Received: 16 August 2023 | *Revised:* 19 September 2023 | *Accepted:* 20 October 2023
DOI: 10.37917/ijeee.20.1.11 *Vol. 20 | Issue 1 | June 2024*

Iraqi Journal for Electrical and Electronic Engineering
Original Article

Open Access
IJEEE University of Basrah College of Engineering

Parallel Search Using Probabilistic DNA Sticker Model to Cryptanyze One Time Pad Polyalphabetic Cipher

Basim Sahar Yaseen*
Computer Science Department, Shatt Al-Arab University College, Basra, Iraq

Correspondance
*Basim Sahar Yaseen
Department of Computer Science,
Shatt Al-Arab University College, Basra, Iraq.
Email: basim17814@gmail.com

Abstract
Nowadays, it is difficult to imagine a powerful algorithm of cryptography that can continue cryptanalyzing and attacking without the use of unconventional techniques. Although some of the substitution algorithms are old, such as Vigenère, Alberti, and Trithemius ciphers, they are considered powerful and cannot be broken. In this paper we produce the novelty algorithm, by using of biological computation as an unconventional search tool combined with an uninhibited analysis method is the vertical probabilistic model, that makes attacking and analyzing these ciphers possible and very easy to transform the problem from a complex to a linear one, which is a novelty achievement. The letters of the encoded message are processed in the form of segments of equal length, to report the available hardware components. Each letter codon represents a region of the memory strand, and the letters calculated for it are symbolized within the probabilistic model so that each pair has a triple encoding: the first is given as a memory strand encoding and the others are its complement in the sticker encoding; These encodings differ from one region to another. The solution space is calculated and then the parallel search process begins. Some memory complexities are excluded even though they are within the solution paths formed, because the natural language does not contain its sequences. The precision of the solution and the time consuming of access to it depend on the length of the processed text, and the precision of the solution is often inversely proportional to the speed of access to it. As an average of the time spent to reach the solution, a text with a length of 200 cipher characters needs approximately 15 minutes to give 98% of the correct components of the specific hardware. The aim of the paper is to transform OTP substitution analysis from a NP problem to a $O(n^m)$ problem, which makes it easier to find solutions to it easily with the available capabilities and to develop methods that are harnessed to attack difficult and powerful ciphers that differ in class and type from the OTP polyalphabetic substitution ciphers.

Keywords
sticker model, probabilistic model, OTP, Polyalphabetic substitution cipher.

I. Introduction

The mathematical explanation of the sticker model was found and described in 1999 by a group of scientists, and this description is divided into two parts. Representation of Information: The sticker model uses two basic groups of single-stranded Deoxyribose Nucleic Acid SSDNA molecules in its picture of a bit sequence. Let a memory strand A Genetic bases in longitude subdivided into B noninterfering areas, each C genetic bases long (thus $A \sim B \times C$). Each area is described with exactly one bit location (or equivalently one logic variable) during the processing. We also design C different sticker strands or simple stickers. Each sticker has B bases long and is complementary to one and only one of the C memory areas. If a sticker is dieted to its matching area on a given memory strand, then the bit congruent to that specific area is on for that strand. If no sticker is dieted to an area, then that area's bit is off. Fig. 1 illustrates this representation scheme. Each memory strand along with its dieted sticker (if any) represents

one bit sequence. Such partial duplexes are called memory complexes. A large set of bit strings is represented by a large number of identical memory strands, each of which has stickers annealed only at the required bit positions. We call such a collection of memory complexes a tube. Operations on Sets of Strings: The basic processing is to combine two strings of bits into one string. The output is a new set containing the multi-set union of all strings in the two input sets. In DNA, this corresponds to the production of a new tube containing all the memory complexes (with their annealed labels undisturbed) from both input tubes. A set of strings can be separated into two new sets, one containing all the original strings containing a given bit, and the other containing all the strings containing a stop bit. This corresponds to isolating those complexes from the collection tube completely using a solid label in the selected bit region. The original insert assembly (tube) has been destroyed. To tune (play) a particular piece in each string of a set, the label for that piece is annealed in the appropriate area on each collector on the set tube (or left in place if already annealed). Finally, to clear (turn off) a small part in each thread of the group, the label for that part (if present) must be removed from each memory pool in the group tube."as discussed elsewhere [1, 2]. Because the sticker model is an efficient model for searching in solution space, it has been used in the search for the solution to many difficult issues "as discussed by Yaseen [3]", including cryptographic issues, and as a new application, the model was implemented to represent, analyze and attack the polyalphabetic cipher "as discussed elsewhere [4, 5]", especially, one time pad (OTP) cipher "as discussed elsewhere [6, 7]", which is a very strong code and difficult to attack. The OTP character-level substitution cipher, such as Vigenère, Alberti, and Trithemius ciphers "as discussed by Bonaaviglis at el [5]", is characterized by the fact that the plaintext, ciphertext, and the key belong to the same alphabet $P_{\{A..Z\}} \times K_{\{A..Z\}} \rightarrow C_{\{A..Z\}}$, but there is a feature that makes it difficult to break, which is that the key segments are varied and different and their number is equal to the number of plaintext segments.

II. Contribution and Novelty

The contribution of the current work lies in several scientific areas. The first is that the current work attacks a previously unattacked type of cipher, namely the case when the length of the key is equal to the length of the plaintext when encrypted (OTP cipher). The second is to present a probabilistic model that harnesses DNA computation for the first time by breaking the OTP cipher. The practical importance of the work in the real world comes from the fact that it presents an attack model for modern OTP ciphers.

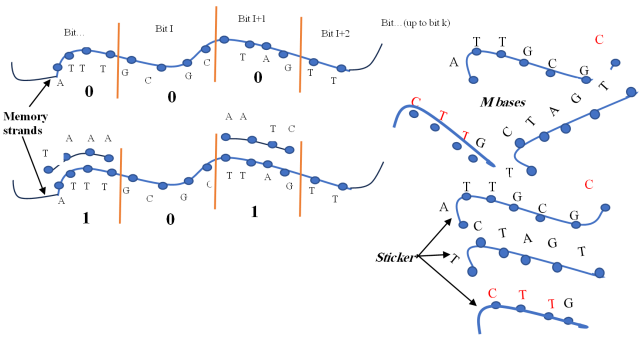

Fig. 1. Representation of Information in the sticker model "as discussed by Roweis and el [1]".

III. Literature Review

Although the concept of One Time Pad (OTP) in substitution cipher systems is old, attempts to cryptanalyze or attack those systems did not take seriousness and did not give the desired results. However, we find that in some of the late attempts, as to the difficulties imposed by the strength of OTP of a previously unexplored probabilistic model, "as discussed by Yaseen [8]", these difficulties are eliminated. As for the use of DNA models in cryptanalyzing and attacking ciphers, we can list the latest serious works in this field as following. The paper in [9] proposes an attacking of a propositional logic-based cryptosystem through DNA algorithm and digital implementation of its operations. The proposed method deals with the possibilities of each binary component of the ciphertext components; therefore, it is difficult to process multiple binaries. The paper in [10] deals with the components of cipher bits, sequence, keystream, and plain text bits, and it transforms their logic from propositional logic to the DNA logic and executes it in polynomial time. The paper in [11] proposes a sticker DNA model to cryptanalyze the cipher generated by the keystream of linear and nonlinear feedback shift registers. The model is based on the logic of creating a binary sequence as a memory strand that represents the possibilities of the plain text sequence, and then it creates all possibilities of paths to the correct solution by linking the stickers to the components of the solution paths that represent the key parts. The paper in [12] creates a Data Base of all solution paths and then searches about the correct path through the technique combined between the GA and the DNA sticker model. The proposed technique implements a parallel search to attack a cryptosystem. The paper in [13] proposes a modified digital simulation of the hypered technique of DNA sticker model with the other technique and uses it to attack linear and nonlinear feedback shift register generators. The paper in [14] proposes a DNA sticker model to cryptanalyze the stream cipher generated from a key stream sequence, that is, the output

of a linear shift register. The cipher sequence is cryptanalyzed by sticker operations at the level of binaries. The paper in [15] proposes a software computer based on the genetic operations of a splicing DNA model with a probabilistic model of the English letter frequency vertically. The formation of genetic bases of the strands, that is, of the letters of a natural language, and their cohesion, probably depends on the occurrence of these letters in the plain text or key. The paper in [16] proposes a splicing DNA model that cryptanalyzes a stream cipher sequence of an unknown generator, but it is known coding for the plain text, and the proposed model exploits the statistics of the plain text with the random properties of the key string segments.

IV. Methodology

The proposed probabilistic sticker model combines the capabilities of the probabilistic model in reducing the search space for polyalphabetic algorithms to a limited space based on a compact statistical hypothesis; it is combined with the capabilities of the sticker model that provides a parallel search within this limited space to reach the correct solution in a specific time that is very much less than the search time in the original search space before the reducing.

A. The Role of Probabilistic Model

According to the hypothesis of the proposed probabilistic model (referred to in Paragraph No. 2), for substitution cryptanalysis, the work on scattering the plaintext statistics as well as the key statistics which belong to the same natural language, is done by encrypting each plaintext block (character or more), by using a different syllable each time from the key blocks. Thus there is no possibility to predict the used key blocks, and this method summarizes the principle of OTP. The role of the probabilistic model is to overcome this hash barrier and to overcome the resulting difficulty. This role depends on integrating the statistics of the frequency of the plaintext letters with the frequency of the key letters, and calculating the frequency of the resulting letters in the ciphertext. Then the resulting statistics are linked to the original statistics. As an important outcome of this hypothesis, with other considerations, it reduces the search space very significantly, and the probabilistic model sorts out the actual probabilities of the occurrence of each pair (plaintext character, key characters). Fig. 2 and Fig. 3 describe the real (plain letter, key letter) pairs and their probabilities. They show the pairs of letters from the plaintext letter and the key and the probability of their occurrence in the ciphertext when they are combined together, where there are letters with high probabilities and others with low probabilities. “as discussed by Yaseen [8]”

B. Representation of Problem’s Information

Level of Probability and codons		Probability of the (letter,key) pair												
		A	B	C	D	E	F	G	H	I	J	K	L	M
High	AAA TTT	AA	IT	CA	DA	EA	AF	TN	HA	IA	SR	RT	AL	IE
Middle	ACA TGT	IS	NO	LR	LS	NR	SN	EC	TC	EE	CH	IC	TS	MA
Low	CAC GTG	HT	BA	OO	MR	TL	MT	AG	ED	BH	FE	GE	EH	TT
Low	CGA GCT	NN	UH	UI	VI	IW	RO	OS	NI	OU	WI	MY	ID	YO
Low	GGC CCG	MO	QL	YE	YF	CC	DC	PR	CF	TP	PU	WO	RU	FH
Low	ATC TAG	EW	WF	PN	PO	MS	EB	LV	PS	RR	DG	SS	YN	RV
Low	CGT GCA	LP	DY	HV	WH	BD	LU	FB	WL	MV	VC	KA	QV	US
Low	CGC GCG	YC	MP	BB	CB	PP	ZG	KW	GI	CG	IB	DH	FG	JD
Low	TTG AAC			GW	KT		YH			QS	MX	NX	WP	
Low	TAG ATC													

Fig. 2. (Plain letter, key letter) pairs and their probabilities. “as discussed by Yaseen [8]”

1) DNA Encoding of Character Pairs

For representing the ciphertext characters and creating the search space for the proposed sticker model, the character format (plain letter, key letter) is converted into DNA code. A triple of the four genetic bases provides 64 DNA codes which are sufficient to encode the English capital letters, although many language characters are not shown in the probability table pairs. In the proposed model, each memory strand region has its own random coding distribution that does not overlap with the distribution of other regions and each region represents one character of the ciphertext. Fig. 3 describes this guess work for the first character from each pair where the basic code has been adopted, and the character associated with it, the complementary code. Each plain letter has a main genetic code and the letter it is associated with appears as the complementary genetic code. For example, the two letters in the Fig. 3 A as a plain text letter are given the main code AAA, and the key letter associated with it is the letter A, which is given the complementary genetic code TTT. The current encoding list of genetic bases is the same and it is for all ciphertext characters because each ciphertext character represents a region; so these encodings do not conflict. The DNA encodings are constant for all columns in the probabilistic table, and for all ciphertext letters, but their interpretation for each is different from one to another.

2) Memory Strands, Complex memory, and Tube

The memory strands are formed from the codes of the triple genetic base of pairs (plain letter, a key letter) consisting of ciphertext characters; each of which represents a region, as well as from its complements that are linked to it to form the

Level of Probability and codons	Probability of the (letter,key) pair												
	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
High AAA TTT	NA	OA	PA	II	AR	SA	TA	UA	ER	SE	TE	YA	IR
Middle ACA TGT	LC	IG	CN	ND	NE	OE	CR	HN	NI	WA	IP	UE	NM
Low CAC GTG	UT	TV	TW	CO	TY	FN	OF	CS	OH	OI	GR	RH	LO
Low CGA GCT	FI	WS	EL	SY	MI	DP	EP	DR	CT	DT	SF	LN	HS
Low GGC CCG	RW	NB	HI	ME	OD	HL	GN	IM	VA	UC	DU	WC	GT
Low ATC TAG	KD	KE	BO	LF	PC	RB	HM	GO	SD	PH	BW	GS	EV
Low CGT GCA	VS	CM	YR	UW	LG	MG	IL	BT	UB	RF	AX	TF	UF
Low CGC GCG	MB	DL	KF		VV	WW		PF	QF			QI	WD
Low TTG AAC	GH	HH	DM									DV	
Low TAG ATC		YQ											

Fig. 3. The DNA encoding for (plain letter,key letter) pairs

memory complex parallel processing of the proposed model, and changing the value of the logical index of the region to 1. Through the parallel processing time, the tubes collect memory complex via the confluence of regions where the logical index equals 1, and the correct path leading to the solution can be inferred from among the paths of possible solutions through the complete tubes. Fig. 4 describes the transition from the ciphertext stage to the formation of memory strands and stickers, passing through the encoding of characters with genetic bases. The ciphertext letters are converted into pairs of plaintext letters and key letters, where the ciphertext branches into multiple paths from these pairs based on the possibilities of Fig. 2. Then each pair is given the genetic code and its complement through Fig. 3. After that, it is created memory strand and complex memory to begin the parallel search process to find the optimal path to the solution.

V. Implementation of Proposed Model

A. Steps of the Proposed Work

The cryptanalyzing and attacking of the proposed model are classified as the ciphertext only attack, and its execution begins by entering the ciphertext with a specific and limited length n. Through use of the computed table in a probabilistic model, its form changes in the first stage to (plain letter, key letter) pairs in the same order of ciphertext characters. These pairs are then swapped with triple genetic random codes on their complements. Parallel processing begins with the creation of each search path formed from the regions of the memory strands as well as the generation of their own stickers. Over the bypassing of the processing time, memory complexes arise; which in turn leads to the formation of tubes

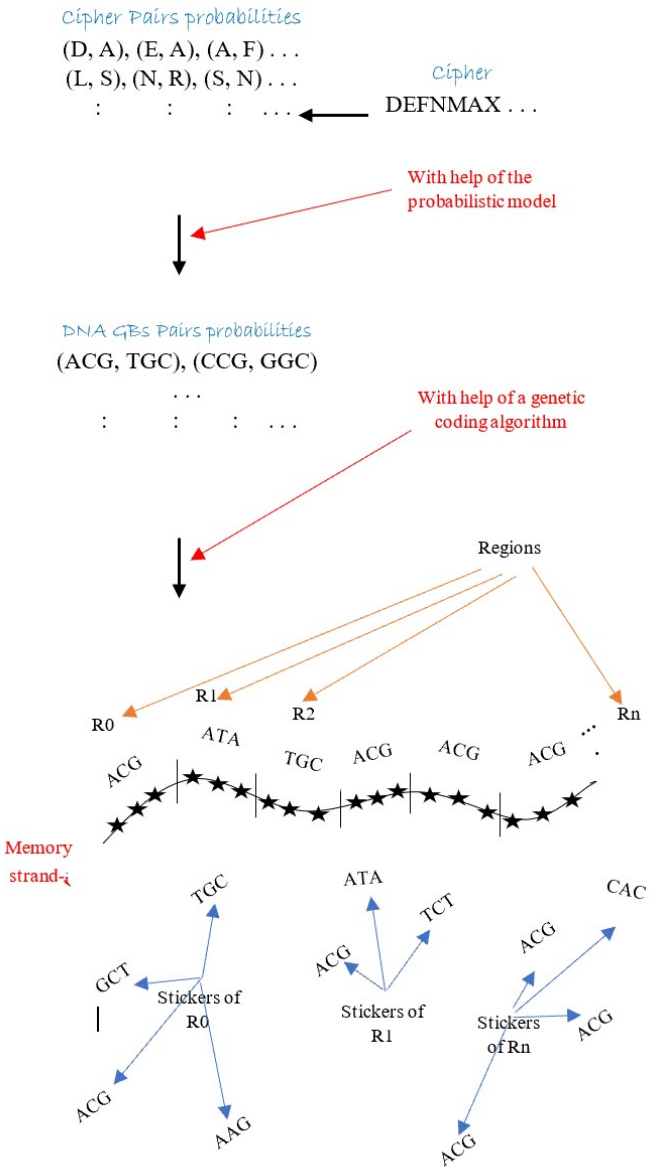

Fig. 4. Convert the cipher text into memory strands and stickers

according to the mechanism of action of the proposed sticker model. Fig. 5 represents the overall scheme of the proposed model, the mechanism of which can be summarized in the following specific steps:

B. Algorithm steps

Input: ciphertext sequence. Output: solution path(sequence of DNA bases). Steps:

1. Input ciphertext sequence of a specific length n.

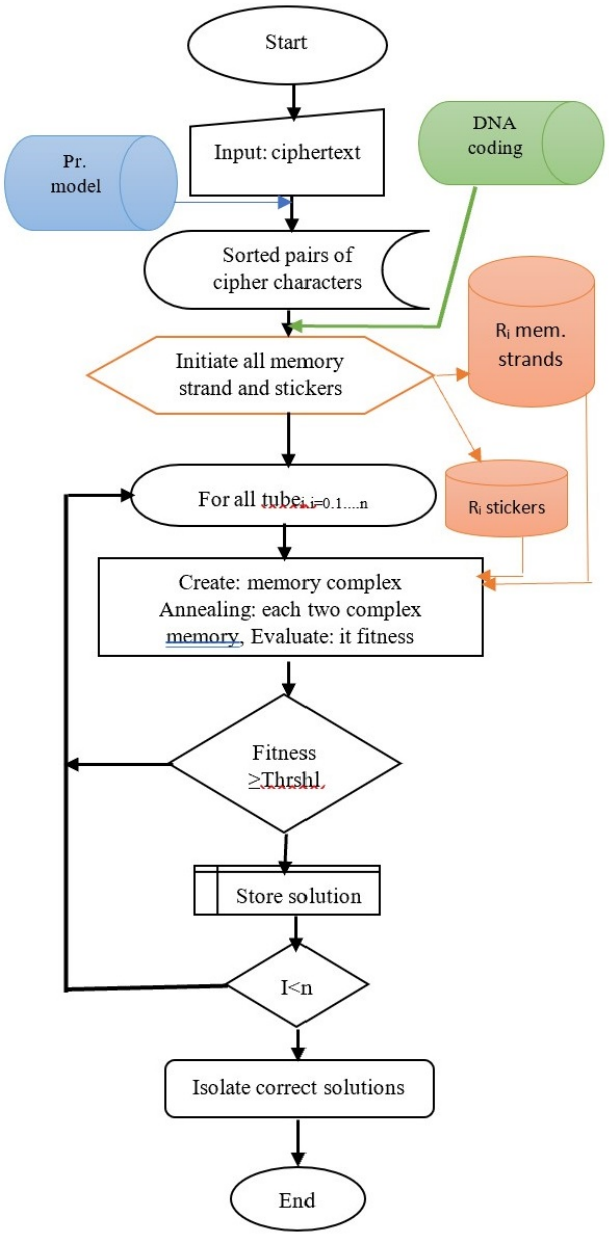

Fig. 5. Overall flowchart of the proposed model

2. Through the probabilistic model, define n (plain letter, key letter) pairs for each encrypted letter.

3. Encode (plain letters, key letter) pairs of creating n DNA codes.

4. Generate all paths (memory strands of n length and their stickers).

5. Through parallel processing, the create of memory complexes and tube structures.

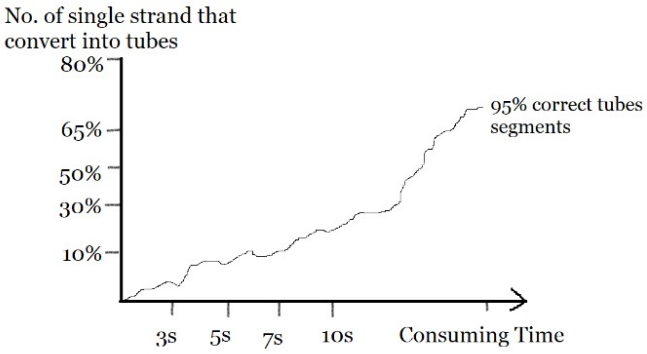

Fig. 6. The curve of developing the proposed model processes

6. Isolate the tubes representing the correct solution paths.

VI. Results of the Proposed Work Execution

The practical aspect of the proposed model was implemented and executed on a computer with the specifications: the computer type is HP-Spectre, the processor is Intel®core™i7-75600 cpu @ 2.4 GHz, 64GB RAM, X64-based processor. When executing several various polyalphabetic ciphers, the final results appeared with an average consumed time of 5-15 minutes, according to the length of the ciphertext. As an example of these ciphers, the ciphertext LOILBSLCFNULVJPXZGV was computed by the Vigenere algorithm using the plaintext THISISACONTRIBUTION and the encryption key SHATTALARABUNIVERSITY. The execution has given several tubes representing the correct solution paths but with varying degrees of accuracy. The result of the best of these tubes which are interpreted as shown in Table I, in which the values of the all-logical variables for the regions are equal to 1. From Table I, for example, the output of area 1, the code ACA and its sticker TGT, has the highest probability in the two letters S and T, especially if the letter T is a plain letter and S is a key letter, and so on for the other strand areas in the table. The process is detailed in the overall flowchart in Fig. 5.

VII. Discussion

After the practical execution of the parallel processing for the proposed model, the process of linking region coded in memory strands with their complement stickers and forming memory complexes begins in a typical time. Then their logical variables are set, which is a very fast process due to their limited number for each region of the memory strand. Throughout, the process of creating memory complexes and then, the tubes begin to appear where the sticker link memory

TABLE I.
THE BEST CORRECT SOLUTION PATH (TUBE)

Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Memory DNA	ACA	GTA	AAA	ACA	AAA	AAA	AAA	AAA	CGA	AAA	GTA	GTA	ACA	GTA
Sticker	TGT	CAT	TTT	TGT	TTT	TTT	TTT	TTT	GCT	TTT	CAT	CAT	TGT	CAT
Letters pair	TS	HH	IA	ST	IT	SA	AL	CA	OR	NA	TB	RU	IN	BI
Plain letter	T	H	I	S	I	S	A	C	O	N	T	R	I	B
Key letter	S	H	A	T	T	A	L	A	R	A	B	U	N	I

TABLE II.
THE COMPARISON BETWEEN THE PROPOSED MODEL WITH OTHER WORKS

Reference of Technique	The Length of the cipher effects the attack time	Solution Accuracy	OTP Cipher Problem	Cipher Text Size
9	Significant	More than 95%	No	Hard to Expand
10	Significant	More Than 95%	No	Hard to Expand
11	Relative	100%	Yes	It may be expanded
12	Relative	100%	Yes	It may be expanded
13	Relative	More Than 95%	Yes	It may be expanded
Proposed model	Constant	More Than 95%	Yes	Expansion has no effect

complexes to another in varied regions. After these creation operations are completed, the remaining execution time is devoted to isolating the formed solution paths (tubes) and searching through them, to achieve the correct solution. Fig. 6 demonstrates the mutation from single regions strands with stickers to the memory complex strands with stickers, and then constructing the tubes, Fig. 6 demonstrates the curve of the developing of proposed model processes, the mutation from single regions strands with stickers to the memory complex strands with stickers, and then constructing the tubes. The rate of the correct solution achieved is 95% , with the rate of interacting single strands to achieve the tubes is 75% .

VIII. CONCLUSION

In any case, the execution time is a typical time compared to the complexity of the paper problem which is a difficult one to solve by known methods. The achieved results favor the proposed method over all known and used methods for cryptanalyzing and attacking such ciphers, especially with known measures; the most important of which are:

• attacking with ciphertext only attack,

• having a valid result in any case,

• going beyond the issue of having a ciphertext with a specified length for methods that rely on the attacked ciphertext for statistical natural language analysis, as in the genetic algorithm, brute force, and correlation attack,

• The proposed method achieves the concept of efficient contribution since its essence was not previously discussed.

When comparing the proposed model with the works mentioned in the Review of Literature paragraph, and through some effective characteristics such as: Effect of the length of the cipher on the attack time, solution accuracy, cipher is OTP or not, and cipher text size, the results were as shown in the Table II.

CONFLICT OF INTEREST

The authors have no conflict of relevant interest to this article.

REFERENCES

[1] S. Roweis, E. Winfree, R. Burgoyne, N. V. Chelyapov, M. F. Goodman, P. W. Rothemund, and L. M. Adleman, "A sticker-based model for dna computation," *Journal of Computational Biology*, vol. 5, no. 4, pp. 615–629, 1998.

[2] K.-H. Zimmermann, "Efficient dna sticker algorithms for np-complete graph problems," *Computer Physics Communications*, vol. 144, no. 3, pp. 297–309, 2002.

[3] B. S.Yaseen, *Stream Cipher Cryptanalysis Using DNA Algorithms*. Doctorate Dissertation, Babylon University, Iraq, 2019.

[4] L. H. Dung, T. M. Duc, and B. Truyen, "Variant of otp cipher with symmetric-key solution," *Journal of Science and Technique-Le Quy Don Technical University*, p. 213, 2020.

[5] P. Bonavoglia, "Trithemius, bellaso, vigenère origins of the polyalphabetic ciphers," 2020.

[6] N. Nagaraj, "One-time pad as a nonlinear dynamical system," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 11, pp. 4029–4036, 2012.

[7] D. Rijmenants, "The complete guide to secure communications with the one time pad cipher," *Cipher Machines & Cryptology*, pp. 1–27, 2010.

[8] B. S. Yaseen, "Constructing probabilistic model for vigenere otp cryptanalysis," *Materials Today: Proceedings*, vol. 60, pp. 1747–1752, 2022.

[9] R. Siromoney and B. Das, "Dna algorithm for breaking a propositional logic based cryptosystem," *Bulletin of the EATCS*, vol. 79, pp. 170–177, 2003.

[10] A. S. Polenov, "The computing of np-complete problems in polynomial time using dna-logic," *World applied sciences journal*, vol. 30, no. 9, pp. 1188–1192, 2014.

[11] S. B. Sadkhan and B. S. Yaseen, "A dna-sticker algorithm for cryptanalysis lfsrs and nlfsrs based stream cipher," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, pp. 301–305, IEEE, 2018.

[12] S. B. Sadkhan-SMIEEE and B. S. Yaseen, "Db based dna computer to attack stream cipher," in *2019 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE)*, pp. 230–233, IEEE, 2019.

[13] S. B. Sadkhan and B. S. Yaseen, "Hybrid method to implement a parallel search of the cryptosystem keys," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, pp. 204–207, IEEE, 2019.

[14] N. A. Abdulmehdi and S. A. Kadum, "Cryptanalysis using dna-sticker algorithm," in *Journal of Physics: Conference Series*, vol. 1818, p. 012088, IOP Publishing, 2021.

[15] B. S. Yaseen, "Cryptanalysis of otp cipher using probabilistic splicing dna computer," *Journal of Design of engineering*, no. 8, pp. 10739–10748, 2021.

[16] B. S. Yaseen, "Splicing dna model for unknown stream cipher cryptanalysis," in *2021 2nd Information Technology To Enhance e-learning and Other Application (IT-ELA)*, pp. 46–51, IEEE, 2021.