JOURNAL OF SOFTWARE ENGINEERING & INTELLIGENT SYSTEMS
ISSN 2518-8739
30th April 2018, Volume 3, Issue 1, JSEIS, CADMEI Copyright © 2016-2018
www.jseis.org

DNA-based cryptography: motivation, progress, challenges, and future

[1] A.E. El-Moursy, [2] Mohammed Elmogy, [3] Ahmad Atwan
[1,2,3]Information Technology Dept.,
Faculty of Computers and Information,
Mansoura University, Egypt
Email: [1]abdullaelsaied@yahoo.com, [2]melmogy@mans.edu.eg, [3]atwan_2@yahoo.com

ABSTRACT

Cryptography is about constructing protocols by which different security means are being added to our precious information to block adversaries. Properties of DNA are appointed for different sciences and cryptographic purposes. Biological complexity and computing difficulties provide twofold security safeguards and make it difficult to penetrate. Thus, a development in cryptography is needed not to negate the tradition but to make it applicable to new technologies. In this paper, we review the most significant research, which is achieved in the DNA cryptography area. We analysed and discussed its achievements, limitations, and suggestions. In addition, some suggested modifications can be made to bypass some detected inadequacies of these mechanisms to increase their robustness. Biological characteristics and current cryptography mechanisms limitations were discussed as motivations for heading DNA-based cryptography direction.

Keywords: DNA; cryptography; encryption; DNA computing; bio-inspired cryptography;

1. INTRODUCTION

Technological development seizes our valuable information including financial transactions are transmitted back and forth in public communication channels, posing a considerably high challenge in confronting with unintended intruders. One suggestion is cryptography that is about constructing protocols built strong mathematically and theoretically by which different security means are being added to such precious information. DNA computing is a new science emerged in recent years clarifying to be very efficient in energy consumption, high information storage capability and parallel processing. Deoxyribonucleic Acid is molecules formed in a certain sequence to construct the information needed for building and maintaining the vital operations of an organism, similar to the way in which binary bits appear in certain order to form different information in our digital world [1].

1.1 DNA computer

DNA computer or biomolecular computer is a computer its input, system, and output is wholly or partially made of DNA molecules, biochemistry, and molecular biology hardware instead of silicon chips technologies. The complexity and ingenuity of living beings are built based on a simple coding system functioning with only four components of DNA molecule similar to the binary coding system of traditional computers. This coding system make DNA is very suited as a medium for data storing and processing.

1.2 DNA biological anatomy

DNA is a blueprint for the living organism; it carries instruction for functioning vital processes. DNA is a collection of molecules stuck together to form a long chain of strands, a certain combination of these DNA strands forms amino acids which are the building block of a living organism. Amino Acids, in turn, combines to form protein, proteins create living cells, and cells create organs.

The bases of DNA nucleotides are of four types (guanine, adenine, thymine, and cytosine) labelled as G, A, T, and C; respectively and usually exists in nature in the form of double-stranded molecules, see Figure 1. Human DNA consists of about 3*109 bases, and more than 99 percent of those bases are quite similar to all people.

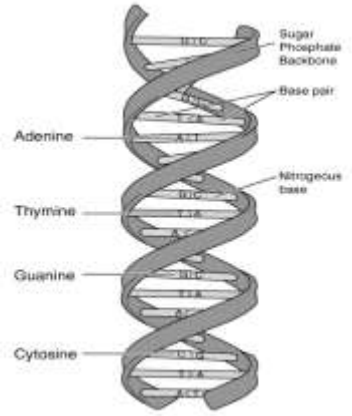

Figure. 1 A short section of a DNA helix and its associated pairs [1]

1.2.1 Central dogma and genetic code

Central dogma is the overall process of transforming DNA nucleotides to synthesize protein (see figure 2), which performs body`s major processes.

Figure. 2 Central dogma process

DNA can be presented as a sequence of nucleotides: AGAGTCTGAGCA. The genetic code is a DNA code written in the form of triplets, named codons. Each triplet uniquely codes one of the Amino Acids, see Figure 3; there are also three excepted codons reserved for 'stop' or 'nonsense' indicating the end of the portion coding. Since there are 3-letters codons combinations with four different bases, this produces 64 possible codons. These encode the 20 standard amino acids, providing redundancy to each amino acid to be encoded by more than one codon.

First base	Second base: U	Second base: C	Second base: A	Second base: G	Third base
U	UUU PHE	UCU SER	UAU TYR	UGU CYS	U
	UUC PHE	UCC SER	UAC TYR	UGC CYS	C
	UUA LEU	UCA SER	UAA STOP	UGA STOP	A
	UUG LEU	UCG SER	UAG STOP	UGG TRP	G
C	CUU LEU	CCU PRO	CAU HIS	CGU ARG	U
	CUC LEU	CCC PRO	CAC HIS	CGC ARG	C
	CUA LEU	CCA PRO	CAA GLN	CGA ARG	A
	CUG LEU	CCG PRO	CAG GLN	CGG ARG	G
A	AUU ILE	ACU THR	AAU ASN	AGU SER	U
	AUC ILE	ACC THR	AAC ASN	AGC SER	C
	AUA ILE	ACA THR	AAA LYS	AGA ARG	A
	AUG MET or START	ACG THR	AAG LYS	AGG ARG	G
G	GUU VAL	GCU ALA	GAU ASP	GGU GLY	U
	GUC VAL	GCC ALA	GAC ASP	GGC GLY	C
	GUA VAL	GCA ALA	GAA GLU	GGA GLY	A
	GUG VAL	GCG ALA	GAG GLU	GGG GLY	G

Figure. 3 DNA to Amino Acids three letters abbreviation coding table [2]

1.2.2 Hybridization

The DNA is a molecule that correlates together to form two long complementary strands (formed by hybridization process defined by Watson-Crick complementary rule, see Figure 4) running anti-parallel to form a double helix structure each of which is made from persistent subunits, called nucleotides.

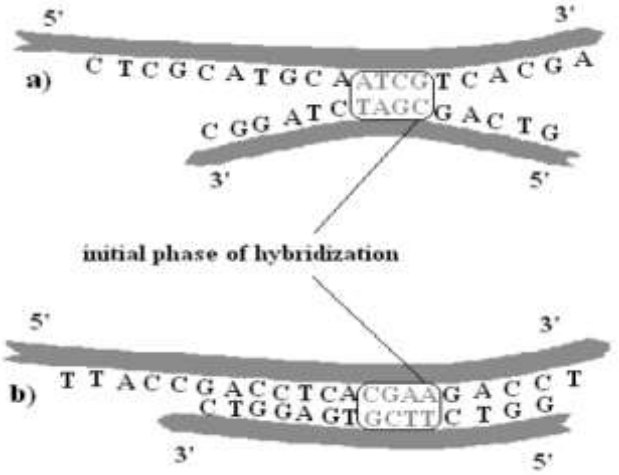

Figure. 4 The Hybridization Process: a) pairing is unstable and cannot go further because sequences have different complementary bases, the strands come apart. b) base-pairing continues because sequences are complementary [1]

1.3 Cryptography

Cryptographic techniques are the kernel of the whole information security field. Cryptography, not only means preventing data from being hijacked, but it also used for authentication. Three types of cryptographic schemes simply achieve these goals: secret key (symmetric) cryptography, public-key (asymmetric) cryptography and hash functions. Two techniques namely a block cipher and stream cipher can be implemented in hardware or software.

1.3.1 Symmetric VS asymmetric

Symmetric encryption is the oldest type and most widely known, it applies to the plaintext to change the content in a particular way; this might be as simple as changing the sequence of the letters or just shifting the letters by a prefixed number of places in the alphabet. As long as both sender and recipient agreed on the same secret key, they can encrypt and decrypt all messages that use the same key easily. The problem with this method is that it needs to exchange this key over the public network while preventing them from falling into the wrong hands. Asymmetric encryption is a suggestion to bypass this, in which there is a mechanism that uses two related keys, one for encryption named as a public key and made freely available to anyone who wants to send a message and the other for decryption and named as a private key and kept secret. The decryption process cannot be completed unless we have the two keys. So the asymmetric key mechanism is more secure, but it lags in speed behind the corresponding symmetric mechanisms. A combination of the speed advantage of secret-key systems and the security advantages of public-key systems in a hybrid system is the best solution. Such a protocol is called a digital envelope.

1.3.2 Block cipher vs stream cipher

A block cipher applies encryption to a block of data at once and requires high computation capability. The key element in the block cipher is the diffusion and the confusion. The Data Encryption Standard (DES) and the Rivest, Shamir, and Adleman (RSA) technique are examples of the block cipher. A stream cipher performs the substitution, bit-wise operations, etc., on each bit of the plaintext with a one-time-pad independently.

1.3.3 Hardware vs software implementation

Cryptographic modules can be executed either by hardware or by software. Whereas, software implementations are known for being easier in maintaining and developing, however, being less secure than their hardware equivalents. The reason lies on

that software solutions make use of shared memory space and running on top of an operation system and more susceptible to be modified.

1.3.4 Cryptography challenges

The more progress in information technology the more challenges it will confront by the mean of insurance, integrity, and security as most of the currently used cryptographic schemes are rely on computational hardness assumptions that cannot keep up with the rapid progress on the information technology development pace.

Cryptography challenges lie on some factors, one of them is based on existing algorithms limitations, and another is on the cryptanalysis attempts to them, in addition to the enormous development of new computer paradigm as Quantum Computing and Molecular Computing.

Quantum Computing: The pace the scientists keep minimizing the transistor size on a silicon chip of a classical computer one day will reach its limit. Quantum Computing may be a replacement. Quantum computers could use quantum bits called qubit by using a single electron instead of digital circuits; it differs that the qubit does not have to be in an exact position, that is called Quantum superposition, the same qubit may be zero, one, any value between zero and one managed by quantum mechanics. The quantum computing is not faster than the classical computer but the key advantage is in the parallelism as each qubit can calculate multiple calculations, in addition to the size of one transistor on a quantum computer can be like the size of one atom. Quantum Computing is one of the threats that undermine the existence nowadays cryptography algorithms, although we are far away from a real implementation.

DNA Computing: DNA Computing takes advantage of massive parallelism embedded in its molecules to try different possibilities at once. If a DNA computer can be realized in a real-world implementation, it can be faster and smaller than any other computer built so far. For instance, the prime factorization problem seems to be realized in an affordable time using a molecular computer.

Secure Channels: There is no secure channel in the real world, but there is at best condition to make insecure channel less insecure. Secure channels are a way to prevent intruders to overhearing or tampering. Confidential channels prevent overhearing but not tampering, and Authentic Channels prevent tampering but not overhearing. Quantum communication channel may be a solution which allows the quantum state to be transmitted as photons through an optical fibre or free space.

Cryptographic Algorithms Limitations: Cryptography, in general, is standard procedure with different keys for each encryption process making the attacker focuses only on one variable which is guessing the key used in the algorithm.

DES has a key of 56 bits which can be brute-forced as demonstrated ten years ago [see Table 1], and the key size also raises some potential challenges in encryption data size of gigabytes which is not that big nowadays. In addition, to its small block size making it sustainable to linear and differential cryptanalysis; it is half the speed of Advanced Encryption Standard (AES) and half the key size.

Table. 1 The DES key strength VS modern computers [3]

Type of Attacker	Budget	Time per key recovered, 40 bits	Time per key recovered, 50 bits
Pedestrain hacker	$400	1 week	Infeasible
Small Business	$ 10.000	12 minutes	556 days
Corporate Department	$300.000	0.18 sec.	3 hours
Big Company	$10.000.000	0.005 sec.	6 minutes
Intelligence Agency	$300.000.000	0.002 sec.	12 seconds

3DES is a trick to keep DES implementation alive some more, by cascading DES three instance; respectively. 3DES is believed to live at least more than 100 years from now, but now it is six times slower than Advanced Encryption Standard (AES) with the same key size especially in dealing with software as it designed for hardware implementation from the beginning. AES is the replacement of DES which done its job very faithfully and never been compromised mathematically.

AES keeps the tradition of DES; it is a symmetric encryption algorithm designed for the use of US federal organization and approved by National Security Agency (NSA). Using 128, 192 or 256 bits keys to bypass the DES challenges, and it is efficient in software and hardware implementation. It is the best till now. However, if AES is now unbreakable, whatever DES was unbreakable to ten years ago. The key advantage of AES over the DES is that it reduced the sub rounds to 10 instead of 16 in DES which simplifies the time complexity, in addition to the parallelism implicated in the double plaintext size. Although, it sustainable to side-channel attacks, which don't attack the AES cipher itself, rather its implementation. Another challenge for making AES infinite in secrecy is the key exchange procedures which make a possibility of the intruder to overhear or to tamper it even if it is being sent over a secure channel. In addition to its complexity, many applications require reduced complexity.

RSA is the most widely used public key cryptosystem as it involves four steps, key generation, key distribution, encryption, and decryption. RSA involves two keys one for encryption and the other for the decryption process, its security based on the factorization problem. The security of RSA exceeded any other cryptography algorithm, but at the expense of the speed, so it used for encrypting the encryption keys in symmetric cryptography algorithms for secure transmission in hybrid encryption techniques.

Cryptanalysis is the field in which hidden or protected information and procedure are being studied to gain access to the content of the encrypted information

Brute-Force Attacks are so named because they do not require much intelligence in the attack process; it is simple as trying each possible key until the correct key is found. It takes 2^n-1 steps in average to get the right key and in the worst case, 2^n steps for a key size of n bits. Parallel and Distributed Attacks is a systematical brute-force attack which works in a parallel and distributed order if we have N processors; we can find the key roughly N times faster than if we have only one processor.

Cryptanalytic Attacks contrary to Brute-Force attacks rely on applications that involve some intelligence ahead of time, a significant reduction of the search space is provided by doing so. While an in-depth discussion of cryptanalytic techniques is beyond the scope of this research. The current known cryptanalytic attacks are discussed briefly:

Differential Cryptanalysis analyzes how differences in plaintext correspond to differences in cipher-text. Linear Cryptanalysis focuses linear approximations on describing the internal functions of the encryption algorithm.

An eternal challenge confronting cryptography is that how to know when an insecure channel worked securely (or, and perhaps more importantly when it did not). In addition, to the static procedures in the encryption algorithms give the eavesdropper a clue of what to do to reverse the encrypted data. The bottom line is: according to last mentioned obstacles, there is no perfect cryptography algorithm as being noticed that the key length is a double-edged sword when dealing with cryptography algorithms. Encryption algorithms themselves maybe not the weak point in an encryption product. However, there are implementation flaws or key management errors. The intrinsic idea of a cryptosystem is a one-way function. A function called a one-way function if it is easy to compute extremely hard to invert but not impossible. In a function $f(x) = y$, it is easy to compute $f(x)$ for all y in the domain range, while it is computationally infeasible to find any x, such $y = f(x)$. A trapdoor one-way function is a function with additional information (trapdoor information) [4].

1.3.5 DNA computing

The organization of organisms is based on a coding system with four components making it very suited to data processing and storing. According to different calculation, one gram of DNA could potentially have the capacity to hold 512 Exabytes, on top on that; the theoretical maximum data transfer speed would be enormous due to massive parallelism of the calculation. DNA computing is the field of science in which biology and computer science merge. The development of biocomputers has been made possible by yielding the science of nano-biotechnology which made by combining the technology of both nanoscale materials and biologically based materials. Though DNA computing still in its primitive stage, it has been applied in many fields and proved efficiency in solving hard problems successfully including, but not limited to, the NP-complete problem (Nondeterministic Polynomial Time), O-1 planning problem, integer planning problem, optimal problem, graph theory, cryptography, database, etc. Utilizing DNA extraordinary characteristics and integrate them in the information technology science will make an incredible leap in the information technology field in the next few years.

1.3.6 The idea

Biological information is very complex and numerous; this emerged the science of bioinformatics to understand and analyze these data as any human brain has not the ability to cope with it. However, when these machines reached its roof capability, scientists needed another theory to deal with information that the machine cannot handle by itself. One suggestion is a bio-inspired computer, a computer inspired by the biological operations. Adleman [5] created the first experiment when he solved a Hamiltonian path graph problem with DNA molecules. He implemented the experiment biologically. Later the idea has extended to computational biology by replacing experiments by computers; known as DNA computer.

Two features of DNA structure amount for its remarkable impact on science. The core idea is as simple as its string nature, and it complementarily resembles the digital structure. The genes themselves are made of the information, stimulating the research in molecular DNA storage.

1.3.7 Molecular computing history

Some researchers show that biological molecules like DNA and enzymes can be built to act like electrical circuits [6]. Someday, these biological circuits could be used, say, to make sensors in cells that would know when to release drugs into the body. The tools of molecular computing start with Adleman [5], a professor of Southern California University for his pioneering work in 1994 setting the stage of biocomputing research combined with the field of mathematics. He made advantage of the biochemical level for solving problems that require an enormous amount of computation or unsolvable by conventional computers. He encoded a small graph by the mean of Hamiltonian path problem in DNA molecules; the computation "operations" were performed with standard molecular protocols and enzymes. Adleman experiment demonstrated the feasibility of fulfillment computations at the molecular level. Adleman mechanism was as follow:

1. Encode all nominees' solutions to the interest computational problem;
2. Generate all possible solution to the computational problem;
3. Keep only paths that start with S and end at E;
4. Keep only paths that have N number of vertices;
5. Keep only paths that visit each Vertex once;
6. Use Polymerase Chain Reaction (PCR) technology to amplify the remaining DNA molecules and get the solution.

The technique later has expanded by various research, including in cryptanalysis, as Boneh et al. [7] explained that the Data Encryption Standard (DES) cryptographic protocol could be broken. From here it leads to, if DNA computing can break codes, it can also be exploited to encrypt data as Pramanik and Setua [1] presented a cryptography technique using DNA molecular structure, one-time-pad scheme, and DNA hybridization techniques. Gearheart et al. [8] were able to demonstrate a novel logic gate design. The design was based on chemical reactions in which observance of the double-stranded sequence indicated a truth evaluation. Ogihara and Ray [9] suggested an implementation to DNA-based Boolean circuits. In 2002, researchers at the Weizmann Institute of Science [10] unveiled a computing storage device made of enzymes and DNA molecules instead of silicon microchips. Finally, in March 2013, researchers created a transcriptor (a biological transistor).

1.3.8 DNA technologies

Polymerase Chain Reaction: PCR is critical in DNA computing as it is the technology that used to extract the problem`s solution. Furthermore, it is an imposition to amplify a sample of DNA over several orders of magnitude and primers. From a cryptographic point of view, PCR can be useful as it requires two primers to accomplish the amplification process. For an adversary, it would be extremely difficult to amplify the message encoded sequence with PCR without these correct primers chosen from about 21023 kinds of sequences.

DNA Fragment Assembly refers to the aligning and merging fragments of DNA sequence to reconstruct the original sequence. The technology is used in DNA sequencing technology and cannot read whole genomes at once.

There are a lot of other DNA technologies as Gel electrophoresis which used to separate mixed DNA fragments and DNA chip technology, a technology used for gene expression and DNA profiling, but they are beyond the scope of this research as the authors here focuses only on technologies that can be used in cryptography.

1.3.9 DNA advantages in computing

Recent research on DNA computing has focused on DNA as information carrier for storing data from ultra-concise information storage and ultra-scale computation. DNA advantages can be listed as follow:

- Parallel processing: 1026 operations/sec
- Low power consumption
- Incredibly light weight
- Data capacity: 2.2 Exabyte per gram
- Imperishable storage

1.3.10 Limitations and challenges

Despite showing a bright future, the research of DNA cryptography still at its initial stage and many scopes still uncovered. Moreover, it confronted with some obstacles the same confronted to DNA computing research that can be summarized as follows:

***Theoretical problems*:** Shannon`s theory illuminated that a powerful tool for generating keys in encryption algorithms should use the complex mathematical procedure, DNA cryptography does not have any mature mathematical background to support this theory.

***Difficult implementation*:** Much extensive material and many biological and lab experiments should be performed to produce a DNA-based cryptosystem; this might be one of the reasons why only a few examples of reliable DNA cryptography mechanisms were exhibited.

A persistent foundation between the biological structure and computer science are required to be the standard to evolve efficient and stable algorithms for DNA computing, therefore, researchers still excavating for mush more theoretical foundation than practical. Further, its slow computing speed and the solution analysis in a molecular computer is a lot harder than a digital one. So, there is a need to create a bridge between existing and new technologies and to open possibilities for a hybrid cryptographic system that provide high confidentiality and stronger authentication mechanism.

1.3.11 DNA digital coding

From a computational point of view or more focused on the eyes of the coder, anything can be coded in binary. Researchers [4] were able using the same tactic to store a JPEG image, and an audio file on DNA digital data storage, this kind of storage device is much more compact than currently used magnetic tapes; this is due to the data density inherent in DNA, in addition to its longevity advantages. When DNA digital coding is being mentioned, it means any scheme to represent digital data in the base of DNA sequence (see figure 5).

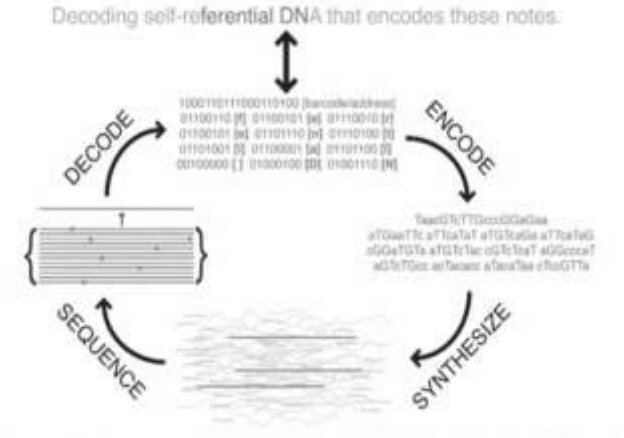

Figure. 5 The DNA digital coding [11]

1.3.12 DNA cryptography

Nowadays, the field of biology and that of cryptography have come to combine (see figure 6). The DNA computing opens a new way for cryptography. The nucleotide bases have the capability for creating self-assembly structures that have excellent means of executing computations. Currently, several DNA computing algorithms are implemented in encryption, cryptanalysis, key generation and steganography, so, from the cryptographic perspective, DNA is very powerful.

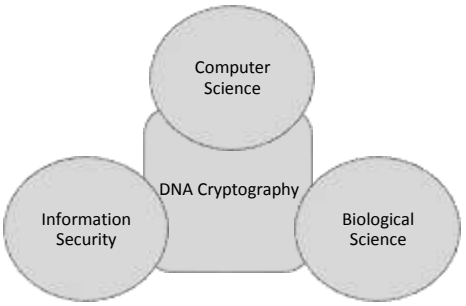

Figure. 6 DNA cryptography

Why DNA encryption instead of digital encryption? Most of the cryptographic systems have been broken at least partially, if not completely or may likely to be a break in the future by the new generation computers. Thus, threats grow exponentially with the growth of technology, from here data transmission and storage have become vulnerable more day after another. As the key element of breaking any encryption mechanism is brute force attacks. One proved when Shamra et al., [12] showed the capability of breaking the Simplified DES (SDES) algorithm at an affordable cost and in a reasonable time. Bhateja and Kumar [13] showed that with the aid of genetic algorithms he could break the Vigenere cipher without the key using elitism with a novel fitness function. Furthermore, about key generation matter, any key is generated in the form of binary code making the exponential power to be two, in contrast to the DNA code exponential power which is 4, making a single bit key eight times stronger.

The bottom line is, the complexity and randomness of DNA structure add an extra layer of security by the mean of cryptography, in addition to its biological capability in high data capacity and parallel processing. From here, the concept of integrating DNA in the field of cryptography has been identified as a possible technology that brings forward a new hope for raising more robust algorithms.

1.3.13 DNA cryptography future

The last few years have witnessed a high leap in this area of DNA cryptography and have seen real progress in applying DNA methodologies into cryptography, and there are a quite number of schemes that perceived the interest of cryptography at the biological level. At present, the work in bio-inspired cryptography, especially from DNA, is focused on applying some technique to encode binary data in the form of DNA sequences. Nevertheless, it still needs much more theoretical and practical implementation; DNA cryptography brings new hope to the future of the information security. The ultimate goal is to enable the discovery of new computational biology insights in addition to creating a global perspective from which unifying principles in biomolecular computing can be discerned.

The rest of this paper is organized as follows: Section 2 lists and categorizes existing cryptography schemes inspired by DNA structure, Section 3 represents current research topics and challenges faced, and the paper is finally concluded in Section 4.

2. RELATED WORK

DNA Cryptography is a new information security branch; it encrypts the information in the form of DNA sequence, making use of its biological properties. In general, existing DNA cryptography techniques use modern biological technologies as implementation tool and DNA as an information carrier. Common biological technologies among recent literature are included DNA Hybridization, PCR amplification, DNA Synthesis DNA digital coding, etc., this section will categorize exertion carried

out in this research area into two main categories, the first is categorized based on the security technique, and the second is categorized based on the algorithm`s inspired procedure.

2.1 Security techniques

Steganography, key generation, and Encryption are the major techniques involved in any cryptosystem. These techniques and their types are shown in Figure 7.

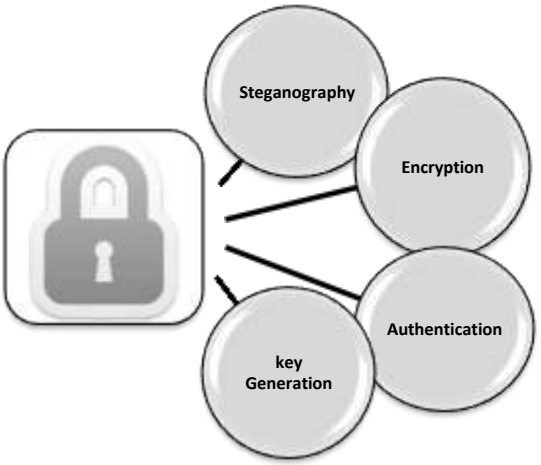

Figure. 7 Security techniques used in cryptography

2.2 Steganography

DNA Steganography is the field of cryptography in which a message is hidden in other messages. DNA can synthesize sequences in any desirable length makes DNA is perfect for data hiding. Therefore, if the advantages of the random trait of DNA could be taken, they can make our cryptography technique in principle invincible.

The principle of DNA steganography is concealing valuable information that needs insurance in a significant number of irrelevant DNA sequence chains. Only the desirable recipient can find the correct DNA fragment based on the conventional information sent and agreed prior the data transferring process. Such a process may not be considered encryption since the plaintext is not encrypted, but it is only disguised within other media.

2.2.1 Encryption

Encryption means converting ordinary information (called plaintext) into an unintelligible form (called ciphertext); decryption is the reverse. The detailed operation of a cipher is controlled both by the algorithm and by a "key," usually a short string of characters. Encryption is the art of protecting data by transforming them into an unreadable form (ciphertext) using a pre-agreed scheme that is publicly available. Only those who possess a secret key can decipher the message back to the plaintext. Since a few years, scientists implemented DNA cryptography using modern biological processes as the tool and DNA as the information carrier for encrypting the DNA information. DNA encryption can be classified into two subfields:

a) *Symmetric Key DNA Cryptography*: In Symmetric key cryptography (also known as secret key cryptography) the receiver and the sender share the same key for encryption and decryption.

b) *Asymmetric Key DNA Cryptography*: Asymmetric key cryptography is also known as the public key cryptography. It uses two keys; one to encrypt the data which labeled as a public key and is distributed freely to everyone and the other to decrypt it which is labeled as a private key, and it is secret to each person and must keep hidden. The two keys are related and generated together.

2.2.2 Authentication

Authentication determines whether someone or something is, in fact, who or what it is declared to be, by adding extra information to the original data. Authentication may be in the form of digital watermarking [14] or in fingerprinting [15]. Digital watermarking is an authentication technique used to authorize the originality of the data by adding extra bits to the original data and designed to be completely unnoticed, Unlike printed watermarks, which are intended to be visible as the reference is the human perception eventually.

2.2.3 Key generation

Key generation is the process of generating keys used in cryptography. The key element in a powerful scheme is the randomness as pseudorandom number generator (PRNG), which is a computer algorithm that produces data that appears random under analysis. Modern cryptography schemes use a combination of symmetric and asymmetric key algorithms since the first tend to be much slower than the last.

2.3 Inspired procedure

DNA cryptography can be implemented using classical computational operations or by processes inspired biologically or a combination of both.

Biologically inspired: Biological operations like PCR, DNA synthesizer, DNA hybridization, DNA fragment assembly, translation, transcription, and splicing can be involved in the process of encryption and decryption. These operations are summarized in Figure 8:

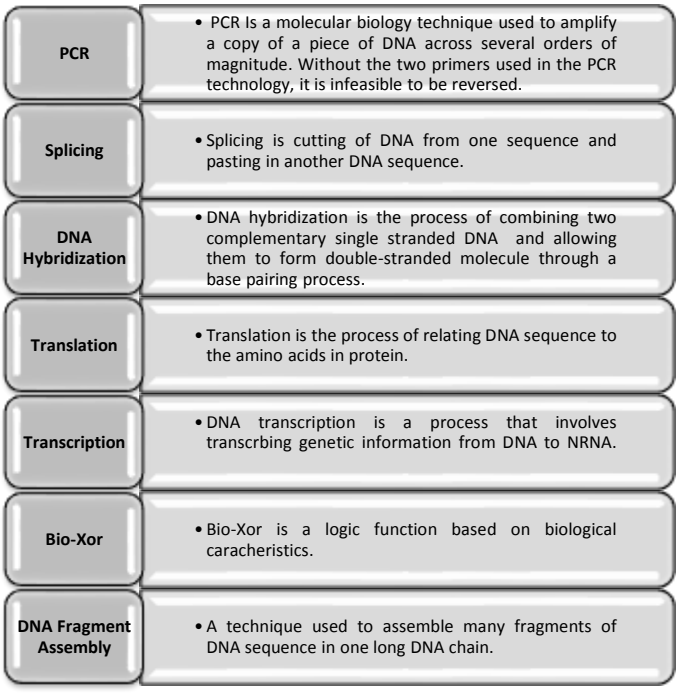

Figure. 8 Biologically inspired procedures involved in the cryptography algorithms

Computationally Inspired: Even computational operations like arithmetical, mathematical, etc., can be involved in the process of encryption and decryption within the DNA cryptography technique [16]. The DNA-type of encryption uses DNA coding scheme and then deals with the DNA sequence as zeros and ones; hence, any arithmetic operation can be applied to produce the encryption algorithm or even uses classical encryption techniques with DNA sequences (see figure 9).

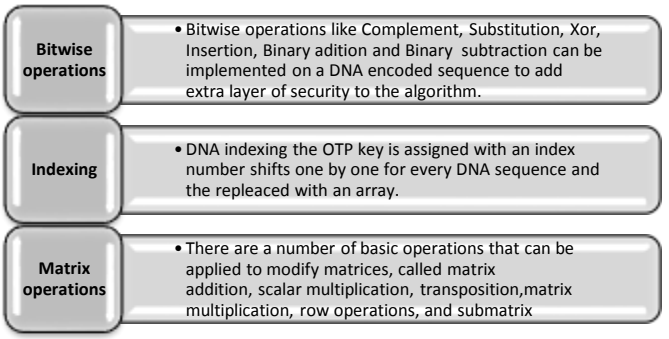

Figure. 9 Computationally Inspired procedures involved in cryptography algorithms

Khalifa and Atito [17] demonstrated a steganography approach using basic biological DNA concepts. The method implemented at two main levels: first a DNA-based play-fair cipher is applied to encrypt the message. The second level uses a two-by-two generic complementary substitution rule to replace the reference to the encrypted DNA. A performance analysis was presented on the hiding capacity as well as robustness against attacks. The proposed technique was tested on different real DNA sequences considering a number of parameters such as time performance and capacity. In conclusion, the proposed scheme can encrypt information into DNA sequences in addition to hiding these data into another reference DNA sequence increasing the security level.

Pramanik and Setua [1] presented a simple encryption method based on DNA sequencing, Watson-Crick complementary rule, and One Time Pad (OTP). A plaintext is first converted to binary by ASCII table conversion, then creating a prefixed length DNA sequence (agreed by sender and receiver) by OTP. The sender scans the binary bits and then the DNA sequence in reverse order, for each one binary bit he takes the WC complementary of the corresponding DNA sequence end, for each 0 binary bit, he does nothing and sends it to the receiver. The technique is distinguished in minimizing the time consumption, although it lacked to the encryption complexity factor and maximize data capacity.

Javheri and Kulkarni [18] proposed a symmetric encryption technique based on mathematical calculations and DNA digital coding. The data is first formed in a matrix; a substitution is performed with a key. The key is generated from another module operation to enhance the complexity, another layer of encryption is added the same way after adding extra information to the cipher to produce *Primary Cipher*. Finally, DNA digital coding is performed to produce the *Final Cipher*. An implementation methodology and experimental results are presented showing an excellence in time over DES. From a cryptographic point of view, the proposed algorithm is eminent, but it lacked to the DNA by depending only on DNA digital coding.

Jain and Bhatnagar [19] proposed a novel symmetric encryption technique based on two security levels; one is based on spiral transposition approach, and the other is based on the DNA sequence dictionary table. Experimental results and analysis showed that the technique overall enhanced security, nevertheless it increases key space complexity.

Wang and Zhang [20] introduced a new way to evolve DNA encryption layer with RSA after converting text into DNA sequence than to numbers and finally using RSA to encrypt the numbers. The technique did not enhance security even using an extra layer of encryption but increased time is consuming as the RSA itself has the highest level of security, in addition, to text only can be encrypted. Vijayakumar et al. [21, 22] proposed an idea to overcome this time limitation by applying the same idea with *Hyper Elliptic Curve Cryptography.*

Cui et al. [23] produced an encryption scheme that is designed by using the concept of PCR amplification as a steganography technique by concealing it to a microdot, in addition to DNA digital coding. The primers and the coding module are used as the key to the scheme.

Clelland et al. [24] showed the capability of performing actual PCR amplification in labs performed to the encrypted data concealed in various mixtures of genomic DNAs from different organisms to block any attempts to use the subtraction technique. The technique also could send individual secret messages to several recipients, a unique set of primers would use to amplify an intended messages to each recipient. The work earns winner of "Junior Nobel Price."

Sabry et al. [25] discussed a significant modification to the old Play-fair cipher by applying it to a DNA-based and amino acids structure. Therefore, these Amino acids pass through a Play-fair encryption process after a pre-processing to the data and converting it to DNA sequence. To bypass the deficit from representing the 20 amino acids to the 25 alphabetic, some additional procedures are implemented. The presented technique enhanced the security of the Play-fair cipher. Performance analysis is presented to evaluate the technique proved the increased security.

Sadeg et al. [26] proposed a new symmetric key block cipher algorithm. The algorithm supports the confusion and diffusion factors. Many of them are computationally inspired as permutation, substitution and XOR operations and many other is inspired biologically as a new bio-XOR technique proposed in this paper. The paper although presented a key generation technique by inventing a Bio-XOR indexing table as a logic function, the paper distinguished by producing a technique that combines biological and computational factors giving the technique more complexity. In addition to the DNA module presented in this work is unpredicted to intruders, the performance analysis showed an increase in time complexity over AES.

Zhang et al. [11] proposed a new symmetric encryption technique based on DNA fragment assembly involved with DNA digital coding. The implanted key is used to hint how the original plaintext was before the fragmentation process. The technique cuts the sequence randomly giving the advance of the difficulty to restore the original text without the key. The following table summarizes DNA-based cryptography research.

Table. 2 DNA-based cryptography

No.	Paper Name	Year	Type	Inspiration	Involved Algorithm – Techniques
1	Hiding messages in DNA microdots [24]	1999	Steganography	Biologically	- DNA digital coding - Microdoting - PCR
2	DNA-Based steganography [27]	2001	Steganography	Biologically	- 2 Primer Keys - DNA Lab - DNA digital coding - Random number generator
3	A Novel Generation Key Scheme Based on DNA [28]	2008	Key Generation	Computationally	- Gene-Bank - Key Expansion Matrix
4	DNA computing based cryptography [20]	2009	Asymmetric encryption	Computationally	- RSA - PCR
5	A Pseudo-DNA Cryptography Method [29]	2009	Symmetric Encryption	Biologically	- Central Dogma
6	An encryption algorithm inspired from DNA [26]	2010	Symmetric Encryption	- Computationally - Biologically	- Transposition - Matrix Permutation - Bio-Xor - Central Dogma
7	DNA Encoding Based Feature Extraction for Biometric Watermarking [2]	2011	Watermarking	Computationally	- DNA Encoding - Biometric watermarking - Discrete wavelet transform
8	Bi-serial DNA Encryption Algorithm (BDEA) [30]	2011	Asymmetric Encryption	- Computationally - Biologically	- PCR - BDEA
9	Index-Based Symmetric DNA Encryption Algorithm [31]	2011	Symmetric Encryption	Computationally	- DNA Encoding - GeneBank - Logistic Map
10	High-Capacity DNA-based Steganography [17]	2012	Steganography	- Biologically - Computationally	- Play-fair - DNA digital coding - Substitution - DNA reference - Complementary rules
11	Integration of DNA Cryptography for Complex Biological Interactions [32]	2012	Symmetric Encryption	- Biologically	- DNA digital coding - PCR - Microdoting
12	DNA Cryptography [1]	2012	Symmetric Encryption	- Computationally	- OTP - DNA hybridization
13	DNA Cryptography Based on DNA Fragment Assembly [11]	2012	Symmetric Encryption	- Biologically	- DNA Digital Coding - DNA Fragment Assembly
14	Three Reversible Data Encoding Algorithms based on DNA and Amino Acids' Structure [33]	2012	Symmetric Encryption	- Computationally - Biologically	- Central Dogma

15	Hardware Implementation of DNA Based Cryptography [34]	2013	- Symmetric Encryption - Key Generation	- Computationally	- AES (modified with DNA key) - DNA code set - OTP
16	A DNA Encryption Technique Based on Matrix Manipulation and Secure Key Generation Scheme [35]	2013	- Symmetric Encryption - Key Generation	- Computationally - Biologically	- Matrix Manipulation - DNA Primers - Central Dogma
17	A Novel DNA Sequence Dictionary Method for Securing Data in DNA using Spiral Approach and Framework of DNA Cryptography [19]	2014	Symmetric Encryption	- Biologically. - Computationally.	- Spiral transposition. - DNA sequence dictionary table
18	Algorithm for Enhanced Image Security Using DNA and genetic algorithm [36]	2015	Symmetric Encryption	- Computationally	- Choastic Function - DNA digital coding - Genetic Algorithm
19	Implementation of DNA Cryptography in Cloud Computing and using Socket Programming [37]	2016	Symmetric Encryption	- Biologically. - Computationally.	- The Bi-directional DNA Encryption Algorithm (BDEA)

Table. 2 Continued

No.	Advantage	Disadvantage	What to enhance	Performance Analysis
1	Implementation simplicity	Security is depending upon the referred DNA sequences that are available on the internet. Microdot increased data capacity	Insert Key generation technique	None
2	Hard to guess according to its biological structure	High tech biomolecular laboratory large capacity	Automated way from data to DNA	None
3	- Improves the independence - Improves the strict avalanche	Increases computation because of the matrix operation	Replace the expansion matrix or reduce the key size	High change in the matrix output by a slight change in the random DNA sequence input
4	Two-level security	Basic biological operations	Increase the security by adding additional mathematical operation	None
5	- Decreased time complexity - reduced cipher capacity	Liable to deferential analysis	Add extra level of security (traditional cryptography)	data recovered successfully
6	- OTP - DNA sequence key - DNA module (bio-Xor) is new and unpredicted	- Key size (Matrix) - time complexity over AES openSSL	Exclude the matrix operation	Outperform over AES but unsurpassed it in open SSL mode
7	- Secret code redundancy - Additional security level	Image encryption only	Add binary encryption	Retrieval without loss
8	- Double Layer Encryption - Asymmetric keys	Increased cipher capacity	- Eliminate PCR amplification - Using compression	None
9	- Reduced encrypted data - Key selected by receiver and sender	Encrypts text only	Add binary encryption feature	High Key change sensitivity
10	Time performance, Capacity, and Payload	Encryption is too simple	Add encryption factors (confusion and diffusion)	Maximum hiding capacity
11	Repeating encryption for the same plaintext will produce a different ciphertext	Rudimentary operations Lacked to DNA encryption factors	Add DNA encryption factors	None
12	- OTP - Minimizes time complexity	- No encryption used. - No DNA encryption involved. - Security depends on upon a key only. - No analysis	Add an extra layer of encryption or involve encryption technique	Security is strong as OTP
13	The cutting is very random	- Increased computation	Substitute the fragment assembly stage	Some errors in the stage of overlap and layout
14	- Serve in biological experiments and DNA computing	- Doesn`t include secret key - Encode English characters only	Insert key generation technique	Encoding is reversible without data loss
15	- Encryption and key generation techniques - Primers increase security. - AES modified increases security	- Text only (64). - The AES modified is not illustrated. - DNA encryption is poor	Extend the 64 DNA set table to cover 256 ASCII table characters.	Increased security.
16	- Different cipher for same key and same data every time	- Encrypts text only - Using matrix manipulation increases time complexity	- Substitute matrix operation	

17	- Increased security at two levels of security. - Binary encryption	- DNA sequence dictionary table increases key space complexity. - Cipher-text is too large. - Binary security level has no key; procedures have to be sent to receivers	- Key generation technique in binary level security - Reduce DNA sequence dictionary table by a key	Reduced complexity
18	- High entropy - Low correlation - sensitive to changes	- Image encryption only	Add binary encryption capability	Retrieval without distortion
19	- Extend encryption to BDEA encryption technique over Unicode characters	- Not Applicable to images and other data types	Applicate on binary data to encrypt images and another data type	Applied and worked on real world web server

3. CURRENT RESEARCH TOPICS AND CHALLENGES

Current research topics are concentrated on two main directions, *one* is on enhancing the security of existing cryptographic techniques by adding an extra layer of security using DNA characteristics and the main challenge in this direction is that the produced mechanism is not applicable to all data types and confined only in particular to images or ASCII code characters. *The second direction* is not to negate the classical but to extend the current cryptographic techniques to apply to new technologies like DNA Computing, and the main challenge in this direction that the classical mechanisms are designed to be applied to the binary data type in the first place.

Researchers concerned involving the concept of DNA structure and bio-computing in their work, by attempting to stimulate biological processes and manipulate them by different means regardless of the main factor of an ideal cryptographic scheme lies on space complexity and time complexity. A common lacking can be analysed as many researchers attempt implementing data encryption on one data type only avoid dealing with other data types. Another common lacking that researchers attempt to add an extra layer of security to a classical encryption technique not accounting time complexity and vice versa. In particular, steganography researches seem to hide specific data in enormous DNA microdots making it harder to storing and transmitting, key generation techniques has no backbone structure for generating and sharing, encryption techniques represented showed that researchers tried to adding an extra layer of security by using DNA structure while failing to preserve processing time factor.

4. CONCLUSION

Combining the new fields of DNA computing in addition to the conventionally used encryption algorithm, moreover, mathematical operation with biological operations in order to increase the concept of confusion, this combination will produce a more robust, more secure algorithm that is hard to decipher without the key; this is DNA cryptography. DNA cryptography field is confronted with some obstacles at the infrastructure level as theoretical problems and difficult implementation as explained in this work. Authors focused on these obstacles aiming to make researchers to take advantage of painstaking effort expanded so far in the DNA cryptography research area and make researchers utilize the analysis and benefits of recent works and try to bypass the limitations found in it. The authors compared the various DNA cryptographic techniques. These parameters would also help the future researchers to design and improve the DNA storage techniques for secure data storage more efficiently and in a reliable manner.

REFERENCES

1. Pramanik, S. and Setua, S. (2012), "*DNA Cryptograph,'*" 2012: Proceedings of 7th International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, pp. 551 – 554.
2. Arya, M.S., Jain, N., Sisodia, J. and Sehgal, N. (2011) *"DNA encoding based feature extraction for biometric watermarking*," 2011 International Conference on Image Information Processing (ICIIP), vol., no., pp.1-6.
3. Leech, D. and Chinworth, M. (2001) *"The Economic Impacts of NIST's Data Encryption Standard (DES) Program,"*, Strategic Planning and Economic Analysis Group, Planning Report 01-2.
4. Jacob, G. and Murugan, A. (2013) *"DNA-based Cryptography: An Overview and Analysis*" International Journal of Emerging Sciences (IJES), vol. 1, pp. 36-42.
5. Adleman, L. (1994) "*Molecular computation of solutions to combinatorial problems*", Science, 266(5187), pp. 1021–1024.
6. Chen, J. (2003) "*A DNA-based, biomolecular cryptography design*" Circuits and Systems, ISCAS '03. Proceedings of the 2003 International Symposium on, vol.3, pp. 822-825.

7. Boneh, D., Dunworth, C. and Lipton, R. (1995), "*Breaking DES Using a Molecular Computer"*, Department of Computer Science, Princeton University, USA, Technical Report CS-TR-489-95.
8. Gearheart, C., Rouchka, E. and Arazi, B. (2012) *"DNA-Based Active Logic Design and Its Implications"* Journal of Emerging Trends in Computing and Information Sciences, VOL. 3, NO. 5.
9. Ogihara, M. and Ray, A. (1996) *"Simulating Boolean circuits on a DNA computer,"* Technical Report 631, University of Rochester.
10. Stefan, L. (2003) *"Computer Made from DNA and Enzymes"*, http://news.nationalgeographic.com/news/2003/02/0224_030224_DNAcomputer.html (Accessed 20 Jan 2017).
11. Zhang, Y., Fu, B. and Zhang, X. (2012) "*DNA cryptography based on DNA Fragment Assembly"* 8th International Conference on Information Science and Digital Content Technology, pp.179-182.
12. Shamra, L., Bhupendra, K. and Ramgapol, S. (2012) "*Breaking of Simplified Data Encryption Standard using Genetic Algorithm*" Global Journal of Computer Science and Technology, Vol. 12.
13. Bhateja, A. and Kumar, S. (2014) "*Genetic Algorithm with elitism for cryptanalysis of Vigenere cipher*" International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, 2014, pp. 373-377.
14. Hamad, S.; Khalifa, A., (2013) *"Robust blind image watermarking using DNA-encoding and discrete wavelet transforms"* 8th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, pp. 221-227.
15. Ghany, K., Hassan, G., Hassanien, A., Hefny, H., and Schaefer, G., (2014) "*A hybrid biometric approach embedding DNA data in fingerprint images"* International Conference on Informatics, Electronics & Vision (ICIEV), pp.1-5.
16. Rakheja, P. (2011) "*Integrating DNA computing in international data Encryption algorithm "IDEA,"*" International Journal of Computer Applications, 26(3), pp. 1–6.
17. Khalifa, A. and Atito, A., (2012) "*High-capacity DNA-based steganography"* 8th International Conference on Informatics and Systems (INFOS), Cairo, pp. 37-43.
18. Javheri, S. Kulkarni, R., (2014) "*Secure Data Communication and Cryptography based on DNA-based Message Encoding*," International Journal of Computer Applications, pp. 35-40.
19. Jain, S. Bhatnagar, V., "*A novel DNA sequence dictionary method for securing data in DNA using spiral approach and framework of DNA cryptography"* International Conference on Advances in Engineering & Technology Research (ICAETR), pp. 1-5.
20. Wang, X. Zhang, Q. (2009) "*DNA computing-based cryptography"* 4th International Conference on Bio-Inspired Computing, pp.1-3.
21. Vijayakumar, P. Vijayalakshmi, V., and Zayaraz, G. (2011) "*DNA Computing based Elliptic Curve Cryptography"* International Journal of Computer Applications, pp. 18-21.
22. Vijayakumar, P. Vijayalakshmi, V., and Zayaraz, G. (2013) "*Enhanced Level of Security using DNA Computing Technique with Hyperelliptic Curve Cryptography*" ACEEE International Journal of Network Security, Vol. 4, No. 1.
23. Cui, G. Qin, L. Wang, Y. and Zhang, X. (2008) "*An encryption scheme using DNA technology*" 3rd International Conference on Bio-Inspired Computing: Theories and Applications (BICTA), pp. 37-42.
24. Clelland, C., Risca, V. and Bancroft, B. (1999) "*Hiding messages in DNA microdots*" in Nature: Vol. 399, pp. 533-534.
25. Sabry, M. Hashem, M. Nazmy, T. and Khalifa, M. (2010) "*A DNA and Amino Acids-Based Implementation of Play-fair Cipher,"* International Journal of Computer Science and Information Security, Vol. 8, No. 3.
26. Sadeg, S., Gougache, M., Mansouri, N. and Drias, H. (2010) "*An encryption algorithm inspired from DNA"* International Conference on Machine and Web Intelligence (ICMWI), vol., no., pp.344-349.
27. Bancroft, F., and Clelland, C. (2001) "*DNA-BASED STEGANOGRAPHY*" United States Patent 6312911.
28. Xin-she, L., Lei, Z., and Yu-pu, H. (2008) "*A Novel Generation Key Scheme Based on DNA"* International Conference on Computational Intelligence and Security, USA, vol.1, no., pp.264-266, 13-17.
29. Ning, K. (2009) "*A Pseudo-DNA Cryptography Method*", Cornell University Libarar: http://arxiv.org/abs/0903.2693 (Last accessed on 20/01/2017).
30. Prabhu, D., Adimoolam, M. (2011) "*Bi-serial DNA Encryption Algorithm (BDEA)*" Cryptography and Security, arXiv:1101.2577.

31. Yunpeng, Z., Yu, Z., Zhong, W. and Sinnott, R. (2011) "*Index-based symmetric DNA encryption algorithm"* 4th International Congress of Image and Signal Processing (CISP), Shanghai, pp. 2290-2294.
32. S. Dhawan and A. Saini, (2012) "*Integration of DNA Cryptography for complex Biological Interactions*" International Journal of Engineering, Business and Enterprise Application, pp. 121-127.
33. Sabry, M., Hashem, M, and Nazmy T. (2012) "*Three Reversible Data Encoding Algorithms based on DNA and Amino Acids Structure*" International Journal of Computer Applications, pp. 24-30.
34. Naveen, J.K.; Karthigaikumar, P.; SivaMangai, N.M.; Sandhya, R.; Asok, S.B., (2013) "*Hardware implementation of DNA-based cryptography*" International Conference of Information & Communication Technologies (ICT), pp.696-700.
35. T. Mandge and V. Choudhary, "*A DNA encryption technique based on matrix manipulation and secure key generation scheme*," International Conference on Information Communication and Embedded Systems (ICICES), Chennai, pp. 47-52.
36. Saranya, M.R.; Mohan, A.K.; Anusudha, K., (2015) "*Algorithm for enhanced image security using DNA and genetic algorithm*" International Conference on Signal Processing, Informatics, Communication, and Energy Systems (SPICES), pp.1-5.
37. Prajapati, B. and Barkha, P., (2016) "*Implementation of DNA cryptography in cloud computing and using socket programming*" International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, pp. 1-6.

AUTHORS PROIFLES