Integrating DNA Computing in Chaos-Based Optimization Algorithm Approach for Cryptography

Rashi Vohra*
Brajesh Patel**

Abstract

Abstract: DNA cryptography is a field which is being in tremendous demand for providing effective electronic security since 1990s. It exploits the extreme randomness and complexity properties of DNA structure for coding and decoding any information. As a new step to enhance the security, this paper presents an alternative security method of chaotic encryption and decryption by embedding DNA security algorithm in it. The presented method of cryptography can transmit highly confidential data efficiently and securely, and complements standard, algorithmic procedures, providing security solutions with novel features.

Keywords: *Cryptography, genetic algorithm, chaotic map, PRNG, DNA.*

Introduction

Over the last few decades, Information security is being to be a most fascinating and interesting process to provide a mean to safeguard the information against any intentional and malicious attacks by any node in path to ensure its CIA triad [1]. The CIA triad stands for three major tenets to information security: confidentiality, integrity and availability. Confidentiality prevents un-authorized disclosure of sensitive information. Integrity prevents unauthorized modification of information thereby assuring the accuracy of information. Availability ensures that the information is available for use when it is needed by preventing loss of access to information.

The outline of the paper is as follows: Section 2; provide a gentle and interesting introduction to cryptography and highlights important concepts of cryptography. The most important development in recent year in cryptography is the adoption of genetic algorithm, chaos theory and DNA computation, Section 3; provide an overview of genetic algorithm, operators in genetic programming., section 4 provides introduction of chaos theory, concepts of chaos theory, discussion on the relation between the chaos and cryptography, and presents some general aspects of DNA computing in cryptography. In section 5, a new approach of DNA based chaotic cryptography is proposed. In Section 6; the conclusions and possible continuations of our work are presented. Finally, in section 7 we point out the references.

Cryptography

In this information age, Cryptography plays a central role in providing information security and is becoming increasingly important as a building block for information security [1]. It has long been used by militaries and governments to facilitate secret communication. Cryptography is a study of design of technique to provide secret communication as it protects the information transmission from the influence of adversaries who may present a threat to information CIA triad. All those who involve in such an art are called as cryptographers. Cryptography is composed of two process encryption and decryption, performed by using a set of codes, termed as cipher.

Cryptography synonymous with encryption, deals with transformation of user information (plaintext) into an unintelligible gibberish (cipher text) that make it unusable by anyone other than an authorized entity and protect it from unauthorized or accidental disclosure while information is in transit and in storage. On the contrary, the plaintext can be restored from the cipher text by the process decryption, opposite to that of encryption. The security offered by cryptographic-based systems depends on both the strength of the cryptographic algorithms chosen for encryption /decryption and the strength of the keys used with those algorithms. Cryptosystems individualize on the basis of: Type of operations used - Substitution/Transposition, Way in which plaintext is processed - Block/Stream, Number of keys used - Symmetric / Asymmetric.

*Student, Shri Ram Institute of Technology, Jabalpur
**HOD, Shri Ram Institute of Technology, Jabalpur

Objective of Cryptography

1. Confidentiality: It is a service which is used to protect identifiable information from forced disclosure to avoid a malicious use of them.

2. Data integrity: Integrity means no data modification, providing an assurance that information can only be accessed or modified by those authorized to do so.

3. Authentication: Authentication gives the ability to know the identity of a user, without saying anything about the access rights of the individual.

4. Non-repudiation: As per non-repudiation neither sender nor the recipient can deny later from sending or receiving the message respectively. It can be viewed as an extension to the identification and authentication service.

The Principal categories of cryptographic algorithms are: private-key cryptography, public-key cryptography and Cryptographic hash functions. Private-key cryptography is sometimes referred to as secret key cryptography or symmetric cryptography because a single key is shared between sender and receiver (key distribution) for enciphering and deciphering by keeping the key secret. The security of the algorithm depends upon how well the key is protected and on the number of bits of the key. For the secret key cryptosystem, with the plaintext X and encryption key K as input in encryption algorithm E. The system can be described as: Y=EK(X) where Y is the cipher-text. The notation for deciphering will be X=DK(Y) where D is the decryption algorithm and Y, K are the input to D.

Some popular encryption algorithms developed using this symmetric cryptography includes DES, 3DES, AES, and RC4. The DES was published by the NIST and is based on Feistel-network version of Lucifer. It takes plaintext (64-bit) and key (56-bit) as input. The plaintext is first passed through initial permutation, followed by 16 round of function (composed of both permutation and substitution function), then swapping is performed between the two half of the output so far generated, the pre-output is then passed through a permutation (inverse of initial permutation function), finally producing ciphertext (64-bit). The decryption algorithm proceeded as encryption algorithm, except with the application of the sub-keys is reversed. With the key length of 56 bits, DES is vulnerable to brute force attack.

To overcome the vulnerability to brute force attack, 3DES is issued by NIST in 1999. The principal drawback of 3DES is the algorithm is relatively sluggish in software, slower (having 3 times rounds as des), use of 64-bit block size. To replace 3DES so as to support block length of 128 bits and key length of 128,192, and 256 bits, AES is published in November 2001 by NIST. In AES, 4 stages are used in each round, one of permutation (shiftrows) and three for substitution (substitute bytes, mix columns, add round key).

The essence of public key cryptography was introduced by Diffie and Hellman in 1976, which eliminate the use of key distribution process. It is more efficient then secret key encryption for concealing the sensitive information. It is also referred as asymmetric cryptography as it uses two different key for encryption and decryption process. The sender locks the data i.e. Encrypt the data by using the public key (known to everyone), whereas the receiver unlock the data i.e. decrypt it to plaintext form by using another key termed as private key (known only to the owner of the key).

The most popular public key cryptosystem is RSA, first published in 1978 and is based on the concept of IFD, which is finding the prime factor of very large integer. It is the most predominant algorithm used today for public-key cryptography. Diffie and Hellman issued by NIST uses the concept of DLP to provide authentication mechanism. Elliptic Curve Cryptography was proposed by cryptographers Victor Miller (IBM) and Neal Koblitz (University of Washington) in 1985, is based on the difficulty of solving the Elliptic Curve Discrete Logarithm Problem (ECDLP). Given two points, P and Q, on an elliptic curve, find the integer n, if it exists, such that P= nQ. It combines the concept of number theory and algebraic geometry

Symmetric encryption algorithms is normally used to encrypt private data for its extremely fast, and low complexity, whereas asymmetric encryption imposes a high computational burden, and tends to be much slower and often used for digital signature and key distribution.

Cryptographic hash functions also known as message digests, because it condenses a message into an irreversible fixed-length hash value. It is a form of cryptographic security but differ from encryption

process. Encryption process completes in two stages: first is to encrypt the data and second is to decrypt the data on the other hand, hash function is used to generate a short fixed length random string from the original message, often used for checking the integrity of the message. Hash function is commonly used to encrypt password in many operating system. It is also termed as a one way function because the original text cannot be recovered from the hash value. Hashing algorithm is generally used for error checking without the use of any secret key.

Genetic Algorithm

In 1960s, I. Rechenberg introduced the concept of evolutionary computation. The genetic algorithm belongs to the family of evolutionary computing as a useful tool for search optimization problems, along with genetic programming, evolution strategies, and evolutionary programming. The genetic algorithm is an adaptive heuristic search algorithm derived from the concept of natural selection and natural genetics. The word “genetics” is derived from the Greek word “genesis” meaning “to grow”. It is the branch of science that differentiates between heredity and variations and account for the resemblances and differences during the process of evolution.

GA handles a population of individuals where each individual represents a possible solution and is represented by a chromosome. The chromosomes can be encoded using bits, numbers, trees, lists, or any other objects, depending upon the type of problem to be solved. Each chromosome is associated with fitness value which corresponds to an evaluation of how good the individual is.

The GA loops over an iteration process containing the stages of selection, reproduction, evaluation and replacement. The algorithm is stopped when the population converges toward the optimal solution. The iteration stops under the various conditions such as maximum generations has evolved, specified time has elapsed, when there is no change in fitness value for a specified number of generations, due to stall generations and stall time limits. The termination stage finally brings the process to a halt[1].

The basic operators in genetic algorithm for performing it operations are: encoding, selection, and recombination and mutation operator. Encoding is process of representing chromosomes. It may be binary, octal, hexadecimal, permutation, value or tree representation. Selection is a process of choosing two individuals from the population to create an offspring for the next generation on the basis of their fitness value.

Higher the fitness value, higher will be the chance of better chromosomes selection. The population fitness over successive generation get improved by the GA selection pressure, as higher the selection pressure, greater will be chance of getting better individual to be favored for reproduction. Recombination also termed as crossover, is a process of producing offspring from the previously selected individuals. It makes clone of good strings but does not create the new ones.

Various crossover techniques are: single-point crossover, two-point crossover, multi-point crossover, uniform crossover, three-parent crossover etc. After crossover, generated offspring's are subjected to mutation, which prevents the algorithm from being trapped in local minimum and maintain diversity in the population.

The advantages of GA are: parallelism, liability, wider solution space, complexity in the fitness landscape, and easy discovery of global optimum. It encounters some limitation too, they are: the problem to identifying the fitness function, computational time, definition of representation of the problem and occurrence of premature convergence.

Chaos Theory and DNA Computing

Chaos theory is a branch in mathematics applications in various disciplines such as physics, engineering, economics, biology and philosophy. In common usage, “chaos" means “A condition or place of great disorder or confusion”. Chaos theory studies the behavior of systems that follow deterministic laws but appear random and unpredictable or we can say a dynamical system that has a sensitive dependence on its initial conditions; small changes in those conditions can lead to quite different outcomes [2][3]. This dependency of a dynamical system on its initial condition is popularly referred to as the butterfly effect. One example of chaotic behavior is the flow of air in conditions of turbulence.

"Chaos is a name for any order that produces confusion in our minds" [George Santayana Dominations and Powers]

A dynamical system must satisfy following chaotic properties, to be referred as chaotic:

1. it must be sensitive to initial conditions;
2. it must be topologically mixing; and its periodic orbits must be dense

In mathematics, a chaotic map is a map that exhibits some sort of chaotic behavior. Chaotic Maps may be classified by a discrete-time or a continuous-time parameter. Discrete maps usually take the form of iterated functions. The Hénon map is a discrete-time dynamical system, takes a point (x_n, y_n) in the plane and maps it to a new point [7]. The two-dimensional dissipative quadratic map (Hénon 1976) is given by the coupled equations as:

$$x_{n+1} = 1 - a x_n + y_n$$

$$y_{n+1} = \beta x_n$$

It depends on two parameters, a and b, which for values of a = 1.4 and b = 0.3 act as canonical Hénon map and is chaotic in nature. For other values it may be chaotic, intermittent, or converge to a periodic orbit. Unlike the logistic map, the canonical Hénon map is interesting because, its orbits defy a simple description [4].They are potential candidate for making a pseudorandom number generator because of their random like, unpredictable dynamics, inherent determinism and simplicity of realization property [4].

DNA: (or deoxyribonucleic acid) is the hereditary material in almost all organisms. DNA is located in the cell nucleus (called as nuclear DNA) which makes up the body, but a small amount can also be found in the mitochondria (called as mitochondrial DNA or mtDNA). DNA is comprised of two long strands of nitrogenous bases arranged in a helix sort of structure [8]. Thus, the design instructions in DNA are spelled out as particular sequences of these bases, known as *genes.*

Each nucleotide in DNA contains one nitrogenous base: *adenine* (*A*), *thymine* (*T*), *cytosine* (*C*), or *guanine* (*G*) and has a sugar component joined to a phosphate group at one point on the sugar, and a nitrogen containing base attached to another point.

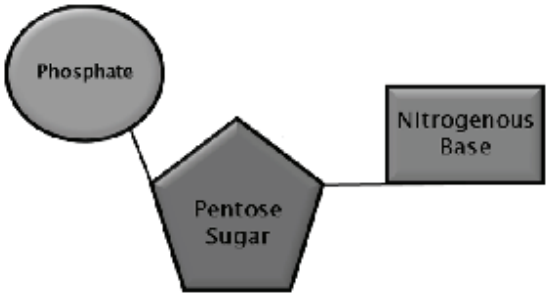

Figure 2. Nucleotide structure

The chains in DNA have the phosphate of one nucleotide linked to the sugar of the next nucleotide to form a strand of alternating sugars and phosphates with dangling nitrogenous bases.

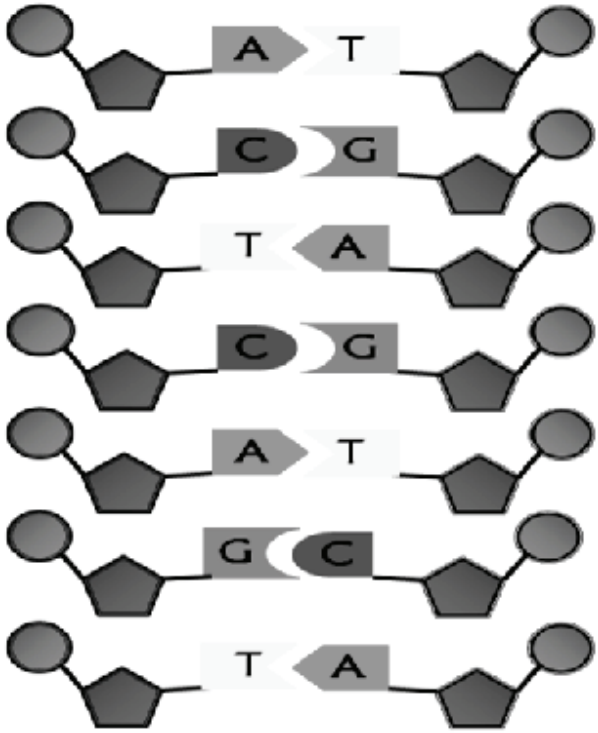

Figure 3. Combination of nucleotide bases in strands

DNA contains two such chains, twisted around each other to form a double-stranded helix with the bases on the inside. Every *A* on one chain forms weak bonds with a *T* on the other strand, and every *C* on a strand bonds weakly to a *G* on the opposite chain. The two strands, held together weakly by the pairing of *A* with *T*, and *G* with *C*, are thus complementary, and the sequence in one can be deduced from the other's sequence.

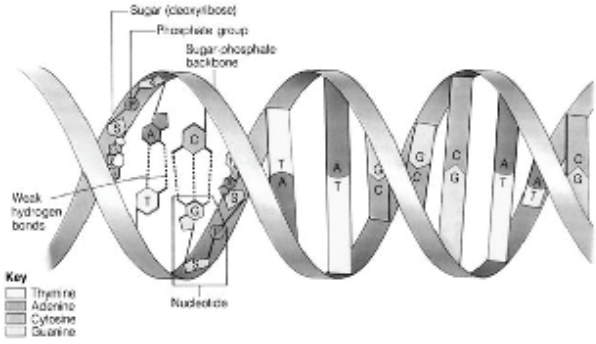

Figure 4. Basic DNA structure

The complementarily of the two DNA strands allows their information to be transmitted as new DNA to new cells during development and growth. Each old strand is used as a template in synthesizing a new complementary one. These complimentary strands have codons as fundamental building blocks. Codons are basically triplets of nucleotide bases. We can use these codons for encoding and decoding of the data [9].

The Proposed Method

Encryption Algorithm: The encrypting process emulates the working of the crossover operator using pseudorandom sequence. The steps for the data encryption as follows:

1. Generate the pseudorandom binary sequence using chaos.
2. Convert the binary pseudorandom sequence into octal pseudorandom sequences ranging from 0-7as O_n.
3. Read the file to be encrypted, and convert each letter entered to its corresponding codon using lookup table-1.
4. Map base nucleotide to a numerical value using another look up table-2.
5. Take two consecutive bytes of the data stream as A1 and A2.
6. Perform the crossover on two consecutive bytes of the data stream as B1 and B2 using Oi.
7. Encrypt data as C1 and C2,where

   X1 = Oi ⊻ (Oi << 4)
   X2 = Oi+1 ⊻ (Oi+1 << 4)
   C1 = B1 ⊻ X1
   C2 = B2 ⊻ X2
8. Save the resulted C1 and C2 in ciphered file.
9. Repeat the step 5 to 8 till end of the file to be encrypted.
10. Convert the resulted ciphered text into DNA sequence using lookup table-2 and save it into another file.
11. The resulted file is the encrypted file.

Decryption Algorithm

The decryption process will work just opposite of encryption.

Table-1: Alphabet to codon mapping table.

a-cga	l-tgc	w-ccg	3-gac
b-cca	m-tcc	x-cta	4-gag
c-gtt	n-tct	y-aaa	5-aga
d-ttg	o-gga	z-ctt	6-tta
e-ggc	p-gtg	_-ata	7-aca
f-ggt	q-aac	,-tcg	8-agg
g-ttt	r-tca	.-gat	9-gcg
h-cgc	s-acg	:-gct	Space-ccc
i-atg	t-ttc	0-act	
j-agt	u-ctg	1-acc	
k-aag	v-cct	2-tag	

Table-2: Mapping nucleotide base to numerical value

a	0
c	1
g	2
t	3

Conclusion

This paper introduces an improved encryption scheme based on genetic algorithm, DNA computing and chaotic map. Here, the key is generated randomly using chaotic map, thus the use of genetic algorithm along with the randomness property of chaos theory resulted in highly reliable and safe approach from the dangerous clutches of message hackers for data encryption [8] [12]. The advantage of this encryption scheme is reduced computationally complexity and retrievable nature. The additional conversions from the string to array of bytes and back make this cipher to require more time for encryption and decryption.

Following are the open subject for future work:

1 Adding some extra features to enhance its performance.

2 Hardware realization for this concept is concerned that can be further used for highly secure data transmission application.

3 Comparative study of other chaotic map performance using the same proposed process.

4 More detail study of security analysis of the proposed scheme.

References

1. Kumar, M. K. Ghose, “Overview of Information Security Using Genetic Algorithm and Chaos”,

Information Security Journal: A Global Perspective, 18:306–315, 2009.

2. V. Patidar, K. K. Sud, N. K. Pareek, “A Pseudo Random Bit Generator Based on Chaotic Logistic Map and its Statistical Testing”, Informatica 33, pp. 441–452, 2009.
3. J.M. Amigo, L. Kocarev, J. Szczepanski,” Theory and practice of chaotic cryptography”, Physics Letters A 366, pp. 211–216, 2007.
4. S. MADHEKAR, “Cryptographic Pseudo-Random Sequences from the Chaotic Henon Map”, Sadhan ¯ a¯ Volume 34, Part 5, pp. 689–701, October 2009.
5. R. A. Joshi, S. S. Joshi, G. P. Bhole, “Improved Image Encryption Algorithm using Chaotic Map”, International Journal of Computer Applications (0975 – 8887), Volume 32– No.9, October 2011.
6. L. Shujuna, M. Xuanqinb, C. Yuanlongc, “Pseudo-Random Bit Generator Based on Couple Chaotic Systems and its Applications in Stream-Cipher Cryptography”, Progress in Cryptology - INDOCRYPT 2001, LNCS, Volume 2247, pp. 316-329, Springer-Verlag, Berlin, 2001.
7. F. Zheng, X. Tian, J. Song, X. Li, “ Pseudo-Random Sequence Generator Based on the Generalized Henon Map”, The Journal of China Universities of Posts and Telecommunications, Volume 15, Issue 3, Pages 64–68, September 2008.
8. Gehani, T. H. LaBean, J. H. Reif, "DNA-based Cryptography", 5th Annual DIMACS Meeting on DNA Based Computers (DNA 5), MIT, Cambridge, MA, June 1999.
9. R. Terec, M. F. Vaida, L. Alboaie, L. Chiorean, "DNA Security using Symmetric and Asymmetric Cryptography", International Journal on New Computer Architectures and Their Applications (IJNCAA) 1(1): 34-51, 2011 (ISSN 2220-9085).