A Novel Approach to an Algorithm for Voice Encryption using DNA based Cryptography

S.Vajjiravelu [1]
A P & Head / MCA
Saraswathi Velu College of Engg
Sholinghur, Vellore District

A.Punitha[2], Ph. D
A P & Head / MCA
Queen Mary's College (Autonomous),
Mylapore, Chennai – 600 004.

ABSTRACT
Global System for Mobile Communications (GSM) is the most widely used cellular technology in the world. Securing voice communication by providing end-to-end security becomes necessary, for this various cryptographic algorithms are evolved significantly with flaws. We identified that the concept of using DNA computing in the field of cryptography as a possible technology that brings forward a new hope for unbreakable algorithms. The voice will first get converted into digital format and then it will be encrypted using DNA Strands. We secretly select a reference DNA sequence **S** and incorporate the secret message **M** into it such that we obtain **S'**. We send this **S'**, together with many other DNA, or DNA-like sequences to the receiver. The receiver is able to identify the particular sequence with **M** hidden in it and ignore all of the other sequences.

General Terms
Securing, Encryption, Algorithm, DNA.

Keywords
GSM, DNA, Strands, Cryptography, Voice Encryption.

1. INTRODUCTION
As speech communications become more and more widely used and even more vulnerable, the importance of providing a high level of security is dramatically increasing. Today, competitors, hackers, or governmental institutions can intercept any GSM cell call with relatively little effort. GSM networks use different security algorithms called A3, A5 and A8. An A3/A8 algorithm is implemented in the SIM cards and in the GSM network authentication centers.[6] The A5 encryption algorithm scrambles the user's voice and data traffic between the handset and the base station to provide privacy; the algorithm is implemented in both the handset and the base station subsystem. As such, a variety of speech encryption techniques have been introduced.

When we talk about voice transmission, voice communication has threat of eavesdropping so in this paper we are proposing a secure way to convert detected voice into secure form, thus protecting it from eavesdropping. The speech detected will first gets converted into digital format and then it will be encrypted using DNA strands.

2. CRYPTOGRAPHY
The fundamental tool for cryptography is the one way function. A function is one-way if it is easy to compute but hard to invert.[5] A one-way function is a function f, such that for each x in the domain of f, it is easy to compute f(x); for essentially all y in the range of f, it is computationally infeasible to find any x, such that y = f(x). A trapdoor one-way function is a one-way function, f with the additional property that given some extra information (the trapdoor information), it becomes computationally feasible to compute for any y in the range of f an x, such that y = f(x).

2.1 Overview of DNA Cryptography
Man has always needed some form of cryptography in order to conceal and protect information. This paper has main aim to facilitate the understanding of principles and some techniques of the new born field of DNA cryptography. There has recently arisen a new area of research known as DNA computing. DNA algorithm is still in its development stage but offers so many possibilities. The history of DNA computing is short, but full of amazing technological achievements. DNA computing and cryptography were introduced in 1990s.[1]

Adleman's pioneering work gave an idea of solving the directed Hamiltonian Path Problem (Travelling Salesman Problem) of size n in O(n) using DNA molecules [1]. The principle used by Adleman lies in the coding of information (nodes, edges) in DNA clusters and in the use of enzymes for the simulation of simple calculations.[2]

It seems that DNA computing is destined to be remembered as a novel idea that was too difficult to implement practically. DNA cryptography is the future of the information security. Its complexity and randomness provides a great uncertainty which makes encoding of data in DNA format better than other mechanism of cryptography. The field of DNA computing is still in its infancy and the applications for this technology are still not fully understood. The research of DNA cryptography is still at the beginning, and there are many problems to be solved. But the vast parallelism, exceptional energy efficiency and extraordinary information density inherent in DNA molecules endow DNA cryptography special advantages over other kinds of cryptography. In this paper, proposed DNA algorithm is encrypting or hiding a data in terms of DNA sequences.

2.2. DNA Computing
DNA stands for Deoxyribo Nucleic Acid. DNA represents the genetic blueprint of living creatures. DNA contains "instructions" for assembling cells. Every cell in human body has a complete set of DNA. DNA is unique for each individual. DNA is a polymer made of monomers called deoxyribo nucleotides. Each nucleotide consists of three basic items: deoxyribose sugar, phosphate group and a nitrogenous base. The nitrogenous bases are of two types: purins (Adenine and Guanine) and pyrimidins (Cytosine and Thymine). They are represented as A, G, C and T. A binds with T and G binds to C. [2] The various operations that can be performed on DNA are ligation, polymerase chain reaction (PCR), gel electrophoresis and affinity purification

We can now easily see one special property of DNA sequences: There is almost no difference between a real DNA sequence and a faked one. This is a property which we shall exploit in our research. There is another fact which is quite useful to us: There are a large number of DNA sequences publicly available in various web-sites, such as . A rough estimation would put the number of DNA sequences publicly available to be around 55 million .

By using the above facts, we designed a DNA based encryption method. This method would 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*.

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)). 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 X 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)). We also assume that the secret message *M* is a binary sequence.

3. PROCEDURES

3.1. Steps for Encryption

To simplify the discussion, we start with the most basic version and give a simple example. The more complicated version of our method will be presented after this basic one is given. Suppose the secret message *M* is 01001100.

Let *S* be ACGGTTCCAATGC.

Our coding steps are as follows:

Step 1. We first code *S* into a binary sequence by using the binary coding scheme. Thus the sequence *S* will now become 0001101011110101000011100 1.

Step 2. Divide *S* into segments whereby each segment contains *k* bits. Suppose *k* is 3. Then we have the following segments: 000, 110, 101, 111, 010, 100, 001, 110, 01.

Step 3. Insert bits from *M*, once at a time, into the beginning of segments of *S*. The result is as follows: 0000, 1110, 0101, 0111, 1010, 1100, 0001, 0110, 01. We should ignore those segments without any secret message inserted. Thus, we will have the following segments: 0000, 1110, 0101, 0111, 1010, 1100, 0001, 0110. Concatenating the above segments, we have the following binary sequence: 000011100101011110101100000 10110.

Step 4. We use the binary code scheme to produce the following faked DNA sequence: *S*'=AATGCCCTGGTAACCG.
As the reader can see, this sequence is quite different from *S*.

Step 5. We send the above sequence *S*' to the receiver, amid many other irrelevant sequences.

The above process is the encryption process. It is easy to see that the decryption process is just to reverse the encryption process. For every received sequence *T*, the receiver extracts a sequence out of it. If the extracted sequence is not a prefix of the reference sequence *S*, ignore *T*. If it is, the receiver knows that he has also successfully extracted the secret message *M* as a by-product.

The above is the basic version of our approach. In a more complicated version, we divide *S* into segments by using a random number generator. That is, *k* is not fixed any more. Instead, it is determined by a random number generator which is known only to the sender and the receiver. Suppose the *k*'s are 6, 3, 2, 4. Then *S* is divided into segments with lengths 6, 3, 2 and 4 respectively. Note that there is also a secret message *M*. We may therefore also use the same random number generator to divide *M* into segments. The parameter used will be denoted as *r*.

3.2. Algorithm -1 Encryption Algorithm

Input: A reference DNA sequence *S*, a secret binary message *M* and a binary coding scheme to code A, C, G and T into binary digits.

Output: An encrypted DNA sequence *S*'.

Step 1. Code *S* into a binary sequence S_1 by using the binary coding scheme.

Step 2. Generate *k*'s by using a random number generator to divide S_1 into segments and generate *r*'s to divide the secret message *M* into segments. Each k_i and r_i is larger than 1 or equal to 1. Denote S_1 by and M by $m_1, m_2, \ldots m_p$

Step 3. insert each m_i of *M* before S_i of S_1 to produce a new binary sequence. Delete $S_{p+1}, S_{P+2} \ldots S_n$. Denote the resulting binary sequence by S_2.

Step 4. Transform sequence S_2 back to a faked DNA sequence S_3 by using the same binary coding scheme used in **Step 1**.

Step 5. Return S_3.

S_3 is sent to the receiver together with many other DNA sequences. The receiver uses the following algorithm to decrypt.

3.3. Algorithm -2 Decryption Algorithm

Input: A set of DNA sequences, one of which has the secret message *M* hidden in it by using Algorithm-1 , a reference DNA sequence *S* and a binary coding scheme.

Output: The secret message *M*.

Step 1. Generate numbers *k*'s and *r*'s denoted as $k_1,k_2 \ldots k_n$ and $r_1,r_2 \ldots r_p$ by using the same random number generator with the same seed of the encoding scheme.

Step 2. For a DNA sequence *S*' of the set, code *S*' into a binary sequence by using the binary coding used by the sender and use $r_{1+}k_{1,}\ r_{2+}k_{2,\ldots}$ to divide the binary sequence into binary segments.

Step 3. For each segment of the first *p* segments of *S*', extract the first r_i bits, called m_i.

Step 4. For each segment of the first *p* segments of *S*', extract the last k_i bits, called s_i.

Step 5. Concatenate all m_i's to be *M* and all s_j's, to be S_1.

Step 6. Transform S_1 to be a DNA sequence by using the same rule. If S_1 is not a prefix of *S*, go to **Step 2**.

Step 7. 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.

4. CONCLUSION

The DNA Algorithm requires a longer time for encryption and decryption, comparatively to the other ciphers. Similarly the algorithm which we are using also have much greater encryption and decryption time than other classical ciphers, but it provides better security than others. The advantage is that DNA has a huge storing capacity, but on the other hand practically using the implementations requires a lot of time. The field of DNA computing is still in its infancy and the applications for this technology have not yet been fully understood. DNA computing is viable and DNA authentication methods have shown great promise in the marketplace of today and it is hoped that its applications will continue to expand. DNA Cipher is the beneficial supplement to the existing mathematical cipher. Sequence can be selected from any web-site associated with DNA sequences. Since there are many web-sites and roughly 55 million publicly available DNA sequences, it is virtually impossible to guess this sequence.

5. REFERENCES

[1] Akanksha Agrawal, Akansha Bhopale, Jaya Sharma, Meer Shizan Ali, Divya Gautam, "Implementation of DNA algorithm for secure voice communication", International Journal of Scientific & Engineering Research Volume 3, Issue 6, June-2012, ISSN 2229-5518.

[2] Grasha Jacob, A. Murugan,"DNA based Cryptography: An Overview and Analysis", International Journal of Emerging Sciences., 3(1), 36-27, March 2013.

[3] H. Z. Hsu and R. C. T. Lee, "DNA Based Encryption Methods", The 23rd workshop on Combinatorial mathematics and Computation Theory, 2012.

[4] Ashish Gehani, Thomas H. LaBean and John H. Reif "DNA-based Cryptography", 5th Annual DIMACS Meeting on DNA Based Computers (DNA 5), MIT, Cambridge, MA, June 1999.

[5] Parwinder pal singh Bhupinder singh Satinder pal Ahuja, " Need of Secure Voice Encryption and its Methods", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 1, January 2012.

[6] D.Ambika, V.Radha, " Secure Speech Communication – A Review", International Journal of Engineering Research and Applications, Vol. 2, Issue 5, September-October 2012, pp.1044-1049.

[7] European Bioinformatics Institute, URL: http://www.ebi.ac.uk/.