ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Techno Research Publishers, Bhopal, India.**
www.computerscijournal.org

ISSN: 0974-6471
April 2015,
Vol. 8, No. (1):
Pgs. 43-48

DNA-based Audio Steganography

RASHMI M. TANK[1], HEMANT D. VASAVA[2] and VIKRAM AGRAWAL[3]

[1]PG Student of Computer Department, B.V.M. Engineering College,Gujarat Technological University,Vallabh Vidhyanagar, Gujarat, India.
[2]Assistant Prof. of Computer Department, B.V.M. Engineering College, Gujarat Technological University,Gujarat Technological University, Vallabh Vidhyanagar,Gujarat, India.
[3]Assistant Prof. of IT Department, B.V.M. Engineering college, Gujarat Technological University, Vallabh Vidhyanagar,Gujarat, India.

(Received: February 16, 2015; Accepted: April 18, 2015)

ABSTRACT

Security is the important criteria relevant to information in transit as well as in storage. Steganography is the technique of hiding secret message in a cover medium in such a way that only the sender and the intended recipient knows the existence of communication. DNA due to its immense storage capacity and high randomness is used now in the field of steganography. Audio steganography is concerned with hiding information in a cover (host) audio signal in an imperceptible way. In this paper, various techniques using DNA sequences and audio files for data hiding is discussed for secure data transmission and reception. It also proposes highly secure method to hide the existence of secret message to prevent unauthorized access. The proposed method has three levels. Single level of encryption and two levels of steganography are used. The main objective of this method is that no one could be able to find the existence of secret message.

Key words: Steganography, Data hiding, Information Security, DNA Sequence, Audio Steganography.

INTRODUCTION

The explosive growth of computer system and their interconnection via computer networks such as internet has led to a heightened need for information and data security. The growing use of internet has led to a continuous increase in the amount of data that is being exchanged and stored in various digital media[1]. In network it is very important to protect the data. In order to achieve this there are several technologies are being used. One of the secured ways to protect the data is encryption and decryption method[5]. Cryptography is the science and art of protecting information by scrambling its content into unreadable format called *ciphertext.* Cryptography system can be broadly classified into *symmetric system* that use a single key for both encryption and decryption, and *asymmetric systems* that use one key for encryption and another for decryption. The Data Encryption Standard (DES) is the most common symmetric cryptography method that is in use today. Examples of asymmetric cryptography algorithm include RSA, Diffie and Hellman and Digital Signature Algorithm (DSA) which is a variant of ElGamal[2]. The objective of secure communication is that actual data should not be revealed to any third party. Steganography techniques are most successful technique in supporting hiding of critical information in ways that prevent the detection of hidden messages. While cryptography scrambles the message so that it cannot be understood[1].

The conventional methods of encrypting are not strong enough today for providing the data security and reliable data transmission. Unauthorized user or intruders may attack and can interrupt or intercept the message for doing some malicious tasks. In order to enhance the data security effective encryption algorithms are required. Recent research has shown DNA as a medium for large scale computation system[4]. DNA computing is a new method of simulating bimolecular structure of DNA and computing by means of molecular biological technology which is a novel and potential growth. Adleman demonstrated the first DNA computing. It marked the beginning of a new stage in the era of information. DNA (Deoxyribonucleic Acid) is the germ plasma of all life styles[1]. Two different kinds of genetic material exist, Deoxyribonucleic acid (DNA) and Ribonucleic Acid (RNA) present in a cell. DNA contains four types of nucleotides like, Adenine (A), Thymine (T), Cytosine (C) and Guanine (G). James D. Watson and Francis Crick were the two co-discoverers of the structure of DNA in 1953 [5]. In a double helix DNA string, two strands are complementary in terms of sequence, that is A to T and C to G according to Watson-Crick rules[1]. There are a large number of DNA sequences publicly available in various domains of biological DNA. A rough estimation would put the number of DNA sequences publicly available in various websites are around to be 55 million[4].

Steganography is the art and science of hiding information such that its presence cannot be detected. The secret information is hidden in some carrier file and then transmitted. The carrier file can be an image, audio file, text file, video file etc[6]. Among different steganography schemes, the image steganography is widely used. But increase use of voice over Internet Protocol (VoIP) and various Peer-to-Peer (P2P) audio services encourage researcher to choose audio steganography[7]. Audio steganography requires a text or audio secret message to be embedded within a cover audio message. Audio files are considered to be excellent carriers for the purpose of the steganography due to presence of redundancy. Due to availability of redundancy, the cover audio message before steganography and stego message after steganography remains same. However, audio steganography is considered more difficult than video steganography because the Human Auditory System (HAS) is more sensitive than Human Visible System (HVS)[8]. A steganography system is more expected to meet three key requirements, namely transparency, capacity and robustness[9]. Least Significant Bit (LSB) is one of the earliest and simpler methods used for information hiding of digital audio. Traditional, LSB is based on inserting each bit from the message in the LSB of binary sequence of each sample of cover digitized audio file[7].

The objective of this paper is to come up with an efficient method to preserve security of secret messages in a text file against unauthorized access by hiding the presence of the text file. The method proposed in this paper works in three levels. The three levels are encryption, DNA steganography and audio steganography. For encrypting the text file, the proposed method uses DNA based RSA encryption algorithm. In the second level, the encrypted secret file is hidden in a DNA sequence taken from publicly available database on NCBI website. In the third level, the DNA sequence which is embedded with the encrypted message is hidden inside an audio file using Least Significant Bit Modification technique.

The motivation of preparing this paper is given in next section. The section III introduces various techniques using DNA sequences and audio file. The section IV gives description of proposed method. Section V enlists applications of proposed method.

Motivation

Information security is one of the important fields in which researches are taking place. Main goal of it is to prevent unauthorized access to the secret message as well as to hide the existence of the secret message. Steganography is the main method to hide the message.

DNA due to its immense storage capacity and high randomness is used now in the field of steganography. This can be considered as recent technique in steganography. A large number of researchers take an initiative for implementing DNA encoding concept in the applications like cryptography, scheduling, clustering, GPU applications, multi-core architectures, forecasting and even trying to apply this in signal and image processing algorithm[1, 4].

Since Human Auditory System (HAS) is more sensitive than Human Visible System (HVS)[8], and since audio files are redundant and highly available, audio steganography is of high importance in the field of steganography. Hence more techniques are to be found out in this field[1].

Literature review

The following is some of the DNA based data encryption and hiding techniques reported recently. Shyamasree C M and Sheena Anees[1] proposed the DNA based Audio Steganography method which works in three levels. First level makes use of DNA based playfair algorithm. The second level hides the secret message in a randomly generated DNA sequence using two-by-two complementary rule. In the third level embedded DNA is hidden inside the Audio file using LSB substitution technique. The indexing technique is proposed by K Menaka[3] uses 3 complementary rules: based on Purine and Pyrimidine, based on Amino and Keto group, based on Strong and Weak H-bonds. Other technique proposed by Bama R, Deivanai S, and Priyadharshini K[4] in the field of DNA steganography is substitution technique which also uses complementary pair rule. The technique proposed by Siddaramappa V[5] uses random function to generate random numbers and uses binary arithmetic to encrypt the message using DNA sequences. The above techniques does not preserve functionality of DNA and for extraction of original message, it needs reference DNA sequence. So sender and receiver must agree on reference DNA sequence. To preserve the functionality of DNA and to make blind algorithm, the latest technique proposed by Amal Khalifa[2] is LSBase key encapsulation scheme using DNA steganography. The session key is hidden inside a chosen DNA sequence and hence can be securely exchanged between parties through public channels such as the internet. This technique uses codon degeneracy to hide the information within DNA sequences without actually affecting the type or structure of the protein they code for.

The steganography hides the secret information inside some carrier files. Various substitution techniques for audio steganography are proposed. The robust substitution technique proposed by Rohit Tanwar, Bhasker Sharma and Sona Malhotra[6] uses the deeper layer bits for embedding and other bits are altered willingly to decrease the amount of error induced. A new steganography technique proposed by Pratik Pathak, Arup Kr. Chattopadhyay and Amitava Nag [7] is based on location selection. It randomizes a bit number of host message used for embedding secret message based on upper three MSB. To improve this technique by randomizing sample number containing next secret message bit, a new technique is proposed by Muhammad Asad, Junaid Gilani and Adnan Khalid[8]. The technique proposed by Anupam Kumar Bairagi, Saikat Mondal and Amit Kumar Mondal[9] hides the secret bit in the deeper layers of the sample based on total number of ones in that sample.

Proposed method

Encryption

In the first level of proposed method, secret message is encrypted using DNA based RSA encryption algorithm. The secret message in a file is converted to DNA sequence using DNA digital coding pattern. From a computational point of view, any DNA sequence can be encoded using a binary coding scheme, in which anything can be encoded by a combination of the two states 0 and 1. Therefore simplest coding pattern to encode the 4 nucleotide bases (A, U, C, G) is: 0(00), 1(01), 2(10), 3(11) respectively. Obviously, there are 4!. So, among these 24 patterns, only 8 kinds of patterns (0123/CUAG, 0123/CAUG, 0123/GUAC, 0123/GAUC, 0123/UCGA, 0123/UGCA, 0123/ACGU and 0123/AGCU) which are topologically identical fit the complementary rule of the nucleotide bases. It is suggested that the coding pattern in accordance with the sequence of molecular weight, 0123/CUAG, is the best coding pattern for the nucleotide bases as illustrated in table 1[1].

The message in DNA form is converted to sequence of codon triplets. These codons are converted to amino acids as illustrated in table 2. The ambiguity numbers are needed to be sent to the receiver for decryption process. The ambiguity numbers used are 0, 1, 2, 3, 4 and 5 corresponding to the first, second, third, fourth, fifth and sixth codon respectively. The amino acid sequence is encrypted using RSA encryption algorithm. The pseudorandom number generator proposed by Ankur, Divyanjali and Vikas Pareek[10] is used in the generation of public key and private key for RSA encryption algorithm. Hence the output of the DNA based encryption is DNA based encrypted sequence and sequence of ambiguity numbers.

DNA Steganography

This section gives the idea about the second level of the proposed method which is the DNA steganography. The encrypted DNA sequence obtained from the first level is hidden inside a reference DNA sequence taken from publicly available database. Randomized LSBase substitution technique is used for hiding a DNA sequence in other. Here the property of codon degeneracy (i.e. multiple codons may produce same amino acid) is utilized in order to change the codon's last base while keeping its type (either Purine or pyrimidine).

Table 1: DNA Digital coding[1]

DNA Nucleotide	Decimal	Binary
A	0	00
C	1	01
G	2	10
U	3	11

Four exceptions should be considered. First the tryptophan (Trp) and methionine (Met) have a single codon, so they can't be used for embedding. The same is true for the stop codon UGA. The fourth case actually appears in the amino acid Isoleucine (Ile), since it is coded by three codons: AUU, AUC and AUA. Therefore, only AUU and AUC can be used while the AUA is neglected.

The codons in the reference DNA sequence are first randomized using Linear Congruential Generator (LCG). Table 3 shows a mapping table for hiding 3 encrypted binary bits at a time inside the least significant base of each codon. The seed and modulus of the LCG need to be sent to the receiver securely. For that purpose it is encrypted using Elliptic Curve Cryptography (ECC).

Here LSBase technique is the blind technique. So the actual reference DNA sequence needs not to be stored for extraction process. The output of DNA steganography is embedded DNA sequence in which actual encrypted DNA sequence is hidden.

Table 2: A Mapping of the DNA Codons Into amino acids

Codons	Character	Codons	Character
GCU,GCC,GCA, GCG	A	AAU, AAC	N
UAA,UAG,UGA	B	UUA,UUG	O
UGU,UGC	C	CCU,CCC,CCA,CCG	P
GAU,GAC	D	CAA,CAG	Q
GAA,GAG	E	CGU,CGC CGA,CGG,AGA,AGG	R
UUU,UUC	F	UCU,UCC,UCA,UCG,AGU,AGC	S
GGU,GGC,GGA,GGG	G	ACU,ACC,ACA,ACG	T
CAU,CAC	H	AGA,AGG	U
AUU,AUC,AUA	I	GUU,GUC,GUA,GUG	V
-	J	UGG	W
AAA,AGG	K	AGU,AGC	X
UUA,UUG,CUU,CUC,CUA,CAG	L	UAU, UAC	Y
AUG	M	-	Z

Audio Steganography

This section illustrates the third level of proposed method. The embedded DNA sequence obtained as the output of DNA steganography is hidden in an audio file to hide the existence of the secret data.

Least Significant Bit (LSB) modification is used for audio steganography. Audio files used are in WAV format. The audio file read in a binary format. The length of the embedded DNA sequence is encoded in lower half of the first 32 audio samples and the sequence itself is encoded in lower half of the remaining audio samples. This will create no distortion.

Extraction of Secret File

This section illustrates how the secret file is extracted from the stego audio file. The extraction process consists of the reverse of all the process that had been done to hide the secret file containing the secret message.

After the encryption phase and two steganography phases, there is a sequence of ambiguity numbers and seed and modulus of LCG as output. Ambiguity numbers from the encryption phase is necessary in the decryption phase. Seed and modulus from DNA steganography is necessary for generating random numbers during extraction process.

As the first step in extraction process, the stego audio file is sampled. From the first 32 samples the length of the embedded DNA sequence encoded in it is decoded. Then from the remaining samples, in which the sequence itself is encoded, is decoded. Using table 3 and ambiguity numbers, the binary encrypted message is extracted from embedded DNA sequence. The seed and modulus are decrypted using elliptic curve cryptography. The DNA based RSA decryption algorithm is used to decrypt the encrypted message and recover original message.

Table 3: Mapping of three binary bits into DNA Nucleotide Base

	Least Significant Base of Codon	
Message Bits	Purine(A/G)	Pyrimidine(T/C)
000	A,0	T,0
001	A,1	T,1
010	A,2	T,2
011	A,3	T,3
100	G,0	C,0
101	G,1	C,1
110	G,2	C,2
111	G,3	C,3

Applications

DNA due to its immense storage capacity and high randomness is used now in the field of steganography.DNA based algorithms can be used in various fields such as job scheduling for clusters, GPU applications, multi-core architectures, etc. Audio files are considered to be excellent carriers for the purpose of steganography due to presence of redundancy[1].

In the business world steganography can be used to hide a secret chemical formula or plans for a new invention. Steganography can also be used for corporate espionage by sending out trade secrets without anyone at the company being any the wiser. Terrorists can also use steganography to keep their communications secret and to coordinate attacks. There are a number of peaceful applications. The simplest and oldest are used in map making, where cartographers sometimes add a tiny fictional street to their maps, allowing them to prosecute copycats. A similar trick is to add fictional names to mailing lists as a check against unauthorized resellers. Most of the newer applications use steganography like a watermark, to protect a copyright on information. Photo collections, sold on CD, often have hidden messages in the photos which allow detection of unauthorized use. The same technique applied to DVDs is even more effective, since the industry builds DVD recorders to detect and disallow copying of protected DVDs[13].

CONCLUSION

Communicating secretly without giving away any kind of crucial information is very important now a days in many fields. In this paper, we proposed a method to hide the secret messages stored in text files from unauthorized access. The method can be applied to text file. The proposed method has three levels. First level is DNA based encryption. Second level is DNA steganography which hides the encrypted message inside the reference DNA sequence. In the third level the embedded DNA sequence is hidden inside the audio file. Applications of proposed method are listed. In conclusion, in the proposed scheme the message has gone through encryption as well as two stages of steganography.

ACKNOWLEDGMENT

I am very grateful and would like to thank my guide for their advice and continued support to complete this paper and help to think beyond the obvious.

REFERENCES

1. Shyamasree C M, Sheena Anees “Highly Secure DNA-based Audio Steganography” International Conference on Recent Trends in Information Technology (ICRTIT) IEEE 2013.
2. Amal Khalifa “LSBase: A key encapsulation scheme to improve hybrid crypto-systems using DNA steganography “ IEEE 2013
3. K. Menaka “Message Encryption Using DNA Sequences “IEEE 2014
4. Bama R, Deivanai S, Priyadharshini K “Secure Data Transmission Using DNA Sequencing” *IOSR Journal of Computer Engineering (IOSR-JCE),* **16**(2), Ver. II (Mar-Apr. 2014)
5. Siddaramappa V “Data Security in DNA Sequence Using Random Function and Binary Arithmetic Operations “ *International Journal of Scientific and Research Publications,* **2**(7); (2012).
6. Rohit Tanwar, Bhasker Sharma and Sona Malhotra “A Robust Substitution Technique to implement Audio Steganography “ International Conference on Reliability, *Optimization and Information Technology, IEEE* (2014)
7. Pratik Pathak, Arup Kr. Chattopadhyay and Amitava Nag “A New Audio Steganography Scheme based on Location Selection with Enhanced Security” IEEE.
8. Muhammad Asad, Junaid Gilani and Adnan Khalid “An Enhanced Least Significant Bit Modification Technique for Audio Steganography” IEEE 2011
9. Anupam Kumar Bairagi, Saikat Mondal and Amit Kumar Mondal “A Dynamic Approach In Substitution Based Audio Steganography” IEEE/OSA/IAPR International Conference on Infonnatics, Electronics & Vision , 2012
10. Ankur, Divyanjali and Vikas Pareek “A New Approach to Pseudorandom Number Generation” Fourth International Conference on Advanced Computing & Communication Technologies, IEEE 2014.
11. Tushar Mandge, Vijay Choudhary “A DNA Encryption Technique Based on Matrix Manipulation and Secure key Generation Scheme” IEEE.
12. Swarnendu Mukherjee, Debashis Ganguly, Swarnendu Bhattacharya, Partha Mukherjee”A Cognitive Study on DNA Based Computation” *International Journal of Recent Trends in Engineering,* **1**(2); (2009)
13. Ronak Doshi, Pratik Jain, Lalit Gupta “Steganography and its Applications in Security” *International Journal of Modern Engineering Research (IJMER)* **2**(6); 4634-4638 (2012).
14. Padma Bh, D. Chandravathi, P. Prapoorna Roja “Encoding and Decoding of a Message in the Implementation of Elliptic Curve Cryptography using Koblitz’s Method” *International Journal on Computer Science and Engineering(IJCSE)* **2**(5); 1904-1907 (2010).