Cryptanalysis Using DNA-Sticker Algorithm

Cryptanalysis Using DNA-Sticker Algorithm

Noora Amir Abdulmehdi[1]*, Sahar Adel Kadum[1]

[1]Department of Computer Science for Women, College of Science, University of Babylon, Babylon, Iraq.

Email: *nooramilee@gmail.com

Abstract. Sticker model is one of the basic computer models of filtering model which is part of Non-Autonomous DNA computing models. This model inputs are represented by encoded one of a double (single – double) strand of DNA molecules , which is annealed with a short fraction of strings at some positions , and therefore these strands called memory complexes . Since its inception, DNA computing has provided a highly efficient and energy efficient technology for solving complete NP problems because of its really high data density and inherent parallels. In this paper, we explain how to use the sticker model, which is one of the types of molecular computations in DNA computing, to design an algorithm for coding analysis of the coded text in an unsystematic cipher method. The cipher text is broken by employing the idea of brute force attack, which works to represent all possible keys and then try it to break the encrypted text until reaching the correct key from among the total keys group representing by solution space. Theoretically, both the cipher text and all possible keys are represented using memory complexes of the DNA strands . Then , using sticker operations , as separation , combination , and proposed operation named length , we can remove the incorrect keys from the solution space of memory complexes , by taking advantage of the parallel processing of DNA computing , therefore this model is classified as filtering model.

1. Introduction

The security of data transmitted over the Internet is the most important concern for researchers in the field of cryptosystems. Therefore, one way to obtain this security is the experience of breaking encryption methods to know the extent of its strength. Since code breaking needs many requirements in terms of the complexity of time, storage and calculations utilizing systems that feature these requirements these days is a challenge that can be overcome with the emergence of new technologies and algorithms.[8]

To protect different data and information, there are some systems designed to be used for this purpose. First of all, the cryptography science is the art of protecting information from unauthorized individuals by converting it into a form that attackers cannot recognize while it is stored and transferred. Basically, scrambling of data content, such as text, image, audio, video, etc. in order to make the data cannot be read , nor clear while the operations for transmission or storage, this activity named data encryption. Reverse data encryption is a data decryption.[7][8]

DNA computing whose signs appeared in the last quarter of the last century and the proposed models that fall within this branch by taken the advantages of DNA computing in terms of storage and processing. [7][15]

The DNA computing can be applied in systems of DNA cryptographic, which are based on DNA. That mean the ability of this study to overcome lots challenges in cryptography field, if it is used correctly. In DNA cryptography DNA is used as information carrier and the modern biological technology is used as an implementation tool. Thus it is a new born cryptography technique, which the important features of it are density of information and extra ordinary, beside the vast parallelism which are inherent in DNA molecules. That features are explored in cryptosystems for all sorts of its techniques. The capability of nucleotide bases-binding (A-T, G-C) offer excellent means of executing computations because the opportunity of generating self-assembly structures. There is additional advantage is the DNA capacity of huge storing. Which is very important to solve the NP-complete problem such as cryptanalysis and find the correct solution among big number of solutions may reach 26^L, where L the length of encryption message.[7][15][3]

DNA computing deals with several modules, one of them is the sticker module as a new approach of DNA computing that makes it easier to search in parallel and get the correct solution from search space of solutions. Because it is one of the filtering model of DNA computing models. This module is the focal point of this thesis in finding the best solution in attacking a ciphered text in less time and in maximum speed.[7]

In this paper, we proposed a sticker model for solving one of NP-complete problem, brute force attack, for attack a cipher text, which encrypted using the unsystematic cipher algorithm.

The bio-computing or molecular computing begin at 1994 [1] using the DNA strand to solve the Hamiltonian bath problem then added to his idea at 1995 [2] the seed for DNA computers depends on molecular interactions may be a viable alternative to computers based on electronics. Depending on these ideas the researching every years are increasing. At 2018, Sarkar [4] attempts to perform arithmetic of complex number, (as addition, subtraction, multiplication, and division of two complex numbers). This work first one represented and perform complex number arithmetic by DNA molecules. The advantages of this work are the easy implementation of bio-molecular operations , and the ability to carried out the operations on arbitrarily large numbers.

In 2018, Yassin[5] proposed DNA-Sticker algorithm in cryptanalysis any stream cipher with LFSRs and NLFSRs. That algorithm required only a limited set of character from the cipher text. In 2018, Zhou[17] proposed algorithms of multifunctional weighting using DNA computing (sticker model), that can better deal with the large data than other algorithms as the parallelism with large-scale of sticker model-algorithm. In their algorithms they used the four basic bio-operations beside (Discard) operation.

In 2020, Dodge[18] proposed DNA-sticker algorithm to find the optimal solution to (TDCSP) the abbreviated of Two-dimensional cutting stock problem. It is proved the ability of sticker model on decreasing time complexity of this algorithm on DNA computers by spending polynomial amount of time considering both length and width of the main board and the number of small pieces. Only four basic bio-operations was used in this model.

The sticker algorithm of Yaseen[5] was deal with cipher text encrypted with LFSR\NLFSR stream method. Also that algorithm depend on correlation, and requited limited number of character to in getting the solutions, but our proposal algorithm deal with cipher text encrypted with unsystematic stream method, and depending on statistical quantities to get the solutions therefore, as the number of character in cipher text increased as that be more helpful in decreasing the possible solutions.

Here we proposed a sticker model for solving one of NP-complete problem, brute force attack , for attack a cipher text, which encrypted using the unsystematic cipher approach.

TABLE 1: RELATED WORKS USED THE STICKER IN THERE RESEARCHES.

Ref.	Year	Researcher	Model	Type of algo.	Solved problem
[4]	2018	Sarker	Sticker model	Theoretical	Complex number arithmetic
[5]	2018	Yaseen	Sticker model	Theoretical	Cryptanalysis (L/NL)FSR
[17]	2018	Zhou	Sticker model	Theoretical	Multifunctional weighting
[18]	2020	Dodge	Sticker model	Theoretical	Two-dimensional cutting stock problem (TDCSP)

The organization of this paper as: section one gave a general idea about the cryptanalysis and the encryption method that decrypt, while second section demonstrated the DNA molecular and the reasons and why used in computations, while the third section illustrated the general idea about the DNA models and demonstrated the sticker system. The last one spoke about the proposal algorithm to decrypt the cipher text, then the conclusions.

2. Cryptanalysis

It is the science which used for breaking cryptosystems. Cryptanalysis is of central importance for modern cryptosystems indeed, because we could evaluate the security level of our crypto methods by people who try to break our methods[2].

The solution of nearly every cryptogram involves important steps which determine both language and the general system that used also reconstruction of both the specific keys to system and plain text. One of the general system sources , the detailed study of the system characteristics aided where necessary by character frequency counts , search for repeated patterns and various statistical test. The frequency matching technique only works if the text is long enough to produce a clear or recognizable frequency count that viewed in the figure (1) [7] .

Generally, there are different types of cipher analysis attacks. Each of them assumes that the cryptanalyst (which attacks the secret message) knows the algorithm used to encrypt the message. They are cipher text-only, known-plaintext, and chosen-plaintext [8].

A cipher text-only attack is the type of attack when an attacker has the text of the encrypted message, and the encryption method in which the message is encrypted. The job of the cipher analyser is to retrieve the plain text of the secret message as much as possible, or to infer the key (or keys) used to encrypt messages, in order to decrypt other messages encrypted with the same keys, and this is the type of our proposal.

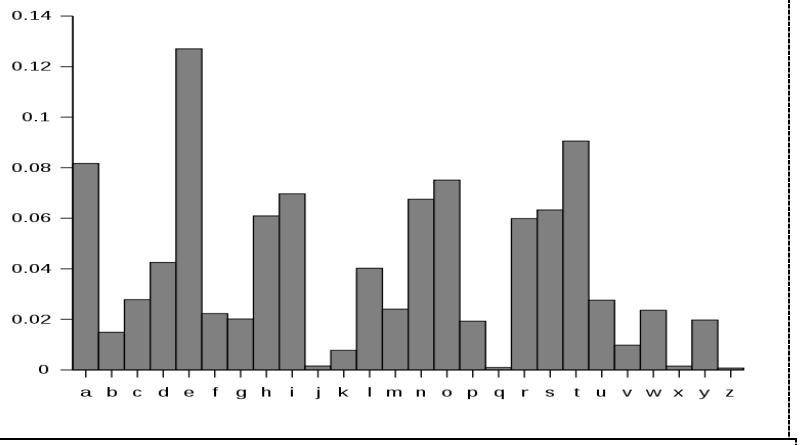

Figure 1 : English Letter Frequency

2.1 Brute Force Attack

Method to attack cipher systems, brute force attack. It is a method of finding the correct key values after trying all the possible key values, and therefore it is considered one of the less intelligent methods. However, this method is more difficult than it appears. Thus, improved encryption algorithms taking into account the brute force is a very delicate task [4][8][15].

An amount of time the breaking a cipher takes to implement is proportional to the size of the key, which means if the number of bits in the key is increased then the number of attempts will be increased ,too. [8]

Therefore, brute force can be easily carried out on parallel computers. Indeed, they are of the embarrassingly parallel kind: it suffices to run n copies of the same program, each searching in a fraction of the space of possible keys to speed up the computation by a factor of n. Although its extreme simplicity, this basic remark should never be forgotten when evaluating the security of a cryptosystem.[4][8]

As a new trend in parallel computation is the use of DNA computing as one of powerful techniques, because it is interesting to find that computations can easily be implemented on such techniques.

2.2. Unsystematic Cipher

This new encryption technique is used to change plaintext in binary form. For characters in the plaintext generates random number (r). This value(r) named added number. It makes the cipher of the character in 2*r different ways. After generate the keys, the second stage is converting the character of the plain text into binary form because this method need to dividing the binary value of the plain text into parts, these parts are equal to (r) value.

As example if r=2 then the all possible key are k={0 (0000 0000), 1 (0000 1111), 2 (1111 0000), 3 (1111 1111)}. And if r=3 then k={0 (000 000 00), 1 (000 000 11), 2 (000 111 00), 3(000 111 11), 4(111 000 00), 5(111 000 11), 6(111 111 00), 7(111 111 11)} and so on for any r value less than the bit represents the one character. This key is encrypt the plain text by xoring its binary values to get the cipher text, at last[9].

3. DNA Cryptography

DNA cryptography is emerged with the research of DNA computing as a new born cryptographic field. It is exploiting exceptional energy efficiency, extraordinary information density and the vast parallelism inherent in molecules of DNA. These features are explored for different digital techniques

as data storage, computing and cryptographic purposes (such as authentication, signature, encryption, and so on).[10]

All over the world, the implementation of many of the research, either to enhance existing methodologies of DNA cryptography or to proposed new idea for an innovative approach in this area. In the years to come when DNA computers are commercially available, modern silicon-based technology will outperform[10].

3.1 Deoxyribonucleic Acid (DNA)

DNA (deoxyribonucleic acid) is the germ plasma of different lifestyles, at least on the earth. It is biological macromolecule. It is built by nucleotides, that every one of them contains a single base. These bases have four types, Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) as in figure 2. Any single-stranded of DNA is constructed with two orientations: one end is called 5′, and the other end is called 3′. Naturally as double-stranded molecules, we can find DNA.[16]

DNA molecule having the capacity to store, process and transmit information. These features are inspired the idea of the concept, DNA cryptography, which depending on the concept of DNA computing. This type of computing uses DNA bases (A, T, C, and G) in order to implement computation. The massive parallelism of DNA molecules is one of the significant advantages of DNA computation. Because of DNA computation-huge parallelism it is in just polynomial time can solving the trapdoor function, basic security secret of most of the conventional cryptosystems. As the mathematical aspect of cryptography is being replaced by DNA chemistry in the domain of DNA cryptography.[10][11][16]

Figure 2: Simple DNA structure

4. DNA Computation

The biological structure of DNA is make useful for representing the securing data as a new technique, that is called DNA Computing (molecular computing or biological computing). It was devised by Adleman at 1994 in order to solving the problem of directed Hamilton path and then implementing on all NP-complete problem alike to The Traveling Salesman problem.[10]

Because DNA ability to use to transmit and store data. The DNA computing as a concept using in the fields of both cryptography and steganography. In addition, it has been identified as a potential technology that may improve new hope in more secret algorithms.[10]

DNA strands consist of millions nucleotides in form of long polymers. Mathematically, this means we can utilize the 4 bases to encode information, that mean the DNA computer is more than efficient an electronic computer ,the needs for 4 bases (A, C, G, T) more efficient than 2 digits (1,0) to encode information[10]

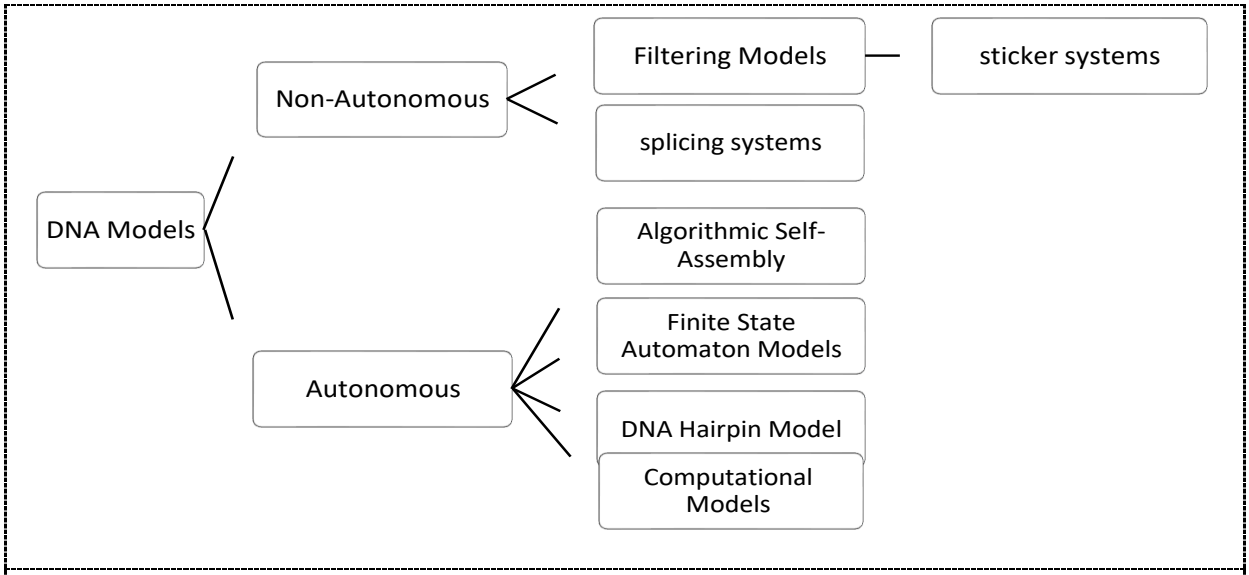

Figure 3: Zimmermann classification for DNA computing.[11]

4.1 DNA Models

In-vitro DNA models are measured and operated by humans, mainly intended at solving complex computational problems. The goal of different DNA models essentially is to execution logical calculations under living cells-limitations.

Researchers in the human-operated field of arithmetic DNA models focus their work on solving complex computational problems. These models generate large combinatorial libraries of DNA, in order to provide a wanted search spaces for algorithms with filtering that space in parallel. There is many variety methods for generating libraries, filtering of solution, and generating output were studied in experimentally researches. One of the basic computationally DNA models is the sticker model. While the Autonomous DNA Models specialize in molecular-scale, autonomous, and partially programmable-models, and therefore the computations are modulated by DNA-manipulating enzymes. Sticker Automata also named Adleman-Lipton's model [12].

It a DNA model for finite state machines ,that proposed by Zimmermann at 2008. That model is called sticker automaton, as short single strands of DNA molecules are encoding the transition rules, which referred to as stickers. A single enzyme or may more ,depending on the solving problem is/are represents/represent the hardware of this model[13]. Figure 3 demonstrate the Zimmermann's-DNA models classification[13].

4.2 Sticker Systems

This model is a type of DNA models, and it is a class of filtering models, which is a part of non-autonomous DNA models. It introduced by L. Adleman and his co-workers[14] at 1998.

4.2.1 Sticker Memory

The most important characteristic of this model is that it contains a random access memory, this memory does not need a strand extension. Its memory consists memory complexes. A memory complex simply, is a DNA strand that is partially double strand and can be viewed as an encoding of an n bit number. Each memory complex is formed by two basic types of single strand of DNA molecules. Figure 4 explain the sticker model and figure 5 gives examples how to represent information with memory complex.

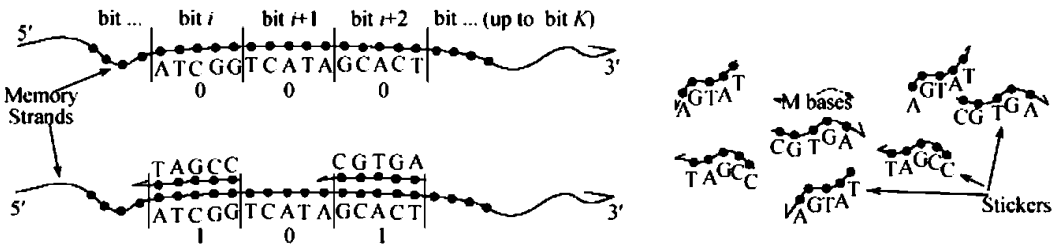

Figure 4: sticker model

Figure 5: different type of memory complexes 00000 ,10100, 11011, 11111

4.2.2 Sticker Model Computation

A sticker computation involves of a determinate sequence of sticker operations. Its input and output are test tubes, named initial test tube and final test tube, respectively. The input of a sticker computation is a test tube, also named initial test tube , and the output is also a test tube called, final test tube. This tube can read in other word that every memory complexes in final test tube is isolated the annealed stickers, or it may be reported that it contains no memory complexes. Many algorithms that using the sticker model require other test tubes, that be empty at the start of a computation.[3][13]

4.2.3 Sticker Model Operations

Adleman announced four basic operations for sticker system model. They are: combination, separation, setting and clearing (14). Then Zimmermann add another operation (discard), that not including in Adleman proposal work(13)(14).We'll explain these processes below. Note, that every T in following definition is mean tube and every i mean position number.

1) Combine (T, T1, T2): the content of both T1 and T2of memory complexes are combined in T, merely the strands of T1 and T2 are poured into T. (T = T1 U T2).
2) Separate (T, i) → (T+, T-): the content of T¬ used to creates T+ and T-, depending on the value of the i bit. If its value one then (T+ = +(T, i)), and T- contains the zeros value (T- = –(T, i)), with respect that T is discard at the end of this operation.
3) Set (T, i): each memory complex in T set to 1 in the i bit, or we can say, turned on. In another ward, the sticker for i bit is annealed to corresponded bit region on all memory complexes in T.
4) Clear (T, i): The bit in i on every memory complex in T set to 0 or turned off. That mean, the stickers of that bit region must be removed (if present) from all memory complexes of T.
5) Discard (T): Empty the contents of T.
6) Count (T) : this is proposal operation. It returns the number of memory complex (solutions) at the test tube.

5. Proposal Method

The proposal system based on DNA-Sticker Model to implement the cryptanalysis cipher text with unsystematic stream method. The proposal system builds cryptanalysis model using the DNA-Sticker. This model is Frequency Sticker Model (FSM). The FSM needs the bio-operations to decrypt any cipher text encrypted with one key and text length of more than 100 characters with analysing and filtering processes to implement their tasks. This model applied on traditional computers to perform the cryptanalysis. The DNA-Sticker has special bio-operations such as (Separate, Combine, Clear, and Set),beside proposal operation (count), as a tools to implement the DNA-logic model.
The proposed system have two steps

First step:

Generating all possible keys is the first step of this system which using four sticker operations (separation, combination, set and clear) Diagram(1) illustrates the general idea of this algorithm. The algorithm as following:

Algorithm 1. Generate all possible keys

```
Begin
  for i=1 to N
            separate(T, i ,Ti+,Ti-);
    if Ti+ is empty then
        Ti+ = set (T,i)
    end if
    if Ti- is empty then
        Ti- = set (T,i)
    end if
    T=combine(Ti+,Ti-);
       end for
 End
```

The size of initial test tubes are as following
S(Ttext)= (B)N , where N is the number of characters in cipher text, and B is binary form of each character in cipher text.
S(Tkeys)=(2r)L , where r=the number of regions that the character divided to encryption or decryption as in unsystematic method (added number), L is the length of key stream.

Second step is by analysing the bits of the cipher text-strand in its tube to decided which are illegal keys in keys tube. This step made useful of the frequency letters of English language. It's implementation is according to table(2). Algorithm 2 explain the mechanism of sticker cryptanalysis.

Table (2) Frequency of English letter-bits

Number of bit	1's frequency		0's frequency	
	Capital letters	**Small letters**	**Capital letters**	**Small letters**
8th bit	0%	0%	**100%**	**100%**
7th bit	**100%**	**100%**	0%	0%
6th bit	0%	**100%**	**100%**	0%
5th bit	32%	32%	**69%**	**69%**
4th bit	37%	37%	**64%**	**64%**
3rd bit	**58%**	**58%**	43%	43%
2nd bit	41%	41%	**60%**	**60%**
1st bit	**55%**	**55%**	43%	43%

Algorithm 2. Cryptanalysis the cipher text

```
Begin
separate(T, 1 ,T+,T-);
separate(Tkey, 1 ,Tkey+,Tkey-);
if T+ is empty then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
end if
separate(T, 2 ,T+,T-);
separate(Tkey, 2 ,Tkey+,Tkey-);
if T+ is empty then
        Tkey=combine(Tkey,Tkey-)
Else
        Tkey=combine(Tkey,Tkey+)
end if
separate(T, 3 ,T+,T-);
separate(Tkey, 3 ,Tkey+,Tkey-);
if T+ is empty then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
end if
separate(T, 4 ,T+,T-);
separate(Tkey, 4 ,Tkey+,Tkey-);
if count(T+) > count(T-) then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
end if
separate(T, 5 ,T+,T-);
separate(Tkey, 5 ,Tkey+,Tkey-);
if count(T+) > count(T-) then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
end if
separate(T, 6 ,T+,T-);
separate(Tkey, 6 ,Tkey+,Tkey-);
if count(T+) < count(T-) then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
end if
separate(T, 7 ,T+,T-);
separate(Tkey, 7 ,Tkey+,Tkey-);
if count(T+) > count(T-) then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
```

```
end if
separate(T, 8 ,T+,T-);
separate(Tkey, 8 ,Tkey+,Tkey-);
if count(T+) < count(T-) then
        Tkey=combine(Tkey,Tkey+)
Else
        Tkey=combine(Tkey,Tkey-)
end if
End
```

The previous analysis is depending on the length of cipher text. It get the correct key with cipher text with number of charaters larger than 90 charaters. The different between this model and brute-force attack is show in table 1. This model can decreasing the expending time at decrypting cipher text by treated all same bit of all characters all at once.
The implementation of this algorithm for any cipher text with length more than 100 character is spend same time, because it need only 7 steps to analysing and filtering the input values and for only 100 characters, which is enough to determine the correct key value by make useful of letter freqyencies. Table 3 explain the different between this algorithm and brute-force algorithm.

Table 3: Comparison between proposed algorithm and brute force algorithm

No	Algorithm	Complexity	Outputs
1	Brute force attack	$O(2^8)$	1 solution
2	Proposed algorithm	O(8n)	1 solution

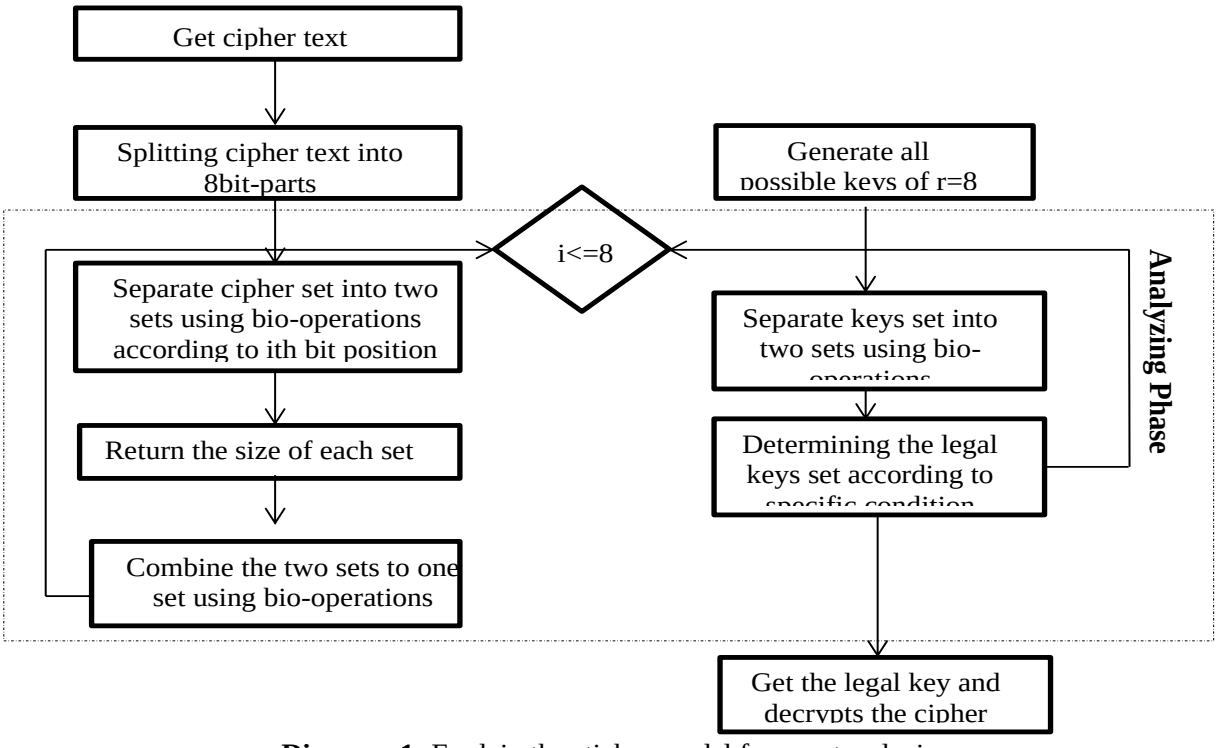

Diagram 1: Explain the sticker model for cryptanalysis

6. Conclusions

In this paper, a sticker model was used to analyse the encoded text. This paper demonstrated the ability to analyse cipher text using the sticker model process (combination, separation). This new method has some advantages in building solution space on other methods, from these features, parallel processing and the ability of memory pools to represent all possible keys as well as generating all possible solutions. The important notification that we have to say, it is not necessary to represent the values of algorithm inputs with DNA bases to be processing in the poster template, because we emulate this model on conventional computers. In fact, we simulate the behaviour of this model on the traditional computer, although we acquire this model feature.

References

[1] Adleman L. Molecular computation of solutions to combinatorial problems. Science, 1994 .
[2] Adleman L. On Constructing a molecular computer. Draft Jan. 11,1995.
[3] Sabari Pramanik, Sanjit Kumar Setua. "DNA cryptography", 2012 7th International Conference on Electrical and Computer Engineering, 2012.
[4] Sarkar M, Ghosal P. Performing Mathematics using DNA : Complex Number Arithmetic using Sticker Model, IEEE Computer Society Annual Symposium on VLSI,2017.
[5] Yaseen B, Sadkhan S. A DNA-Sticker algorithm for cryptanalysis LFSR and NLFSR based stream cipher, IEEE, 2018.
[6] Jiron I, Soto S, Marín S, Acosta M, Soto I. A new DNA-based model for finite field arithmetic . Heliyon 5, 2019.
[7] Mandrita Mondal, Kumar S. Ray,2015, Review on DNA Cryptography.
[8] Lasry G. Methodology for the Cryptanalysis of Classical Ciphers with Search Metaheuristics , Kassel university press,2018.
[9] Tas O, Alatas B, Akin E. A new approach to stream cipher: unsystematic cipher, Journal of Electrical and Electronics Engineering, Istanbul University,2004 .
[10] Sajisha K, Mathew S. An encryption based on DNA cryptography and steganography. International Conference of Electronics, Communication and Aerospace Technology (ICECA),2017.
[11] Singh S, Sharma Y. A Review on DNA based Cryptography for Data hiding. International Conference on Intelligent Sustainable Systems (ICISS),2019.
[12] Polenov A. The computing of NP-complete problems in polynomial time using DNA-logic, World applied sciences journal,2014.
[13] Ignatova Z, Mart´ınez-P´erez I, Zimmermann K. DNA Computing Models 2007.
[14] Roweis S, Winfree E, Burgoyne R, Chelyapov N, Goodman M, Rothemund P, Adleman L. A sticker based model for DNA computation. J. Computational Biology, 1998.
[15] J. Katz and Y. Lindell,2009, Introduction to Modern Cryptography Antoine Joux, Algorithmic Cryptanalysis
[16] Tuam S O, Kadum S A, Rashad F A. Text Encryption Approach using DNA Computation and Chaotic Indexing. International Journal of Emerging Trends in Engineering Research. Volume 8. No. 8, August 2020.
[17] Chunyan Zhang, Weijun Zhu, Qinglei Zhou,2018, The Algorithms of Weightening Based on DNA Sticker Model, ICPCSEE 2018, CCIS 902, pp250-262.
[18] M. Dodge,S. A. MirHassani, F. Hooshmand. Solving two-dimensional cutting stock problem via a DNA computing algorithm.2020.