Diyala Journal of Engineering Sciences

ISSN 1999-8716
Printed in Iraq

Vol. 10, No. 01, pp. 21-30, March 2017

A NEW ALGORITHM FOR ENCRYPTING ARABIC TEXT USING THE MATHEMATICAL EQUATION

Basim Najim al-din [1], Saad Abdulazeez Shaban [2]
[1, 2] Assistant lecturer, computer science department, College of education for pure sciences, University of Diyala
(Received: 7/12/2015; Accepted: 27/3/2016)

ABSTRACT: - Most organizations today use two levels of security: Data security level (encryption) and Network security level (firewall) etc. Therefore, encryption plays a vital role in securing daily operations of organizations. Securing data at first level aims to hide data characteristics. So, different algorithms and methods are being researched and developed to cover this gap for securing different data types especially text messages and make it very difficult for unauthorized parties to break it. From these, the major no. of algorithms are specialized to encrypt English texts, but no one for Arabic language texts although there were many attempts to invent such algorithms.

As a result, this research proposed a new encryption method to encrypt Arabic text by using the principle of integration to provide better security and increasing the complexity of guessing the correct keys and correct plain text.

At the end, the results show that the new cryptosystem is inevitable to cryptanalysis attack**.**
Keywords: *encryption, decryption,* security.

1- INTRODUCTION

As the digital computer systems invented at the twentieth century along with computer networks, the need to secure these resources and their communication channels arise. Unauthorized parties should not have ability to access to data in the way toward final destination.

Cryptography (originally came from Greek word “kryptos” which means “hidden”) [(1)], is a method of storing and transmitting data between sender and receiver over a communication media (network) secretly in such a way insures that only the intended recipient is allowed to read and modify data [(2)].

The cryptography terminology involves two terms, plain-text (original message to be sent) and the cipher-text (the encrypted message). The cryptanalysis is the art interested with attempting to break ciphered texts, while ‘cryptology’ is a part of mathematics science that studies both cryptology and cryptanalysis [(3)]. Decryption is a process of extracting the original plain-text from cipher-text by the authorized recipient.

From all the mentioned above, cryptosystem invented by employing both encryption and decryption methods [(1)].

Cryptosystems classified into classical cipher and modern cipher. Examples of classical ciphers are substitution and transposition. While symmetric and asymmetric are examples of modern ciphers.

Ciphers have two essential types: transposition and substitution [(4)]. Transposition cipher takes group of letters from the plain text and scrambles them between rows and columns depending on a special key used to determine letters locations that will be scrambled [(5)]. Myszkowski cipher, Route cipher, Rail fence cipher and Columnar cipher are examples of transposition cipher [(6)].

On the other hand, substitution cipher involves replacement of each letter in the plain-text with a corresponding letter from the alphabet. Pigpen, Beaufort, Caesar and Vigenere ciphers are examples of substitution type [4].

There are two basic types: (1) symmetric and (2) asymmetric cryptography.

- . In a symmetric cryptography, both sender and receiver are sharing similar keys to encrypt and decrypt the plain text. These keys should be kept secretly between the two parties [4]. The method of symmetric encryption considered secure according to the security of the single key, and the complexity and size of key determines the complexity of ciphering/deciphering processes [4].
- . In an asymmetric key cryptography, a single secret key is used between two parties, which is called a public key systems. So, each user has different keys, or asymmetric keys (private keys), using two keys one for encryption and another for decryption, the public and private keys cannot be derived from each other. Some examples of asymmetric key algorithms are: RSA, Elliptic Curve Cryptosystem (ECC), Diffie-Hellman, El Gamal, Digital Signature Standard (DSS) [4].

Cryptanalysis is the art concerned with attempting to break the cryptosystem by breaking cipher-text to get plain-text information contained in the original message [7].

This can be done by using various mathematical methods like frequency and brute force attack. To attack the intended message, attacker in advance should has decided which one of these methods depending on the information can be obtained from cipher-text like length of cipher-text, cipher-text itself and other information related to it [7].

The most common types of cryptographic attacks are:

1. Known plain text: involve breaking successfully one cipher-text and obtaining plain-text to be used for finding relation between the new coming cipher-text [3].
2. Chosen plain-text: involve encryption a chosen-plain text by cryptanalyst then analyzing the result to obtain more information related to plain-text like public key [3].
3. Cipher text only: involve accessing only the cipher-text by cryptanalyst, so the attacker will try to decrypt cipher-text to get its plain-text by other means like using frequency attack [4].
4. Chosen cipher-text: involve choosing any cipher-text by cryptanalyst and searching for some matching plain-text with a well-known and analyzed cipher-text [3, 4].
5. Adaptive chosen cipher-text: involve adapting the attack by cryptanalyst through sending alternate cipher-text to be decrypted. Depending on the result, further cipher texts can be chosen [3, 4].

Problem statement

The use of Internet and network is growing rapidly. This growth forces the Internet community to protect the data that is transmitted through the Internet. Therefore, to provide secure environment, different encryption methods start to appear. For instance, the substitution ciphers are the simplest form of encryption that prevent unauthorized party to get access to the original text, hence, protecting data from being misused or interpreted. However, all types of substitution ciphers can be broken easily these days due to the fact that these encryption methods utilize a single key in their crypto system process [8]. The Arabic text that transmitted over the internet need to be protect from un authorized parties to read and know the content of these information , so there is a need to find and develop different algorithms to protect and transmit the Arabic text securely over the internet [8].

2- RESEARCH OBJECTIVES

The objective of this research is to find a new proposed algorithm which is depend on the mathematical equations such as integration principles to make the Arabic text more secure when transmit over the internet and make guessing the correct keys and plaintext more difficult when applying an cryptanalysis on this text.

To evaluate the inevitability of the proposed encryption technique against berlekampmessay and linear feedback shift register, because the cipher text of the proposed algorithm is stream cipher of bits and the XOR operation used in the proposed algorithm .

3- LITERATURE REVIEW

There are two types of symmetric encryption algorithm, block cipher and stream cipher; the block ciphers encrypt the plain text in chunks, such as DES and AES, and common block sizes are 64 and 128 bits. Where the stream ciphers encrypt plain text one byte or bit at a time such as one – time pad, Caesar cipher and RC4 [4]. In the literature there are many researches focused on the block and stream ciphers and they are.

(1) John Justin M, Manimurugan S (2012) this paper focused mainly on the different kinds of encryption techniques that are existing, and framing all the techniques together as a literature survey. Aim an extensive experimental study of implementations of various available encryption techniques. Also focuses on image encryption techniques, information encryption techniques, double encryption and Chaos-based encryption techniques. This study extends to the performance parameters used in encryption processes and analyzing on their security issues [10].

(2) Prakash Kuppuswamy, Saeed Q Y Al-Khalidi (2012) proposed new symmetric key algorithm using modular 37 and select any number and calculate inverse of the selected integer using modular 37. The symmetric key distribution should be done in the secured manner [11].

(3) Ayushi (2010) proposed symmetric key algorithm using ASCII characters. Message in plain text can be understood by anybody knowing the language as long as the message is not codified in any manner [12].

(4) Ragheb Toemeh, Subbanagounder Arumugam (2008) discussed the Cryptanalysis of polyalphabetic by applying Genetic algorithm is presented. The applicability of Genetic algorithms for searching the key space of encryption scheme is studied. In Vigenere cipher, guessing the key size is done by applying Genetic Algorithm. The frequency analysis is used as an essential factor in objective function [9].

From above literature review observed that, all these research uses the one random number to generate the key and the key generating mechanism because of using one format for generating the key, but in the proposed algorithm two random numbers is used to generate the key, moreover the mechanism of generating the key is not constant and doesn't use constant format for the equation to generate the key and encrypt the message, but different equations is used as in the methodology section.

4- METHODOLOGY

4.1 key generation phase

4.1.1 Choosing an equation $x^n \pm a$

4.1.2 Choosing a random numbers x ,a and n

4.1.3 Where n ,x are represent the keys used in the encryption and decryption process

4.1.4 Choosing random key k

4.2 Encryption phase

4.2.1 Choosing the Arabic plain text to be encrypt

4.2.2 Compute the integration for the equation $\int x^n \pm a$ where x represent the random number and (a) represent any real number.

4.2.3 Find the value of integration $\frac{x^{n+1}}{n+1} \pm ax + c$ where c is the character

4.2.4 Convert the value of integration that obtained from previous step into binary format

4.2.5 Compute XOR for each message character with the random key

4.2.6 Apply all above five steps to all characters in the message

4.2.7 Send the cipher text over the internet to the receiver

As illustrated in the Figure (1).

4.3 Decryption phase

4.3.1 receive the cipher text from the internet

4.3.2 Compute XOR between each message character and the key k.

4.3.3 convert the binary format of the message into numeric values

4.3.4 substitute x value and compute $\frac{x^{n+1}}{n+1} \pm ax + c = v$

4.3.5 Find c value which represent numeric character value through solve the following function $c = v \mp \frac{x^{n+1} \mp (n+1)ax}{n+1}$ where (v) is value of cipher text characters.

4.3.6 Convert numeric value obtained from previous step to its corresponding character

4.3.7 Apply all above six steps to all characters in the message to obtain the plain text.

5- IMPLEMENTATION

5.1. Key generation phase

1. choosing the equation x^n+a
2. choosing randomly the values of n and a
3. Let n=2 , a=3
4. choosing randomly x value
5. Let x= 3
6. Let k=11

5.2. Encryption phase

1. Let the plain text is "**تكنولوجيا المعلومات**"
2. Compute integration $\int x^2 + 3 = \frac{x^3}{3} + 3x + c$
3. Taking the first character in the message which is **ت** where the value of **ت**=3 which represent the (c) value.
4. Compute the integration value for the character **ت** as follows

$$\frac{(3)^3}{3} + 3(3) + 3 = 21$$

5. Convert the numeric value (21) obtained from the previous step into binary format 21=0010101
6. Compute XOR as follows:

0010101
XOR
0001011
0011110

7. Apply all above six steps to all characters in the message to obtain the cipher text
8. The cipher text is

   0011110 0011101 0010010 0010000 0011100 0010010 0001110 0010000 0001010 0001010 0010000 0010011 0000110 0010000 0010010 0010011 0001010 0011110

5.3- Decryption phase

1. Taking the first character in the cipher text which is

   0011110
2. Compute XOR between the first character in the cipher text and the key = 11 in binary format as follows:

   0011110
   XOR
   0001011
   0010101
3. Convert the binary format value obtained from XOR operation above into numeric value as follows :

   0010101=21
4. Substitute the x value which is (3) in the $\frac{x^3}{3}+3x+c=21$ to find the numeric value of the character c.
5. Compute the function value to find the c value $\frac{(3)^3}{3}+(3\times 3)+c=21\rightarrow 9+9+c=21$

   $c=21-18\rightarrow c=3$
6. Convert the numeric value obtained from the previous step into corresponding character 3="ت"
7. The plain text is "تكنولوجيا المعلومات"

4. RESULT AND DISCUSSION

The results shows the encryption and decryption time are less than the stream and block cipher and also show the performance of the proposed method is better comparing with the stream and block cipher as shown in the table (2) and figures (3) , (4) . Moreover, through using berlekamp massey cryptanalysis against the proposed method, the results shows failing of this cryptanalysis method to break the Arabic cipher text for the proposed method and didn't success to guess the correct keys and correct equation used in encryption the message, also using linear feedback shift register (LFSR) cryptanalysis lead to same results as in berlekamp massey, as shown in table (3).

5. CONCLUSION AND FUTURE WORK

This research propose a new algorithm to encrypt and decrypt Arabic text using an integration technique to encrypt and decrypt the Arabic text , by using different types of cryptanalysis methods such as berlekamp massey cryptanalysis and linear feedback shift register (LFSR)on the proposed algorithm, the results showed that this method investable against this types of cryptanalysis, and couldn't guess the correct integration formula and doesn't guess the correct keys, which shows the strength of proposed algorithm and strength of the key generation technique. Moreover, the key generation technique is easy to compute but hard to invert which means that this algorithm is a one way function which means that P≠NP, and this leads to the fact that this problem is NP-hard problem. In the future we can use the second order equation to encrypt and decrypt the Arabic text.

9. REFERENCES

1. Mishra, A., Enhancing Security of Caesar Cipher Using Different Methods. International Journal of Research in Engineering and Technology, 2013. 2: p. 332.
2. Luciano, D. and G. Prichett, Cryptology: From Caesar ciphers to public-key cryptosystems. The College Mathematics Journal, 1987. 18(1): p. 2-17.
3. Schneider, B., Applied Cryptography: Protocols, algorithms, and source code in C. 1996: John Wiley & Sons.
4. Wong, C., Security Metrics, a Beginner's Guide. 2011: McGraw Hill Professional.
5. Carter, B. and T. Magoc, Classical Ciphers and Cryptanalysis. Space, 2007. 1000: p. 1.
6. Vobach, A., Pseudo-random transposition cipher system and method. 1996, Google Patents.
7. Dhavare, A., R.M. Low, and M. Stamp, Efficient cryptanalysis of homophonic substitution ciphers. Cryptologia, 2013. 37(3): p. 250-281.
8. Saroha, V., S. Mor, and A. Dagar, Enhancing Security of Caesar Cipher by Double Columnar Transposition Method. International Journal of Advanced Research in Computer Science and Software Engineering, 2012. 2(10): p. 86-88.
9. Toemeh, R. and S. Arumugam, Applying Genetic Algorithms for Searching Key-Space of Polyalphabetic Substitution Ciphers. Int. Arab J. Inf. Technol., 2008. 5(1): p. 87-91.
10. John Justin, M. and S. Manimurugan, A survey on various encryption techniques. International Journal of Soft Computing and Engineering (IJSCE) ISSN, 2012. 2231: p. 2307.
11. Kuppuswamy, P. and Y. Alqahtani, New Innovation of Arabic language Encryption Technique using new symmetric key algorithm". International Journal of Advances in Engineering & Technology, ISSN, 2014. 22311963.
12. Srivastava, V.K., A.K. Srivastava, and M. Khan. A Symmetric Key Cryptographic Algorithm. International Journal of Engineering Research and Technology. 2012. ESRSA Publications.

Table 1: ASCIIA Code for Arabic Alphabet

1	2	3	4	5	6	7	8
أ	ب	ت	ث	ج	ح	خ	د
9	10	11	12	13	14	15	16
ذ	ر	ز	س	ش	ص	ض	ط
17	18	19	20	21	22	23	24
ظ	ع	غ	ف	ق	ك	ل	م
25	26	27	28	29	30	31	32
ن	ه	و	ي	فراغ	ء	ؤ	ئ
33	34	35	36	37	38	39	40
,	.	:	"	؟	آ	ة	ى
41	42	43	44	45	46	47	48
لا	لأ	0	1	2	3	4	5
49	50	51	52	53	54	55	56
6	7	8	9				

Table 2: encryption / decryption time and performance

Algorithm	Encryption time	Decryption time	performance
Stream	77 sec	77 sec	2.20
Block	80 sec	80 sec	2.40
New algorithm	55 sec	57 sec	0.745

Table 3: berlekamp massey and LFSR cryptanalysis.

Character	Berlekamp massy Input sequence	Berlekamp massy Minimal length	Berlekamp massy Feedback polynomial	LFSR feedback polynomial	LFSR output stream
ت	0011110	4	x^4 + x + 1	$x^{4}+x^{3}+x^{2}+x+1$	1101111011
ك	0011101	3	x^3 + x + 1	X^5+x^4+x^3 + x^2 +x+ 1	0110111110
ن	0010010	3	x^3 + 1	X^5+x^2+1	0110100100
و	0010000	3	1	X^5+1	0110101101
ل	0011100	4	x^4 + x^3 + x + 1	X^5+x^4+x^3 + 1	0110100011
و	0010010	3	x^3 + 1	X^5+x^2 + 1	0110100100
ج	0001110	4	x^4 + x^3 + x + 1	x^4+x^3 + x^2 + 1	1101001110
ي	0010000	3	1	X^5+ 1	0110101101
ا	0001010	4	x^4 + x^2 + 1	x^4+ x^2 + 1	1101101101
ا	0001010	4	x^4 + x^2 + 1	x^4+ x^2 + 1	1101101101
ل	0010000	3	1	X^5+ 1	0110101101
م	0010011	4	x^4 + x^3 + 1	X^5+x^2+x + 1	0110111100
ع	0000110	5	x^5 + x^2 + x + 1	x^3+ x^2 + 1	1011100101
ل	0010000	3	1	X^5+1	0110101101
و	0010010	3	x^3 + 1	X^5+x^2+1	0110100100
م	0010011	4	x^4 + x^3 + 1	X^5+x^2+x + 1	0110111100
ا	0001010	4	x^4 + x^2 + 1	x^4+ x^2 + 1	1101101101
ت	0011110	4	x^4 + x + 1	x^4+x^3 + x^2 +x+ 1	1101111011

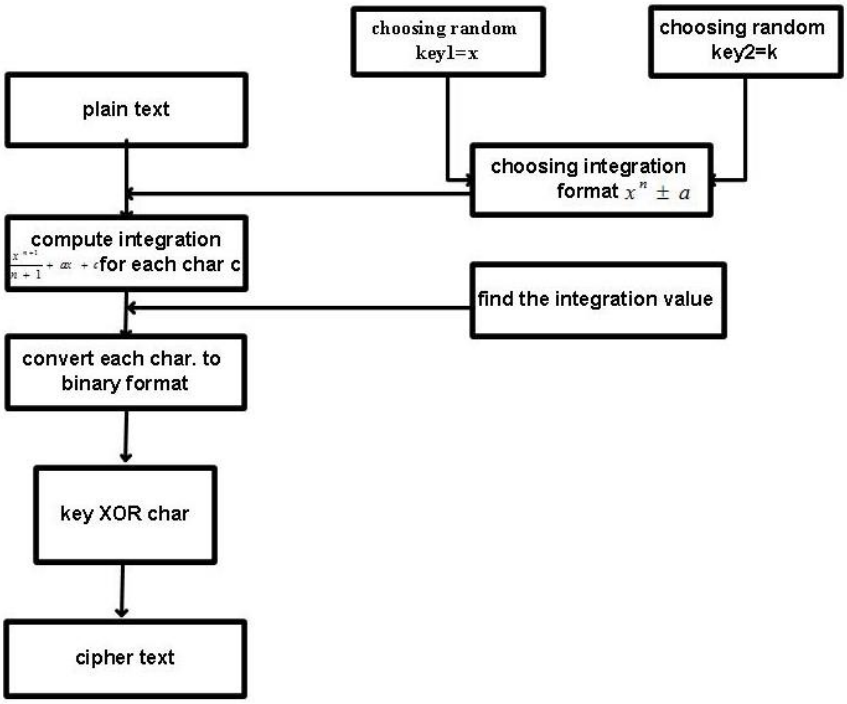

Figure 1: Block Diagram for Encryption Phase

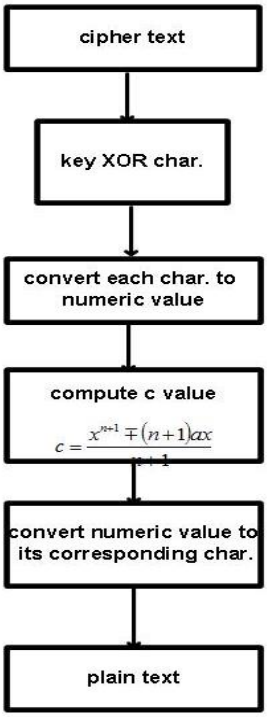

Figure 2: Block Diagram for Decryption Phase

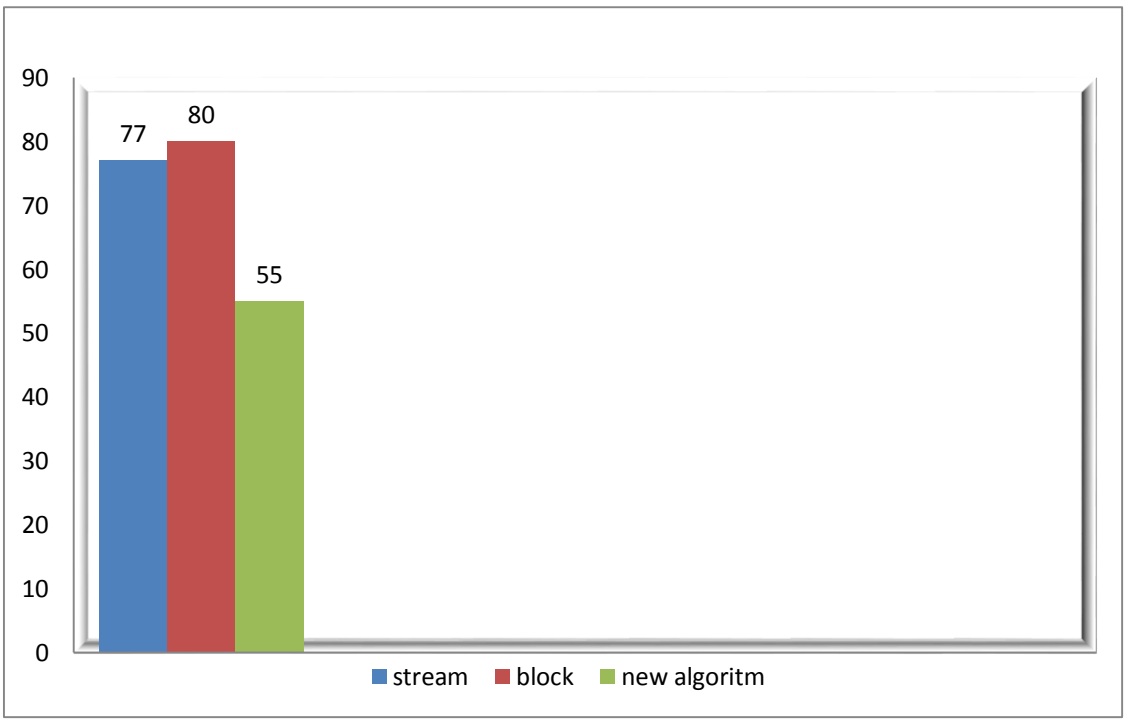

Figure 3: Encryption Time

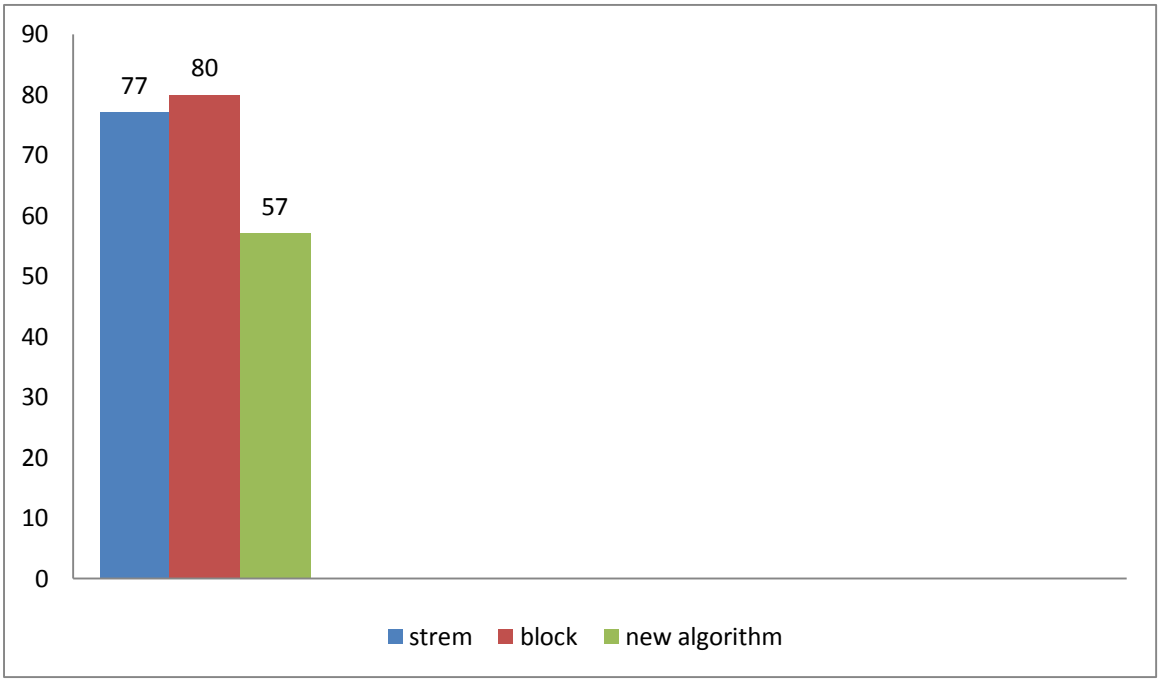

Figure 4: Decryption Time

خوارزمية جديدة لتشفير النص العربي باستخدام المعادلات الرياضية

باسم نجم الدين، سعد عبدالعزيز شعبان
مدرس مساعد، كلية التربية للعلوم الصرفة – قسم علم الحاسوب – جامعة ديالى

الخلاصة

ان معظم المؤسسات في هذه الايام تستخدم مستويين من الحماية: الاول مستوى حماية البيانات (التشفير) والاخر مستوى حماية الشبكات (الجدار الناري).

لذلك, فان التشفير يلعب دورا كبيرا في تامين العمليات اليومية لهذه المؤسسات. ان حماية البيانات في المستوى الاول يهدف الى اخفاء خصائص البيانات. لذلك, فان مختلف الخوارزميات والطرق يتم بحثها وتطويرها لغرض تغطية هذه الفجوة لحماية مختلف انواع البيانات وبالاخص الرسائل النصية وجعل من الصعب جدا على الاشخاص الغير مرخصين التمكن من كسر هذه الشيفرات. ومن بين هذه , فان العدد الاكبر من طرق التشفير مخصصة لتشفير النصوص الانجليزية, لكن ولا اي واحدة منها مخصصة للنصوص المكتوبة باللغة العربية على الرغم من انه كانت هنلك محاولات لغرض ابتكار مثل هذه الطريقة.

ونتيجة لذلك، فان هذا البحث يقترح طريقة تشفير جديدة لتشفير النص العربي باستخدام مبدأ التكامل لغرض توفير حماية بشكل أفضل وزيادة تعقيد التخمين للمفاتيح الصحيحة اللازمة لفك التشفير وكذلك النص الصريح الصحيح.

في النهاية، فان النتائج تظهر بان نظام التشفير الجديد منيع ولا يمكن تجاوزه في هجمات تحليل الشيفرات.