iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4809

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Securing Data through DNA Computing: A Comprehensive Study

Pragya Agarwa[1], Swati Deshmukh[2]
[1, 2]*Assistant Professor, Computer Engg Department, AISSMS COE, Pune,* MH, India

Abstract: ***DNA computing is a new computational paradigm by harnessing the potential massive parallelism, high density information of bio-molecules and low power consumption, which brings potential challenges and opportunities to traditional cryptography. In this paper, on the basis of reviewing the principle of DNA computing and the development situation of DNA computing briefly, I analyze some schemes and presents alternative security methods based on DNA.***

Keywords: ***DNA Cryptography, DNA Computing, Encryption, Decryption, DNA Stegnography, DNA Certification.***

I. INTRODUCTION

In today's world data is of extreme importance. Everyone wants to protect the data and keep it secured. To secure the data we need to encrypt it before transmitting it to receiver.

But all the encryption algorithms existing now a days are broken, so the world of information security looks in new directions to protect the data. DNA Cryptography which is a new branch of cryptography, uses DNA Computing technology and has been identified as a new hope for designing unbreakable cryptographic algorithms.

DNA works as a basic computational tool in this field and has massive storage capacity. One gram of DNA can store 10^8 tera bytes.

In this study, plain text messages and images are transformed using DNA as a information carrier. Researchers have designed many cryptographic algorithms using DNA cryptography, which can break even the modern algorithms like DES, RSA etc.

In this paper I've summarised the techniques which are currently being used for encryption or decryption using DNA cryptography. Some other applications of DNA Computing like stegnography and certification are also discussed in this paper. I've also analysed these algorithms and techniques and have given an insight to the benefits which can be achieved with the help of DNA Cryptography.

II. DNA & DNA COMPUTING

A. DNA

DNA stands for Deoxyribonucleic Acid. It represents the genetic blueprint of living creatures. It contains "instructions" for assembling cells. Every cell in human body has a complete set of DNA which is unique for each individual. It is a double-stranded polynucleotide.

The two strands of a DNA molecule are anti parallel where each strand runs in an opposite direction. Two strands are held together by weak hydrogen bonds between the complementary base pairs Adenine-Thymine and Guanine-Cytosine. Instructions are coded in a sequence of the DNA bases.

A segment of DNA is exposed, transcribed and translated to carry out instructions.

Complementary strands (for example, ATCGAACT complements TAGCTTGA), anneal and twist around to form a double-helix.

The data density of DNA is impressive. Just like a string of binary data is encoded with ones and zeros, a strand of DNA is encoded with four bases, represented by the letters A, T, C, and G.

The bases (also known as nucleotides) are spaced every 0.35 nanometers along the DNA molecule, giving DNA a remarkable data density of nearly 18 Mbits per inch. In two dimensions, if you assume one base per square nanometer, the data density is over one million Gbits per square inch. Compare this to the data density of a typical high performance hard drive, which is about 7 Gbits per square inch –a factor of over 100,000 smaller.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

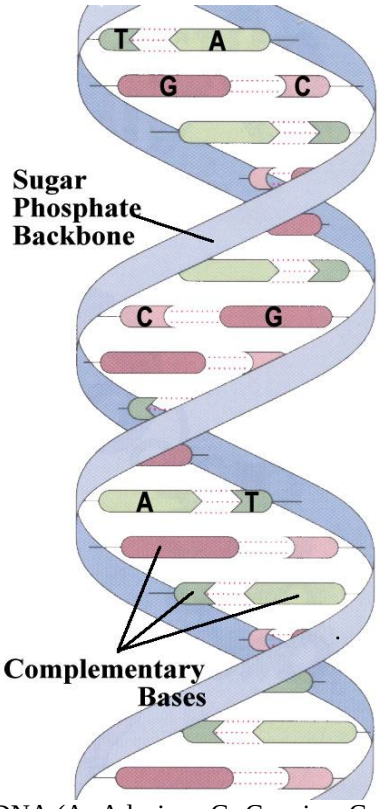


Fig. 1. Structure of DNA (A- Adenine, G- Guanine, C- Cytosine and T- Thymine)

B. DNA Computing-

DNA computing is a form of computing which uses DNA and molecular biology, instead of the traditional silicon-based computer technologies. The goal of the DNA computing field is to create a device that can work independent of human involvement. DNA computing is fundamentally similar to parallel computing in that it takes advantage of the many different molecules of DNA to try many different possibilities at once. DNA computing is multidisciplinary in nature with contribution from Biochemistry, Molecular Biology, Genetics, Chemistry, Mathematics, Computer Science, Biology and Physics. The idea behind blending biology with technology is due to the limitations faced by the semiconductor designers in decreasing the size of the silicon chips, which directly affects the processor speed. Biocomputer consists of biochips unlike the normal computers, which are silicon-based computers. This biochip consists of biomaterial such as nucleic acid, enzymes, etc. The power of a biocomputer is that it acts as a massively parallel computer and has immense data storage capability. Thus, it can be used to solve NP-complete problems with higher efficiency.

DNA computers could help researchers to answer complicated mathematics problems that other types of computers have thus far been unable to solve. It is hoped that DNA computers will be able to mimic the way that our current electronic computers think and perform. In fact, one experiment found that a DNA computer works more effectively in comparison to a digital computer for the application of data analysis provided by logic gates. A lot of researchers are focusing on solving many different problems by using DNA Computing. The categorization of different areas in which DNA computing can be used is as follows:

Scheduling
Clustering
Forecasting
Signal and Image Processing
Security field-

1) Encryption/Decryption (cryptography),
2) Steganography,
3) Certification

III. DNA CRYPTOGRAPHY

A. *Advantages of DNA Cryptography-*

The DNA cryptography method may be considered in the conventional cryptographic methods for at least the following advantages that it possesses:

1) The sender needs not to have much key information to encrypt the information. Initially a part of the key is sufficient to encode the information.
2) It requires a little information (only the private keys or part of the key) to be communicated through the secure channel.
3) The information that are to be transferred through public channel, small in size than that of original information.
4) One-time pad can be possibly used as the key giving enough storage, since almost one key for one piece of information. It provides lots of strength in encoding techniques

Table:1
Comparision Between Traditional and DNA CRYP To Graphic Methods

	Traditional Cryptography	DNA Cryptography
Security	One Fold	Two Fold
Time Complexity	>= Few Seconds	>= Few hours
Storage Medium	Silicon Chips	DNA strands
Storage Capacity	1 Gram of silicon chip carries 16 MB	1 Gram of DNA carries 10^8 TB
Stability	Dependent on implementation environments	Dependent on Environmental conditions

This paper investigates a variety of bimolecular methods for encrypting and decrypting data. We assume the plaintext message data is encoded in DNA strands. For example, the DNA strands in solution in a test tube may form a "wet" data base of biological data (e.g., the DNA of personnel of an organization) which might be needed for security.

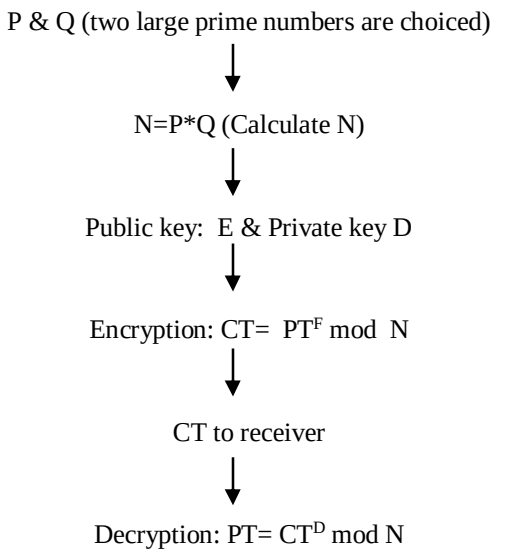
B. *DNA Cryptography and its various techniques-*

DNA cryptography is the new field of interest in the common scenario, where it is possible to exploit the inherent massively-parallel computing properties of DNA bonding to perform the encryption and decryption of the public and private keys. The resulting encryption algorithm used in the transaction is much more complex than the one used by conventional encryption methods

1) *Implementation of RSA Algo :* The RSA algorithm is the most important and proven asymmetric key cryptographic algorithm. The RSA algorithm is based on the mathematical fact which is easy to get and multiply large prime numbers together. The private and public keys in RSA are based on extremely large prime numbers. The algorithm is quite easily. Whereas, the real challenge for RSA is the selection and generation of the public keys and private keys or else the attacker can crack it .

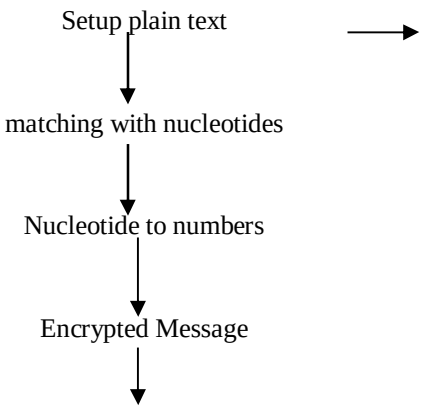
The following description of figure. shows how the RSA algorithm works:

a) P and Q are two large random prime numbers.
b) E is public key, D is private key. When A send a message to B, A encrypts the message using B's public key,then just B can decrypt the cipher text to plaintext, using B's private key.
c) CT is cipher text and PT is plaintext.



RSA Algorithm

When we implement RSA Algorithm with DNA Computing, as explained in [32], the plain text message is encrypted as Cipher. The steps for DNA implementation of this algorithm are as follows:



RSA Algorithm using DNA Computing

It has a requirement to set up a list for corresponding plaintext. It is shown as Table. 2.

TABLE. 2 Encode Plaintext

A-CCA	B-GIT	I-AGT	J-CGA
C-ITG	D-GGT	K-GAA	L-CGT
E-TIT	F-TCG	M-CCT	N-TCT
G-CGC	H-ATG	O-CGG	P-ACA
Q-CAA	R-ACT	S-GCA	T-CIT
U-GTC	V-TCC	W-GCC	X-ATC
Y-AAA	Z-TCA		

TABLE.3 . Mapping Process from base nuc1eotides to numbers

A-0l C-03 G-07 T-20

The experimentation encrypts a message for security consideration, which aims at not being read by others. In this way designing of a new encryption scheme is possible with the combination of RSA algorithm and DNA Computing. The well effectiveness of the method has been verified by simulation.

2) *DNA One Time pad Cryptosystem using Substitution :* The input to a substitution one-time- pad system, is a plaintext binary message of length n, partitioned into plaintext words of fixed length, and a substitution one-time-pad consisting of a table randomly mapping all possible strings of plaintext words into cipher words of fixed length, such that there is a unique reverse mapping. The plaintext is encrypted by substituting each its block of the plaintext with the cipher word given by a table, and is decrypted by reversing these substitutions as in [22]. DNA Implementation, in the case of encryption by substitution, one test tube of short DNA strands (the plaintext messages) is converted into another set of entirely different strands (the encrypted messages) in a random yet reversible way. DNA encoded messages are manipulated such that the plaintext is converted to cipher strands and the plaintext strands are removed. This substitution method requires one-time-pad DNA sequences to accomplish this conversion. The overall scheme involves long DNA pads containing many segments, where each segment contains a cipher word followed by a plaintext word. The cipher word acts as a hybridization site for binding of a primer, which is then specifically appended with a plaintext word to produce word-pairs. The word-pair DNA strands can be used as a type of lookup table in the first step of the conversion of plaintext into cipher text and accomplish the remaining steps to complete encryption and decryption. DNA one-time pads, an ideal library of one-time pads would contain a huge number of pads and each pad would provide a perfectly unique, random mapping of plaintext to cipher word pairs. Construction of the libraries of codebook pads can be approached using segmental assembly or build-up procedures used successfully in previous gene library construction projects. The design of sequence words for the plaintext and cipher lexicons represents the first technical challenge. We would like the lexicons to be distinct, or disjoint in the mathematical sense. We also need essentially complete coverage of the lexicon on each pad, as well as unique word mapping, so a single plaintext word pairs with a single cipher word and vice versa.

3) *DNA PCR Technology:* PCR is an amplification and quantification process of DNA. The purpose of designing PCR is to increase the amount of DNA, as it is very difficult to deal with small amount of DNA strands. The name Polymerase chain reaction comes from the enzyme (biological catalyst) known as polymerase used in the technique and chain represents that this amplification process occurs in many cycles one after another. By performing PCR, short sequences of DNA can be analysed even in samples con- taining only minute quantities of DNA [29]. PCR can select small strands of DNA and amplifies those. In practice, amplification of DNA involves cloning of segments of interest into vectors for expression. PCR is highly efficient so that untold numbers of copies can be made from small selected DNA. Moreover, PCR uses the same molecules that nature uses for copying DNA. To perform PCR, one should know the sequence of DNA to be amplified to design the right primer for it, where primer is a sequence containing few numbers of nucleotides complimentary to the specific sequence of DNA which is to be amplified [6]. In short, we can identify the PCR process into two phases:

- Two ”primers”, short single-stranded DNA sequences to correspond to the beginning and ending of the DNA stretch.
- Polymerase enzyme that moves along the segment of DNA, reading its code and assembling the copy.

The encryption key, in this case, is compound, consisting of both PCR primers pair and public key. Similarly decryption key consists of complementary primer pairs and private key.

Plain Text ⟶ Cipher Text ⟶ Cipher DNA ⟶ Amplified DNA

Figure-2

Encryption starts with the exchange of two primers (forward and reverse) between sender and receiver via a secure channel]. For Encryption, pre-processing can be done, that the whole algorithm like RSA can be applied first.Then cipher text can be converted into DNA sequence by coding scheme. By performing this, entirely different cipher text can be obtained. In literature, cipher DNA refers the term cipher text which is in the form of DNA sequence, and plaintext DNA denotes the plaintext which is in the form of DNA. The prepared cipher DNA is then flanked by the secret primers and mixed with a number of other unknown DNA. Sender sends this DNA mixture to Recevier. For decryption, Bob can retrieve the cipher DNA by performing PCR using its secret primer [51], and reverse the whole process which is done for encryption. Anyone without knowing the two primers cannot retrieve target cipher DNA. Cipher DNA can be retrieved by using DNA decryption key (secret primers)

and is converted into cipher text by using coding scheme. Finally, in the figure denotes that the cipher text is decrypted by RSA private key.

4) *DNA Chip Technology:* DNA chips enable researchers to manipulate the vast amounts of data from genome-sequencing .DNA chip technology is very important for the manipulation of biological data. It is commonly used to find expression of many genes in parallel. These chips like silicon chips can be used to handle and store the data in the form of DNA sequences.

DNA chips consist of large number of spots embedded on a solid surface, most commonly used is a glass slide.

Each spot consists of different kind and number of probes, where probes are small nucleotide sequences, which are able to bind to the complimentary nucleotides. Nucleotides, which bind to these probes, are fluorescent labeled, whenever any DNA sequence binds to these probes, it is observed under a laser dye and data is calculated electronically depending upon the ratio of the binding of probe with the DNA in each spot.

Technique, considering the typical cryptographic scenario, has following steps:

a) Encryption key is a collection of particular probes where decryption key is a collection of corresponding probes having complimentary sequence. The decryption key is then sent to the receiver in a secure manner.

b) Plaintext is converted into a binary format. This binary format is then embedded into DNA chip as a cipher text (cipher DNA). Without knowing the decryption key one cannot read the plaintext from the DNA chip

c) Receiver uses the decryption key and hybridizes the cipher DNA. With the help of computer software he can retrieve plaintext.

5) *DNA Tiles Assembly Method :* Utilizing the macromolecular building blocks which are called DNA tiles based on the branched DNA strands, DNA nano structures provide a programmable methodology for bottom-up nano-scale construction of patterned structures. These tiles have sticky ends termed pads, which can match the corresponding sticky ends of other DNA tiles, facilitating further assembly into larger structures known as DNA tiling lattices. The self-assembly process is characterized by numerous beneficial attributes. It is cost-effective, versatile and facile. The process occurs towards the system's thermodynamic minima, resulting in stable and robust structures. Simple self-assembly schemes are already widely used in chemical syntheses. It has been suggested that the more complicated schemes will ultimately be useful for circuit fabrication, nano- robotics, and amorphous computing,. Winfree et al. first came up with the idea of computing by self-assembly tiles. Because DNA tiles can be more easily "programmed" to incorporate the constraints of a given problem, it is possible to exercise some degree of control over avoiding the considerable waste of material. At the same time, parallel computation can be enhanced by self-assembling process where information is encoded in DNA tiles. Winfree , Nadrian Seeman , Reif, and Rozenberg have done large theoretical and experimental work to research the relation of DNA computation to self-assembling structures from the mid-1990s. Currently, the self-assembly computation systems have been demonstrated in both 1-D arrangements called "string tiles", and 2-D lattices of DNA. Other stable forms of nucleic acids include Z- DNA, non-migrating Holliday junctions, and duplexes with triple crossovers. For 2-D self-assembly, Winfree has proposed the tile assembly model, and has demonstrated that it is turning universal by showing that a tile system can simulate Wang tiles [53], which Robinson has shown to be universal . The tile self-assembly model, which is a formal model with the self-assembly molecules constrained to self-assemble on a square lattice, is an extension of the tiling theory by Wang tiles that include a specific mechanism for growth based on the physics of molecular self-assembly. The first experimental demonstration of computation using DNA tile assembly [52], gave a two layer, linear assembly of TX tiles that performed a bit-wise cumulative *XOR* computation. Barish et al. have demonstrated a tile assembly system implementation that could copy an input bit and count it in binary. So DNA nanostructures provide a novel methodology for us to resolve all kinds of mathematical and combinatorial problems. Now the applications of DNA tiles are more and more universal. To implement XOR OTP algorithm with DNA, first the operational difficulty was overcome: the non-reusability of the key and the requirement that the key must be random. In [40], a DNA self-assemble system is designed to achieve the non- reusability and randomness of keys. There are four systems: encrypting system, ciphertext extracting system, key extracting system, and decrypting system to implement the encrypting, ciphertext extracting, key extracting and decrypting processes. Whole encrypting process of OTP is explained as follows:

First, we describe the basic input tiles. As we know, when we want to send a string of binary message *m* with *n* bits, we compute cipher text c_i using formoula $c_i = m_i$ xor k_i , i=1……n, in which the k_i is the secret key kept in the sender and receiver before hand.

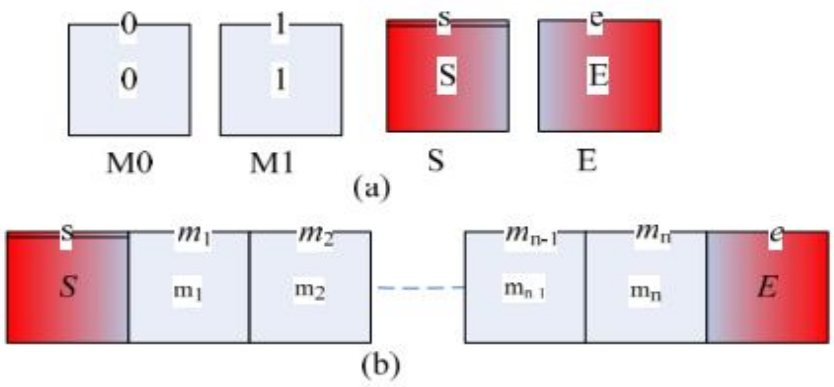


Figure 3.

A string of binary input can be encoded by input tiles. We will use four types of tiles to encode the input binary message. We denote the set of input tiles as, that are shown in Fig 3 a, which is a graphical representation of each tile has four sides, expressed as <north, east, south, west>, so the set of input tiles can be denoted as T= {*S=<s, null, null, null>,E=< e, null, null, null>*, *M=<m, null, null, null>*}. The value in the middle of each tile *t* represents that tile's value, denoted as *v(t).* These tiles have one binding domains, which will stick to the corresponding tiles with the same binding domains. Let *m* be *n-bit* input message. The configuration for *n-bit* input message *m* is shown in Fig. 3 b where m_i is {0,1} . Two special tiles are called boundary tiles and denoted as '*S*' (with one *strong* bond and three *null* bonds) and '*E*' (with one *weak* bond and three *null* bonds) respectively. They are used to denote the start and the end of the input plaintext. They set limits on the extent of the calculation, and will facilitate a modular approach to the process.

6) Implementing XOR One Time Pad (Vernam Cipher) using DNA based Addition : There are a number of possible methodologies for construction of cipher words used for the cryptosystems. One methodology is the random assembly of one-time-pads in solution. We view such methods less favorably due to the difficulty of achieving both full coverage and yet still avoiding possible conflicts by repetition of plain- text and/or cipher words.

Other methods like DNA chip technology and Polymerase chain reaction also have their own problems. Therefore a general algorithm for DNA-based modulo-2 addition of any two non-negative rational binary numbers is presented in [42], [54]. DNA representation of all possible pairs of input non-negative binary bits as shown in figure 4.

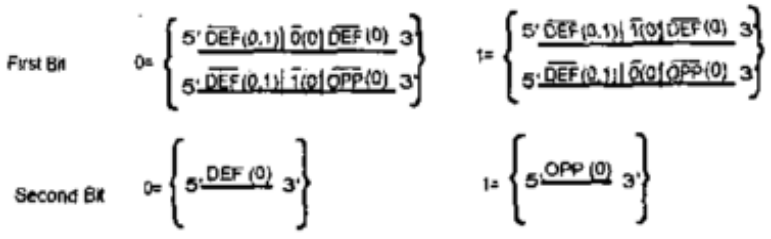


Figure 4: DNA representation of all possible non-negative binary bits to be added

Position refer to the value of either 0 or 1 of bit to be added in the original message and in the cipher word, respectively. DNA sequences are single-stranded, unique, and non-complementary, except that overlining indicates a complementary DNA sequence, for example, *DEF (0,*1) is complementary to *DEF(0,*1). A number in parentheses refers to a position, whereas a number not in parentheses refers to the value of the digit at that position. Here the position information provides CNT-probe accessibility for data write-in/read-out. The first bit is represented by two DNA strands, each containing (from the 5' end) a "position transfer operator". The second digit is represented by a single DNA strand with the sequence *DEF* or *OPP* if the digit is either 0 or 1, respectively. This strand represents an operator that serves as a primer in a primer extension reaction.

As a schematic example, we illustrate how to add binary 1 + 1. The bio-chemical reaction is shown in Figure 5,

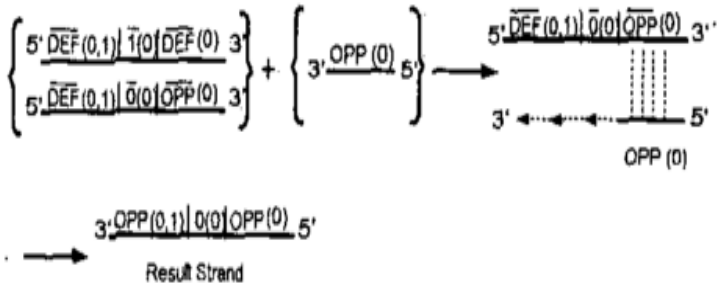


Figure 5 : Illustration of the operation 1+ 1 as an example of a DNA-based addition of two binary bits.

Vertical dotted lines represent hybridization between complementary DNA elements, and reiterated arrows represent primer extension.

In the reaction, the operator (primer) representing the bit i in original message sequence (Wi) hybridizes to the appropriate DNA strand representing the coded bit i (G;). Primer extension yields result DNA strand. This successive reaction represents an example of a process of horizontal chain reaction, in which input DNA sequences serve as successive templates for extension of result strands.

In classical cryptography, the Vernam cipher (now known as the XOR one-time-pad cryptosystem) is deployed by generating a sequence, S of R independently distributed random bits known as a one-time-pad. The one-time-pad is replicated, and stored one copy at the source and one at the destination. Let L be the number of bits of S that remain unused, where initially L = R.

XOR is the operation that essentially is Boolean modulo-2 addition. When a plaintext binary message M which is n < L bits long needs to be sent, each bit M; is XOR'ed with the message bit K, to produce encrypted bits

$$C_i = M_i + K_i \text{ for } i=n$$

The n bits of S that have been consumed are then destroyed at the source.The encrypted sequence C = *(C1, C~, , ,., Gn)* is then dispatched to the destination,

At the destination, the identical process is repeated. The sequence C is used in place of M.

Our DNA encoded messages are modified in this case by a bit-wise moduio-2 addition computation, so that fragments of the plain text are converted to cipher strands using the one time pad DNA sequences and the plain text strands are removed afterwards.

In order to take advantage of the massive parallel processing capabilities of biomolecular computation, the following method for basic operations such as arithmetic (addition and subtraction) permit chaining of the output of these operations into the inputs to further operations, and to allow operation to be executed in massive parallel fashion.

Generalization of the addition to two non-negative vectors with n-digit binary numbers is straightforward, The two bits in each of the positions 2° through 2^n are represented as shown in Figure 4 with the following modification. At a position **i** other than 1, unique DNA sequences represent the values O(i) and l(i), and operators are replaced appropriately; for example, *DEF(I)* and *DEF(1,2)* by DNA sequences representing *DEF(i)* and *DEF(i,* i+ 1). The modulo-2 addition operation is in theory exactly as described above. This operation yields a final result strand longer than that shown in Figure 6 but with the same basic structure. This more general algorithm can readily be extended to the addition of any two n-digit positive rational numbers.

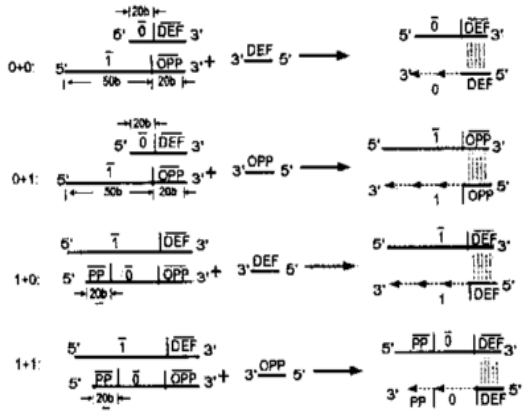


Figure 6: Illustration of a general DNA-based algorithm for adding two binary digits.

Combining test tube primer extension reagents performs addition plus DNA one-time-pad strands that appropriately represent two numbers to be added. The addition is then followed by a primer extension reaction.

7) *Encryption using DNA Fragment assembly:* DNA fragment assembly is a technology which attempt to reconstruct a large number of DNA fragments into the original long chain of DNA. In order to solve the limit of length of the sequence, this technology was developed [41] and the measures are as follows: First, amplifying the DNA chain and getting lots of backup; Second, Obtain a large number of short DNA fragment by cutting the DNA long chain at the random locations; At last, recombination the DNA fragment, which have the overlapping part, back into the original DNA chain. This strategy is also called shotgun sequencing. By the help of software, first the encryption staff will translate the plaintext into binary code sequence, and then keep on translating it into plaintext long chain DNA. After that, the encryption staff will obtain a large number of DNA fragments by cutting the long-chain DNA randomly by biotechnology. Then the encryption staff can prepare the key of short-chain DNA. When he finished these steps, he can start the encryption operations -- implanting the key of short-chain into plaintext DNA fragment, and send it to the receiver. After the receiver has obtained the DNA fragment which is already processed, firstly, he will move the key of short-chain, adding by the encryption staff, and then starting fragment assembly to restore the long-chain of plaintext DNA. Finally, the receiver use the software to translate the long-chain DNA into binary code sequence, and then keep on convert them to plaintext.

By the use of comparison program of DNA chain, it is found that the resulting DNA chain has the similarity of 8.0% with the plaintext DNA long-chain and there are some errors in the stage of overlap and layout.

8) *Encryption using different structures of DNA Molecule:* Design of controllable DNA structures was first introduced by Nadrian Seeman in 1980s [23]. He invented DNA nanotechnology - the science and technology of building devices using DNA molecules. In nature DNA plays the role of genetic information carrier, but in this branch of technology it is used just as a structural material. Attraction of complementary DNA strands is used for building different nanoscale structures which are described below.

Holliday junction is one of the simplest forms of DNA building blocks. It is a junction between 4 complementary to each other DNA strands (Fig. 7). It is not recommended to be used in the nano engineering because of its instability induced by strong electrostatic repulsion.

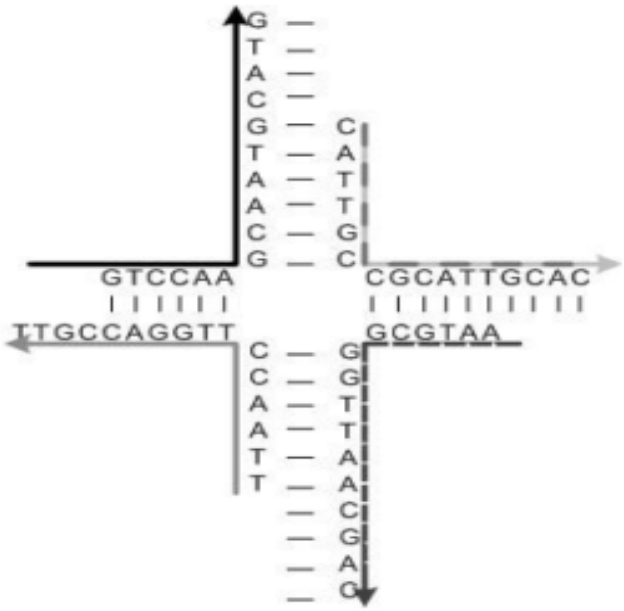


Figure.7: Holiday Junction

Double crossover molecule (DX) is a more stable structure which consists of two DNA helixes connected by two Holliday junctions. There are five different structures of DX molecules [figure : 8]. Three of them have parallel helical domains: DPE, DPOW, DPON and the other two antiparallel helical domains: DAE, DAO.

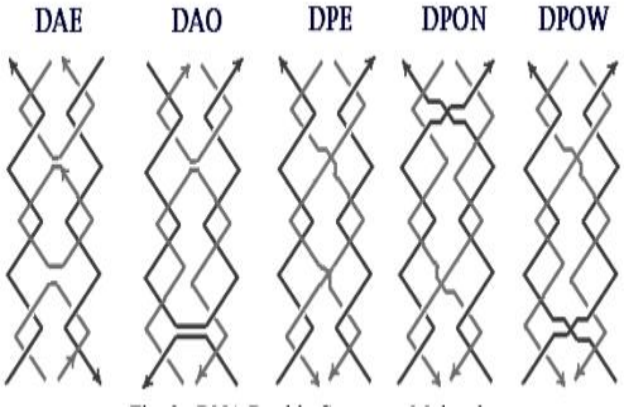


Figure.8: DNA Double Crossover Molecules

DX molecules with antiparallel domains are more stable than those with parallel domains . Therefore the DAE and DAO structures present more interest. The difference between them is that DAE molecule has even number of double helical half-turns between crossovers and DAO molecule has odd number of half-turns.

The method described in [33] consists of using one long single DNA strand, in order to create a basic structure, and many short DNA sequences that come as complementary parts to the basic structure forming in this way double stranded DNA in desired shape. The same principle was used in crossovers that appear between DNA helixes are incorporated for stability of the structure.

An important aspect is finding a long single-stranded DNA. Certain synthesizers allow the synthesis of long oligonucleotides up to 250 bases. Longer single-stranded DNA can be founded in viruses like M13mp18..

Long DNA strands can be synthesized as oligonucleotides of 210 bases in length and short strands as oligonucleotides of 42, 46, 50, or 54 bases in length. The length of short oligonucleotides depends on the number of single-stranded (sticky) terminations of the structure. These terminations are used for binding other structures with complementary ends.

Steps of the implementation for this structure are the following:

1) Generation of random sequence from DNA alphabet (A, C, T, G) of length 210 bases.

2) Generation of complementary sequences to the first one. For example for the first middle circular sequence (Fig. 9) complementary sequences were generated in intervals: 22-42 and 169-189.

At the edges complementary sequences can have different lengths depending on the number of “sticky end” of the structure. For such sequences is performed concatenation of the terminations like: “ACTG”, “TACC”, or “TGAC” (Fig.9)

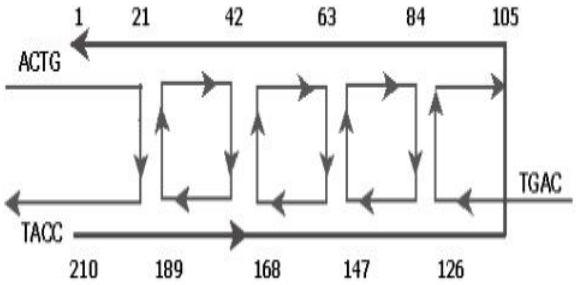


Figure.9: DNA Structure for a message binary ‘1’.

9) *DNA Cryptography using two level keys:* A symmetric key cryptography method in [31] has introduced a new format of cipher text, where the primary cipher text obtained after encoding is being divided into three unequal parts and then extra parameters such as primer code, file type code, integrity code, and authentication code are added in between parts of the cipher text to obtain the final cipher text.

Format of Cipher Text-

From plain text (PT) the primary cipher text (CT) is obtained by using the encryption algorithm and the 1st level key (PK1).

Abbreviations used are: CT-PRIMARY CIPHER TEXT, AUT-AUTHENTICATION CODE (ENCRYPTED FORM), INTR-INTEGRITY CODE (ENCRYPTED FORM), FT-FILE TYPE CODE (ENCRYPTED FORM), SPM-STARTING PRIMER

(GARBAGE), EPM-ENDING PRIMER (GARBAGE), OCT-ORIGINAL CIPHER TEXT FORMAT. The following steps are to be followed to obtain the final cipher text.

Step-1: Encrypt the plain text with 1st level key (PK1).

Step-2: Divide the primary cipher into three unequal parts.

step-3: Attach AUT, INTR, FT, SPM, EPM with the above CTB's after encrypting CTB's using level 2 private keys(which include the information about the introns (AUT,FT ,ETC) positions and the length of the SPM and EPM).

IV. DNA STEGANOGRAPHY

Steganography is a class of techniques that hide secret messages within other messages. In a steganography system, also explained in [55], [22], the original plaintext is not actually encrypted but is instead disguised or hidden within other data. Historical examples of steganography systems are the use of grills that mask out all of an image except the secret message, microphotographs placed within larger images, invisible inks, etc. The cryptography literature generally consider conventional steganography methods to have low security (in fact there is disagreement whether steganography is actually encryption, since the plaintext is not actually encrypted but instead disguised within other media) and there are numerous cases where steganography methods have been broken in practice . However, it is very appealing due to it's simplicity. There are a number of techniques for applying steganography in the context of biomolecular computation. One method is to take one or more input DNA strands (considered to be the plaintext message) and append to them one or more randomly constructed "secret key" strands. The resulting "tagged plaintext" DNA strands are then hidden by mixing them within many other additional "distracter" DNA strands which might also be constructed by random assembly. Given knowledge of the "secret key" strands, the resulting solution of DNA strands can be decrypted by a number of possible known recombinant DNA separation methods. For example, the plaintext message strands may be separated out by hybridization with the complements of the "secret key" strands might be placed in solid support on magnetic beads or on a prepared surface. These separation steps may combined with amplification steps.

In a true cryptographic cipher the security is dependent on a secret key. The s ecurity of the above system is expected to derive from - as is the case with all steganographic systems- the fact that the adversary is unaware of the existence of the message in the medium of transmission, and/or can not distinguish the plaintext message from the medium. As soon as this assumption is no longer valid, the system can generally be compromised. In particular, this DNA steganography system's security is entirely dependent on the degree that the message DNA strands are indistinguishable from the "distracter" DNA strands.

V. DNA CERTIFICATION

Strictly speaking, DNA certification doesn't deal with much DNA computing techniques, but mainly employ the biological characteristics of DNA [55]. Currently, the DNA certification is broadly applied in the field of justice, finance, and so on, which will certificate biological individuals accurately.

Celland apply their methods to the purpose of certification and security. In 2000, DNA Technology Company of Canada also using the DNA sequence to the product certification of the Sydney Olympic game in those years. Nearly 50,000,000 keepsakes were all marked with a special ink from Olympic T-shirts to coffee cups. This kind of DNA segment in the inky mark was randomly extracted an athlete' genome from nearly a hundred, then it was rather difficult to fabricate. This kind of way to utilize a portable scanner to scan the information in the ink marking would distinguish whether the keepsakes were authentic which were much cheap than the whole interest trademarks but the increasing cost was only one nickel.

If we make use of the basic principle of DNA steganography to the appraisal of DNA, we can carry on much wider certification. At present, there have been masses of biological genetic engineering are under way. The above technologies will make the researchers add the DNA certification information to the organ tissue, and by identifying the information of DNA certification to validate the customer identity and the copyright information.

Currently, in these kinds of DNA techniques, the development of DNA certification technique is most mature and the application is most wide. However, the introduction of the DNA computing into the DNA steganography and certification techniques will improve the complication of algorithm and the level of security.

VI. PROBLEMS WITH DNA COMPUTING

Although the DNA computing is a fire-new computing mode, it can't get away from the influence of Turing in the corresponding theoretical computing model. The DNA computing is still placed in a theoretical stage, its computing model was mostly just using

molecular technique to resolve a certain problems, and put on an experiment to resolve a certain problem, the varieties of problems lead to the discrepancy of computing schemes, there still haven't an uniform computing and coding model currently. Under the existing DNA computing mode, the time complexity of DNA computing compared to the space complexity doesn't increase with the computational complexity remarkably. That is, DNA computing only converts the time complexity into space complexity. Then, once the complication of problems breaks the physical limit of DNA segment, which operated by the bio-chemical technique, DNA computing is still too far away to reach. Boneh spend nearly 4 months to construct DES(E) solution, however, the quantities of cipher key of AES algorithm utilized by the US federal government is 21 times compared to DES algorithm. Therefore, according to the Boneh's way, it will cost several years if we construct AES(E) solution. So we can say that Boneh's method can only break the symmetric system under 64 bits. Mathematical cryptography can be easily increasing the length of the cipher, thereby it'll prevent the cryptography from powerful attack using DNA computer. Therefore, in terms of existing DNA computing mode, though DNA computers greatly improve the ability of the cipher break of people, it is not good to construct genuine intimidation to the security of cryptography. DNA cipher is the beneficial supplement to the existing mathematical cipher, it is a good prior choice especially to the lower demand real-time encryption system.

VII. CONCLUSION

After discussing the various DNA based encryption/decryption algorithms, we can say that these algorithms provide the better and double layer security than the traditional cryptographic algorithms. But these algorithms impose two fold computing difficulty as well and due to this they posses high confidential strength.

The application technology of DNA computing in cryptography is comparatively not mature enough and some DNA computing methods and models could not be used in laboratory. The nanotechnology and biological technology hold tremendous promise, but many technical hurdles will have to be overcome before complex methods can be developed into a practical commercial technology. If the molecular word can be controlled at will, it may be possible to achieve vastly better performance for information storage and information security.

VIII. FUTURE SCOPE

In future research, scientists should do further research to provide a theoretical proof of DNA cryptosystem's validity to make it be provable security level, and perfect the algorithm's security model. Also, they need to make full use of DNA computing and biological characteristics to eliminate the disadvantages.

In theory OTP cipher is absolutely secure. But practically, key distribution and key generation are critical issues to be resolved for the use of OTP ciphers. Key space should be large enough so that keys can only be used once. DNA having huge storage capacity, can be manipulated to generate key space to be used for OTP cipher.

There are also some areas that need to be improved. For example, Time and computational complexity are two of the most important parameters for any kind of cryptographic systems, DNA cryptography dealing with the manipulation of DNA sequences takes a lot of time to deal and work out with DNA sequences as compared to time taken by many very efficient algorithms of traditional cryptography.

In steganography field, there is a strong need to prove and show that whether DNA stegnography systems with natural DNA plaintext input can or cannot be made to be unbreakable.

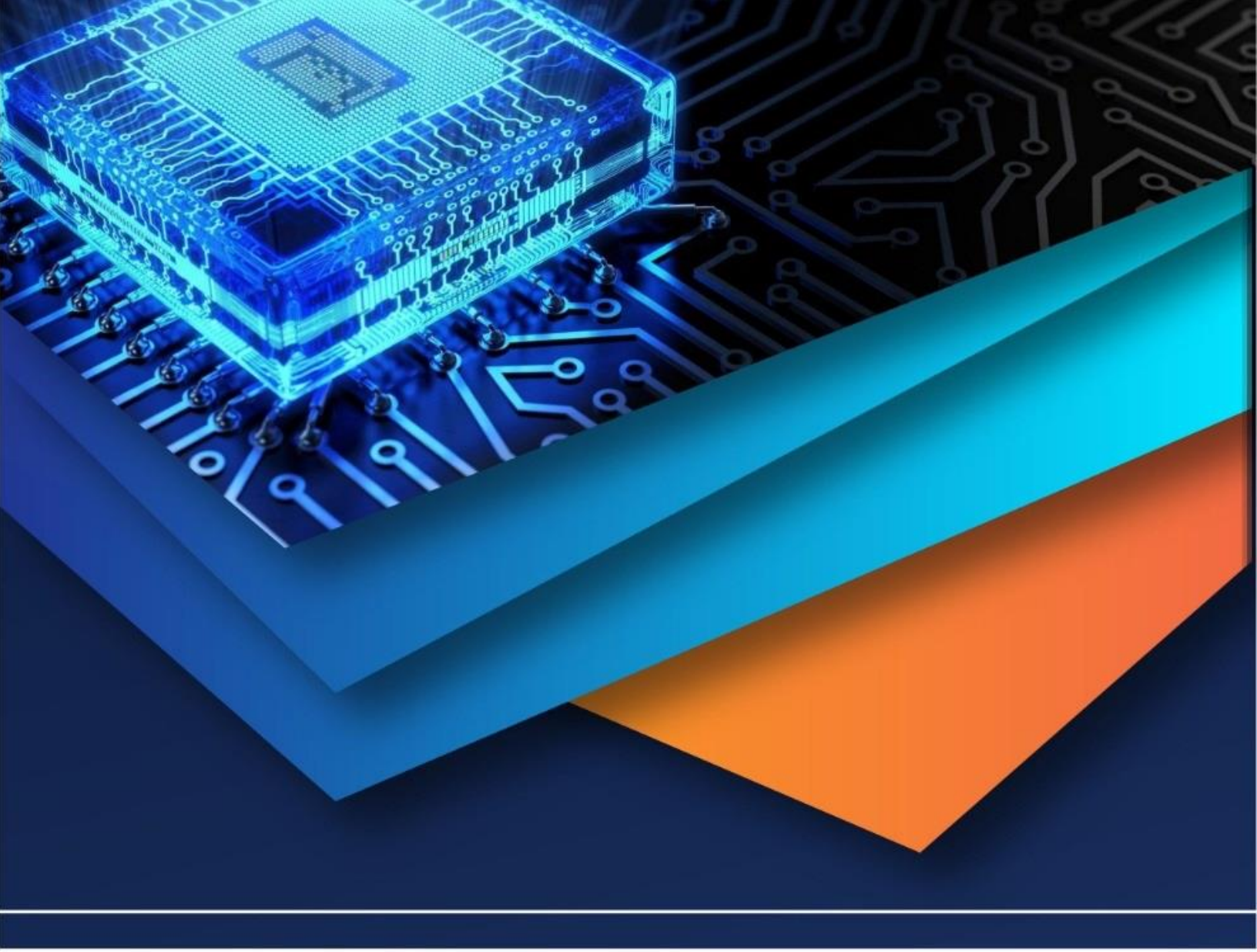
In short we can say that DNA Cryptographic field requires a lot of research and work to have a position in which it can be implemented and used for practical purposes. There is a need that people from traditional cryptography and DNA technology should exchange knowledge among each other and cryptosystems should be devised in such a way that they can enjoy benefit from both the fields

REFERENCES

[1] L. Adleman, Molecular Computation of Solutions to Combinatorial Problems",Science266:1021-1024 (Nov. 11) 1994.
[2] T. Kazuo, O. Akimitsu, and S. Isao, "Public-key system using DNA as a one-way function for key distribution," Bio Systems, Elsevier Science, vol. 81, no. 1, pp. 25–29, 2005.
[3] M. Yamamoto, S. Kashiwamura, A. Ohuchi, and M. Furukawa, "Large-scale dna memory based on the nested pcr," Natural Computing, an international journal, vol. 7, no. 3, pp. 335–346, 2008.
[4] R.I. Lipton. DNA Solution of Hard Computational Problems, Science, 268, 1995, pp. 542-545.
[5] J.Chen, H.Li, K.Sun, and B.Kim, "How will bio- informatics impact signal processing?" IEEE Signal Processing Mag., vol. 20, no. 6, pp. 16-26, Nov. 2003.
[6] A. Suyama. DNA chips- integrated chemical circuits for DNA diagnosis and DNA computers", 1998
[7] J. M. Lehn, "Sopra molecular Chemistry," Science, vol. 260, pp. 1762–1763, 1993.

[8] L. M. Adleman, Q. Cheng, A. Goel, M. Huang, D. Kempe, P. Moisset, P. Rothemund, "Combinatorial optimization problems in self-assembly," in Annual ACM Symposium on Theory of Computing (STOC), p. 23–32, 2002.
[9] H. Abelson, D. Allen, D. Coore, C. Hanson, G. Homsy, T. Knight, R.Nagpal, E. Rauch, G. Sussman, R. Weiss, "Amorphous computing," Communications of the ACM, vol. 43, pp. 74–82, 2002.
[10] E. Winfree, T. Eng, G. Rozenberg, "String tile models for DNA computing by self-assembly", in proceedings the 6th international meeting on DNA-based computers, p. 65, 2000.
[11] E. Winfree, "Algorithmic self-assembly of DNA," Ph.D. Eng. dissertation, California Institute of Technology, Pasadena CA, 1998.
[12] N. C. Seeman, "DNA nanotechnology: novel DNA constructions,"Annu. Rev. Biophy. Biomol. Struct., vol. 27, pp. 225–248, 1998.
[13] J. H. Reif, "Computing: successes and challenges," Science, vol. 296, pp. 478–479, 2002.
[14] G. Rozenberg, H. Spaink, "DNA computing by blocking," Theoretical Computer Science, vol. 292, pp. 653–665, 2003.
[15] E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, "Design and self- assembly of 2D DNA crystals," Nature, vol. 394, pp. 539–544, 1998.
[16] C. Mao, W. Sun, N. C. Seeman, "Designed two dimensional DNA holliday jumction arrays visualized by atomic force microscopy," J. Am. Chem. Soc, vol. 121, pp. 5437–5443, 1999.
[17] C. Mao, T. H. LaBean., J. H. Reif, N. C. Seeman, "Logical computation using algorithmic self-assembly of DNA triple-crossover molecules," Nature, vol. 407, pp. 493–496, 2000.
[18] E. Winfree, "Algorithmic self-assembly of DNA," Ph.D. Eng. Thesis, Caltech, Pasadena, CA, 1998.
[19] H. Wang, "Proving theorems by pattern recognition I," Bell System Technical Journal, vol. 40, pp. 1– 42, 1961.
[20] R. M. Robinson, "Undecidability and nonperiodicity for tilings of the plane," Inventiones Mathematicae, vol. 12, pp. 177–209, 1971.
[21] R. Barish, P. Rothemund, E. Winfree, "Two computational primitives for algorithmic self-assembly: copying and counting," Nano Letters, vol. 5, pp. 2586–2592, 2005.
[22] Ashish Gehani, Thomas H. LaBean and John H. Reif, 5th Annual DIMACS Meeting on DNA Based Computers (DNA 5), MIT Cambridge, MA, June 1999.
[23] N.C. Seeman, "Nucleic Acid Junctions: Building Blocks for Genetic Engineering in Three Dimensions", Adenine Press, pp. 269-277, 1981.
[24] T.J. Fu, N.C. Seeman, "DNA double crossover molecules",Biochemistry, Vol. 32, pp. 3211– 3220, 1993.
[25] N.C. Seeman, H.Wang, X. Yang, and J. Chen, "New Motifs in DNA Nanotechnology", Nanotechnology, Vol. 9, No 3, pp. 269-277, 1998
[26] M. Levitt, "How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations", Biochemistry, Vol. 75, No. 2, pp. 640-644, 1998.
[27] W.M. Shih, J. D. Quispe, G.F.A Joyce, "1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron", Nature,Vol. 427, pp. 618-621, 2004.
[28] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns", Nature, Vol. 440, pp. 297-302, 2006.
[29] Beenish Anam, Kazi Sakib, Md. Alamgir Hossain and Keshav Dahal, "Review on the advancements of DNA Cryptography", SKIMA 2010, arXiv:1010.0186v1 [cs.CR], 1 oct 2010.
[30] Deepak Kumar and Shailendra Singh, "Secret Data writing using DNA sequence",International Conference on "Emerging Trends in network and computer communications" (ETNCC), Pages-402-405, 22-24 April, 2011
[31] Bibhash Roy, Gautam Rakshit and Pratim Singha, "An Improved Symmetric key cryptographywith DNA based strong Cipher", International Conference on "Devices and Communications(ICDECOMM), Pages 1-5, 24-25 Feb 2011.
[32] Xing Wang, Qiang Zhang, "DNA computing-based cryptography", Fourth international conference on " Digital Object Identifier", Pages 1-3, 2009.
[33] O. Tornea, M.E. Borda, V. Pileczki and R. Malutan , "DNA Vernam Cipher", 3rd International Conference on "E Health and Bio Engineering", 24th-26th November,2011, Iasi, Romania.
[34] Monica Borda, Olga TORNEA,"DNA Secret Writing Techniqes", 8th International Conference on communications, 10-12 June 2010.
[35] Amal Khalifa, Ahmed Atito, " High capacity DNA based Steganography", 8th international conference on Informatics and Systems (INFOS2012), 14th-16th may.
[36] Junzo Watada, Rohani binti abu Bakar, " DNA Computing and its Applications", 8th international conference on intelligent systems design and applications,2008.
[37] Christy M.Gearheart, Benjamin Arazi, Eric C. Rouchka , " DNA based random number generation in security circuitry" , Biosystems 100, 2010, pp 208-214.
[38] Zhang Yunpeng, Zhu yu, Wang Zhong, Richard O. Sinnott, " Index based Symmetric DNA Encryption Algorithm", 4th international conference on Image and Signal processing, 15-17 oct 2011, vol. 5, pp. 2290-2294, Xi'an, China.
[39] Raghava Nallanthighal, Vijeta Rani, "Pseudorandom binary key generation from binary files", International conference on Computational Intelligence and Communication Networks (CICN),2011, pp. 631-634.
[40] Zhihua Chen, Jin Xu, "One-Time-Pads Encryption in the Tile Assembly Model",Bio Inspired Computing: Theoris and Applications,2008, 3rd international conference on Digital Object Identifier, pp. 23-30.
[41] Yunpeng Zhang, Bochen Fu, " DNA Cryptography based on DNA Fragment Assembly , Information Science and Digital Content Technology (ICIDT), vol 1, 8th International conference, 2012, pp. 179-182.
[42] M. Shyam, N. Kiran,V Maheswaran, " A novel encryption scheme based on DNA Computing".
[43] Er. Ranu Soni, Er. Vishakha Soni, Er. Sandeep KumarMathariya, " Innovative Field of Cryptography: DNA Cryptography, ITCS, SIP, JSE-2012, CS & IT 04, pp. 161-179, 2012.
[44] D.Prabhu, M.Adimoolam,"Bi-serial DNA Encryption Algorithm (BDEA)", Computer Science, arXiv :1101.2577[cs.CR]
[45] Guangzhao Cui, Cuiling Li, Haobin Li, Xiaoguang Li, "DNA Computing and Its Application to Information Security Field",fifth International Conference on Natural Computation, henan Key Lab of Information-based Electrical Appliances, Zhengzhou, 2009.
[46] Yunpeng Zhang* and Liu He Bochen Fu," Research on DNA Cryptography", In Tech Publications, March 2012.

[47] Miki Hirabayashi, Akio Nishikawa, Fumiaki Tanaka, Masami Hagiya, Hiroaki Kojima, Kazuhiro Oiwa, "Analysis on Secure and Effective Applications of a DNA-Based Cryptosystem", Sixth International Conference on Bio-Inspired Computing: Theories and Applications, 2011.
[48] Shihua Zhou ,Qiang Zhang, Xiaopeng Wei, "An Image Encryption Algorithm Based on DNA Self-Assembly Technology", IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), vol 2, pp 315-319, 2010 .
[49] Shuhong Jiao, Robert Goutte, "Code for Encryption Hiding data into Genomic DNA of Living Organisms", ICSP Proceedings, 2008.
[50] Lidia Ogiela, "Biological Modelling in Semantic Data Analysis Systems", Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, 2012.
[51] G. Cui, L. Qin, Y. Wang, and X. Zhang, "An encryption scheme using DNA technology," in IEEE 3rd International conference on Bio- Inspired Computing: Theories and Applications (BICTA08), Adelaid, SA, Australia, 2008, pp. 37–42.
[52] C. Mao, T. H. LaBean., J. H. Reif, N. C. Seeman, "Logical computation using algorithmic self-assembly of DNA triple-crossover molecules," Nature, vol. 407, pp. 493–496, 2000.
[53] H. Wang, "Proving theorems by pattern recognition I," Bell System Technical Journal, vol. 40, pp. 1–42, 1961.
[54] Jie Chen, " A DNA based Biomolecular Cryptography Design", IEEE, pp 0-7803-7761, 2003.
[55] Guangzhao Cui, Limin Qin, Yanfeng Wang, Xuncai Zhang, " Information Security Technology based on DNA Computing", IEEE, pp. 1-4244-1035, 2007.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY