Proposed Approach for Key Generation Based on the RNA

Assit Prof. Dr.Alia Karim Abdul Hassan
Computer Science Depart- University of Technology

Abstract

Due to fundamental role of random key in the design of cryptography algorithms, a new method is proposed for generating random key based on the bases of RNA translation to protein chain. The proposed method accept a key sequence size(9,18,27,…) byte and generate key sequence with extended length appropriate with plain message length. The security strength of generated key is acceptable according to the results of statistical tests of randomness.

General Terms

cryptography, key sequence, randomness, RNA, DNA

Introduction

Cryptographic is learning mathematical technology related aspects of information security such as confidentiality, data integrity, and authentication entity, and data origin authentication, encryption is not the only tool to provide protection of information [1]. It is not surprising, came new types of Cryptographic shortly after the large-scale development of computer-mediated communication, and the growth of Internet technologies. In data and telecommunications, and there is a need Cryptographic when connecting through any broker era, which includes just about any network, on top of the Internet [2].

The security of encrypted data is entirely dependent on the strength of the encryption algorithm and the secret of the key [3]. In cryptography, the key is part of the information which determines the yield practical algorithm for Cryptographic cipher. Without the key, the algorithm does not have a result. In Cryptographic, key to making a special on from plain text to determine the encrypted text, or vice versa during decryption. Also used in other keys encryption algorithms, such as digital signature systems and message authentication codes [4].

To prevent the key from being guessed, keys need to be creat, in fact randomly and contain sufficient universe. Since random number one value can not be predicted, and the computers are not very good at producing truly random data. Instead, they rely on the pseudo-random number generator (PRNG), so it must be strong encryption with random seed values really [5].

Deoxyribo nucleic acids (DNA) computing is to solve computational problems with the help of biological and chemical processes on the method of DNA strand [6]. Since then compounded more and more researchers of a promising future for this area and start working on it. The use of computing DNA on the other hand is far from reality and the world of information security is the focus better on other encryption technology for new promising methods [7]. Ribo Nucleic Acid (RNA) is a copy of the DNA to come out to the cytoplasm to tell the cell what needs to be done in order to survive[8]. In this paper a proposed method to generate a random key using RNA computing technology.

Related Work

In [9] use chemical properties of the DNA sequences of the cipher text to encrypt data over public channel to add key extension and complexity. In[10] develop a secured symmetric key generation scheme which generates primary cipher and this primary cipher is then converted into final cipher using DNA sequences, so as to make it again more complicated in reading.

Key generation

Key generation is the process of generating keys for cryptographic. Keys can be created by a different techniques, such as using the output of random bit generator. Cryptographic keys needed to be generated within strong cryptographic modules. Random numbers required for key generation must be generated within the module that generates the key. The security strength (*randomness*) that can be supported by a key depends on the algorithm with which it is used, the size of the key , the process that generated the, and how the key was handled[11].

Deoxyribonucleic Acid (DNA)

The DNA is a double-anti similar stranded helix of nucleotides is responsible for carrying the cell’s genetic feature, “code” which generates proteins. The DNA strands contain a huge number oflinked nitrogen-based polymer nucleotides. The nucleotides consist of Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). The nucleotides only combine in C-G and T-A pairs. DNA chain is formed in a free phosphate (**5’**end) and in a free hydroxyl group(**3’**end) with **3’** end of one strand pairs with the **5’** end of the other [7]. For This reason**5’** CTGA **3’**should not be confused with AGTC.

Ribo-Nucleic Acid (RNA)

A DNA gene hasthe information for makingthe right polypeptide by the process:

DNA ➔ RNA ➔ Protein

For cell decoding, DNA is copied into **messenger RNA**(mRNA), called **transcript**.The ribonucleic acid,RNA, is distinguished from DNA for having different chemical composition, where pyrimidine uracil is usedinstead of thymine. [12].RNA is a single-stranded moleculewhich contains the bases adenine (A), cytosine (C) and guanine (G) but not the thymine; instead,it contains uracil (U) base. The messenger RNA (mRNA)andtransfer RNA (tRNA) are two different types of RNA.

Messenger RNA (mRNA)

The information stored in ***mRNA***carries is utilized to make proteins; a copy of a gene(figure 1).

Transfer RNA (tRNA)

TRNA, an adaptor in protein synthesis,turns around to formribonucleotideswhich combine withothers within the same chain tomake3 loops (Fig. 1b tRNA[13,14].

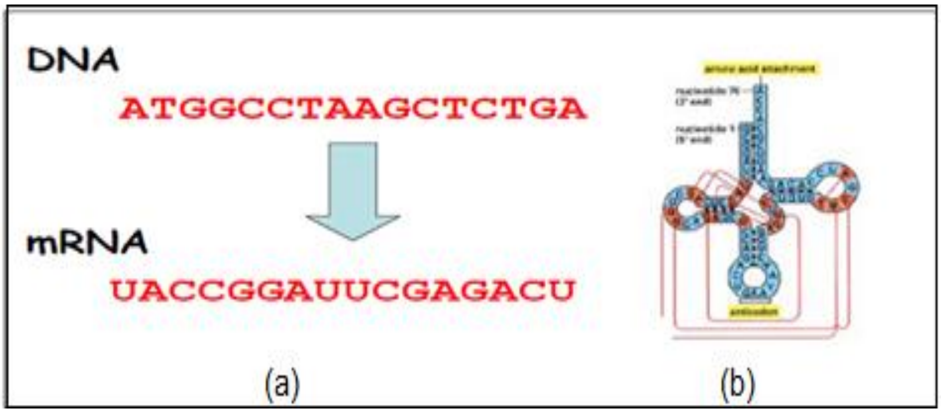


Figure 1: a) mRNA b) tRNA

Protein Synthesis

Protein synthesis procedure as the following :First, the information to code for a single amino acid, are made of three nucleotides (a triplet). This information is first *transcribed* into the messenger RNA (mRNA), which has a series of bases complementary with DNA, from which it is copied. In fact, mRNA, like DNA has only four bases, whereas proteins may contain up to 20 amino acids. Permutation of the 4 bases yield 4^3 or

64 triplets. The mRNA in turn serves as an intermediary that contains the same genetic information and *translates* this information into the amino acid sequence of the protein as shown in figure 2.

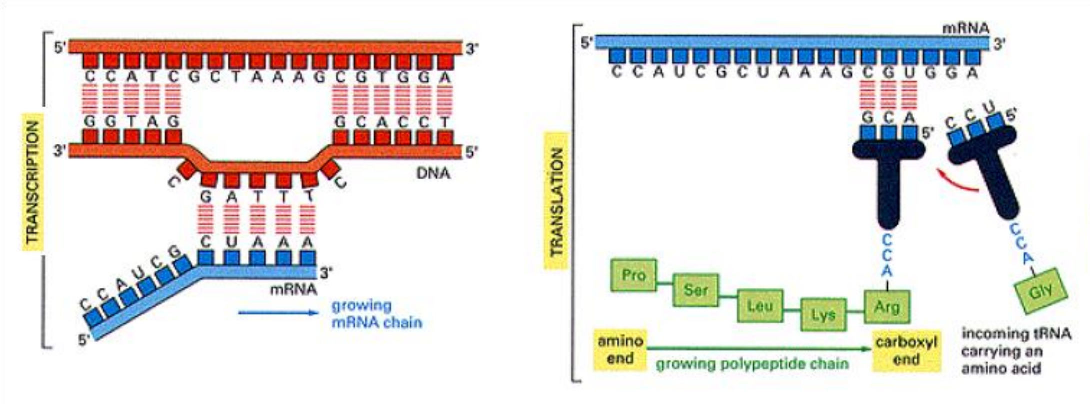
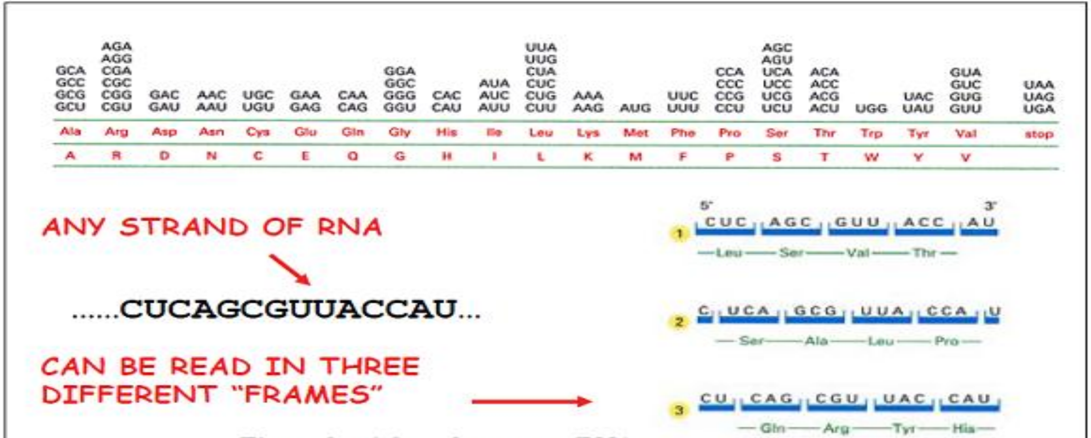


Figure2: informaion flow in protein synthesis

Figure3 illustrates many RNA codons which may code for a single amino acid. The leucine codon for instance could be coded by CUU, CUC, and CUA. The base occupying the triplet's third position is the only difference in synonymous codons. For coding, the codon's first two bases are the important ones in coding. For this reason, synchronous third base codons exchange could take place unnoticed.



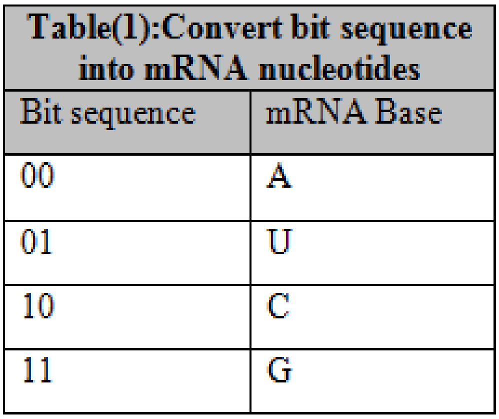
GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Figure 3: triple codones on mRNA

The mRNA translation to protein starts on the 5' end of the mRNA and codes for the beginning of a methionine amino acid. The starting action finishes when tRNA occupies one of the two binding ribosome sites. In figure 2, the next tRNA binds to the A site and all available tRNAs will go toward the site but only thetRNA whose anticodon is complementary to the codon of the A site will attach on the mRNA. Joining the amino acids together is described as peptide connection and this will go on until a "stop" codon (UAA, UGA, or UAG) endsthe process (Figure 2) [13,14].

The Proposed Approach for Key Generation

a proposed method for secure key sequence generation based on the bases of RNA translation to protein chain. The proposed method accept a key sequence size(9,18,27,…) byte these bytes are first converted binary code then every 2 bit are converted to one nitrogenous base, in order to get the RNA strand using table (1), split RNA strand to a group of codons (3 nitrogenous bases). As describe each RNA codon can be coded to different amino acide this property is employed to expand the mRNA key sequence. In a proposed table (table 2) each anticodon have a sequence with 6-bit length because the number of codons in RNA is 64(2^6bits). In the table(2) several RNA codons may code for a single amino acid are grouped in a parenthesis(}), group size from 2 to 4. This property is used for expansion the key sequence. If the group size is 2 use both codons to extend the cipher text from one codoin to 2 and the bit sequence 12bit. If the group more than 2 codons for example if there is a GGU with code number 60 it can be choose one of its alternative of same group as GGC, GGA, GGG to be expand the cipher sequence (each 6-bit will expanded to 12-bit). Choosing one of the group is made by agreement between the sender and receiver.



Table(1):Convert bit sequence into mRNA nucleotides

Bit sequence	mRNA Base
00	A
01	U
10	C
11	G

The process of protein chain is begins with AUG that is used to start protein synthesis and stop when find end is (UAA or UAG or UGA) and the translation of all the (codons) found them to protein. to take advantage of this feature in this work by identifying (start and end codon) and calculate the number of the (codons), and take this number as the number and benefit from the work rotate right shift of) RNA strand, for example, if the number of codons equal to 4, between the start and end meaning that shift right rotate will 4 codons. This step provide the generated key sequence more randomness. The proposed method for key generation is described in the algorithm1, followed with example describe the algorithem1 steps trace.

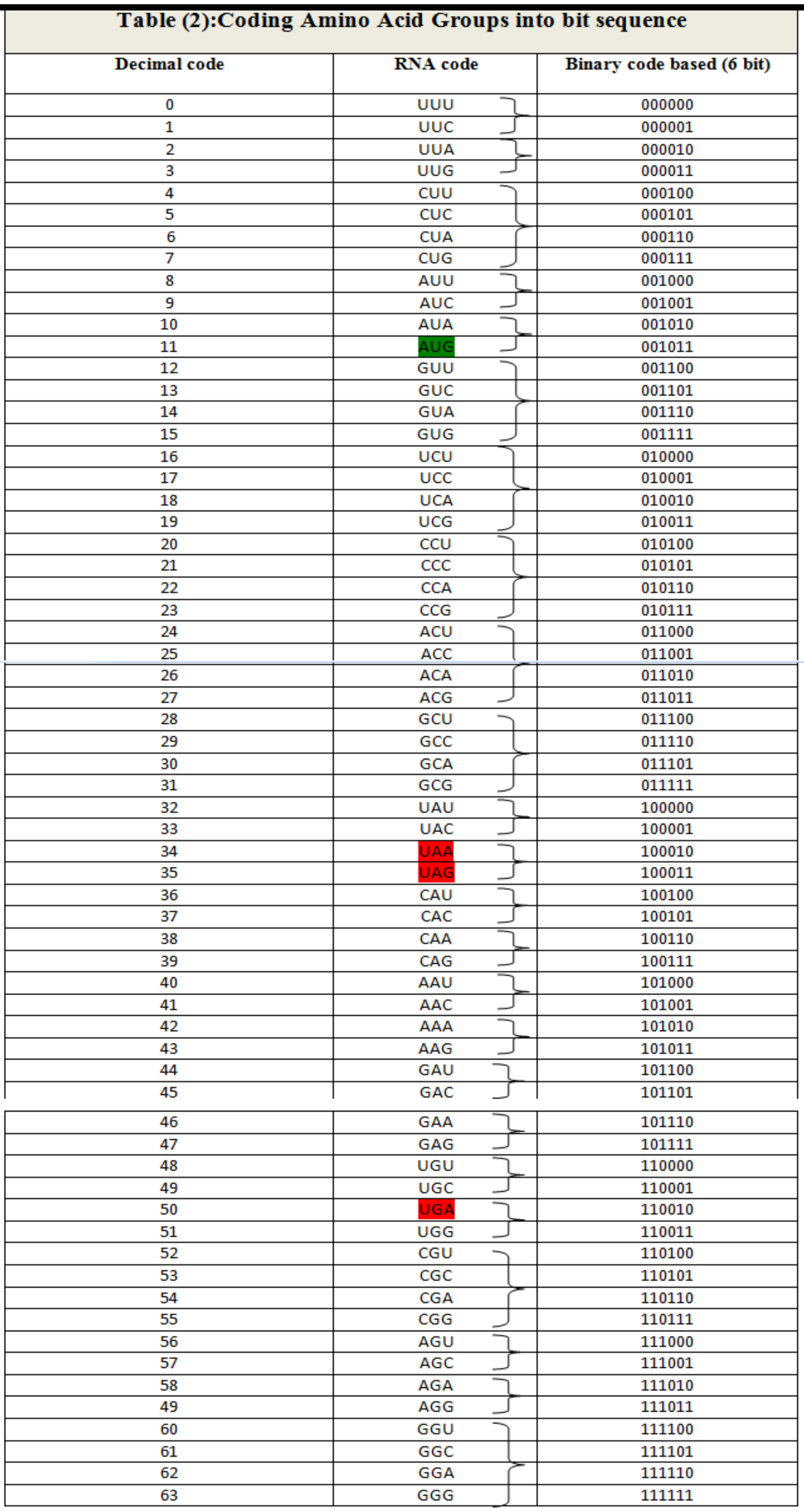


Table (2):Coding Amino Acid Groups into bit sequence

Decimal code	RNA code	Binary code based (6 bit)
0	UUU	000000
1	UUC	000001
2	UUA	000010
3	UUG	000011
4	CUU	000100
5	CUC	000101
6	CUA	000110
7	CUG	000111
8	AUU	001000
9	AUC	001001
10	AUA	001010
11	AUG	001011
12	GUU	001100
13	GUC	001101
14	GUA	001110
15	GUG	001111
16	UCU	010000
17	UCC	010001
18	UCA	010010
19	UCG	010011
20	CCU	010100
21	CCC	010101
22	CCA	010110
23	CCG	010111
24	ACU	011000
25	ACC	011001
26	ACA	011010
27	ACG	011011
28	GCU	011100
29	GCC	011110
30	GCA	011101
31	GCG	011111
32	UAU	100000
33	UAC	100001
34	UAA	100010
35	UAG	100011
36	CAU	100100
37	CAC	100101
38	CAA	100110
39	CAG	100111
40	AAU	101000
41	AAC	101001
42	AAA	101010
43	AAG	101011
44	GAU	101100
45	GAC	101101
46	GAA	101110
47	GAG	101111
48	UGU	110000
49	UGC	110001
50	UGA	110010
51	UGG	110011
52	CGU	110100
53	CGC	110101
54	CGA	110110
55	CGG	110111
56	AGU	111000
57	AGC	111001
58	AGA	111010
49	AGG	111011
60	GGU	111100
61	GGC	111101
62	GGA	111110
63	GGG	111111

Algorithm (1): proposed Key generation using RNA

```
Input: Key sequence (characters, numbers)size (9 byte,18 byte,27 byte,…….)
Output: random key bit sequence with expanded size
Begin
Step1 :convert key sequence to binary sequence
Step2: code each 2bit from the message binary sequence to RNA 4base using table(1 ).
Step3: split RNA strand to a group of codons ( 3 nitrogenous bases)
Step4: Extend each codon in RNA sequence by selecting another codon that belong to the same amino acid and appending them, according to table (2).
Step5: Read RNA strand until find the AUG that is used to begin protein synthesis, count the number codons and stop when find end codon is (UAA or UAG or UGA)
Step6: Apply rotate right shift on RNA strand based on the number of codons between start and end codon
Step7: End
```

Implementation and Experiment Results

This section illustrates the implementation of the proposed approach. The proposed approach was programmed using Visual c#.net 2008. Several examples executed :

Exampl1:

Key=baghdad@

Step1: Size of key =72 bit

010001101000011011100110000101100010011010000110001001100000001000101111

Step2: RNA CODON=UAUCCAUCGCUCAUUCACUCCAUCACUCAAACACGG

Step3: Extend RNA CODON=AUAAUGGGUGGAAGCAGUGAGGAAUAAUGAGUGGUUAGGAGAUAGUAAUGAUAGGUUGUCUGUUGCGCCGCU

Step4: Determine start and end codon =
AUAAUGGGUGGAAGCAGUGAGGAAUAAUGAGUGGUUAGGAGAUAGUAAUGAUAGGUUGUCUGUUGCGCCGCU

Step5: SHIFT NUMBER 7

UAGGUUGUCUGUUGCGCCGCUAUAAUGGGUGGAAGCAGUGAG
GAAUAAUGAGUGGUUAGGAGAUAGUAAUGA

Step6: Target key (extended key generation)

0011000111011111011111110111111010110100111000010000001010111
1101111101010101010001111001110101001101001000110111011011101
00110100110001010011010 0

<u>Exampl2:</u>

Key=Computer@

Step1 :Size of key =72

110001101111011010110110 00001110 10101110 00101110
101001100100111000000100

Step2: RNA
CODON=GAUCGGUCCGUCAAGCCCGCACGCCCUCUAGCAAUA

Step3: Extend RNA CODON=
CUACUUGCCGCUAGGAGACAGCAAUUCUUUGGGGGCCGUCGA
GCGGCCGGAGGGGAUGACCGUCGAUAUUAC

Step4: Determine start and end codon =

Step5: SHIFT NUMBER = 0

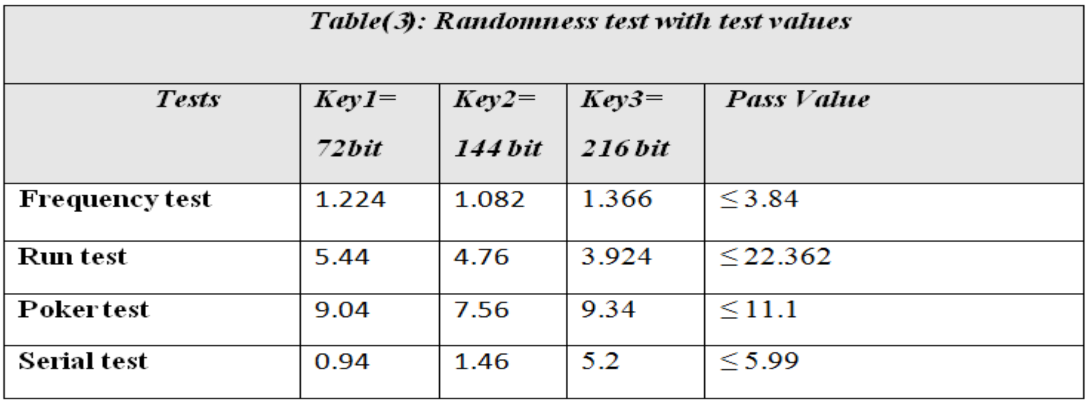
CUACUUGCCGCUAGGAGACAGCAAUUCUUUGGGGGCCGUCGA
GCGGCCGGAGGGGAUGACCGUCGAUAUUAC

Step6: Target key (extended key generation)

1000100111001010110100111011010011011001100001101000000000000
0011111011110101101101100010111010111101110011110110111111011
010110110110000100100100

Key strength evaluation

Statistical Tests of Randomness are used to check random property that a random sequence is likely to have. Useful statistical tests are four basic tests, and they are: Frequency test, Serial test, Poker test, Runs test [11]. The output of tests must be compared with passes values that illustrated in Table (3) to decide if the outputs of randomness tests are good for the sequences to pass. Randomness tests are applied on different key sizes: 72, 144, and 216 bits.



Table(3): Randomness test with test values

Tests	*Key1=* *72bit*	*Key2=* *144 bit*	*Key3=* *216 bit*	*Pass Value*
Frequency test	1.224	1.082	1.366	≤3.84
Run test	5.44	4.76	3.924	≤22.362
Poker test	9.04	7.56	9.34	≤11.1
Serial test	0.94	1.46	5.2	≤5.99

Conclusions

a proposed method for random key sequence generation based on the bases of RNA translation to protein chain. The proposed method accept a key sequence size(9,18,27,…) byte. Key sequence input the algorithm was significantly increased. This was because each character of the input message converted to a binary code of length 8 word from which pairs corresponding to RNA bases were created addition extension when each codon in RNA sequence by choosing another codon that belong to the same amino acid and appending them, according to a proposed table (2). The security strength of generated key is acceptable according to the results of statistical tests of randomness.

References

1. S.G. Srikantaswamy, H.D Phaneendra.," ***A New Approach for Designing Cryptographic Systems based on Feistel Structure***" , International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397 Vol. 5 No. 01 Jan 2013. http://www. Ivsl.org.

2. J. Hsiao, T. Gulliver, "***MARC – A New Block Cipher Algorithm***", Contemporary Engineering Sciences, Vol. 5, 2012, no. 6, 281 - 286.

3. W. Stallings, "***Cryptography and Network Security Principles and Practices***", third addition, 2003, Pearson Education, Inc.

4. H. Bidgoli, J. Wiley, "***Key (cryptography)***", The Internet Encyclopedia, 2004.

5. J. Knudsen, "*Java Cryptography*", O'Reilly Media, Inc., First Edition May 1998.

6. K. Hammed, ' ***DNA Computation Based Approach for Enhanced Computing Power***', International Journal of Emerging Sciences ISSN: 2222-4254 1(1) April 2011.

7. **Future in Our Genes?** ", SANS Institute 2000 - 2002 As part of GIAC practical repository.www.giac.org/paper/gsec/1617/dna.../102969

8. Mrs. Rebello's ,'DNA notes'. www.nclark.net/**DNA**_RNA

9. B. Bazli, D. Llewellyn-Jones, M. Merabti,'***Data Encryption in Communication Using DNA Sequences'***, the 15th annual PostGraguate symposium on the convergence on telecommicationa, networking, and broad casting, Liverpool,June, 2014.

10. S. Javheri, R.Kulkarni,'***Secure Data communication and Cryptography based on,DNA based Message Encoding'***,International Journal of Computer Applications (0975 – 8887), Volume 98– No.16, July 2014

11. E. Barker, A. Roginsky, ' ***Recommendation for Cryptographic Key Generation*** ', National Institute of Standards and Technology Special Publication 800-133 Natl. Inst. Stand. Technol. Spec. Publ. 800-133,

26 pages (December 2012)

12. P. Akkara**, "Applying DNA Self-assembly in Formal Language Theory",** MSc, University of Cincinnati, Engineering and Applied Science: Computer Engineering,2013

13. S. Kumar **,'Concept of DNA & RNA',** Anthropological Survey of India Manav Bhavan, Bogadi, Mysor http://nsdl.niscair.res.in/.../PDF+5.2**Concept**+of+**DNA**

14. http://www.nobelprize.org/educational/medicine/dna/b/translation/translation_process.html

طريقة مقترحة لتوليد مفتاح بالاعتماد على RNA

الخلاصة

بسبب الدور الأساسي لل مفتاح العشوائي في تصميم خوارزميات التشفير ، ويقترح طريقة جديدة ل توليد مفتاح عشوائي على أساس قواعد الترجمة RNA ل سلسلة البروتين. الطريقة المقترحة تقبل حجم سلسلة مفتاح (27،18،9 ، ...) بايت وتوليد سلسلة مفاتيح مع بالطول المناسب لطول الرسالة . قوة أمن المفتاح المتولد بهذه الطريقة مقبولة وفقا ل نتائج الاختبارات الإحصائية العشوائية .