



# A New DNA-Based Approach for Enhancing AES using Sequence Alignment and Dynamic Multi Modeling Techniques

Abdulla Mahmoud  
Information Technology Dept.  
FCIS, Mansoura University,  
Mansoura, Egypt.  
Abdullaelsaied@yahoo.com

Mohammed Elmogy  
Information Technology Dept.  
FCIS, Mansoura University,  
Mansoura, Egypt  
melmogy@mans.edu.eg

Ahmad Atwan  
Information Technology Dept.  
FCIS, Mansoura University,  
Mansoura, Egypt  
atwan@mans.edu.eg

Hassan Soliman  
Information Technology Dept.  
FCIS, Mansoura University,  
Mansoura, Egypt  
hsoliman@mans.edu.eg

## ABSTRACT

The pace of growth in information technology stimulates the information security field to rival to keep up at the same pace. One of the most common encryption techniques is the Advanced Encryption Standard and, despite its good reputation, the analysis in this work showed some limitations in its procedures. Indeed, as a response to this scarcity, many approaches have been proposed, and many other exertions have been spent. One promising consequence is bio-inspired cryptography that emerged with the progress of the DNA computing field. In this paper, a new enhanced method will be proposed using properties of bio-inspired cryptography, bioinformatics, and conventional cryptography to produce a more complex, robust, and less prone to attacks procedures. The contribution presented in this work can be epitomized in three main aspects. First, making the AES internal steps dynamic and key dependent. Second, modifying the Shift-Row step in the AES by involving the Sequence Alignment algorithm. Third, a key exchange technique inherited from the process of DNA fingerprint. The presented algorithm has been built on a DNA basis to demonstrate the capability of applying such a complex system on a promising biological environment such as a molecular computer. It must be clarified that the perception is not to negate or replace the mathematical and theoretical basis of the traditional cryptography, but to combine it with the sophisticated biological characteristics to produce a more powerful cryptographic technique, also to create a bridge between the existing and the new technologies. The experimental analysis showed remarkable enhancement in testing the modified Shift-Row function detached and the encryption strength added by the multi-model layer. Theoretically, the overall modified algorithm showed a significant advance over the classical AES in the resistance against linear and differential cryptanalysis which reveals a promising future of bio-inspired cryptographic techniques.

## General Terms

Security, Cryptography, AES, Bioinformatics, DNA

**Keywords:** Bio-inspired Cryptography, Advanced Encryption Standard, DNA-Based Cryptography, Encryption, Security.

## 1. INTRODUCTION

A cryptographic system is a combination of a key generation technique and a set of encryption procedures. A cryptographic system applies these encryption procedures to the information using the encryption key to produce an encrypted output, which will be meaningless to anyone but the intended recipient who has the decryption key. Knowledge of the key used in the cryptosystem is essential for decryption [1]. The intrinsic idea of a powerful cryptosystem is to produce the key in a one-way function. A function called a one-way function if it is easy to compute  $f(x)$  for each  $x$  in the domain  $f$ , essentially for all the values in the domain range, and it is computationally infeasible to invert to find  $x$  such that  $y = f(x)$ . Giving some extra information (the trapdoor information), it becomes attainable to compute  $y$  value, such that  $y = f(x)$  [2], this trapdoor information is given only to the intended recipient.

Cryptographic challenges lie on the factor that currently used cryptographic schemes are rely on computational hardness assumptions. In many cases, these assumptions are misplaced, which raises a concern, what if the current encryption technologies have been penetrated but still in an undeclared way? what if a breach has been suddenly discovered with no chance or time to rectify? While fully relying on it in all our financial transactions and personal information. Further, The vast development of new computer paradigms as Quantum Computing and Molecular Computing constitutes a challenge to current cryptographic mechanisms [3].

DNA computing is an emanating approach of computing in which DNA and molecular biology are used instead of traditional silicon technologies. DNA is a molecule in the organisms' body that carries the genetic instructions used for functioning and developing [4]. It is working as a storage medium for genetic information. For every living cell, DNA represents the basic storage medium, and its primary function is to absorb and transmit biological instructions. The DNA is made from persistent subunits, called nucleotides [5]. About  $10^{11}$  of DNA molecules could be confined into a space of a small marble size [6]. Therefore, theoretically in a DNA Computer, DNA molecules are used as a storage medium, and biological functions are used to implement the desired operations [7]. As the core idea and component of the classical computer is the transistor, the core component of molecular computing is DNA molecules [8]. More precisely, in a silicon-based computer, the basic function is logic functions. The logic function handles two electrons as inputs and extracts one electron as output based on pre-specified rules. By the same tactic, scientists discovered a similar process to the classical logic function in biological computers. The electronic computers use digital signals known as binary digits (represented as 0 and 1), whereas a DNA strand uses biological molecules (expressed in 4 alphabets characters A, C, T, and G), which can ingest much more data than its predecessor. Primarily, the centrality has been on the fulfillment of aspects of biological systems in computational devices. Examples are artificial neural networks and genetic algorithms. Since Adleman's first experience [9], advances have been made: DNA logic gates. SAT problems [10] and DNA storage device [11].

The main priority of DNA computation over conventional silicon-based machines are miniaturization and parallelism. A DNA computer can perform an enormous number of operations simultaneously [12], which allow the processing rate of DNA strands to increase exponentially, introducing a nascent data structure and calculating method [5], other than that these operations are handled by photons instead of electrons, which is extremely

faster and has a low power consumption [13]. The high storage capacity and the incredibly lightweight are also key advantages of DNA computing, as DNA has an enormous density, which proved its capability of high information storage density, moreover according to variant calculations, a single DNA gram could potentially have the capacity to hold data up to 512 Exabytes [13]. DNA has the potential for long-lived storage as it is considered as an imperishable storage medium [13].

Theoretically, DNA computing can attack different regions of the computing paradigm simultaneously, putting forward challenges to conventional information security[14]. Many proposals have been submitted for breaking traditional cryptosystems using new computer paradigms as DNA computing. It has indicated that conventional cryptosystems were perhaps insecure [15].

Challenges confronting DNA computing may include the slow processing speed on a DNA-computer regarding implementing a single process, making the system take the same time to solve a complex calculation equal to a simple one. Also, it is much harder to analyze the answers given by a DNA-computer than by a digital one [11].

From a cryptographic point of view, DNA can be very effective, using complex biological characteristics as the main security foundation and contemporary biological techniques as implementation tools [14] for maximum utilization of the innovated features. The potential power in DNA based computing paradigm will strengthen the current cryptosystems by opening the way for a hybrid cryptographic system using bio-inspired processes [16]. The development of a new computing paradigm requires adapting the existing encryption techniques to have the capability to be applied to a biological environment. Moreover, biological processes have potential characteristics to make existing encryption techniques more robust. From this point on, the researchers contend to adapt existing encryption techniques adaptable to a promising biological environment, also the researchers inspired by biological characteristics to make existing encryption techniques more powerful.

The rest of the paper is organized as follows: Section Two discusses the basic concepts of the techniques involved in the research. Section three discusses the related work in the research area, their contributions in the field, limitations, and advantages are also discussed. Section four discusses the proposed mechanism in detail. Section five discusses the experimental results and analysis. Finally, section six discusses the conclusion and future work.

## 2. BASIC CONCEPTS

This section discusses the basic concepts related to the research area in a brief attitude to simplify the proposed idea. The involved algorithms and techniques are also discussed to analyze their limitations to illustrate the reason that inspired the authors to make their contribution.

### 2.1 AES Design

AES is based on a design principle known as a substitution-permutation network, and it is proved to be fast in both hardware and software implementations. Unlike its predecessor DES, AES isn't based on a Feistel network [17]. AES is a diverse of Rijndael cipher, which has a key size of 128, 192, or 256 bits with 10,12 and 14 rounds respectively, with a fixed block size of 128 bits. AES operates on a 4x4 matrix order of bytes, termed the state. Most AES operations are calculated in a special finite field. The key size used in AES specifies the number of transformation rounds repetition that transforms the input, called the plaintext, into the final output, called the ciphertext [17].

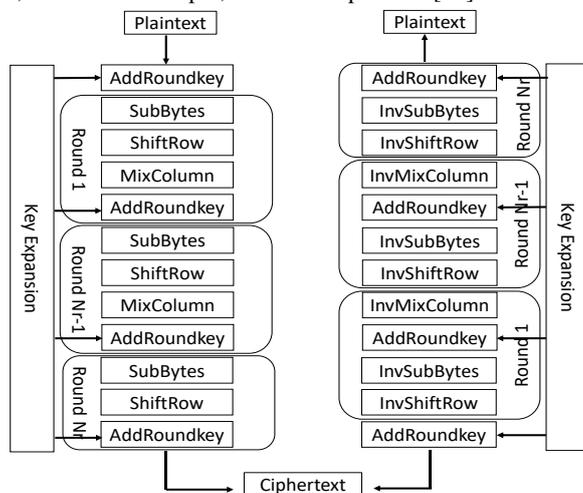


Fig 1 AES overall Procedures [18]

The binary form of plaintext is divided into  $N_b$  bits blocks [STATE]; each block goes through  $N_r$  rounds. At each round ( $N_r$ ), the plaintext is manipulated through four operations to be encrypted. A key scheduler expands the main key educing a unique key for each round with the same block size.

Pesoude code for 128 AES [17]:

**Algorithm 1:** AES-128 (Round Operation)

- 1: create an input variable
- 2: input ← Original Data (PT)
- 3: store PT in the input variable
- 4: convert PT to Binary (B)
- 5: if  $B \bmod 128 = 0$
- 6: divide B into 128 segments S
- 7: else
- 8: padding and divide into 128 Segments S

- 9: end if
- 10: create round 1 input (n1) by Xor S with 128-bit key (K)
- 11: store n1 in the output variable
- 12: create matrix m
- 13: input← (n1)
- 14: store n1 in the matrix (m1) as an input variable
- 15: create S-box (ip)
- 16: create (sn1) from (m1) by performing substitution with (ip)
- 17: create (sh1) from (sn1) by shifting rows
- 18: create a state (st)
- 19: create (mc1) from (sh1) by mixing with the state (st)
- 20: create the round key (k1) from the 128-bit key (K)
- 21: Xor (mc1) with round key (k1) to create x1
- 22: create an output variable
- 23: store x1 as the output variable
- 24: end

The detailed operations of the rounds are as follow:

### 2.1.1 AES Round Operations

- Add round key (Introductory step):

Before the repetitive rounds take place, an initial step is performed to ensure the most utilization of every single step of the algorithm. As shown in figure 1 [19], the rounds start with manipulating the plaintext with a fixed substitution box called SubBytes, and without the introductory step, the SubBytes operation will be useless. The preparatory operation is called “key whitening”.

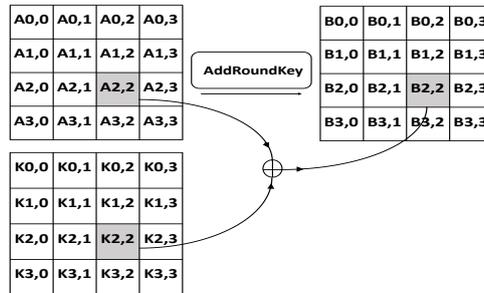


Fig 2 AES Add Round Key Step [19]

- Repetitive Steps:
  - SubBytes step:

In the SubBytes step, the 16 input bytes are substituted by a fixed lookup table included in the design called S-box. The result is a matrix of four columns and four rows as shown in figure 3 [17].

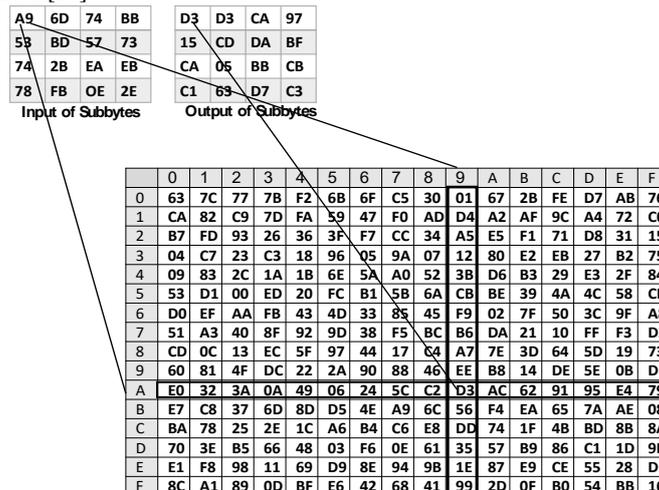


Fig 3 AES SubBytes Step [17]

- ShiftRows Step:

Each row of the 4\*4 matrix is shifted starting from the second row by 1,2 and 3 shifts respectively to the left. Any entries that ‘dropped off’ are re-inserted to the right side of the same row.

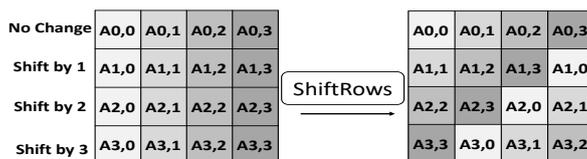


Fig 4 AES Shift Row Step [20]

o MixColumn Step:

The output of the ShiftRow Step is transformed using a special mathematical function. By taking as input the four bytes of each column and outputs four completely new bytes finally to produce a new 16-bit matrix as shown in figure 5 [21]. It should be noted that this step is missing in the last round to maintain the symmetry between the encryption and decryption process.

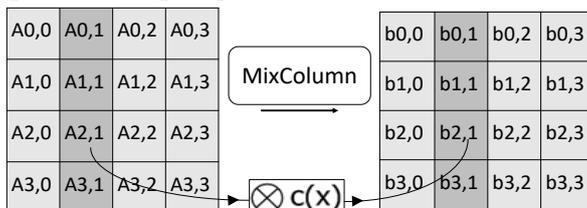


Fig 5 AES MixColumns Step [21]

• AddRoundKey:

The 16 output bytes of the MixColumn step are now reconsidered as 128 bits and are XORed with the 128 bits round key [19]. The resulting 128-bit output is interpreted as 16 bytes again through another similar round unless if this is the last round, the result is the ciphertext

2.1.2 Security of AES

The security of an encryption algorithm is measured based on its resistance against known attacks. AES has proved to be efficient against attacks except:

- Related-key attacks: performed on reduced-round versions and 192-bit, 256-bit versions. The attack outdoes the exhaustive search attack. The Related-Key attack is based on the idea of knowing how the plaintext is being encrypted with more than one key [22].
- Side-channel attacks: Which doesn't attack the AES algorithm structure itself, but attacks its implementation with a power differential analysis.

2.2 Sequence alignment

In bioinformatics, a sequence alignment is a method of aligning two or more sequences of DNA to identify regions of proximity that may be a consequence of functional or evolutionary relationships between the sequences. Sequence alignments are also used for non-biological purposes, such as scoring the edit distance cost between strings.

2.2.1 Alignment Methods:

Very short or very similar sequences can be aligned by the human factor. Nevertheless, in most cases, the problem requires the alignment of highly variable, lengthy sequences that cannot be aligned merely by human effort. Even so, human knowledge is applied to construct a high-quality sequence alignments algorithms.

- Global and Local Alignment
  - o Global alignments, attempt to align every fraction in the given sequences, they are most useful when the sequences in query set are nearly similar and of nearly equal size. A general global alignment technique is the Needleman–Wunsch algorithm [23], which is based on dynamic programming.
  - o Local Alignments are more useful when the sequences in the query set are dissimilar and are suspected of containing regions of similarity within their larger sequence context. The Smith-Waterman algorithm [24] is a general local alignment method based on the same dynamic programming scheme with the additional capability to choose the start and end place.
- Pairwise Sequence Alignment
 

Pairwise sequence alignments are methods of aligning two query sequences to find the best-matching piecewise (local or global). Pairwise sequence alignments are often used when the query does not require extreme precision (such as searching a database for sequences with high similarity to a query). The three primary methods of producing pairwise alignments are dynamic programming, dot-matrix methods, and word methods.

3. RELATED WORK

Modern cryptographic techniques are built strong mathematically and theoretically, this may not be the problem at the time being, but soon it might be, given the growth of the computing power and new technology invention, as the DES was considered unbreakable twenty years ago. As 2014 was “The Year of Data Breaches”, 2015 is off to a fast start with several prominent data breaches. As per statistics gathered by ‘BreachLevelIndex’ [25]. It is clear that researchers are in urgent need of a new data security approach if organizations want to stay ahead of attackers to protect their important information more effectively. These elements stimulate the researchers' interest in how to take advantage of the biological processes characteristics and the potential advances in DNA based computing to produce more sophisticated encryption techniques. The researchers also need to develop a new data encryption technique to be applicable to a new computing paradigm as bio-computers. To fulfill the gap, many approaches have been proposed. Sabry et al. [16] designed a DNA-based Advanced Encryption Standard (AES) with all its implementations. The authors built their algorithm with all its specifications (data, functions, and algorithms operations) on a DNA basis instead of bits. The authors aim to prove the ability to build such a

complex system on a DNA basis in the way of making it a suitable candidate for implementation in a biological environment or on DNA computers. The proposed mechanism maintains the same security strength and robustness of the standard algorithm.

Pramanik and Setua [5] proposed a computationally inspired symmetric encryption technique. The technique characterized in time consumption minimization using a technique implemented in parallel order, and the technique is based on DNA molecular structure, OTP, and DNA hybridization. The key generation technique is based on OTP. The proposed technique overcomes the limitation of encryption processing cost by implementing a technique characterized in parallel processing.

In the same year, Audey et al. [26] presented a new approach by generating key-dependent shift rows transformation for enhancing the security of the AES algorithm. The approach employs methods inspired by DNA processes and structure. The reverse complement process will be used for altering the Shift-Rows step by making it to be key-dependent, the cipher key  $K_r$  at round  $r$  is used as the key for applying the reverse-complement over state bytes (byte level). This technique makes the byte transposition process dynamic rather than static, while the newly created Shift-Rows have identical characteristics to the original algorithm, in addition to the resistance against attacks has been increased. The authors stated that a coefficient correlation test had been performed for static and dynamic independence validation. Also, they used the NIST test suite to test the randomness of the new transformation.

Pushpa [27] presented a new encryption technique based on DNA sequence reference in which both sender and receiver will share a common DNA sequence for the encryption and the decryption process. The authors aimed to produce a more secure encryption technique by hiding the data in a DNA form. The proposed technique converts any type of data into a binary form then into a DNA sequence form according to a DNA coding table, complementary role will then applied to the DNA sequence. The DNA reference sequence is assigned with an index according to a fixed table. The ciphertext is represented as an equivalent assigned number of the DNA sequence. The DNA sequence reference is considered as the main security backbone and sent over a secure communication to ensure the quality of the encryption process.

Bahubali and Latha [28] proposed a two-stage security approach for enhancing classical encryption techniques. The authors contemplate adding an extra layer of security by adding a random key generation technique based on Dynamic DNA coding. The proposed technique generates a random substitution table called the Key Combination table in which the binary data have an equivalent value in a DNA sequence. The generated Key Combination table will be securely transferred to the receiver for the decryption process. The authors did not manipulate the characteristics of classical encryption, but only proposed to add an extra layer of security to the traditional encryption techniques. Also, the proposed technique can handle any data type by converting it into binary data and then into the DNA sequence.

Bahig and Nassr [30] also proposed a DNA based AES by manipulating the functions of the classical AES to deal with DNA sequences besides its ability to encrypt any type of data. The proposed technique aims to make the classical technique applicable to a biological environment. The technique adds the feature of hiding data in the DNA form. The proposed technique also has the same security level as the classical AES.

Challita and Bakhache [31] presented an enhanced version of Simplified AES (S-AES) by using PWLCM (Pair Wise Linear Chaotic Map). The proposed enhanced algorithm is particularly adequate for use in small and low power networks such as medical devices. The authors proposed modifying the ShiftRows and SubBytes functions by adding some extra characteristics based on a chaotic generator. The proposed scheme bypass the limitation of the static procedures for the ShiftRow and SubBytes functions by making them depending on the initial state of the chaotic generator.

Undoubtedly, many researchers have concerned about utilizing molecular computing and biological processes in the field of cryptography. Even so, there is no significant impact on cryptography has been perceived. This shortcoming could be due to no work has focused on overcoming the limitations of traditional cryptography or tried to exploit the characteristics of biological processes to produce more sophisticated algorithms. Many of them have involved their contributions in encryption techniques that no longer in use; even more, most of them didn't present the final effect of their contribution mathematically to prove how their contribution improved the technique. These shortcomings urged the research to overcome these limitations to achieve the maximum utilization of the characteristics inherited from the biological processes and to produce an encryption technique applicable to a biological environment. The proposed work in this paper overcomes these limitations by selecting an on-duty encryption technique, and also the most secure and most widely used Encryption Technique (AES). Analyzing AES flaws and proposes an amendment by using bio-inspired processes and DNA sequencing to inject the solution in the right flaw to ensure maximum optimization.

The limitations of the used research methods are confined to the lack of the literature research with regard to the DNA computing field and the related DNA based cryptography technique as a result of that the DNA-based PC does still not realize, making some of the proposed contributions tend to be theoretical and hypothetical more than practical real-time representations.

#### 4. PROPOSED MECHANISM

The proposed scheme is composed of 3 stages; each stage exhibits some procedures to prepare the data for the next stage. First, data need to be preprocessed; the pre-processing stage takes the original data and produces a stream of 128-bits sequence without any manipulating. The bitstream of data will then transformed in the form of a DNA sequence by DNA Digital Coding. The outcome of the preprocessing stage comes in turn to be input to the Dispersal Stage in which the data passes through several substitution and permutation procedures, the same as in the conventional AES algorithm with an additional feature of the mode selection process. The mode selection layer adds an extra layer of security by making the order of the round four internal steps dynamic rather than static based on the last 3 bits of the round key. Subsequently, the encrypted DNA sequence will go through another layer of encryption level verification named the Confirmation Stage to validate the irrelevancy between the input and the output. The stage uses sequence alignment to check the similarity level between the input of the dispersal stage and its` output. If the similarity result is higher than a certain threshold, the data go through another dispersal stage until the output achieves the desired dissimilarity level. The data go through another layer of encryption called the Reverse Sequence Alignment to perform an additional similarity reduction. The presented Reverse Sequence Alignment algorithm will manipulate the output data from the similarity check process to deface any relation remained after the threshold set. After the Reverse Sequence Alignment process, the final encrypted output will be produced. It must be addressed that all the round internal operations will be done in a DNA sequence form [16].

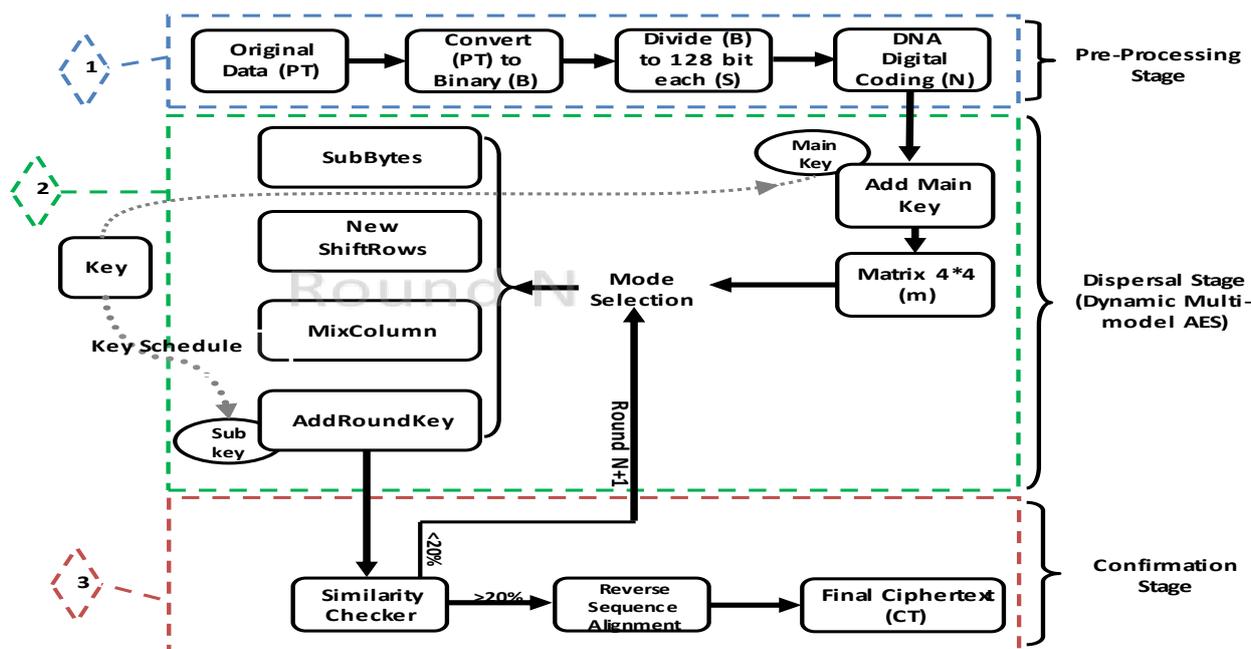


Fig 6 The proposed mechanism flowchart

### 4.1 Pre-Processing Stage

Similar to the classical AES, the proposed technique can treat any type of data, thanks to the preprocessing stage in which any type of data will be handled to produce data unified in type and length. The stage takes the original data in any form to generate a 128-bit binary sequence through first converting the data (PT) into binary code (B) then divide this code into a unified 128 segments. Unlike AES, the proposed technique has an additional step to convert the binary stream of data into a DNA sequence by DNA digital coding as shown in table 1.

Table 1 DNA digital coding table

Bit 1	Bit 2	DNA
0	0	A
0	1	C
1	0	T
1	1	G

#### Algorithm 2: Pre-Processing Stage

- 1: create an input variable
- 2: input ← Original Data (PT)
- 3: store PT in the input variable
- 4: convert PT to Binary (B)
- 5: if  $B \bmod 128 = 0$
- 6: divide B into 128 segments S
- 7: else
- 8: padding and divide into 128 Segments S
- 9: end if
- 10: create DNA digital coding table (DC)
- 11: convert S to equivalent DNA sequence N from (DC)
- 12: create an output variable
- 13: store N in the output variable
- 14: end

### 4.2 Dispersal Stage

The main objective of the dispersal stage is to achieve the confusion and diffusion factors required for idealistic encryption, which is almost achieved in the classical AES. The 128 bitstream output of the preprocessing stage is treated as input in the dispersal stage but all manipulated in the form of DNA sequence [16]. First, a key-whitening preparatory step will be performed to ensure the maximum utilization of the algorithm's internal steps by adding the round key to the data by an Xor bitwise operation. The Xor operation in the form of a DNA sequence is called Bio-Xor. The output of the Bio-Xor operation is formed in a 4\*4 matrix (M). By default, ten repeated rounds of substitution, permutation, mix columns, and bitwise operations come after. The user also has the ability to set the number of rounds up to 10 rounds based on the value of the information in hand taking into account the desired encryption time, as more rounds will produce more powerful encryption but on time complexity.

#### Algorithm 3: Dispersal Stage

- 1: create an input variable
- 2: input ← Data (N)

- 3: store N in the input variable
- 4: Xor N with the main key (Km) to generate (DK)
- 5: create matrix M
- 6: input ← (DK)
- 7: store DK in the matrix (M) as an input variable
- 8: select mode
- 9: create S-box (IP)
- 10: perform the substitution on (M) with S-box (IP)
- 11: create (SH) from (IP) by applying the modified ShiftRows
- 12: applying the reverse sequence alignment algorithm on (SH) and (IP) to produce (SA)
- 12: create a state (ST)
- 13: create (MC) from (SA) by mixing with the state (ST)
- 14: generate a round key (K1) from (Km)
- 15: Bio-Xor (MC) with round key (K1) to create X1
- 16: store X1 as an output variable
- 17: end

To assure a fully achieving of the confusion and the diffusion factors, the proposed technique provides a multi-mode feature, in which the order of the internal round operations will be manipulated based on the last 3 bit of the round key (table 2). The additional multi-mode feature will produce a 128 bit key with 8 different combinations ( $2^3$ ) for each of the 10 rounds as the mode is based on the round key making the mode being altered for each round increasing the level of security to the power of  $2^{128 \times 3 \times 10}$ .

**Table 2:** The last 3 bits of the 128-bit key and their corresponding mode.

Last 3 bits of AES 128 Round Key	Mode selection
000	Mode 1
001	Mode 2
010	Mode 3
011	Mode 4
100	Mode 5
101	Mode 6
110	Mode 7
111	Mode 8

The difference between one mode and the other is in the arrangement of the internal procedures, with eight different modes (table 3). The proposed algorithm has the power of eight different encryption algorithm combined, as the intruder has to obtain which mode has been behaved for each round, these last 3 bits are not limited to the mode selection process but also used for the encryption and the decryption process as in the classical AES for maximum utilization of the encryption key.

Table 3: Multi-model product selection.

MODE	ROUND PROCEDURES ORDER			
	Step 1	Step 2	Step 3	Step 4
CLASSIC AES	SubBytes	ShiftRows	MixColumns	AddRoundKey
MODE 1	SubBytes	MixColumns	ShiftRows	AddRoundKey
MODE 2	AddRoundKey	SubBytes	MixColumns	ShiftRows
MODE 3	MixColumns	SubBytes	ShiftRows	AddRoundKey
MODE 4	ShiftRows	AddRoundKey	SubBytes	MixColumns
MODE 5	SubBytes	AddRoundKey	MixColumns	ShiftRows
MODE 6	AddRoundKey	SubBytes	ShiftRows	MixColumns
MODE 7	MixColumns	ShiftRows	AddRoundKey	SubBytes
MODE 8	ShiftRows	MixColumns	AddRoundKey	SubBytes

The proposed algorithm also has a modified ShiftRows step named the New ShiftRows. The modification is made to increase the security of the overall AES algorithm as the ShiftRows step in the classical algorithm is merely a rearrangement of the rows' cells to avoid encrypting the columns separately, in which case the AES would be split into a block cipher of four independent columns. Simply giving an input matrix to the ShiftRows step in the original AES, the output will be as in figure 7.

ACGT	CTAG	TAGC	ATCG	ACGT	CTAG	TAGC	ATCG
TACG	ATGC	CTGA	CGTA	ATGC	CTGA	CGTA	TACG
AGCT	TGCA	CATG	GTAC	CATG	GTAC	AGCT	TGCA
CAGT	GCAT	TCGA	TCAG	TCAG	CAGT	GCAT	TCGA

**Fig 7** Applying the ShiftRows step in AES with DNA sequence code

The conventional sequence alignment algorithm attempts to align two sequences and increase the similarity between these two sequences by performing insertions, deletions, and substitutions in any mismatch to find an alignment that minimizes the total cost. The proposed new ShiftRows will exploit the characteristics of the sequence alignment algorithm by making the Sequence Alignment Algorithm to execute in reverse order. The proposed approach takes the input and the output of the conventional ShiftRows step as the two sequences to manipulate in the proposed reverse sequence alignment algorithm. The proposed approach will perform the same

insertions, deletions, and substitutions but in any match position leaving any mismatch position without changing. The output of the proposed approach will erase any similarity between the two sequences. Figure 8.a shows the initial aligning between the input and the output of the original ShiftRows step preparing to apply the reverse sequence alignment approach.

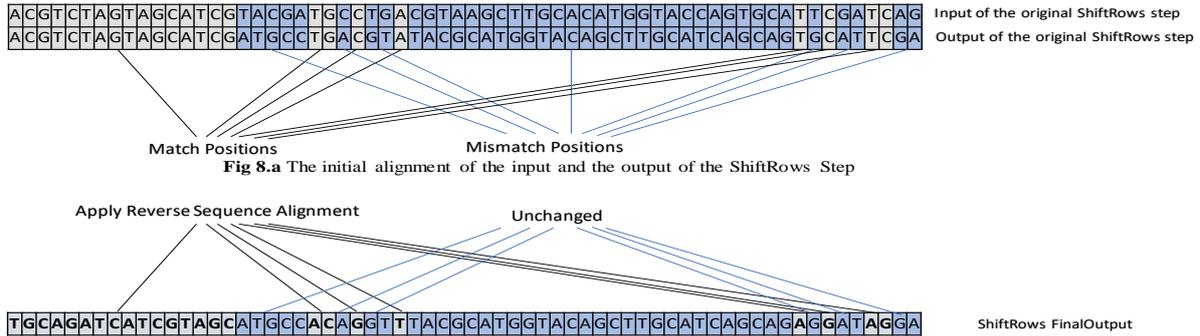


Fig 8.b Applying the proposed Reverse Sequence Alignment Approach on the output of the ShiftRows step

It turns out that a large amount of matching position exists between the input and the output of the ShiftRows step, especially there is no sequence alignment method has been used yet and the alignment is in the initial stage. The analysis of using sequence alignment algorithms to align the two sequences in the best match will be discussed in the experimental results section. The results induce that any linear cryptanalysis method used will easily obtain the linear relationship between the input and the output of this step. Figure 8.b shows using the proposed sequence alignment algorithm to reduce the similarity between the input and the output of the original ShiftRows. As the two given sequences (the input and the output of the ShiftRows step) are equal in length, so in the proposed approach it is preferred not to use insertion or deletion to simplify the process and to reduce the processing time and just adhere to substitutions. Therefore, in every match position, the output will be substituted with the reverse complement of the nucleotide.

### 4.3 Confirmation Stage

The overall output of the dispersal stage will go through another layer of dissimilarity confirmation called the confirmation stage. The confirmation stage is not for encryption, it only makes sure that the encrypted data in the last stage have no relevance to the original data tacitly. Based upon the stage is considered as the main box of the proposed scheme as it evaluates the processed data of the previous stage and validates its security robustness by checking the similarity between the input plaintext of the round and its` output ciphertext to a certain threshold percentage after applying a sequence alignment algorithm. The confirmation stage adds the feature of bypassing additional rounds if the results are satisfying already, counterbalancing the required encryption time. The stage evaluates the output from each round of the dispersal stage to measure the security level robustness, this is done by using similarity check to a certain similarity threshold, if the assigned threshold is realized, then the output passes throw a reverse sequence alignment process to obliterate any remaining regions of similarity, but if the similarity exceeds the assessment value, then it passes through round N+1 in which another layer of confusion and diffusion is performed until the assigned threshold is realized, and finally produces the final ciphertext.

---

#### Algorithm 4: Confirmation Stage

---

- 1: create two input variables
  - 2: input1 ← plaintext (X1)
  - 3: input2 ← ciphertext (PT)
  - 4: store (X1) in the input variable
  - 5: store (PT) in the input variable
  - 6: use Local alignment to align (X1) with (PT)
  - 7: use similarity checker to obtain the similarity between (X1) and (PT)
  - 8: if the similarity is <20%
  - 9: apply Reverse Sequence Alignment on (X1) to produce final ciphertext
  - 10: else
  - 11: return to round N+1
  - 12: end else
  - 13: end
- 

## 5. EXPERIMENTAL ANALYSIS

This section is about the experimental analysis and results of the proposed algorithm with a comparison with the Classical AES and recent researches. The detailed instructions of the proposed Key Transmission based on DNA Paternity will also be discussed. The proposed approach will produce a more powerful encryption technique based on three main criteria:

### 5.1. Multimodeling (Dynamic procedure)

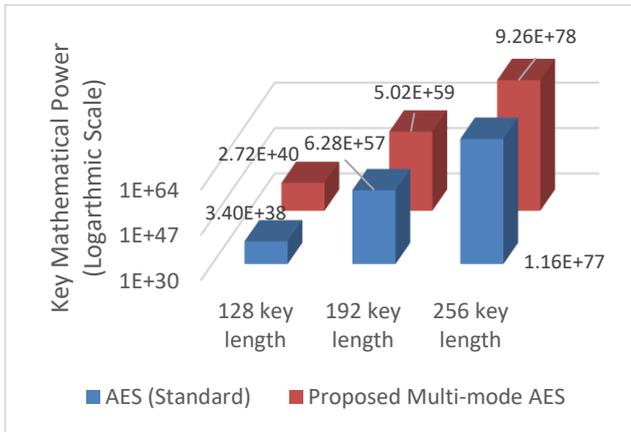
The hardness for the intruder to compromise any cryptographic system is based on many factors. First, the intruder has to specify the technique used for encrypting the desired data; once it is specified and it is very primitive, he has the procedures used in the encryption process. In AES, the internal procedures are fixed and proclaimed whereas the same operations and substitution tables are typically used in every encryption process, that makes the intruder concerns to obtain the relationship between the input data and the output by analyzing

the internal encryption procedures which are already announced, regardless of obtaining the encryption key. To bypass this limitation, a multi-model scheme is proposed in which the procedures are dynamically selected depending on the encryption key. In a 128-bit key AES, the last 3 bits are used to select a specific encryption mode providing eight different modes for each round separately as discussed before. It is substantial to demonstrate that choosing the last 3 bits to represent the dynamic multi-model is based on the truism that the intruder must expose the whole key to expose the multi-model procedures. The intruder has no clue about which mode is selected for each of the ten rounds making it impossible to obtain a linear or differential relationship between the plaintext and the ciphertext. The intrinsic idea behind the dynamic multi-model technique proposal is that the approach will strengthen the security of the overall algorithm without enlarging the key size, as the larger the key size the larger the computational cost and encryption time. The proposed multimodel technique also produces a more powerful encryption technique without any significant additional procedures which could also significantly affect the computation cost and the required encryption time. Table 4 shows how the proposed multimodel approach will produce a more powerful encryption technique without any significant impact on the processing cost.

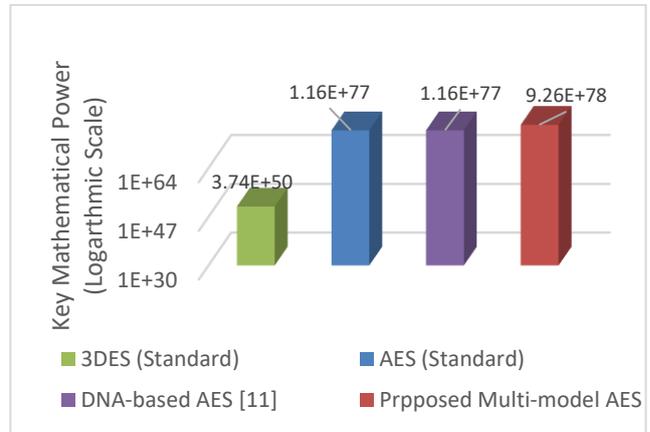
	Classical AES – 128-bit key	Proposed multi-model AES – 128-bit key
<b>Number of rounds</b>	10	10
<b>Mode type</b>	Single-mode	Multi-modes (eight modes)
<b>Number of total modes</b>	One mode	Eighty modes (8 possible combinations for each of the 10 rounds)
<b>Performance</b> (on a single-core Intel Core i7 Processor Extreme Edition, i7-980X) – in parallel mode	4.2 cycles/byte	4.2 cycles/byte

The classical AES requires 4.2 cycles to process the encryption of a single byte of data and 1.3 cycles for decryption [29]. With the additional multimode feature, there are no additional significant processes required, as the encryption is done through one mode based on the round key, not the all modes combined producing no additional processing cost. The security of the proposed technique will be equivalent to the security of the classical AES but with eighty rounds represented in the 8 different modes for each round without actually performing the eighty rounds, instead of classical AES ten rounds. Although the resistance against linear and differential cryptanalysis will be increased. Furthermore, it will also raise the cost of testing each key, as the intruder has to test the key for each different model for an intruder.

**The graphical representation of the dynamic multi-model key strength:** The representation aims to illustrate the increased security in the encryption power through the leap in the key combinations added by the multiple model approach. As clarified in the analysis that the proposed multi-model scheme will strengthen the encryption power of the AES as the security of the new proposed algorithm is based on key-dependent dynamic procedures. To represent this enhancement, the authors will compare the encryption key mathematical power for the proposed new multi-model AES with the classical AES for each of the three key length versions. The authors will also compare the 128-bit version of the proposed technique with recent encryption techniques among them the classical AES.



**Fig 9a** Comparison between the mathematical power of the Encryption key for the conventional AES and the proposed Dynamic Multi-Model AES



**Fig 9b** Comparison between the mathematical power of the Encryption key for the 3-DES, conventional AES, DNA-based AES, and the proposed Dynamic Multi-Model AES

Figure 9a showed an exponential elevation in the mathematical power regarding 128 bit, 192 bit, 256 bit based encryption key respectively. All without any significant impact in speed performance, as all the modifications are being done without any additional procedures. The power of the proposed algorithm key has not only increased by the 8 different mode combinations ( $2^{128*3}$ ) but also to the power of  $2^{128*3*10}$  as the mode selection process is based on the round key (sub-key) which make the mode being altered for each round independently.

Figure 9b showed a comparison between 3-DES, conventional AES, DNA-based AES [16], and the proposed modified AES to represent the mathematical power of the key corresponding to each algorithm. The 3DES has a 56 bit key with a  $2^{112}$  and  $2^{168}$  possible combinations. The standard AES has a 128 bit key with a  $2^{128}$  possible key combinations. The DNA-based AES presented by Sabry et al. [16] kept the same security strength of the standard AES with no increasing or decreasing with a possible key combination of  $2^{128}$ . Finally, the proposed Multi-model AES showed a great enhancement in the strength over the standard AES with a  $2^{128}$  possible key combination for each selected mode for each round producing a mathematical power for the encryption key equivalent to  $2^{128*3*10}$ .



% (similarity) = 0 percent

**Fig 11** Applying Reverse sequence Alignment on ShiftRows step

The results of using Reverse Sequence Alignment to modify the original ShiftRows step produced a zero similarity ratio between the input of the modified ShiftRows step its` output. As the proposed modified technique obviated naïvety of just shifting the rows linearly by presenting a more sophisticated technique by reducing any similarity between the input and the output.

### 5.3. Key Generation based on DNA Fingerprint

The proposed technique is inspired by the DNA Paternity fingerprint in which the offspring`s DNA Sequence is a derivative from the parents` DNA. The proposed key generation technique supposes that giving an offspring`s DNA inherited from two parents, this offspring`s DNA sequence will be unique and related only for those two parents. If one parent changes, the offspring`s DNA will differ. The main idea is based on combining the two DNA sequences; one from the sender and the other from the receiver who wants to communicate securely to produce a unique DNA sequence. The sender`s and the receiver`s synthesized DNA will work as a backbone for generating the offspring's DNA, which will be the key for encrypting the message. The data cannot be decrypted without the generated offspring`s DNA sequence. The attacker cannot obtain the offspring`s DNA sequence used for encryption without the two parents' synthesized DNA (the sender and the receiver). Figure 12 shows how the encryption key in a DNA sequence form will be generated using two synthesized DNA chosen randomly by the sender and the receiver.

TAGCATAGTTTSCCTAGTACGATGTCATGCTAGTTTCTAGATGGTACCACTGCCATTGGTACAG	Sender synthesized DNA (Parent 1)
CAGTATAGCAGTSCATTTTSCCAGCAGTATAACGCATGSSTAAGCTGGCATCCGCAGTGCATTCAA	Receiver synthesized DNA (Parent 2)
TAGATAGTTTSCCATTTGGCAGTSCATATACTAGTGTAAAGATGGGCATCCGTGCCATGCCATACAA	Generated Encryption Key (Offspring's DNA)

**Figure 12** Generating the encryption key from two synthesized DNA

For each match between the sender and the receiver synthesized DNA, the generated encryption key will be the same, and for each mismatch, the generated encryption key will be a combination of both sender`s and receiver`s DNA sequence respectively, one after one for each gap. The overall key generation process simulates the process of DNA fingerprinting in a DNA paternity test as the offspring` DNA has characteristics of both parents` DNA.

The generated offspring synthesized DNA from the two parents' DNA is a one-way function, which cannot be reversed to produce the parents' DNAs from the offspring DNA.

## 6. Conclusion and future work

The presented work showed how the combination of computationally and biologically inspired procedures can improve the diffusion and confusion factors required for ideal cryptography, thus reveals a promising future in a bio-inspired cryptographic technique. The proposed technique showed an enhancement over the conventional AES technique by adding the multi-model layer which will strengthen the security of the algorithm and the key without enlarging the key size to avoid adding any computation cost, as the larger the key size the more expensive computation costs. The proposed scheme is also distinguished by implementing an algorithm in which encrypted data can be examined before sending it in a public communication channel to ensure the irrelevancy of the ciphertext to the corresponding plaintext. The adaptable similarity threshold and the adaptable number of dispersal stage rounds may be very useful, giving the user the capability to interact to set the required encryption level based on the data importance level especially when the data in hand has a variety priority level considering the time complexity. Further, the presented work showed the capability of building a complex algorithm like the AES on DNA bases instead of bits, which prove the applicability to an underdevelopment DNA-based computing system and the potential biological environments.

The analysis showed a potential improvement in the modified Shift-Row step detached, and generally, the mathematical power of the overall modified algorithm added by the proposed dynamic multi-mode. The value of the proposed multi-model approach lies in producing a more powerful version of the AES without the need for significant manipulation of the hardware and software environment and implementations as the proposed approach kept the same standards, steps, key size, and procedures.

In future works, the authors will discuss applying the proposed system on a real-time application to show that the modifications applied don` t negatively affect the performance of the encryption and the decryption procedures. Further, The research directions are on how to apply the multi-model scheme in corresponding cryptography algorithms.

## References

[1] Biggs, N.: Codes: An Introduction to Information Communication and Cryptography. Springer, London, doi: 10.1007/978-1-84800-273-9\_4 (2008).

[2] Jacob, G. Murugan, A.:DNA based Cryptography: An Overview and Analysis. International Journal of Emerging Science, vol. 3, pp 36-42 (2013)

[3] Mosca, M.: Cybersecurity in an Era with Quantum Computers: Will We Be Ready?. IEEE Security & Privacy, vol. 16, no. 5, pp. 38-41 (2018)

[4] Church, G. Gao, Y. Kosuri, S.: Next-Generation Digital Information Storage in DNA. Science, New York ( 2012).

[5] Pramanik, S. Setua, S.: DNA Cryptography. 7th International Conference on Electrical and Computer Engineering. Dhaka, Bangladesh, pp. 551– 554 (2012).

[6] Tagore, S. Bhattacharya, S. Islam, A. Islam, L.: DNA Computation: Applications and Perspectives. Journal of Proteomics & Bioinform, vol. 3, pp. 234-343 (2010).

- [7] 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), Unnao, pp. 1-5 (2014).
- [8] George, A. Singh, H.: Design of Computing Circuits using Spatially Localized DNA Majority Logic Gates. IEEE International Conference on Rebooting Computing (ICRC), Washington, DC, pp. 1-7, (2017).
- [9] Adleman, L.: Molecular computation of solutions to combinatorial problems. *Science*, 266 (5187), pp. 1021–1024 (1994).
- [10] Liu, W. Sun, S. Guo, U.: A DNA Computing Model of Perceptron. Pacific-Asia Conference on Circuits, Communications and Systems, Chengdu, pp. 726-729 (2009)
- [11] Oghihara, M. Ray, A.: DNA computing on a chip. *Nature*, pp 143-144 (2000)
- [12] Paspula, R. Chiranjeevi, K. Kishor, K. Ramysree, B.: A Symmetric Key Encryption Method with a Reference DNA Sequence Technology. International Journal of Engineering Trends and Technology (IJETT), vol 5, pp. 223-231 (2016).
- [13] Rajni, A.: DNA Computing. International Journal Of Engineering And Computer Science, India, vol 6 issue 1, (2017).
- [14] Prabhu, D. Adimoolam, M.: Bi-serial DNA Encryption Algorithm(BDEA). Computing Research Repository - CORR (2011).
- [15] Boneh, D. Dunworth C. Lipton, R.: Breaking DES using a molecular computer. In Proceedings of DIMACS Workshop on DNA computing (1995).
- [16] Sabry, M. Hashem, M. Nazmy, T. Khalifa, M.: Design of DNA-based Advanced Encryption Standard (AES). IEEE 7th International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, pp. 390-397 (2015).
- [17] Daemen, J. Rijmen, V.: The Design of Rijndael: The Advanced Encryption Standard (AES) 2<sup>nd</sup> edition. Information Security and Cryptography, National Institute of Standards and Technology (2020).
- [18] Dalakoti, N. Mehra, A.: Hardware efficient AES for image processing with high throughput. 1st International Conference on Next Generation Computing Technologies (NGCT), Dehradun, pp. 932-935 (2015).
- [19] Mahendra, L. Santoso, Y. and Shidik, G.: Enhanced AES using MAC address for cloud services. International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, pp. 66-71 (2017).
- [20] Chauhan, Y. and Sasamal, T.: Enhancing Security of AES Using Key Dependent Dynamic Sbox. International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 468-473 (2019).
- [21] Nadjia, A. and Mohamed, A.: AES IP for hybrid cryptosystem RSA-AES. 2015 IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15), Mahdia, pp. 1-6 (2015).
- [22] Biryukov, A., Khovratovich, D.: Related-Key Cryptanalysis of the Full AES-192 and AES-256. *Advances in Cryptology - ASIACRYPT*. LNCS, Springer, Heidelberg, vol. 5912, pp. 1–18 (2009).
- [23] Needleman, B. Wunsch, D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, vol: 48 (3), pp. 443–453 (1970).
- [24] Smith, F. Waterman, S.: Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, vol. 147 (1), pp. 195–197 (1981).
- [25] Gemalto. (2015). Breach Level Index. 2014 Year of Mega Breaches & Identity Theft. <https://www.infopoint-security.de/medien/breach-level-index-annual-report-2014.pdf>
- [26] Auday, W. Ramlan, M. Zuriati, Z. Nur, N.: A New DNA Based Approach of Generation Key-Dependent Shift Rows Transformation. International Journal of Network Security & Its Applications (IJNSA) Vol.7, No.1 (2015).
- [27] Pushpa, B.: A new technique for data encryption using DNA sequence. 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, pp. 1-4 (2017).
- [28] Akiwate, B. Parthiban, L.: A Dynamic DNA for Key-based Cryptography, International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, pp. 223-227 (2018).
- [29] Akdemir, D. Dixon, M. Feghali, K. Fay, P. Gopal, V. Guilford, J. Ozurk, E. Wolrich, G. Zohar, R.: Breakthrough AES Performance with Intel @ AES New Instructions, (2010).
- [30] Bahig, H. Nassr, D. DNA-Based AES with Silent Mutations. *Arabian Journal for Science and Engineering*, vol 44, pp. 3389-3403 (2019).
- [31] Challita, N. Bakhache, B.: Enhancement of S-AES using chaos for the support of biomedical applications, 2nd International Conference on Advances in Biomedical Engineering, Tripoli, pp. 175-178, (2013).