An Enhanced DNA Sequence Table for Improved Security and Reduced Computational Complexity of DNA Cryptography

Maria Imdad, Sofia Najwa Ramli(✉), Hairulnizam Mahdin, Boppana Udaya Mouni, and Shakira Sahar

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Malaysia
maria.imdad123@gmail.com, {sofianajwa,hairuln}@uthm.edu.my, mouniudaya@gmail.com, shakirasahar@gmail.com

Abstract. Recently, DNA cryptography rejuvenates the art of secret writing by combining biological information and cryptography. DNA's double-helical structure serves as a template for encoding decoding information, vast storage and randomness. The structure includes DNA encryption that uses a DNA sequence table to substitute plaintext into the DNA sequence. However, this encoding table can result in leakage of information about the plaintext, character frequency, and key, by carefully examining the ciphertext through frequency analysis attack. Therefore, this paper proposes an enhanced DNA table for all 96 printable ASCII characters which are created to improve the entropy so that the probability of each encoding base (A, T, C, G) is equally likely and to reduce the computational complexity of DNA cryptography. An algorithm has been selected to implement both tables for performance measurement. The results show that encoding and encryption time is reduced, high entropy ciphertext, better frequency distribution ciphertext is obtained. Information leakage in terms of conditional entropy is also reduced by the proposed table. In conclusion, the proposed table can be used as a DNA sequence table in DNA cryptography to improve overall system security.

Keywords: DNA cryptography · DNA sequence table · Entropy · IoT application

1 Introduction

Novel encryption techniques tend to ensure system security more than traditional cryptographic methods either by combining two or more traditional techniques or by taking the advantage of biological characteristic of DNA encryption [1, 2]. With advancements in technology, more efforts are to ensure system security from attacker's perspective. A recent study sheds light on the fact that traditional cryptographic solutions either symmetric or asymmetric encryption are not secure any longer and cannot be used directly

M. M. Alam et al. (Eds.): BODYNETS 2020, LNICST 330, pp. 106–120, 2020.
https://doi.org/10.1007/978-3-030-64991-3_8

as a standalone solution [3]. One of the presented solution is using DNA Cryptography only or in combination with traditional cryptography, where DNA computing, encoding, decoding, and biological simulation processes yield better security [4]. Recent advancements in DNA cryptography present the solutions based on symmetric and asymmetric cryptographic systems, which when deployed showed improved usability [5]. These advancements are primarily because of DNA computing being fast and secure than existing technique and indicating that in near future DNA chips will replace silicon chips in computer systems for highly fast processing. A single gram of DNA has 1021 bases which are equal to 108 of data, due to the compactness in the double-helical structure of DNA [6].

DNA computing was started by L. Adleman in 1994 to solve complex computational problems primarily [7]. Recent development shows that it uses DNA cryptography to ensure system security for the next generations. Specifically, proposed encryption algorithms apply DNA cryptography with a DNA table that serves as a foundation for encoding and decoding the ASCII characters using DNA bases. DNA has four bases which are known as Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) for encoding. Initially, researchers in [8] have designed a base table using those bases and most techniques use the table either with or without iterations currently. This means that in one iteration code, “TTTT” sequence is representing A in the plaintext characters but in other iterations, this code may be used to represent Y. Only the code positions against characters changes but the sequence remains the same, and this is done to ensure that same plain text is retrieved at the other end after passing through this table. However, the analysis of the DNA encryption technique shows that frequency analysis reveals information about the key [9]. Thus, this paper proposes an enhanced DNA encoding table that is designed keeping in view the frequency count, randomness in codes and entropy of the table. Later, the proposed table is compared with the table in [8], followed by a complete encryption-decryption process and performance analysis.

The rest of the paper is organized as follows: Sect. 2 has a review of the work done so far. Section 3 gives a detailed insight into the proposed table. Section 4 has an experimental analysis, statistics followed by conditional entropy and its calculation and lastly Sect. 5 concludes this paper.

2 Related Work

DNA cryptography is an umbrella having technologies that are inspired by genetic entity DNA, ranging from Polymerize Chain Reaction (PCR) of DNA synthesis to digital coding using the same bonding and stimulation patterns as defined in DNA by nature [9]. This paper focuses on the study of techniques where DNA coding has been used to improve system security. DNA coding has four basic nitrogenous bases A,T,C,G and their representation using binary bases 0 and 1. Several DNA based cryptographic techniques have been proposed where DNA bases are used in combination with one-time pad (OTP) [10–12]. The technique in [10] uses microdots to save ciphertext while PCR is used at the decryption end. Whereas researchers in [11] use basic mathematical operations like DNA addition, DNA subtraction combined in a Feistel structure for system security. Meanwhile, the technique in [13] uses a lookup table to rearrange DNA bases, where a dynamic table for 256 ASCII characters is created. Then, it applies iterations to change the positions of characters which followed by a mathematical series before it uses asymmetric encryption for the encryption-decryption process. The combination of the output from the asymmetric encryption process with chunks of dynamically encoded text produces the ciphertext.

Later, researchers in [14] introduce a dynamic ASCII table where random ASCII characters are assigned to DNA bases initially. The dynamic ASCII table brings a new insight into DNA cryptography that random table results in different ciphertexts with the same plaintext making it challenging for an attacker to get access to the table along with iteration. A mathematical series is used for iteration purposes in which every iteration changes the position of the characters dynamically. For example, plaintext "A" may be encoded as "AAAA" in one iteration but in the next iteration, it may be "ACCT" to provide the randomness to the ciphertext. The same plaintext is encoded over different iteration to yield different ciphertexts. This encoded text is modified using OTP and is followed by genomic conversion. The final ciphertext is a compressed form of the genomic conversion into an amino acid table.

A biological simulation-based technique is proposed in [15] whereby a unique DNA based encoding table has been introduced. In this technique, a random encoding table is introduced after every session resulting in session-based output. The approach encodes the same plaintext that has different outputs in every different session. On the other hand, authors in [16] remove limitations of OTP ciphertext using DNA and amino acid coding, followed by randomness evaluation using NIST tests. A biotic-DNA oriented secret key mechanism is introduced in [17] and they use genetic information gathered from biological systems. The technique in [18] is a combination of digital coding, traditional cryptography as well as PCR amplification. Digital coding and traditional encryption are used to encode the plaintext followed by PCR for key generation. The technique in [19] has the underlying foundation of the signature method and asymmetric encryption with DNA. Initially the plain text is converted into ASCII codes followed by binary code, and is transformed into a matrix. This data is transmitted physically as a biological molecule in DNA.

DNA cryptography is not limited to text encryption but is equally applicable in image encryption as well [20]. A technique proposed by Zhang et al. [21] is image-based encryption comprised of map lattices of linear, as well as non-linear coupled with spatiotemporal chaos. A technique proposed in [22] is a combination of the hyperchaotic system along with a genetic recombinant, for image encryption where the system proved good security for image encryption. Zhang et al. proposed an image encryption system [23] based on permutation algorithms. Mix chaotic mapping in addition to Josephus traversing is used in [24] for image encryption. A combination of the chaotic system has been proposed in [25] which gave good encryption results than the previous techniques. DNA cryptography has been implemented in cloud computing to improve system security by enabling socket programming [26, 27]. In [28] an architectural framework has been proposed where digital signatures have been used in combination with DNA. Robust DNA codes based on DNA sequence has been proposed in [29, 30].

A technique proposed in [31] has used the same base table as in [8] and it uses DNA computing for intrusion detection. In this technique, DNA encoding is used to convert the network traffic data into DNA sequences. The idea behind selecting DNA computing is that it follows the same mechanism to detect diseases as an intrusion detection system does. The results of the technique show that DNA can be used for intrusion detection and can give better results using a better encoding method.

Analyzing the existing work proves that all these techniques use the same base table as in [8] with or without iterations. This base table has a unique code for all ASCII printable characters, alphabets, capital, small, numbers and special characters. This table has 96 codes for 96 different characters which are 26 capital, 26 small, 10 numbers and 34 special characters. The bases contain binary coding, A = 00, T = 01, G = 10, C = 11. Each character is coded using DNA bases such as "y" is coded as "AAAA". These codes are unique, indicating no two characters can have the same code. Where primarily this table is used to introduce randomness in cipher text. This table being randomly generated can yield more or less randomness across iterations, so by carefully examining this fact a new table has been created, which will always yield better randomness. New table is static in nature but can improve security as it has more random encoding as compare to base table. Frequently occurring characters have high random codes than less frequently occurring. Where a detail description is provided in Sect. 3. Keeping these facts into consideration, a new encoding table has been designed and these two tables will be compared based on different parameters.

3 Proposed DNA Table

The proposed table has unique codes for all 48 * 2 = 96 matrix as of base table as in Table 1. These codes are not assigned randomly but have the underlying foundation as follows:

- Frequency count of "71,013,156" characters from [32], and additional "5747" characters online, to rank these characters according to the number of occurrences.
- A character that has high-frequency count will have codes with all four bases without being repeated in a particular order, similarly going down to least frequency characters with repetition of DNA bases. High-frequency character "t" is coded as "ACTG", having all four bases and when it is converted into its binary code is "00110110". Meanwhile, the same character is coded as "TCCC", 01111111 in [8] having more repeating bits in a sequence. This is due to the repetitions of English language basic characteristics. Thus, the code should have all four bases to reduce repetitions so that the bits are more random and less predictable for the attacker.
- Each DNA base is counted exactly 96 times as in Table 1. This has been carefully selected and designed having the probability 96/384 for each base and the final entropy factor.

Table 1. The proposed DNA encoding table

Rank	Character	Frequency	DNA code	Rank	Character	Frequency	DNA code
1	e	7,741,972	ATCG	49	H	123,634	CGTC
2	t	5,507,785	ACTG	50	x	123,585	CTGC
3	a	5,263,861	ACGT	51	7	120,193	AGAC
4	o	4,729,276	ATGC	52	W	107,223	ACAG
5	n	4,535,686	AGTC	53	L	106,998	ATAG
6	i	4,527,428	AGCT	54	O	105,776	AGAT
7	s	4,186,244	GCTA	55	F	100,951	ACAT
8	r	4,137,989	TACG	56	Y	94,312	ATAC
9	h	2,955,955	TCAG	57	G	93,618	TATC
10	l	2,553,528	TGCA	58	J	78,794	TCCA
11	d	2,369,920	TGAC	59	z	66,509	TGTA
12	c	1,960,612	TAGC	60	j	65,894	TATG
13	u	1,613,333	TCGA	61	U	57,512	TCTG
14	m	1,467,476	GATC	62	q	54,288	TGTC
15	f	1,296,945	GCAT	63	:	54,102	GACC
16	p	1,255,599	GTCA	64	)	53,753	GCGA
17	g	1,206,847	GTAC	65	(	53,472	GTGA
18	y	1,062,140	AGCG	66	$	51,586	GAGT
19	w	1,015,755	GACT	67	K	46,612	GCGT
20	,	985,065	CATG	68	;	36,839	GTGC
21	.	946,186	CTGA	69	V	31,104	CGCA
22	b	866,356	CTAG	70	*	20,772	CACG
23	v	653,397	CAGT	71	?	12,481	CTCG
24	0	546,333	CGTA	72	Q	11,872	CGCT
25	1	461,006	CGAT	73	/	8,198	CTCA
26	k	460,798	CATC	74	X	7,682	CACT
27	5	374,503	ACGC	75	&	6,539	AGAG
28	2	333,599	ATCA	76	Z	5,672	ACAC
29	T	325,562	ACTA	77	!	2,201	ATAT
30	S	304,999	AGTA	78	%	2,005	TGTG
31	“	284,771	ATGA	79	+	324	TATA
32	9	282,397	AGCA	80	>	89	TCTC

(*continued*)

Table 1. (*continued*)

Rank	Character	Frequency	DNA code	Rank	Character	Frequency	DNA code
33	A	280,987	ACGA	81	<	84	TCGC
34	M	259,574	TCGT	82	=	24	GAGA
35	–	252,382	TGCT	83	#	12	GTGT
36	C	229,383	TAGT	84	@	3	CACA
37	I	223,370	TGAT	85	{	2	CCCG
38	N	205,465	TCAT	86	}	2	CGCC
39	‘	204,593	TACT	87	[	2	TGGT
40	4	192,545	GATG	88	]	2	GTTG
41	3	187,640	GTAG	89	ˆ	2	GAAG
42	8	182,681	GTCC	90	_	2	GTTG
43	B	169,490	GCTG	91	|	1	TAAT
44	6	153,881	GACG	92	~	1	TCCT
45	R	146,455	GCAG	93	\	1	CAAT
46	P	144,300	CTAC	94	‘	1	GAAT
47	E	138,459	CGAC	95	€	1	CGGT
48	D	129,645	CAGC	96	£	1	GTAA

The subsequent section gives a detailed insight into how this entropy is calculated and what is the ideal value for this entropy.

3.1 Entropy

Entropy is the measure of uncertainty in bits and this concept was introduced by Shannon in 1948 [32, 33]. The uncertainty of the cipher is the number of plaintext bits that must be recovered from scrambled ciphertext to get the message back, and this is measured via entropy. The entropy of a variable is the weighted average of optimal bit representation size such as the average size of an optically encoded message. Mathematically, entropy can be defined as in (1) [9, 34].

$$H(X) = -\sum_{x \in X} \Pr[X] log_2(\Pr[X]) \tag{1}$$

Meaning, higher the probability of an event less the uncertainty. Here we are calculating the entropy of X with four bases as $X = \{A, C, T, G\}$. Now, the probability of each DNA base multiplied by log of its probability as modified for four values as in (2):

$$H(X) = -\big[P(A)log_2(P(A)) + P(C)log_2(P(C)) + P(G)log_2(P(G)) + P(T)log_2(P(T))\big] \tag{2}$$

The highest uncertainty is only achieved when the values are equally distributed. Figure 1 explains that the probability of an event that ranges from 0 to 1 and the entropy can range from 0 to 1. This graph gives an insight into the entropy of an event where two outcomes are considered:

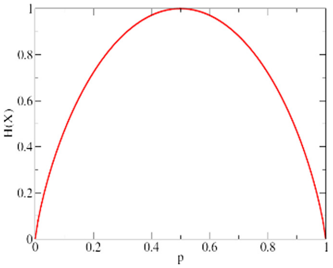


Fig. 1. Entropy and probability distribution

- The value of entropy is 0 for both the least and highest probability, which proves that if the probability of occurrence is 0, an event entropy will be 0, indicating that event will never happen. Similarly if the probability of an event is 1 means this event will always happen against the entropy is 0 because, in this scenario, there is no uncertainty about the information.
- The entropy of the system is maximum "1" when the probability is "1/2 = 0.5". This clearly indicates that all events have the same chance to occur. If the probability increases from "0.5" then entropy decreases and similarly if the probability decreases the entropy also decreases, because in a former event is less likely to occur whereas in later the event is more likely to occur.
- For a system where number of events increases, for example, it has four possible outcomes and the probability range from 0 to 1, its distribution differs. Each event has equal probability 1/4 = 0.25 only then maximum entropy will be achieved. Here the entropy reaches a maximum value which is 2.

3.2 Entropy Calculation for DNA Table

This section calculates the entropy of the base table that originally described in [8] and followed by the entropy of the proposed table.

Entropy of Base Table

The entropy of the base table is calculated by examining the number of occurrences of each DNA base table in [8], whereby A = 83, C = 106, G = 96, and T = 99. And the total number of bases in the table is 384. By substituting these values in Eq. (2),

$$H(X) = -\left[\frac{83}{384} log_2\left(\frac{83}{384}\right) + \frac{106}{384} log_2\left(\frac{106}{384}\right) + \frac{96}{384} log_2\left(\frac{96}{384}\right) + \frac{99}{384} log_2\left(\frac{99}{384}\right)\right]$$

$$H(X) = 1.9164$$

Hence the minimum number of bits needed to encode all possible meaning of the table or number of bits of information per character is 1.92.

Entropy of the Proposed Table

The entropy of the proposed table is calculated by examining the number of occurrences of each DNA base in the table, where A = 96, C = 96, G = 96, and T = 96. And the total number of bases in the table is 384. By substituting these values in Eq. (2),

$$H(X) = -[\frac{96}{384}log_2\left(\frac{96}{384}\right) + \frac{96}{384}log_2\left(\frac{96}{384}\right) + \frac{96}{384}log_2\left(\frac{96}{384}\right) + \frac{96}{384}log_2\left(\frac{96}{384}\right)$$

$$H(X) = 2$$

The minimum number of bits needed to encode all possible meaning in the proposed table or number of bits of information per character is 2. Higher entropy makes conducting frequency analysis harder in DNA cryptography. The DNA table with higher entropy introduces more uncertainty about the ciphertext when an attacker does not have any information about the plaintext. Meanwhile, a lack of good entropy can leave a cryptosystem vulnerable and unable to encrypt data securely. Later, Sect. 4 discusses the experimental result of the entropy for the proposed DNA table.

4 Experimental Results

To compare the performance of both DNA tables, this section describes the encryption algorithm [14] of DNA cryptography as follows..

1. Read input (plaintext)
2. Create a DNA sequence using dynamic DNA table
3. Convert sequence into 2 bit binary
4. XOR binary sequence with random binary key of equal length
5. Convert the sequence of step iv into DNA sequence
6. Use mRNA table to convert sequence from step 5
7. Transfer mRNA to tRNA
8. Divide tRNA into two and interchange their positions
9. Apply the reverse simulation (U to T)
10. Generate cipher text using amino acid table

This algorithm is implemented in MATLAB R2019b. The next subsections describe the performance measurements, where outputs of base and proposed DNA table are compared in terms of frequency count, time entropy, and conditional entropy of key given cipher.

4.1 Frequency Analysis of Encoded Text

After taking the input from the user, the DNA table is used to encode the plain text. Four DNA bases A, C, T, and G replace each plaintext character so an input string of length, m becomes m * 4 in the encoded text. Given below are the graphs, indicating the frequency count of the input string of 3000 characters and 6000 characters as in Fig. 2 and Fig. 3 respectively. An input string of 3000 characters yields an encoded string 3000 * 4 = 12000 DNA bases. Figure 2 gives the insight of DNA bases frequencies. From Subsect. 3.2 with the calculation of the proposed table entropy, the encoded count of all DNA bases should be equally distributed, thus making a string of 12000 bases has 12000/4 = 3000 frequency per base. Figure 2 proves that the frequency count of the base table is not uniformly distributed, ranging from 1938–4830 when compared to the proposed DNA table with the range in between 2933 to 3036.

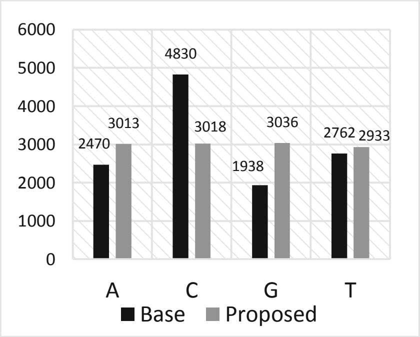


Fig. 2. Frequency count input 3000 characters

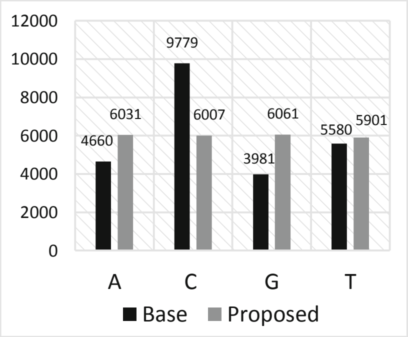


Fig. 3. Frequency count input 6000 characters

Figure 3 shows the frequency count using another input of 6000 characters that creates 6000 * 4 = 24000 DNA bases. Ideally, the frequency of each base should be 6000 or nearly equal to it, but the frequency count of the DNA base table ranges from 3981–9779 while the proposed DNA table creates the value near to 6000, ranging from 5901–6061.

4.2 Computational Time of Encoding Process

Figure 4 illustrates a time comparison to substitute the same plaintext of equal length using both DNA tables. The graph proves that time difference is negligible in case of small plaintext, but as the size of plaintext increases, the time required to encrypt the plaintext using the proposed DNA table has a significant difference as compared to the base table. Based on the graph, for an input of 50 characters, the time taken by the base table is 0.0539 ms while the proposed table is 0.0522 ms with the time difference is 0.0017 ms. But, as the size of plaintext increases to 1750 characters, the time difference is 0.0646 ms, and for 2500 characters, the time difference has increased to 0.9931 ms. This difference is mainly because of the frequency analysis performed prior to table creation, as most occurring characters are at the start so the loop does not iterate through the whole table. Instead, it immediately encodes the character and exits but in the base table, the characters are randomly placed so the time increases with the size of the input.

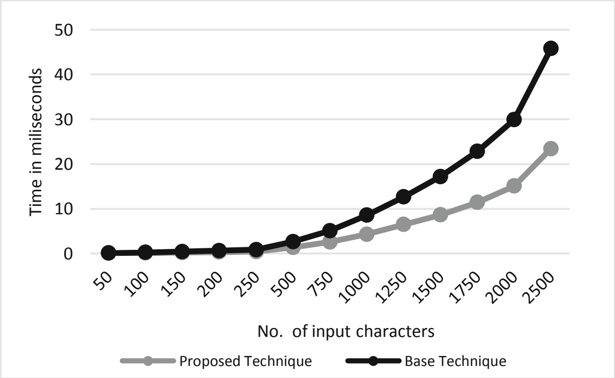
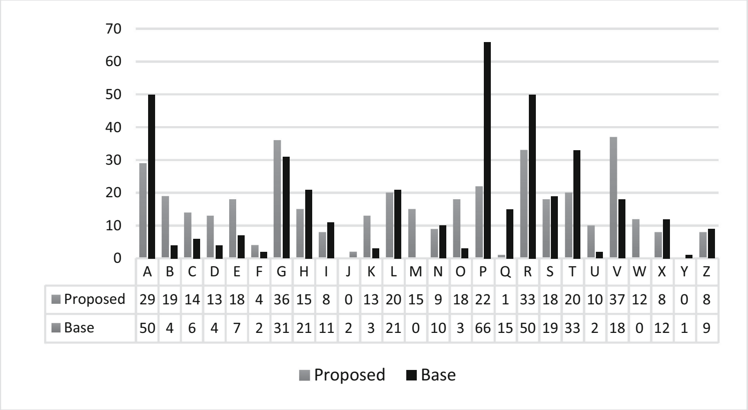


Fig. 4. Time comparison of encoding process using the DNA base and proposed table

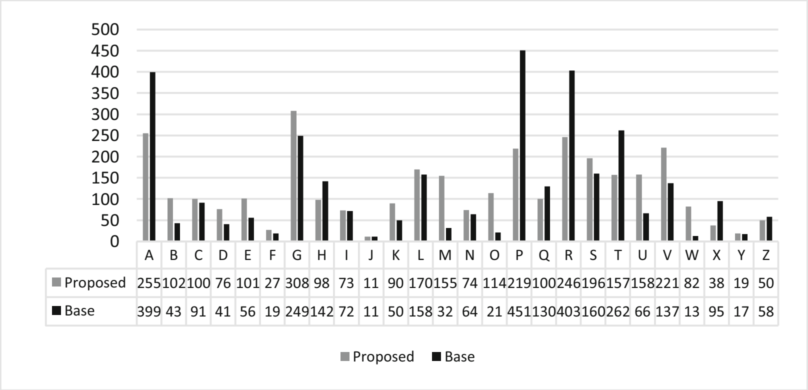
4.3 Frequency Analysis of Ciphertext

At the end of this algorithm, the DNA cryptography generates the ciphertext. The text DNA sequences are mapped to the amino acid table. The final ciphertext is in the form of 26 English characters. Figure 5 implies the frequency analysis of the ciphertext with 300 characters plaintext as the input of the encryption algorithm. Based on both DNA tables, it is obvious that frequency ranges from 0–66 for the base technique whereas it ranges from 0–36 for the proposed technique. Meanwhile, Fig. 6 gives a frequency analysis of 2400 characters plaintext as the input. For the base technique, the frequency count ranges from 11–451, while for the proposed table, it ranges from 11–308. This frequency difference is important from an attacker's perspective to conduct a frequency analysis attack due to the randomness of the ciphertext. This way, he cannot extract meaningful information about the plaintext from when he does not know the secret key [9, 35].



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
▨ Proposed	29	19	14	13	18	4	36	15	8	0	13	20	15	9	18	22	1	33	18	20	10	37	12	8	0	8
▨ Base	50	4	6	4	7	2	31	21	11	2	3	21	0	10	3	66	15	50	19	33	2	18	0	12	1	9

Fig. 5. Frequency analysis of ciphertext (with 300 characters)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
▨ Proposed	255	102	100	76	101	27	308	98	73	11	90	170	155	74	114	219	100	246	196	157	158	221	82	38	19	50
■ Base	399	43	91	41	56	19	249	142	72	11	50	158	32	64	21	451	130	403	160	262	66	137	13	95	17	58

Fig. 6. Frequency analysis of ciphertext (with 2400 characters)

4.4 Computational Time of the DNA Encryption Technique

Time is an important factor when it comes to the computational complexity performance of the algorithm. Figure 7 shows the encryption time that is calculated for multiple inputs. The inputs are the number of characters in the plaintext and computational time is the total encryption time in seconds. The same input is provided to both experiments using the base and the proposed DNA tables. The graph indicates that the encryption time for the proposed table is significantly less than using the base table. Based on the graph, for an input of 600 characters, the time taken by the proposed table is 0.1624 s while and the base table is 0.248 s. One of the factors for this time difference is because the table is designed by taking into account that the characters with high rank or frequency must be in start so that loop traverse time reduces.

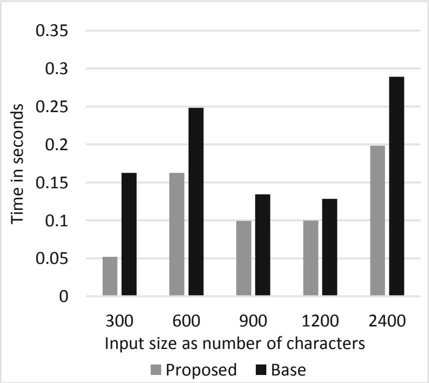


Fig. 7. Encryption time comparison

4.5 Entropy Analysis

The entropy of the cryptosystem varies as the number of inputs or sample space varies. Figure 8 gives the entropy of different ciphertexts, provided the same plaintext and the same random key. The X-axis is the input size or the number of characters, whereas the Y-axis is the entropy of ciphertext H(C). For input size ranging from 300 to 2400, it is obvious that the entropy of the proposed technique is more than the entropy of the base technique. This difference is because at the time of encoding, at a very early stage of encryption algorithm the frequencies of encoded text are nearly equally distributed, and highly ranked characters have most random codes. As explained in [9], the entropy of English characters is $26log_2 26 = 4.7$ which an ideal entropy value for the proposed cipher is. In the case of 2400 characters as the input, the entropy of the proposed technique is approximately 4.4 which is not very less than ideal value, while for the base table, the entropy is in between 4.09 to 3.91 which is less than ideal.

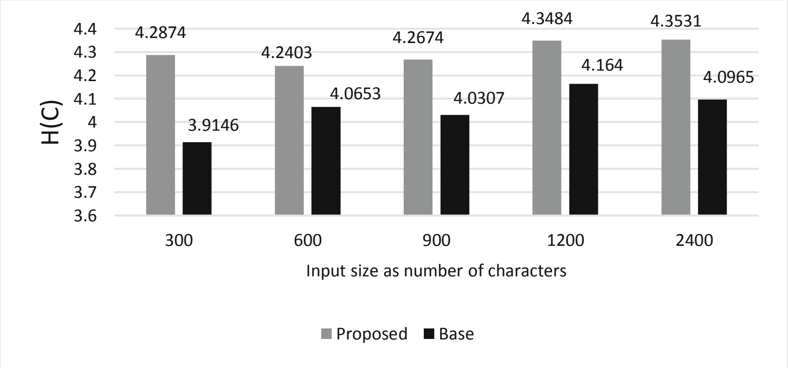


Fig. 8. Entropy of ciphertext

4.6 Conditional Entropy

According to Kerchoff's Principle, the security of the cryptosystem depends primarily on having strong keys and keeping them secret, but not in the encryption-decryption algorithm because it can be accessed [35]. Thus, the system must ensure that there is no information leakage about key or plain text from the ciphertext. Conditional entropy is also called key equivocation when it comes to cryptography. Key equivocation of a cryptosystem can be described as in (3) with M^N = plaintext with length N, C^N = ciphertext with length N and K = random key [24].

$$H(K/C^N) = H(K) + H\left(M^N\right) - H\left(C^N\right) \tag{3}$$

Key equivocation is a process in cryptanalysis where the attacker has only access to ciphertext and he tries to infer some information about the key. It is also known as a ciphertext-only attack. Given below is the calculation of conditional entropy in both scenarios whereby the encryption uses the base and the proposed table. Key and plaintext are the same for both cases with the number of input characters of 1200. Using the equation in (3) to calculate the conditional entropy for base table, $H(K/C^N) = 0.9994 + 4.7615 - 4.164 = 1.5969$ and the conditional entropy for the proposed table, $H(K/C^N) = 0.9994 + 4.47615 - 4.3484 = 1.1272$.

This conditional entropy is the information being leaked or it can be stated as the amount of information of the key with the given cipher. The result shows that the proposed technique reveals less information about the key than the base technique. Hence, the proposed table technique can serve as a good substitute for the base technique for improvements in terms of frequency analysis, computational time, and entropy.

5 Conclusion

Traditional cryptographic techniques are designed based on substitution, and transposition operations. With advancements of technology however compromise the security of the cryptographic algorithms. Researchers propose new security solutions to overcome security issues including DNA cryptography which involves PCR, DNA synthesis and digital coding. DNA coding is mostly used for encryption-decryption techniques with a basic table of 96 ASCII characters to encode the plaintext into DNA bases. This research work has improved that table, by carefully constructing the table based on frequency analysis, randomness in code and entropy of the table. Later, the paper compares the performance of the proposed table with the base table which originally designed for DNA cryptography. The results indicate that the proposed table gives a balanced frequency of occurrence in the encoded text and ciphertext, reduces encoding time based on DNA bases and encryption time, better entropy of the ciphertext and finally the conditional entropy is less than the base technique. The entropy of the proposed table is slightly less than the base table but it has a huge impact on the output. Hence, this table can be used for better security and computational time of DNA encryption techniques.

Acknowledgement. The authors would like to thank all reviewers for their helpful comments. The authors would also like to thank the Ministry of Higher Education Malaysia for supporting the research under Fundamental Research Grant Scheme Vot No. FRGS/1/2019/ICT03/UTHM/03/1 and partially sponsored by Universiti Tun Hussein Onn Malayisa.

References

1. Lu, M.X., et al.: Symmetric-key cryptosystem with DNA technology. Sci. China Ser. F: Inf. Sci. **50**(3), 324–333 (2007)
2. Anam, B., Sakib, K., Hossain, M.A., Dahal, K.: Review on the advancements of DNA cryptography. arXiv preprint arXiv:1010.0186 (2010)
3. Kabir, K.S., Chakraborty, T., Alim Al Islam, A.B.M.: SuperCrypt: a technique for quantum cryptography through simultaneously improving both security level and data rate. In: Proceedings of the 2016 International Conference on Networking System and Security, pp. 25–33 (2016)

4. Pelletier, O., Weimerskirch, A.: Algorithmic self-assembly of DNA tiles and its application to cryptanalysis. arXiv preprint arXiv:cs/0110009 (2001)
5. Galbraith, S.D.: Mathematics of Public Key Cryptography. Cambridge University Press, New York (2012)
6. Zhang, Y., He, L., Fu, B.: Research on DNA cryptography. In: Applied Cryptography and Network Security. InTech, Rijeka (2012)
7. Adleman, L.M.: To combinatorial problems. Science **266**(5187), 1021–1024 (1994). https://doi.org/10.1126/science.7973651. Bibcode: 1994Sci…266.1021A. CiteSeerX 10.1.1.54.2565
8. Noorul Hussain, U., Chithralekha, T.: Inventors; assignee. A novel DNA encoding technique and system for DNA cryptography. India Patent 5107, CHE, 2012, 7 December 2012
9. Othman, H., Hassoun, Y., Owayjan, M.: Entropy model for symmetric key cryptography algorithms based on numerical methods (2015). https://doi.org/10.1109/arcse.2015.7338142
10. Borda, M., Tornea, O.: DNA secret writing techniques. In: Proceedings of the 8th International Conference on Communications (COMM), pp. 451–456 (2010)
11. Zhang, X., Zhou, Z., Niu, Y.: An image encryption method based on the feistel network and dynamic DNA encoding. IEEE Photonics J. **10**(4), 3901014 (2018)
12. Zhang, Y., Liu, X., Sun, M.: DNA based random key generation and management for OTP encryption. Biosystems **159**, 51–63 (2017)
13. Biswas, R., Alam, K.M.R.: A technique for DNA cryptography based on dynamic mechanisms. J. Inf. Secur. Appl. **48**, 102363 (2019)
14. Hossain, E.M.S., Alam, K.M.R., Biswas, M.R.: A DNA cryptographic technique based on dynamic DNA sequence table. In: Proceedings of the 19th International Conference on Computer and Information Technology, 18–20 December 2016, North South University, Dhaka, Bangladesh. IEEE (2016)
15. Hussain, N., Rahman, U., Balamurugan, C., Mariappan, R.: A novel DNA computing based encryption and decryption algorithm. In: International Conference on Information and Communication Technologies (ICICT 2014), Procedia Computer Science, vol. 46 pp. 463–475 (2015)
16. Ibrahim, F.E., Moussa, M.I., Abdalkader, H.M.: A symmetric encryption algorithm based on DNA computing. Int. J. Comput. Appl. **97**(16), 41–45 (2014)
17. Babu, E.S., Raju, C.N., Prasad, M.H.M.K.: Inspired pseudo biotic DNA based cryptographic mechanism against adaptive cryptographic attacks. Int. J. Netw. Secur. **18**(2), 291–303 (2016)
18. Cui, G., Qin, L., Wang, Y., Zhang, X.: An encryption scheme using DNA technology. In: Proceedings of the 3rd International Conference on Bio-Inspired Computing: Theories and Applications, USA, pp. 37–42. IEEE (2008)
19. Lai, X., Lu, M., Qin, L., Han, J., Fang, X.: Asymmetric encryption and signature method with DNA technology. Sci. China Inf. Sci. **53**(3), 506–514 (2010)
20. Kulsoom, A., Xiao, D., Aqeel-ur-Rehman, Abbas, S.A.: An efficient and noise resistive selective image encryption scheme for gray images based on chaotic maps and DNA complementary rules. Multimed. Tools Appl. **75**, 1–23 (2016). https://doi.org/10.1007/s11042-014-2221-x
21. Zhang, Y.Q., Wang, X.Y.: A symmetric image encryption algorithm based on mixed linear-nonlinear coupled map lattice. Inf. Sci. **273**, 329–351 (2014)
22. Wang, X.Y., Zhang, H.L.: A novel image encryption algorithm based on genetic recombination and hyper-chaotic systems. Nonlinear Dyn. **83**, 333–346 (2016)
23. Zhang, W., Yu, H., Zhao, Y.L., Zhu, Z.L.: Image encryption based on three-dimensional bit matrix permutation. Signal Process. **118**, 36–50 (2016)
24. Wang, X.Y., Zhu, X.Q., Zhang, Y.Q.: An image encryption algorithm based on Josephus traversing and mixed chaotic map. IEEE Access **6**, 23733–23746 (2018)

25. Parvaz, R., Zarebnia, M.: A combination chaotic system and application in color image encryption. Opt. Laser Technol. **101**, 30–41 (2018)
26. Prajapati Ashishkumar, B., Barkha, P.: Implementation of DNA cryptography in cloud computing and using socket programming. IEEE (2016)
27. Hammami, H., Brahmi, H., Yahia, S.B.: Secured outsourcing towards a cloud computing environment based on DNA cryptography, pp. 31–36. IEEE (2018)
28. Chouhan, D.S., Mahajan, R.P.: An architectural framework for encryption & generation of digital signature using DNA cryptography, pp. 743–748. IEEE (2014)
29. Keerthana Priya, S.V., Saritha, S.J.: A robust technique to generate unique code DNA sequence, pp. 3815–3820. IEEE (2017)
30. Sadeg, S., Gougache, M., Mansouri, N., Drias, H.: An encryption algorithm inspired from DNA. In: Proceedings of the International Conference on Machine and Web Intelligence, pp. 344–349. IEEE (2010)
31. Rashid, O.F., Othman, Z.A., Zainudin, S.: A novel DNA sequence approach for network intrusion detection system based on cryptography encoding method. Int. J. Adv. Sci. Eng. Inf. Technol. **7**(1), 183–189 (2017)
32. Jones, M.N., Mewhort, D.J.K.: Case-sensitive letter and bigram frequency counts from large-scale English corpora. Behav. Res. Methods Instrum. Comput. **36**, 388 (2004). https://doi.org/10.3758/BF03195586
33. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
34. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(4), 623–656 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb00917.x
35. Mohan, M., Kavitha Devi, M.K., Jeevan Prakash, V.: Security analysis and modification of classical encryption scheme. Indian J. Sci. Technol. **8**(S8), 542–548 (2015). ISSN (Print): 0974-6846 ISSN (Online): 0974-564