Volume 1, No. 5, December 2010
Journal of Global Research in Computer Science
ISSN-2229-371X
RESEARCH PAPER
Available Online at www.jgrcs.info

SECURE AND ERROR FREE TRANSMISSION OF AUDIO/VISUAL CONTENT

Deo Brat Ojha*[1], Ajay Sharma[2], Abheshek Dwivedi[2], Bhupendra Kumar[3] and Amit Kumar[3]

[1]Professor, Department of Mathematics,Raj Kumar Goel Institute of Technology,
5th K.M. Stone Delhi – Meerut Road,Ghaziabad, U.P.201003,India
e-mail: ojhdb@yahoo.co.in
[2]Research Scholar Singhania University, Jhunjhunu, Rajsthan, India
[3]Research Scholar Mewar University, Chittorgarh, Rajsthan, India

Abstract: Human beings are always in search of process through which transmission of audio/visual content will become authentic, secure, speedy, compact, integrated and error free between two communicators. In this paper, the requirement specially to obtain the error free message with the help of error correction function with qualities above said.

Keywords: List Significant Bit (LSB), Cryptography, McEliece public-key cryptosystem, Steganography , SEQUITUR algorithm.

INTRODUCTION

Fast and secure diagnosis but error free is obligatory in the medical world to save the life of world creature. Nowadays, the transmission of images is a daily routine and it is necessary to find an efficient way to transmit them over the net [3,4,5]. For content transmission, two different approaches of technologies have been developed. The first approach is based on content protection through encryption [1], [2]. In this approach, proper decryption of data requires a key. The second approach bases the protection on digital watermarking or data hiding, aimed at secretly embedding a message into the data. In the current era, the transmission of data over internet is so much challenging over the internet. In this manner, the better way to transmit the content over internet is encryption. Using the cryptography we secure the content as well as also better utilization of the communication channel with compression technique.

Cryptography is a tool of security that aims to provide security in the ciphers of any kind of messages. Cryptographic algorithms use encryption keys, which are the elements that turn a general encryption algorithm into a specific method of encryption. The data integrity aims to verify the validity of data contained in a given document. [2]

McEliece proposed the first public-key cryptosystem (the McEliece Scheme) based on algebraic coding theory in 1978[1] . The idea behind McEliece public-key cryptosystem is based on the fact that the decoding problem of an arbitrary linear code is an NP-hard problem [2].The McEliece scheme has the advantage of high speed encryption and decryption and this system employs probabilistic encryption [1,2], which is better than other type of deterministic encryption[9] in preventing the elimination of any information leaked through public-key cryptography.

It is point of remark [9] that the security comparison is made here for classical attackers. The picture changes drastically to the advantage of the McEliece system if we consider two systems to offer the same level of security if breaking them requires quantum computers with the same number of qubits.

In this article, we introduced a content transmission with compression and encryption. This arrangement distributes in six different modules, and each module plays an important role in their manner.

PRELIMINARIES

Stegnography

Steganography is a technique used to embed secret information into non-secret information, preventing the message from being detected by non-authorized people.[15]

The purpose of steganography is to hide the very presence of communication by embedding messages into innocuous-looking cover objects, such as digital images. To accommodate a secret message, the original cover image is slightly modified by the embedding algorithm to obtain the stego image. The embedding process usually

Incorporates a secret stego-key that governs the embedding process and it is also needed for the extraction of the hidden message [16].

There are three basic views behind hiding information. The first is capacity, which is the amount of information that can be embedded within the cover file. An information-hiding algorithm has to be able to compactly store a message within a file. Next is security, which refers to how a third-party can detect hidden information within a file. Intuitively, if a message is to be hidden, an ideal algorithm would store information in a way that was very hard to notice. High security layers have been proposed through three layers to make it difficult to break through the encryption of the input data and confuse steganalysis too. Various encryption techniques like cryptography, digital watermarking, steganography etc have already been introduced in attempt to address these growing concerns [17].

Steganography have four application areas:

- Copyright Protection. It has security, invisibility and robustness requirements. Watermark techniques fit in this area.
- Authentication. It has security and invisibility requirements. Digital signature fits in this area.
- Secret and Invisible Communication. It has requirements for security, invisibility and insertion of high volumes of secret data. [18]

McEliece Public-Key Cryptosystem

Secret Key: W is a random $(k \times k)$ non-singular matrix over $GF(2)$, called the scrambling matrix, T is a $(k \times n)$ generator matrix of binary Goppa code T with the capability of correcting an n-bit random error vector of weight less than or equal to α, and Q is a random $(n \times n)$ permutation matrix.

Public Key:

$V = WTQ$

Encryption: $c = mV + e$, where m is a n-bit message, c is n -bit ciphertext, and e is an n-bit random error vector of weight α.

Decryption: The receiver first calculates $c' = cQ^{-1} = mWT + eQ^{-1}$,where Q^{-1} is the inverse of Q. Because the weight of eQ^{-1} is the same as the weight of e, the receiver uses the decoding algorithm of the original code T to obtain $m' = mW$. Finally, the receiver recovers m by computing $m = m'W^{-1}$, where W^{-1} is the inverse of W [1,2].

Data Compression

A compression scheme can be employed what is known as lossless compression on secrete message to increase the amount of hiding secrete data, a scheme that allows the software to exactly reconstruct the original message [6].

The transmission of numerical images often needs an important number of bits. This number is again more consequent when it concerns medical images. If we want to transmit these images by network, reducing the image size is important. The goal of the compression is to decrease this initial weight. This reduction strongly depends of the used compression method, as well as of the intrinsic nature of the image. Therefore the problem is the following:

1. To compress without lossy, but with low factor compression. If you want to transmit only one image, it is satisfactory. But in the medical area these are often sequences that the doctor waits to emit a diagnostic.

2. To compress with losses with the risk to lose information. The question that puts then is what the relevant information is to preserve and those that can be neglected without altering the quality of the diagnosis or the analysis. The human visual system is one of the means of appreciation, although subjective and being able to vary from an individual to another. However, this system is still important to judge the possible causes of degradation and the quality of the compression [7].

The SEQUITUR Algorithm [12]

The SEQUITUR algorithm [8] represents a finite sequence as a context free grammar whose language is the singleton set {σ}. It reads symbols one-by-one from the input sequence and restructures the rules of the grammar to maintain the following invariants:

(A) no pair of adjacent symbols appear more than once in the grammar, and

(B) every rule (except the rule defining the start symbol) is used more than once. To intuitively understand the algorithm, we briefly describe how it works on a sequence 123123. As usual, we use capital letters to denote non-terminal symbols. After reading the first four symbols of the sequence 123123, the grammar consists of the single production rule S ➔ 1, 2, 3, 1 where S is the start symbol. On reading the fifth symbol, it becomes S ➔ 1, 2, 3, 1, 2 Since the adjacent symbols 1, 2 appear twice in this rule (violating the first invariant), SEQUITUR introduces a non-terminal A to get

S ➔ A, 3,A A ➔1, 2

Note that here the rule defining non-terminal A is used twice. Finally, on reading the last symbol of the sequence 123123 the above grammar becomes

S ➔ A, 3, A, 3 A ➔ 1, 2

This grammar needs to be restructured since the symbols A, 3 appear twice. SEQUITUR introduces another non-terminal to solve the problem. We get the rules

S ➔ B,B B ➔ A 3 A➔ 1 2

However, now the rule defining non-terminal A is used only once. So, this rule is eliminated to produce the final result.

S ➔ B, B B ➔ 1, 2, 3

Note that the above grammar accepts only the sequence 123123.

Error Correction Code:

A metric space is a set C with a distance function dist: $C \times C \rightarrow R^+ = [0, \infty)$, which obeys the usual properties (symmetric, triangle inequalities, zero distance between equal points)[12].

Definition: Let $C\{0,1\}^n$ be a code set which consists of a set of code words c_i of length n. The distance metric between any two code words c_i and c_j in C is defined by

$$dist(c_i, c_j) = \sum_{r=1}^{n} \left| c_{ir} - c_{jr} \right| \qquad c_i, c_j \in C$$

This is known as Hamming distance [12].

Definition: An error correction function f for a code C is defined as $f(c_i) = \{c_j / dist(c_i, c_j) \text{ is the minimum, over } C - \{c_i\}\}$. Here, $c_j = f(c_i)$ is called the nearest neighbor of c_i [12].

Definition: The measurement of nearness between two code words c and c' is defined by nearness $(c, c') = dist(c, c') / n$, it is obvious that $0 \leq$ nearness $(c, c') \leq 1$ [12].

Definition: The fuzzy membership function for a codeword c' to be equal to a given c is defined as [12]

$FUZZ(c') = 0$ if nearness$(c, c') = z \leq z_0 < 1$

$= z$ otherwise

OUR APPROACH

In our proposed scheme, we encrypt the original text message letter by letter applying a function, which involves certain mathematical operation using corresponding letters and also numbers from the original image. We use sequitur as a compression technique and McEliece as an encryption technique, which give us improved result. The McEliece scheme has the advantage of high speed encryption and decryption and this system employs probabilistic encryption.

Sequitur is a single-pass hierarchical algorithm that builds a context-free grammar for a string. The resulting grammar compactly represents the original structure and has the interesting property that the compressed format itself contains useful information about the string. Then Hide compressed and encrypted text into cover image using Stegnography algorithm i.e List Significant Bit (LSB) coding is the way to embed information in cover image file. In this LSB technique is applied on compressed encrypted message. It is really

appreciable method to provide high security to the high confidential image.
The proposed method is enhanced or characterized by robustness, larger amount of secrete data, less time complexity and especially high security.
Input an audio/visual content follows these phases:

Module1. Process for Convert Medical Image File into Message Bit

Phase 1.: Generating $n \times n$ blocks
In RGB space the image is split up into red, blue and green images. The image is then divided into 8×8 blocks of pixels and accordingly the image of $w \times h$ pixels will contain $W \times H$ blocks. Where, $W = w/8, H = h/8$.

Phase 2: DCT
All values are level shifted by subtracting 128 from each value. The Forward Discrete Cosine Transform of the block is then computed. The mathematical formula for calculating the DCT is:

$$\mathbf{T(u,v) = \sum_{x=0}^{n} \sum_{y=0}^{n} f(x,y), g(x,y,u,v)}$$

Where,

$$\mathbf{g(x,y,u,v) = \frac{1}{4}\alpha(u)\alpha(v)\cos\left[\frac{(2x+1)u\pi}{2n}\right]\cos\left[\frac{(2y+1}{2n}\right.}$$

Where

$$\mathbf{\alpha(u) = \begin{cases} \frac{1}{\sqrt{2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{for. u} = 0 \\ 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{for. u} = 1, 2, \ldots . N - 1 \end{cases}}$$

Phase 3: Quantization
Quantization is the step where the most of the compression takes place. DCT really does not compress the image, as it is almost lossless. Quantization makes use of the fact that, the high frequency components are less important than the low frequency components. The Quantization output is

$$\mathbf{Q_{DCT} = round\left(\frac{T(u,v)}{Z(u,v)}\right)}$$

The $\mathbf{Z(u,v)}$ matrix could be anything, but the JPEG committee suggests some matrices which work well with image compression.

Module2. Algorithm for Encrypting the Confidential Message

Encryption using McEliece Cryptosystem
Secret Key: W is a random $(k \times k)$ non-singular matrix over $GF(2)$, called the scrambling matrix, T is a $(k \times n)$ generator matrix of binary Goppa code T with the capability of correcting an n-bit random error vector of weight less than or equal to α, and Q is a random $(n \times n)$ permutation matrix.
Public Key : $V = WTQ$
Encryption : Let $g : m \to mV$,where m is a 64-bit message. Then after for the sake of secrecy add error e , which is a n-bit random error vector of weight α.
then $E = g(m) + e$, E is a n-bit ciphertext

Module 3. Algorithm to Embed Confidential Message into Cover Image File

Algorithm to embed confidential message into cover image file named inFile generate new file with embedded message file named outFile.
Encoded-Message (msg,inFile on input-mode, outFile on output-mode)
Step 1:
Read offset bytes from input inFile and writes to output File outFile
Step 2:
Calculate message length and write it into output file by embedding using XOR function it in last two bits for every byte. Suppose, Message length being 16 bits, will be stored in 8 pairs of 2 bits.
Step 3:
Embed each byte of message in 4 pairs of 2 bits each is embedded in 4 byte of input file and written into output file named outFile.
Step 4:
Write the remaining bytes of the input file into output file.

Module4. Algorithm for Compress the Confidential Message

Perform the lossless compression technique (SEQUITUR) on cipher text to increase the amount of hiding secrete data.

Compression using SEQUITUR

After quantization, the scheme uses a filter to pass only the string of non-zero coefficients. By the end of this process we will have a list of non-zero tokens for each block preceded by their count.
DCT based image compression using blocks of size 8x8 is considered. After this, the quantization of DCT coefficients of image blocks is carried out. The SEQUITUR compression is then applied to the quantized DCT coefficients.

Module 5. Algorithm for Generate of Message from Image

The picture is received at receive side. This function decode message from a file named outFile open on output mode. Decode Message (outFile on Input-mode)
Step 1:
Read offset bytes from the input file and apply again XOR function, Generate message bit.
Step 2:
Read last 2 bits of consecutive 8 bytes and concatenate them to get the message length.
Step 3:
Read last 2 bits from input file in pairs of 4 and concatenate them to get message of 1 byte.
Step 4:
Repeat step 3 until the message is extracted of calculated length.
Step 5:
Decompress & Decrypt the message.

Module6. Procedure for detecting and correcting error

If any error occurred during the transmission of message, we can detect and correct using fuzzy error correcting code.

Receiver check that $dist(t(c)c') > 0$, he will realize that there is an error occurs during the transmission. Receivers apply the error correction function f to $c' : f(c)$.

Then receiver will compute nearness $(t(c), f(c')) = dist(t(c) f(c')) / n$

$$FUZZ(c') = 0 \quad \text{if nearness}(c, c') = z \le z_0 < 1$$
$$= z \quad \text{otherwise}$$

CONCLUSIONS

In the Health Insurance Portability and Accountability Act (HIPAA) [11] requires that medical providers and insurance companies implement procedures and policies to protect patient's medical information.

In this paper, we used McEliece scheme due to its probabilistic property. The efficiency and security of McEliece cryptosystem comparatively better than the RSA cryptosystem also[9,10]. Here we use LSB as stegnography, SEQUITUR as a compression technique and SEQUITUR has the ability to read a stream in reverse also. So our approach is more appropriate, secure and futuristic than previous literature of medical data transmission over un-secure channel. .

REFERENCES

[1] R.J.McEliece, " A public-key cryptosystem based on algebraic coding theory," DSN Progress Report,42-44,1978,pp.114-116.

[2] E.R.Berlekemp, R.J.McEliece, and H.C.A. vanTilborg, " On the inherent intractability of certain coding problems," IEEE Transactions on Information Theory, vol.24, No.5, 1978, pp.384-386.

[3] G. Lo-varco,W. Puech, and M. Dumas. "Dct-based watermarking method using error correction codes", In ICAPR'03, International Conference on Advances inPattern Recognition, Calcutta, India, pages 347–350, 2003.

[4] R. Norcen, M. Podesser, A. Pommer, H.P. Schmidt, and A. Uhl. "Confidential storage and transmission of medical image data", Computers in Biology and Medicine, 33:277–292, 2003.

[5] Diego F. de Carvalho, Rafael Chies, Andre P. Freire, Luciana A. F. Martimiano and Rudinei Goularte, "Video Steganography for Confidential Documents: Integrity, Privacy and Version Control" , University of Sao Paulo – ICMC, Sao Carlos, SP, Brazil, State University of Maringa, Computing Department, Maringa, PR, Brazil.

[6] Nameer N. EL-Emam, "Hiding a Large Amount of Data with High Security Using Steganography Algorithm" Applied Computer Science Department, Faculty of Information Technology, Philadelphia University, Jordan

[7] Borie J., Puech W., and Dumas M., "Crypto-Compression System for Secure Transfer of Medical Images", 2nd International Conference on Advances in Medical Signal and Information Processing (MEDSIP 2004), September 2004.

[8] N.Walkinshaw, S.Afshan, P.McMinn "Using Compression Algorithms to Support the Comprehension of Program Traces" Proceedings of the International Workshop on Dynamic Analysis (WODA 2010) Trento, Italy, July 2010.

[9] Johannes Buchmann, Carlos Coronado, Martin D¨oring, Daniela Engelbert,Christoph Ludwig, Raphael Overbeck, Arthur Schmidt ,Ulrich Vollmer, Ralf-Philipp Weinmann , "Post- Quantum Signatures",eprint.iacr.org/2004/297.

[10] Canteaut and N. Sendrier, "Cryptanalysis of the original McEliece Cryptosystem, Advances in Cryptology" - ASIACRYPT '98 Proceedings, Springer-Verlag ,1998, pp.187–199.

[11] "Health Insurance Portability and Accountability Act (HIPAA) and Its Impact on IT Security," Regulatory Compliance Series 3 of 6, Apani Networks White Paper

Compliance Series. May 12, 2005. http://www.apani.com.

[12] V.Pless, " Introduction to theory of Error Correcting Codes", Wiley , New York 1982.

[13] Jessica Fridrich and Miroslav Goljan, Digital image steganography using stochastic modulation, Department of Electrical and Computer Engineering, SUNY Binghamton, Binghamton, NY, 13902-6000, USA.

[14] Swarnendu Mukherjee, Swarnendu Bhattacharya, Amlan Chaudhury Triple Layer Data Security ACM Ubiquity, Volume 9, Issue 17, April 29-May 5 ,2008

[15] Zhao, J. In business today and tomorrow, ACM Communications of the ACM, 1998.

[16] Diego F. de Carvalho, Rafael Chies, Andre P. Freire, Luciana A. F. Martimiano and Rudinei Goularte, Video Steganography for Confidential Documents: Integrity, Privacy and Version Control , University of Sao Paulo – ICMC, Sao Carlos, SP, Brazil, State University of Maringa, Computing Department, Maringa, PR, Brazil.

