IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. II (Mar-Apr. 2014), PP 19-22
www.iosrjournals.org

Secure Data Transmission Using Dna Sequencing

Bama R[1], Deivanai S[2], Priyadharshini K[3]
[1]*Associate Professor, CSE, Sri Sai Ram Engineering College, Anna University, Chennai*
[2]*PG Scholar Sri Sai Ram Engineering College, Anna University, Chennai*
[3]*PG Scholar Sri Sai Ram Engineering College, Anna University, Chennai*

Abstract: *The medical record system includes substantial information such as diagnoses of diseases, treatment undergone, patient's health condition, medication information, examination results and nursing actions. Consequently, these information helps medical staff understand the patient's medical record instantly and can provide accurate diagnoses for the disease. Traditional paper based medical report system has many problems including illegible handwriting, conversation difficulty, slow data transfer, easily damaged and rampant . To overcome such drawbacks, we reformed it into electronic patient records. With the increased development of internet, sharing the medical reports over a public domain is not much secured that it can be accessed by attackers and destroy them. To protect these information from attackers, existing system provides access control through Lagrange Interpolation in which the key we use is randomized, no relationship between each key so that the key can be managed effectively. To provide access control and to enhance security Elliptic Curve Cryptography was used in which the attackers encounter Elliptic Curve Discrete Logarithmic Problem. Once the prime number is big enough, attackers will have trouble deciphering the key but it increases the size of the encrypted message, is difficult, increases implementation errors and not much secured too. So that, we proposed DNA sequencing which ensures secured data authorization, storage and transmission.*

Keywords*: Access Control , Mobile Agent, DNA Sequence*.

I. Introduction

Access control has applied like database management system, online pay-tv system and electronic subscription system, etc. Mobile agent is a self distributed computing program between each host and switch information host to host on Internet. Here, Mobile agent represent the doctor, through remote connection it will be secure to exchange medical information. Also it is autonomy that can decrease delays of transmission, reduce network traffic and applies to different platforms. Mobile Agent's characteristics includes fault tolerance, adjustment and personalization, it helps to send message and can exchange with other individual systems or different mobile agents.

Mobile agent's function is to acquire the information given by the patients and transmit the information through the Internet or other related services and platforms in order to search or deal with information. When mobile agent receive the medical information of the patient, they will provide the patient with necessary medication and the update the medication details in the database[3]. These records can be transmitted or exchanged from a particular hospital information system to another hospital host.

Although the mobile agent technology is convenient for medical network system or other businesses, transmitting the information through a public domain (internet) is not much secured and also its necessary to have access control and security to prevent illegal behaviors from attackers.

At present Key Management and Access control is provided by means of Lagrange Interpolation Polynomial and Elliptic Curve Cryptography. But it is not considered to be much secured, since most of the information system has been used in the network environment and can be accessed by the public. And in the transmission process, the exchange of information is more likely to be stolen or destroyed. The process of encryption and decryption helps in transmitting the confidential information, so that only the authorized person can able to access those information.

Mobile agent is an important application of access control mechanisms, and brings great convenience to medical institutions, but it still has a lot to improve in aspect of security and performance. Concerning security problems in public network of mobile agent we encounter some defects, to correct them, what we mentioned here is to establish the completely safe access control mechanism. We applied DNA sequencing for encryption and decryption and try to keep the access control mechanism in medical environment with mobile agent.

II. Cryptography And Dna Sequencing

In today's world, security is one of the most significant issues of data transmission, researchers are working on the evolvement of new cryptographic algorithms. Cryptography is the process of providing security

of data transmission via public network by encrypting the original data or message. The plain text will be converted to the message which cannot be read by human. An efficient direction of providing data security can be termed as DNA based Cryptography. The encryption and decryption process proposed in this paper will use the DNA sequencing property of the DNA. We have proposed here how the DNA sequencing can be utilized in cryptographic algorithms and how the message can be made more secure and reliable for transmitting effectively via networks.

DNA sequencing is also dependent on our ability to use gel electrophoresis to separate strands of DNA that differ in size by as little as one base pair[4].In order to obtain complex computation in the process of achieving the cipher text recent trends are focused on DNA computing and DNA based encrypting algorithms. One of widely used process of secret writing is called cryptography which provides data and information security and protects that information from several malicious attacks. Security is concerned with the protection and providing security on network and data while transmitting over the network. But to achieve complete security against attacks is a challenging issue of data communications.

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. One potential key application of large scale computation system is DNA based cryptography. A large number of researcher take an initiative for implementing DNA encoding concept in the applications like cryptography, scheduling, clustering, forecasting and even trying to apply this in signal and image processing application . From few years back, most of the research works have been going on DNA based encryption schemes. Biological properties of DNA sequences are used in almost of the cryptographic works. In this paper, we have proposed a new technique where biological properties are not directly used.

The first one is a segment of DNA sequence of Litmus, its real length is with 2856 nucleotides long:

ATCGAATTCGCGCTGAGTCACAATTCGCGCTGAGTCACAATTCGCGCTGAGTCACAATTGTGACTC
AGCCGCGAATTCCTGCAGCCCCGAATTCCGCATTGCAGAGATAATTGTATTTAAGTGCCTAGCTCG
ATACAATAAACGCCATTTGACCATTCACCACATTGGTGTGCACCTCCAAGCTCGCGCACCGTACCG
TCTCGAGGAATTCCTGCAGGATATCTGGATCCACGGAAGCTTCCCATGGTGACGTCACC.

The second one is a segment of DNA sequence of Balsaminaceae, its real length is with 2283 nucleotides long:

TTTTTATTATTTTTTTTCATTTTTTTCTCAGTTTTTAGCACATATCATTACATTTTATTTTTTCATTACT
TCTATCATTCTATCTATAAAATCGATTATTTTTATCACTTATTTTTCTAATTTCCAATATTTCATCTA
ATGATTATATTACATTAAAGAAATCGGTTAAAAGCGACTAAAAATCAATCTGGAACAAGGCTTAG
TTTATTTAATATATTATTTTATGTAATTTCTATTGAAAAATTAGTTAAAAGGCAAGTATTTGAGAT.

Instead, we have used different properties of DNA sequences in our proposed encryption scheme. A DNA sequence is a sequence consisting of four DNA bases namely: A, C, G and T. Each of the bases is related to a nucleotide. 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 web sites are around to be 55 million.

In the encryption we secretly select a reference sequence *S* from publicly available DNA sequences. Only the sender and the receiver are aware of this reference sequence. The sender would transform this selected DNA sequence *S* into a new sequence *S*' by incorporating the DNA sequence *S* with the secret message *M*. This transformed sequence *S*' is sent by a sender to the receiver together with many other DNA sequences. The receiver would then examine all of the received sequences, identify *S*' and recover the secret message *M*.

Table 1. Substitution of Alphabet

Character	Codon	Character	Codon
A	CTAG	a	CCCA
B	ATCG	b	ACGG
C	TTAG	c	AGGC
D	GTAC	d	CCAG
E	TAGC	e	ATTG
F	CATG	f	CGCG
G	GATC	g	ATAT
H	ATGC	h	GAAC
I	CCAT	i	TAGC
J	TTCG	j	CGAT
K	TACG	k	GCGC
L	AAAT	l	GTTA
M	CGCT	m	GACC
N	GCTA	n	CGGA

O	AGCT	o	CGGT
P	TTCA	p	GGCA
Q	CTAA	q	ACCC
R	AATG	r	GTAT
S	GGTA	s	TTAC
T	CTTA	t	TCGC
U	AGGT	u	ACTT
V	TTGA	v	TTAC
W	GATT	w	GCAT
X	CATT	x	CGTA
Y	ATTC	y	TACC
Z	TATG	z	GCTT
0	AAAG	!	GAAT
1	TTTG	@	CTTG
2	GCCC	#	AAAA
3	GTTT	$	GTAT
4	AACC	%	GAAA
5	AATC	^	ATGG
6	TCCC	&	TGGC
7	GTAA	*	TTAA
8	TAAA	(	TTAG
9	CTTT	)	CGGG
-	TAAG	}	AAGG
`	ATTA	[	CCCC
+	TTTC	]	TAAT
=	TATA	|	ATTT
{	CCCG	\	GCGC
;	AAAT	:	GGGG
“	ATTC	‘	AATT
<	GTTC	>	CCGG
.	ACTT	?	CCAA
/	TTTA	_	GGCC
,	CCCT	~	GGAA
space	TTTT		

III. Encryption & Decryption In Dna Sequencing

We assume that there are two schemes used by the sender and the receiver which are kept secret. The first one is a binary coding scheme which transforms alphabets A, C, G and T into binary codes and vice versa. For instance, the following may be a binary coding used: ((A 00) (C 01) (G 10) (T 11))[2]. It should be noted that more digits may be used. The second scheme is a complementary pair rule. That is, we shall assign each alphabet a complement, denoted as C(x). We stipulate that C(C(*x)) ≠x*. The following may be such a rule: ((A C) (C G) (G T) (TA)). The various operations that can be performed on DNA are ligation, polymerase chain reaction (PCR), gel electrophoresis and affinity purification.

The Substitution Approach

For this method, we also will use a reference sequence *S*. Let us assume that *S*=ACGGAATTGCTTCAG and the secret message $M=m_1 m_2 \ldots m_p$ is 0111010. The length of *S* is 15 and is larger than the length of *M*, *p*,which is 7 in this case. For illustration, assume that the complementary rule is the same as given in the above sections. That is, the rule is as follows: ((A T) (C A) (G C) (T G)).

Our main idea is as follows:

Step 1. Suppose the length of the reference sequence *S* is 15. Select *p* distinct numbers randomly from 1 to 15, *p* is equal to 7 in this case. Assume that they are sorted as 2, 3, 5, 10, 12, 13 and 15. Let $A=A_1, A_2, \ldots, A_p=\{2,3,5,10,12,13,15\}$.

Step 2. Transform *S* into *S*’ by the following rule:

For all *i* from 1 to 15,

if *i* is equal to some A_j and m_j is 1, $1 \le j \le p$, set S_i to $C(S_i)$,

if *i* is equal to some A_j and m_j is 0, do not change S_i, and

if *i* is not equal to any A_j, set S_i to $C(C(S_i))$.

Thus *S*’=GCCATGCCAACTAGG.

Step 3. Send *S*’ to the receiver with many other sequences.

The receiver would not need to generate the set *A*. After receiving a set of sequences, he will check all positions of each sequence *S*' in the set. There are only three possible cases:
(1) S'_i is the same with S_i (The secret bit m_j is equal to 0).
(2) S'_i is $C(S_i)$(The secret bit is equal to 1).
(3) S'_i is $C(C(S_i))$. If there exists one *j* such that S'_j and S_j are not of the above three cases, it means that the sequence should be ignored.

Encryption Algorithm for Substitution Approach
Input: A DNA sequence *S*, the secret message $M=m_1 m_2 \ldots m_p$ and a complementary rule.
Output: An encrypted DNA sequence *S*".
Step 1. Use a random number generator to generate a set of available integer sequences, called set *A*. The number of *A* is *p*, the length of *M*.
Step 2. Initialize *i* and *j* to 1.
Step 3. For each element S_i of *S*, do the following operations:
if *i* is equal to some A_j and m_j is 1, $1 \le j \le p$, change S_i to $C(S_i)$,
if *i* is equal to some A_j and some m_j is 0, do not change S_i,
if *i* is not equal to any A_j, set S_i to the double-complement of itself.
Step 4. Return *S*'.

Decryption Algorithm for Substitution Approach
Input: A set of DNA sequences, the reference sequence *S* and the complementary rule.
Output: Secret message *M*.
Step 1. Initialize *i* and *j* equal to 1.
Step 2. For the next sequence *S*' of the sequence set, do the following operations:
For each S_i:
if there exists a *j* such that such that $S'_j \neq S_j$, $S'_j \neq C(S'_j)$ and $S'_j \neq C(C(S_j)$, ignore *S'*; otherwise,
if S'_i is the same with S_i, set m_j=0 and increment *j*,
else if S'_i is the same with $C(S_i)$, set m_j=1 and increment *j*.
Step 3. Concatenate all m_j's to be *M* and return *M*.

For an intruder to find out the secret message, he must be equipped as follows. (1) He must know precisely the reference DNA sequence *S*. Since there are roughly 55 millions DNA sequences available publicly, it is extremely hard to guess one. (2) He has to know the random number generator and the two seeds used. (3) He has to know the binary coding scheme.

IV. Conclusion

DNA Sequencing for a Electronic Medical Record System has been introduced to access the patient's medical record securely and instantly. This medical record system provides accurate medication at anywhere in case of emergency. The method of DNA Sequencing for encryption and decryption has more than 55 millions of publicly available sequences and is impossible for the attackers to obtain the records of the patients. The proposed scheme of DNA Sequencing is more reliable, efficient and secured. The improved concept proposed in this system is providing high security over the network.

References

[1] Tsung-Chih Hsiao, Tzer-Long Chen, Chih-Sheng Chen, Fu-Sheng Xu, Starlition Tsui, Yu-Fang Chung and Tzer-Shyong Chen, "Secure Data Transmission for Controlling Access via Key Management Scheme", IEEE 2013.
[2] Jin-Shiuh Taur, Heng-Yi Lin, Hsin-Lun Lee and Chin-Wang Tao, "Data Hiding In Dna Sequences Based On Table LookUp Substitution", International Journal of InnovativeComputing, Information and Control, Volume 8, Number 10(A), October 2012
[3] M. H. Kao, "The Study of Agent-based Secure Schemes on ElectronicMedical Records System", Master Thesis, Tunghai University,Taichung, 2010.
[4] Monica BORDA, Olga TORNEA, "DNA Secret Writing Techniques", Technical University of Cluj-Napoca,2010.
[5] Ban Ahmed Mitras , Adeeba Kh. Aboo, "Proposed Stenography Approach using DNA Properties", College of Computer & Mathematic Science-Mosul Univ./ IRAQ, 2013.
[6] Kritika Gupta, Shailendra Singh, "DNA Based Cryptographic Techniques:A Review" , PEC University of Technology, 2013.
[7] M. I. Youssef, A. Emam and M. Abd ELghany, "Multi-Layer Data Encryption Using Residue Number System in DNA Sequence", International Journal of Security and Its Applications, Volume 6, No. 4, October, 2012.
[8] European Bioinformatics Institute, URL: http://www.ebi.ac.uk/.
[9] Grasha Jacob, A. Murugan,"DNA based Cryptography: An Overview and Analysis", International Journal of Emerging Sciences., 3(1), 36-27, March 2013.