DATA SECURITY USING PRIVATE KEY ENCRYPTION SYSTEM BASED ON ARITHMETIC CODING

Ajit Singh[1] and Rimple Gilhotra[2]

Department of Computer Science & Engineering and Information Technology
BPS Mahila Vishwavidyalaya, Khanpur Kalan, Sonepat-131305 Haryana (India).
[1]E-Mail: ghanghas_ajit@rediffmail.com
[2]E-Mail: rimple.gilhotra@gmail.com

ABSTRACT

Problem faced by today's communicators is not only security but also the speed of communication and size of content.In the present paper, a scheme has been proposed which uses the concept of compression and data encryption. In first phase the focus has been made on data compression and cryptography. In the next phase we have emphasized on compression cryptosystem. Finally, proposed technique has been discussed which used the concept of data compression and encryption. In this first data is compressed to reduce the size of the data and increase the data transfer rate. Thereafter compress data is encrypted to provide security. Hence our proposed technique is effective that can reduce data size, increase data transfer rate and provide the security during communication.

KEYWORDS

Arithmetic coding, cryptography, floating point number, one time pad, compression-crypto

1. INTRODUCTION

The present network scenario demands exchange of information with more security and reduction in both the space requirement for data storage and the time for data transmission. This can be accomplished by compression and encryption, such kind of system is called compression –crypto system. Encryption is indeed a secure coding technique and data compression is also a coding technique, whose purpose is to reduce both the space requirements for data storage and the time for data transmission. In proposed system i.e data security using private key encryption system encoded string is produced by a model from an input string of symbols and based on arithmetic coding that can be used to achieve the present network scenario for exchange of information with more security and compression.

2. DATA COMPRESSION

Data compression offers an attractive approach for reducing communication costs by using available bandwidth effectively. Compression algorithms reduce the redundancy in data representation to decrease the storage required for that data. Over the last decade there has been an unprecedented explosion in the amount of digital data transmitted via the Internet, representing text, images, video, sound, computer programs, etc. [3]

In the more modern model-based paradigm for coding, where, from an input string of symbols and a model, an encoded string is produced that is a compressed version of the input. The decoder, which must have access to the same model, regenerates the exact input string from the encoded string .The model, is a way of calculating, in any given context, the distribution of probabilities for the next input symbol. It must be possible for the decoder to produce exactly

DOI : 10.5121/ijnsa.2011.3305

the same probability distribution in the same context. Compression is achieved by transmitting the more probable symbols in fewer bits than the less probable ones. More complex models can provide more accurate probabilistic predictions and hence achieve greater compression. The effectiveness of any model can be measured by the entropy of the message with respect to it, usually expressed in bits/symbol. [4] Shannon's fundamental theorem of coding states that, given messages randomly generated from a model, it is impossible to encode them into less bits (on average) than the entropy of that model. [5]

A message can be coded with respect to a model using either Huffman or arithmetic coding. It's well known that the Huffman's algorithm generates minimum redundancy codes compared to other algorithms. [6] But the disadvantage of Huffman is that, all codes of the encoded data are of different sizes. Therefore it is very difficult for the decoder to know that it has reached the last bit of a code. Arithmetic coding can be viewed as a generalization of Huffman coding. It efficiently represents more frequently occurring sequences of pixels values with fewer bits. [7] Arithmetic coding typically has a better compression ratio than Huffman coding, as it produces a single symbol rather than several separate codeword and can be use in compression based encryption system. [9]

2.1 Arithmetic Coding

Today, for data compression there exist many techniques. The most popular one is Arithmetic Encoding. This encoding techinque has been developed extensively since its introduction several decades ago and is notable for offering extremely high coding efficiency. Arithmetic coding, invented by Jorma Rissane and turned into a practical method by Witten, Neal and Cleary, achieves superior Compression to the better-known Huffman algorithm. [1] In fact, arithmetic coding is a method to ensure lossless data compression. It is indeed a form of variable length entropy encoding. In the case of other entropy encoding techniques, the input message is separated into its component symbols and each symbol is replaced by a code word. But arithmetic coding encodes the entire message into a single number, a fraction n where ($0.0 \leq n < 1.0$) [2]. The coding algorithm is symbol wise recursive; i.e., it operates upon and encodes (decodes) one data symbol per iteration or recursion. On each recursion, the algorithm successively partitions an interval of the number line between 0 and 1, and retains one of the partitions as the new interval. Thus, the algorithm successively deals with smaller intervals, and the code string, viewed as a magnitude, lies in each of the nested intervals. The data string is recovered by using magnitude comparisons on the code string to recreate how the encoder must have successively partitioned and retained each nested subinterval. [8]

3. CRYPTOGRAPHY

In the 21st century the importance of information and communication systems for society and the global economy is intensifying with the increasing value and quantity of data that is transmitted and stored on those systems. At the same time those systems and data are also increasingly vulnerable to a variety of threats, such as unauthorized access and use, misappropriation, alteration, and destruction. Cryptography addresses the above issues. It is the foundation of all information security aspects. The techniques employed to this end have become increasingly mathematical of nature. Classical cryptosystems is very easy to understand, easily implemented and very easy to be broken. New forms of cryptography came after the widespread development of computer communications. In data and telecommunications, cryptography is necessary when communicating over any entrusted medium. In the last few decades, however, the trend has been on placing cryptography onto a sound mathematical framework. This modern focus has initiated the evolution of the field from an art into a science, which includes just about any network, particularly the internet. This evolution comes with modern cryptography (MC) really begins with Claude Shannon arguably the father of mathematical cryptography. He published a related paper, **"**Communication Theory

of Secrecy Systems"**,** in 1949. These, in addition to his other works on information and communication theory established a solid theoretical basis for cryptography and for cryptanalysis. And with that, cryptography more or less disappeared into secret government communications organizations such as the NSA and equivalents elsewhere. [11]Today's cryptographic techniques have become the immediate solution to protect information against third parties. These techniques required that data and information should be encrypted with some sort of mathematical algorithm where only the party that shares the information could possible decrypt to use the information. [12]

Within the context of any application to application communication, there are some specific security requirements including:-

- Authentication: - The process of providing one's identity.
- Confidentiality: - Ensuring that no one can read the message except the intended receiver.
- Integrity: - Assuring the receiver that the received message has not been altered in any way from the original.
- Non-repudiation:- A mechanism to prove that the sender really send this message.

There are two types of cryptographic schemes: symmetric (private key) cryptography, and asymmetric cryptography, each of which described below [10].

3.1 Symmetric Key Cryptography [12]

In symmetric key cryptography (also known as private-key cryptography), a secret key may be held by one person or exchanged between the sender and the receiver of a message. If private key cryptography is used to send secret messages between two parties, both the sender and receiver must have a copy of the secret key. However, the key may be compromised during transit. If you know the party you are exchanging messages with, you can give them the key in advance. However, if you need to send an encrypted message to someone, you have never met; you will need to figure [15] out a way to exchange keys in a secure way. One method is to send it via another secure channel.

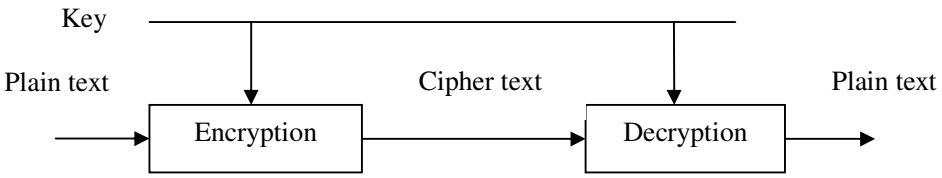

Figure 1: Symmetric Cryptosystem

Two types of symmetric encryption algorithm are:-

3.1.1 Stream ciphers [18]

Stream ciphers encrypt plaintext one byte or one bit at a time. Examples of Stream ciphers are RC4 cipher and the one-time pad.

3.1.2 Block ciphers [18]

Block ciphers encrypt plaintext in chunks. Common block sizes are 64 and 128 bits. Examples of Block ciphers are DES and the AES.

3.2 Asymmetric Key Cryptography [12]

In the two-key system, (also known as the public key system) one key encrypts the information and another, mathematically related key decrypts it. The computer sending an encrypted message uses a chosen private key that is never shared and so is known only to the sender. If a sending computer first encrypts the message with the intended receiver's public key and again with the sender's secret, private key, then the receiving computer may decrypt the message, first using its secret key and then the sender's public key. Using this public-key cryptographic method, the sender and receiver are able to authenticate one another as well as protect the secrecy of the message.

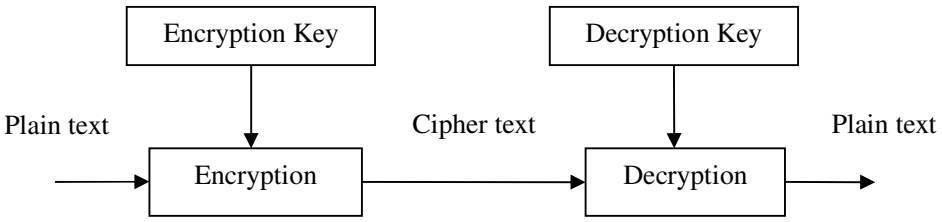

Figure 2: Asymmetric Cryptosystem

Some examples of popular asymmetric encryption algorithms are:-

RSA
PGP
DSA

However, public key cryptography by itself is not the solution to all secure communication needs. The main problem is that it tends to be very slow. Instead, public key methods are generally used in combination with traditional, symmetric key methods since they are much faster. Provably secure symmetric key method will result in a great breakthrough in modern cryptography.

4. COMPRESSION-CRYPTO SYSTEM [13]

Information security protecting the contents of information stored on standard disks, tape, is an ever growing concern among enterprise IT organizations, in companies of all sizes. To keep up with this growing concern, more and more IT organizations are turning to encryption techniques to protect their valuable information assets. In addition to concerns over securing stored information, IT organizations are faced with ever-increasing costs to ensure that there is enough capacity in their storage solutions to meet the organization's present and future demands. To keep up with the challenges associated with tight IT budgets and increasing storage capacity needs and, some companies are beginning to turn to data compression techniques. The solution to both the density and the security problems has been to simply use software-based compression techniques to reduce the number of tapes increasing storage density and encryption to secure the data on the tapes. The compression cryptosystem can be understood with the help of figure [13].

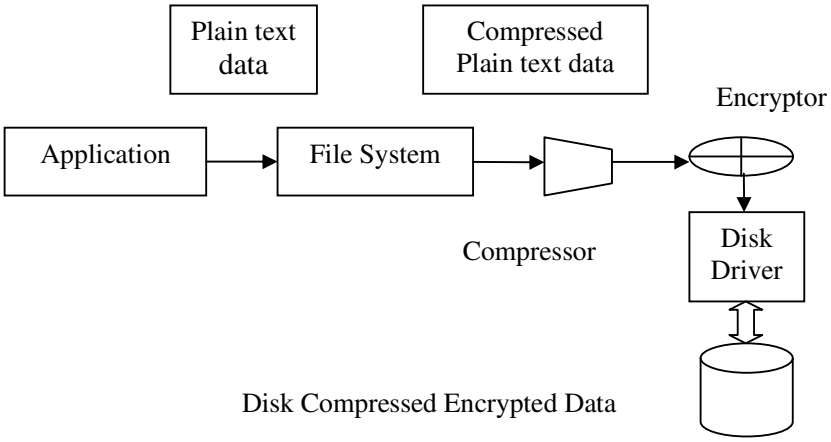

Figure 3: Compression and Encryption processing flow

5. EXISTING SYSTEMS

One of the existing system used compression along with encryption using RSA algorithm. This system is basically used for mobile communication. This system provides a solution to this SMS security problem. The approach that is used in this system is to secure the SMS message using Hybrid Compression Encryption (HCE) system. This system compresses the SMS to reduce its length, then encrypts it using RSA algorithm [16]. But this system is using RSA. RSA is a Public Key Encryption method. A disadvantage of using public-key cryptography for encryption is speed. One more exiting system is presented, which provide us a errorless integrated secure transmission of medical information data like Image, Audio, Video etc. This system also used lossless compression technique like Sequitur for efficient utilization of communication channel. The combination of encryption (McEliece public-key cryptosystem) with compression provides confidentiality in the transmission. [17]. But this system has a limitation with length. Its efficiency drops as the length of data increases. Also, the system requires a very large public key which makes it very difficult to use in many practical situations.

Now a day's Arithmetic coding is in its full form. It is a statistical method and its compression ratio is very good. So, we used Arithmetic coding with private key encryption system .Because private key encryption system is fast and its mathematical implementation is easy.

6. PROPOSED COMPRESSION-CRYPTO SYSTEM

The proposed technique is based on the concept of arithmetic coding in which a word of text is converted into floating point number that lie in range between 0 and 1. This floating point number is converted into binary number and after that one time key is used to encrypt this binary number. Finally after encryption, result is again a binary number; this number is converted into decimal number again and sends to the receiver.

6.1 Requirements

6.1.1 Table

A table contains symbols along with percentage of probability of occurrence. According to the requirements of user the table is designed that may contain lower case alphabets, upper case alphabets and symbols like %,@ etc.

6.1.2 One Time Pad

It's a one-time key provided by user for securing the content. This is a random key that is as long as message. In addition the key to be used to encrypt and decrypt a single message, and then discarded. Each new message requires a new key of same length as a new message.

6.2 Implementation

This technique converts a word of text into floating point number by using arithmetic coding. The result is compressed data. Obtained floating point number is encrypted using private key encryption technique i.e. one time pad key. The output is secure and compressed data.

6.3 Explanation

6.3.1 Compression and Encryption

Firstly input symbol is compressed using arithmetic coding after that a private key is used to encrypt the result of arithmetic coding.

Algorithm

To compress and encrypt the message Algorithm includes following steps:-

Step 1:- using table encodes the input symbol.

a) Initialize lower_ bound=0, upper_ bound=1

b) While there are still symbols to encode

Current _range = upper _bound - lower _bound

Upper_ bound = lower _bound + (current _range * upper _bound of new symbol)

Lower_ bound = lower_ bound + (current _range * upper_ bound of new symbol)

End while

Step 2:- The string may be encoded by any value within the probability range and after that convert the output decimal number into binary format.

Step 3:- limit the number of bits by using the formula:-

No_ of_ bits=log ⌐2/upper_ bound_ last encoded symbol - lower_ bound_ last encoded symbol¬

Step 4:- No _of _bits is used to reduce the number of bits obtained in step2.

Step 5:- Select any one time pad and Xor it with result of step4.

Step 6:- Rotate 2 bits right.

Step 7:- Convert the result of step 6 into decimal format again.

Output: - output is floating point no that is corresponding to the inputted symbol.

6.3.2 Decompression and Decryption

Getting the floating point it's time now that we convert it into original text.

Algorithm:-

Step 1:- Convert the received data into binary format.

Step 2:- Rotate 2 bits to left.

Step 3:- Selected one time pad is Xored with the result of step2.

Step 4:- Convert the result back into decimal form.

Step 5:-Encoded_ value=Encoded input

While string is not fully decoded

Identify the symbol containing encoded value within its range
current_ range = upper _bound of new symbol - lower _bound of new symbol
encoded value = (encoded _value - lower_ bound of new symbol) ÷ current_ range

End while

Output: The output is the original symbol.

6.3.4 Example

Table1: symbols along with probability of occurrence

Symbol	Probability	Range(lower_ bound, upper_ bound)
a	30%	[0.00,0.30)
b	15%	[0.30,0.45)
c	25%	[0.45,0.70)
d	10%	[0.70,0.80)
e	20%	[0.80,1.00)

Compression and encryption

Data to be encoded and encrypted is “abd”

Step1:-

Encode 'a'
current_ range = 1 - 0 = 1
upper bound = 0 + (1 × 0.3) = 0.3
lower bound = 0 + (1 × 0.0) = 0.0

Encode 'b'
current range = 0.3 - 0.0 = 0.3
upper bound = 0.0 + (0.3 × 0.45) = 0.135
lower bound = 0.0 + (0.3 × 0.3) = 0.09

Encode 'd'
current range = 0.135-0.09 = 0.045
upper bound = 0.09 + (0.045 × 0.8) = 0.126
lower bound = 0.09 + (0.045 × 0.7) = 0.1215

Step2:

The string "abd" may be encoded by any value within the range [0.126, 0.1215).

Now output is 0.12375 and its binary equivalent= .00011$\overline{111101011100001010001}$

Step3:- No_ of_ bits= log ⌈2/0.0045⌉ = ⌈log444.44⌉ =8bits

Step4:- So after reducing number of bits binary value is 0.00011111.

Step5:- Our One time pad is – 11010100

Data- 00011111 from step 4.

After Xoring the output is 11001011

Step6:- Rotate 2 bits right the result is 11110010

Step7: .11110010 in decimal is 0.9453125

Decompression and Decryption

Step 1:- Received data is 0.96875and binary format of received data is .11110010

Step2:- Apply 2 left shifts to result of step1 the result is 11001011

Step3:- Apply selected one time pad and Xored it with the result of step2 the result is 00011111

Step4:- convert .00011111 into decimal i.e. 0.1210

Step5:- Using the probability ranges from table decodes the three character string encoded as 0.1210.

Decode first symbol
0.1210 is within [0.00, 0.30)
0.1210 encodes 'a'
Remove effects of 'a' from encode value
Current _range = 0.30 - 0.00 = 0.30
Encoded _value = (0.1210 - 0.0) ÷ 0.30 = 0.4033

Decode second symbol
0.4033 is within [0.30, 0.45)
0.4033 encodes 'b'

Remove effects of 'b' from encode value
current range = 0.45 - 0.30 = 0.15
encoded value = (0.4033 - 0.30) ÷ 0.15 = 0.6886

Decode third symbol
0.6886 is within [0.70, 0.80)
0.6886 encodes'd'

7. KEY FEATURES

The proposed technique having the following key features:

- It is a Private Key Encryption Technique
- Used both data Compression and Cryptography concept.
- Use Less Bandwidth of Secure Channel
- Highly secure
- Cipher text generated for same information always different due to one- time pad during encryption.
- In proposed system generated cipher text takes very less bandwidth of secure channel.
- Provides precision control to convert entire string or file.

8. CONCLUSION

The proposed technique provides an excellent integration of data compression with the cryptography to increases the data security and transfer rate during data communication. In this technique we can reduce the size of data using the arithmetic encoding data compression technique and after that compressed data can be encrypted to provide the security. The Present network scenario demands exchange of information with reduction in both space requirement for data storage and time for data transmission along with security. Our proposed technique fulfils all such requirements as this technique use the concept of data compression and encryption.

9. REFERENCES

[1] V.Kavitha & K.S Easwarakumar, (2008) **"**Enhancing Privacy in Arithmetic Coding" ICGST-AIML journal, Volume 8, Issue I

[2] J.A Storer, (1988) "Data Compression: Methods and Theory" Computer Science Press.

[3] Dr. V.K. Govindan & B.S. Shajee mohan *"*An intelligent text data encryption and compression for high speed and secure data transmission over internet"

[4] I.H. Willen, Randford M.Neala & John G.Cleary,(1988) , "Arithmetic Coding for Data Compression", Communications of the ACM Volume 30 Issue 6.

[5] SHANNON C. E, (1948). "A Mathematical Theory of Communication". The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656

[6] Jagdish H.Pujar & Lohit M.Kadlaskar "A new lossless method of image compression and decompression using Huffman coding technique" Journal of Theoretical and Applied Information Technology.

[7] Mamta Sharma, (2010) "Compression Using Huffman Coding" IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.5, May 2010

[8] Glen G. Langdon, (1984) "An introduction to arithmetic coding", IBM Journal of Research and Development Volume 28, No.2

[9] Amir Said, (2004) "Comparative Analysis of Arithmetic Coding Computational Complexity," Imaging Systems Laboratory HP Laboratories Palo Alto HPL-2004

[10] Whitfield Diffie & Martin E. Hellman,(1979) "Privacy and Authentication: An Introduction to Cryptography"proceedings of the IEEE, vol.67, no.3

[11] http://www.garykessler.net/library/crypto.html#intro

[12] Onwutalobi Anthony-Claret Department of Computer Science University of Wollongong "Using Encryption Technique"

[13] http://www.wwpi.com

[14] http://michael.dipperstein.com/arithmetic/

[15] www.feldstein.net/images/figure1.gif

[16] Tarek M Mahmoud, Bahgat A. Abdel-latef, Awny A. Ahmed & Ahmed M Mahfouz "Hybrid Compression Encryption Technique for Securing SMS", International Journal of Computer Science and Security (IJCSS), Volume (3): Issue(6)

[17] Deo Brat Ojha, Ajay Sharma, Abhishek Dwivedi, Nitin Pande,& Amit Kumar(2010) "Space-Age Approach To Transmit Medical Image With Code Base Cryptosystem Over Noisy Channel," ublished in International Journal of Engineering Science and Technology Vol. 2(12), 7112-7117

[18] http://www.tech-faq.com/block-and-stream-ciphers.html

Author1

Dr. Ajit Singh is presently working as Reader and Head of CSE & IT Department of School of Engineering & Sciences in BPSMV, Khanpur Kalan (Sonepat). He is also having the additional charge as a director of university computer center (UGC). He posses qualifications of B.Tech, M.Tech, ph.D (p). He is a member of BOG (Board of Governors) of Haryana State Counselling Society, Panchkula and also member of academic council in University. He published approximate 20 papers in National/ International journals and conferences and holds a teaching experience of 10 years. He holds the membership of Internal Quality Assurance cell, UG-BOS & PG-BOS and the NSS advisory committee. He is also an associate member of CSI & IETE. His research interests are in Network Security, Computer Architecture and Data Structure.

Author2

Ms. Rimple Gilhotra has completed her B.Tech degree in Information Technology from JMIT Radaur,KurushteraUniversity,Kurushtera ,India in the year 2008, and She is pursuing M.tech in Computer Science and Engineering, Bansathali university Banasthali from June2009. Currently she is Doing Internship from B.P.S.M.V Khanpur Kalan, Sonipat. Her research Interests are in Cryptography & Steganography.