DES keys have been broken within 24 hours. 3DES reuse the implementation of DES is secure and not breakable with today’s technology but it is slow. AES is the successor of DES and accept keys of 128,192,256 bits. Symmetric block cipher algorithm like AES and DES, works fast and very secure and used to encrypt and decrypt the message using the same key. Hence there is a key exchange problem arise. Key exchange problem is solved by public key cryptography RSA. In RSA message is encrypted with public key and decrypted with private key. Asymmetric algorithms are quite slow hence key exchange work done by asymmetric algorithm and encryption is performed by symmetric algorithm. Both are easy to implement.

One time pad with the chaotic map is used which have Chaotic systems have many interesting features such as sensitivity on initial condition and system parameter, ergodicity and mixing properties. Chaotic map is used for key generation in one time pad. e.g. the plain text is a string as abc and key is generated from chaotic map as .2. Then generated cipher text is qyj and if we apply little changes in key then the whole cipher text will be changed. Same plain text is encrypted with key .3 then a lot of modification in cipher text occur hence cipher text is vob.

The table below which gives us the comparison between various parameters of AES, 3DES, DES and RSA [11].

Table2: Comparison between AES, 3DES, DES and RSA

Factors	AES	3DES	DES	RSA
Key Length	128,192 or 256 bits	(k1,k2and k3)168 bits (k1,k2 is same)112 bits	56 bits	1024 to 4096 bits
Cipher Type	Symmetric block cipher	Symmetric block cipher	Symmetric block cipher	Asymmetric block cipher
Block Size	128,192 or 256 bits	64 bits	64 bits	----
Developed	2000	1978	1977	1977
Cryptanalysis resistance	Strong against differential, truncated differential, linear, interpolation and square attacks	Vulnerable to differential, Brute force attacker could be analyze plain text using differential cryptanalysis	Vulnerable to differential and linear cryptanalysis; weak substitution tables	Vulnerable against linear attack
Security	Considered Secure	One only weak which is exit in DES	Proven inadequate	Breaking RSA encryption is as hard as factoring
Possible Keys	$2^{128}.2^{192}$ or 2^256	2^112 or 2^168	2^56	2^{1024}
Secrecy	Based on sharing secrecy	Based on sharing secrecy	Based on sharing secrecy	Based on personal secrecy
Operation	Substitution or Permutation of symbols	Substitution or Permutation of symbols	Substitution or Permutation of symbols	Applying mathematical functions to numbers

Conclusion

The proposed scheme provides some advantages over other existing methods. The process of joint compression and encryption is an effective tool to protect the large volume of data. This process of joint compression and encryption under PGP should be able to compete with other schemes. Future works may include modification in stronger encryption algorithm PGP which provide more security using public and private key. In last, this paper gives a good foundation work for text compression and encryption based on PGP.

References

[1]. A Anil Kumar and Anamitra Makur, "Lossy Compression of Encrypted Image by Compressive Sensing Technique", 2009.
[2]. Q. L. J. C. Kwok-Wo Wong, "Simultaneous Arithmetic Coding and Encryption Using Chaotic Maps," Circuits and Systems II: Express Briefs, vol.57, pp.146,150, Feb. 2010.
[3]. A. G. a. E. M. M. Grangetto, "Selective encryption of JPEG 2000 images by means of randomized arithmetic coding ," Multimedia Signal Processin ,6th Workshop on, pp. 347–350, 29 Sept.-1 Oct 2004.
[4]. E. M. a. G. O. M. Grangetto, "Multimedia selective encryption by means of randomized arithmetic coding,"Multimedia," vol.8, pp. 905–917, Oct 2006.
[5]. J. V. a. J. W. H. Kim, "Secure Arithmetic Coding Using Interval Splitting,"Signals, Systems and Computers, in Conference Record of the Thirty-Ninth Asilomar Conference on, , 2005.

[6]. K.-H. W., Hung-Min Sun, "Designing an Arithmetic Code for Multimedia Files Compression and Encryption," in Second International Workshop on Computer Science and Engineering., 2009.
[7]. S. B. R. J., Dr. V. Kavitha, "RMAC-A New Encryption Scheme for Arithmetic Coding to evade CCA Attacks," 2011.
[8]. Yuh-Ming Huang,Yin-Chen Liang, "A secure arithmetic coding algorithm based on integer implementation," in Communications and Information Technologies (ISCIT), 2011 11th International Symposium on ,2011.
[9]. A. M. M. S. B. Med Karim Abdmouleh, "A New Method Which Combines Arithmetic Coding with RLE for Lossless Image Compression," Journal of Software Engineering and Applications, 2012.
[10]. V. Mahnaz Sinaie, "A Low Complexity Joint Compression-Error Detection-Cryptography Based On Arithmetic Coding," in 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA), 2010.
[11]. C. Y. J. Q. a. X. F. Tao Xiang, "Cryptanalysis of Secure Arithmetic Coding Based on Interval Swapping".
[12]. P. Kumbhar and S. Krishnan, "SMS compression using arithmetic coding modification," in International Conference on Devices,Circuits and Systems (ICDCS), 15-16 March 2012.