A new Algorithm for Encrypt Arabic Text by using first Order Equation for Three Variables

Basim Najim al-din abed al-Obaidi

Diyala university-college of education for pure sciences-computer science department

Received: 27 October 2015 **Accepted: 18 May 2016**

Abstract

In this days many data exchanged over the Internet, so finding the best solution that offer the necessary protection against the information hackers becomes the basic goals of many researches. Many researches focus on the Encryption algorithms that play a main role in information security systems. The goal of every encryption algorithm is to make it as hard as possible. If a good encryption algorithm is used, there is no technique considerably better than trying every possible key to break the cipher text. It is difficult to define the quality of an encryption algorithm. Sometimes algorithms look strong and complicated but turn out to be very easy to break.In this research, a new encryption method is proposed to encrypt Arabic text by using the standard of first order equation for three variables. The sender and the recipient will share a first order equation for three variables and two randomly constants represent the values of y, z which represent the keys of used in encryption\decryption process. The result of the equation xored with randomly shared value between both sides which is represent the third key for the proposed method to get final cipher text of the proposed method. By applying different cryptanalysis techniques such as berlekamp Massey cryptanalysis, linear feedback shift register (LFSR), autocorrelation attack, brute force attack , frequency attack, m-138 cipher text only attack and side – channel attack to test the inevitability of the proposed method, the results showed that the proposed method is hard to be broken by the crypt analytics and

attackers. Moreover, comparing the timing and performance of the proposed method with the block and stream ciphers showed the proposed method is better than the block and stream ciphers in these measures.

Keywords: First order equation, berlekamp Massey, linear feedback shift register, autocorrelation attack, side-channel attack.

الخوارزمية الجديدة لتشفير النصوص العربية باستخدام معادلة من الدرجة الاولى لثلاث متغيرات

باسم نجم الدين عبد العبيدي

جامعة ديالى – كلية التربية للعلوم الصرفة

الخلاصة

في هذه الايام فان العديد من البيانات يتم تناقلها عبر الانترنيت ، لذلك فان ايجاد افضل الحلول لحماية المعلومات من قراصنة المعلومات اصبح من الاهداف الاساسية للعديد من الابحاث . العديد من الابحاث تركز على خوارزميات التشفير التي تلعب الدور الرئيسي في انظمة حماية المعلومات . ان هدف كل خوارزمية تشفير هو جعلها اصعب قدر الامكان. فاذا تم استخدام خوارزمية تشفير جيدة ، فأنه لاتوجد تقنية افضل من تجربة كل المفاتيح الممكنة لكسر النص المشفر. انه من الصعب تعريف كفاءة خوارزمية التشفير . ففي بعض الاحيان فأن الخوارزمية تبدو قوية ومعقدة ولكنها سهلة الكسر،في هذا البحث تم اقتراح طريقة تشفير جديدة لتشفير النص العربي بواسطة استخدام مقياس معادلة من الدرجة الاولى في ثلاث متغيرات . بحيث المرسل والمستقبل يتشاركون بمعادلة من الدرجة الاولى لثلاث متغيرات بالاضافة الى عددان ثابتان يتم اختيارهما بطريقة عشوائية ويتفق عليهما الطرفان يمثلان قيم ص، ع في المعادلة واللذان يمثلان المفاتيح المستخدمة في عملية التشفير وفك الشفرة . بعد ذلك فأن ناتج المعادلة يغمل له اكس اور مع قيمة عشوائية مشتركة بين الطرفين والتي تمثل المفتاح الثالث للطريقة المقترحة للحصول على النص المشفر النهائي للطريقة المقترحة . وبتطبيق تقنيات تحليل الشفرة المختلفة مثل تحليل شفرة البيركامب ماسي ، سجل ازاحة ردود الفعل الخطية ، هجوم معامل الارتباط الذاتي ، هجوم القوة الغاشمة ، وهجوم التردد ، وهجوم م-138 للنص المشفر فقط و هجوم حانب – القناة لاختبار قوة الطريقة المقترحة في مواجهة تحليل الشفرة . ومن هذاه الاختبارات فأن النتائج اظهرت بأن الطريقة المقترحة من الصعب كسر التشفير فيها من قبل محللي الشفرة والمهاجمين . بالاضافة الى ذلك ، فان مقارنة الوقت المستغرق للتشفير وفك الشفرة والكفاءة لشفرة الطريقة المقترحة مع الشفرات الكتلية و الحزمية اظهرت بأن الطريقة المقترحة هي الافضل من الشفرات الكتلية والحزمية من ناحية هذه المقاييس

الكلمات المفتاحية : معادلة من الدرجة الاولى ، بيرلكامب ماسي ، سجل ازاحة ردود الفعل الخطية ، هجوم معامل الارتباط الذاتي

Introduction

Cryptography is the art of hiding the meaning of data and transmitting information over a communication channel securely in which only the recipients are allowed to read/interpret it and others should not be able to read/interpret it even though they get access to it [6].

The word Cryptography is came from the Greek word, “kryptos” which means ‘hidden’ and “graphein” that means ‘to write’. Therefore, cryptography is the art and the science of making any information unintelligible to all except the recipients [5]. In the terminology of cryptography, the sending data is called "the plain-text", while the encrypted data is called the "cipher-text". On the other hand, the art and the science of breaking Cipher text is called cryptanalysts, and the branch of mathematics that study both cryptography and cryptanalysis is called "cryptology" [7].

The system that uses encryption and decryption methods is called cryptosystem [6]. These cryptosystems are classified into two types: classical ciphers and modern ciphers. The classical ciphers are further divided into two types of ciphers: substitution and transposition ciphers. Where, the modern ciphers are also further divided according to the key into two cipher types: symmetric and asymmetric ciphers.

Types of Ciphers

There are two basic types of ciphers according to encryption/decryption mechanisms : substitution and transposition ciphers [11]. In the substitution cipher the mechanism of encryption depending on replaced each letter in the alphabet with a corresponding letter from the alphabet, Beaufort cipher , Caesar Cipher, Vigenère cipher, pigpen cipher are some examples on substitution cipher [11]. Where, transposition cipher permutes letters in the same message [1]. Rail fence cipher, route cipher, columnar cipher, myszkowski cipher are some examples on transposition cipher [10]. Moreover, types of cryptography are divided into two types according to the key generation mechanisms: symmetric and asymmetric cryptography.

Symmetric Cryptography

The mechanism of sharing the same key for the encryption and decryption process between both parties is called symmetric cryptography. Symmetric keys are also called secret keys because they must be kept as a secret key. The security of the symmetric encryption method

depending on key protection [11]. In addition, the key size and it is complexity determine the complexity of the encryption/decryption process. However, it requires more efficient and save way to delivering the keys securely to the communicating parties [11].

Asymmetric Cryptography

In a symmetric key cryptography, the key used in the encryption process differ from the key that used in the decryption process, and there are two keys in this type of cryptography public and privet keys. Using two keys one for encryption and the other for decryption. The public and private keys cannot be derived from each other. There are many examples of asymmetric key algorithms such as: RSA, Elliptic Curve Cryptosystem (ECC), Diffie-Hellman, El Jamal, Digital Signature Standard (DSS) [11].

Cryptographic Attack Methods and cryptanalysis

Cryptanalysis is the art and the science to break a cipher text to get the security information that is contained in the original message (plain text). There are many types of cryptanalysis methods according to the cryptographic methods through employing some mathematical methods; such as frequency and brute force attack …etc. [2]. Cryptographic attacks as part of the cryptanalysis attempts to decrypt the cipher text without knowing the any information about the key. The following is the most common five related types of cryptographic attacks: Known plain text, Chosen plain text, Cipher text only, Chosen cipher text, Adaptive chosen cipher text.

Problem statement

The use of Internet and network is growing rapidly. This growth showed the need of protect the Arabic text that is transmitted over the Internet. Therefore, many attempts start to appear to provide secure environment to protect the Arabic text transmitted over the internet, and since there is no enough algorithms to fill this gap in this field and this is in addition to the enormous development in the cryptanalysis methods, So it became necessary to find new ways to encrypt/decrypt the messages written in Arabic language to transmit it securely over the internet.

Research Objectives

The objective of this research is to find a new algorithm which is depend on the first order equation for three variables to add more security to the Arabic text when transmit over the internet. And add more complexity to guess the correct text when applying a cryptanalysis on

this text. Moreover, to evaluate the inevitability of the proposed encryption technique against berlekamp Massey and linear feedback shift register.

Literature Review

Because of the huge and great development of encryption and cryptanalysis systems, many research which was carried in this area to improve or develop the cryptographic methods to make the encrypted texts most secret.

Ragheb Toemeh, Subbanagounder Arumugam (2008) the Cryptanalysis of polyalphabetic by applying Genetic algorithm is discussed, and the applicability of Genetic algorithms for key space searching of the encryption method has been studied. By applying Genetic Algorithm in Vigenere cipher, the key size guessing is done. The frequency analysis is applied as an extremely important factor in objective function [9].

John Justin M, Manimurugan S (2012) the paper focuses basically on the different types of encryption techniques that are existing, and wording all these techniques together in a literature survey. The study aimed to experimental study of the implementations of various encryption techniques. Also the study focuses on the information encryption techniques, image encryption techniques, Chaos-based encryption techniques and double encryption techniques. This study Have expanded to the performance of the parameters that used in the encryption processes [3].

Prakash Kuppuswamy, Saeed Q Y Al-Khalidi (2012) a new symmetric key algorithm using modular 37 and select any number and calculate inverse of the selected integer using modular 37 is proposed. The symmetric key distribution should be done in the secured manner [4].

Ayushi (2010) symmetric key algorithm using ASCII characters is presented. Message in plain text can be understood by anybody knowing the language as long as the message is not codified in any manner [8].

From literature above we observed that, all these research uses the one random number to generate the key and the key generating mechanism is clear and using one format for generating the key, but in our proposed algorithm we use two random numbers to generate the key and 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 we use different equations as in the methodology section.

Methodology

Key generation phase

Choosing a first order equation of three variables such as $3x + 2y - z$ where x represent the character of the message and y, z are two random numbers

Choosing first random number key1=y

Choosing second random number key2=z

The format of the equation and the two random numbers key1, key2 are secret and only the sender and receiver know it.

Encryption phase

Compute the value of the equation above for each character in the Arabic text

Use the absolute value for the obtained equation to avoid the negative value

Convert the obtained character to the binary format

Compute XOR between key1 and the character in the odd position, and between key2 and the character in the even position

Convert all message to the binary format and send it to the receiver over the internet

Decryption phase

Using key1 and key2 to decrypt the message

Compute XOR between key1 and the character in the odd position and key2 with the character in the even position

Convert the binary format to the numeric value for each message characters

Compute the inverse for the first order equation to find the value of each character in the message such as for the equation that we took as an example $x = \frac{z - 2y}{3}$ where the value of the x represent the character that we want to decrypt and y represent key1 and z represent key2

At the end we convert the message from binary format to the character format

Implementation

For each Arabic letters we use the synthetic specific value to do the mathematical calculation as shown in the Figure (1):

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	هـ	و	ي	٠	١	٢	٣
33	34	35	36	37	38		
٤	٥	٦	٧	٨	٩		

Figure 1: Synthetic Specific Value for Arabic Alphabet

The key generation, encryption and decryption algorithm mentioned in the following:-

Key generation phase

1. **Let we choose the equation** $2x-3y+4z$
2. **Let we choose y=7**
3. **Let we choose z= 20**
4. **Let plain text = " تكنولوجيا المعلومات"**

Encryption phase

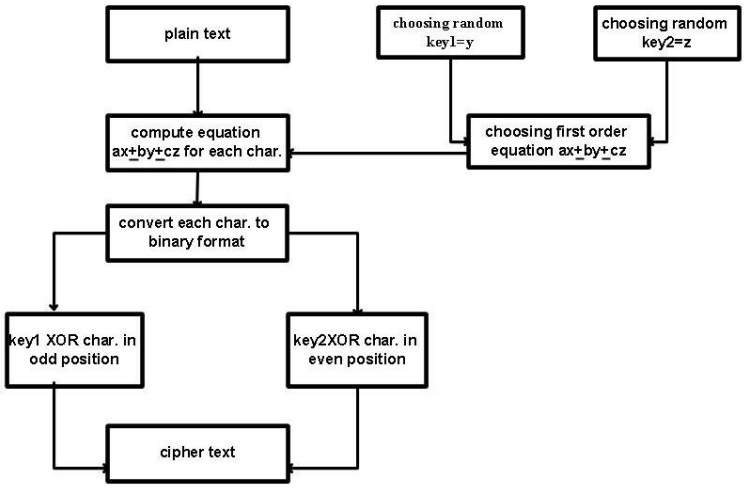


Figure 2: Encryption Phase

1. **We take the value of the first character (odd) from the table above**

ت=3

2. **Compute the equation**

2(3)-3(7) +4(20) =67

3. **Convert the obtained character to the binary format**

1000011

4. **Compute XOR between the character value (odd position) and key1=7 in binary format**

1000011

XOR

0000111

1000100

5. **Do all the four steps above to all message and then concatenate all obtained characters in binary format and send it over the internet.**
6. **The cipher text is:**
   1000100 0000010 0011110 0001111 0010000 0001111 0000010 0001101 0000110 0010101 0010000 0001100 0010101 0000011 0011100 0001100 0000110 0010111

Decryption phase

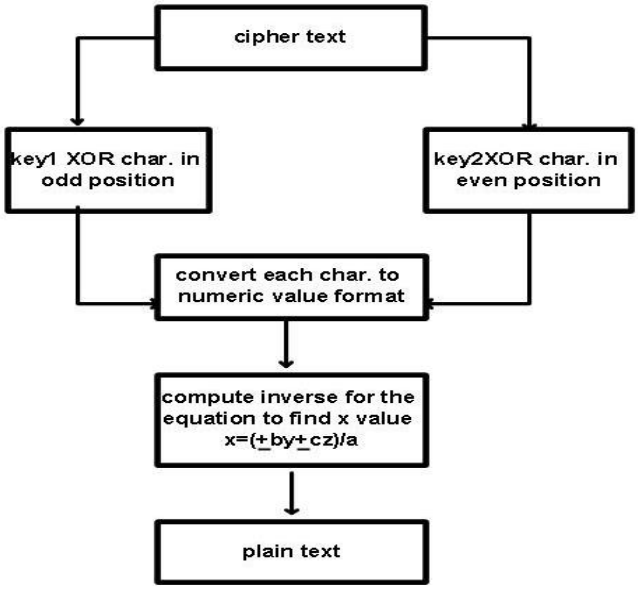


Figure 3: Decryption Phase

1. **Receive the message from the internet and decrypt the message begin with the first character which is 0111011**
2. **Compute XOR between first character in the cipher text (odd position) and the key1**

   0111011

   XOR

   0000111

   0111100

3. **Convert the obtained character from binary format to the numeric value**

   1000011=67

4. **Compute x value from the equation as follows:**

   2(x)-3(7) +4(20) =67

   2x+61=67

   $$x = \frac{6}{2}$$

   X=3

5. **Convert the numeric value to the character that represent this value which is:**

   X=ت

6. **Do all five steps above to obtain the plain text**

   Plain text =”تكنولوجيا المعلومات”

Result and discussion

The results show the encryption and decryption time is faster comparing to the stream and block cipher and also show the performance of the proposed method is better comparing with the stream and block cipher. 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. Table (1) show the comparison of encryption /decryption time and performance between block ,stream cipher and the proposed algorithm for the message size of 1000bit.

Table 1: Encryption/Decryption Time and Performance Table

Algorithm	Encryption time	Decryption time	performance
Stream	77 sec	77 sec	2.20
Block	80 sec	80 sec	2.40
New algorithm	65 sec	65 sec	0.9

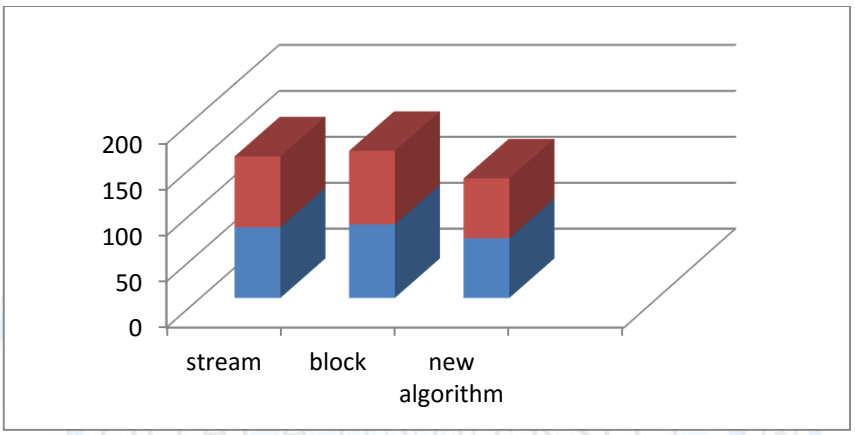


Figure (4) show encryption/decryption timing

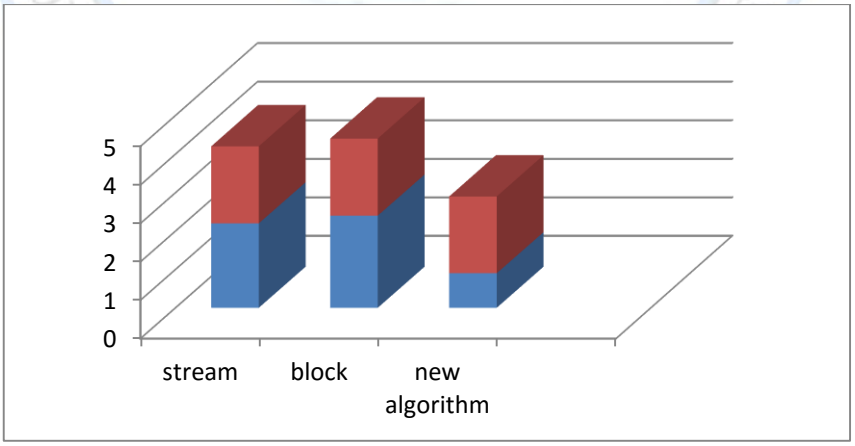


Figure (5) show the performance comparison between three block, stream and proposed method

Cryptanalysis tests

By applying the brute force attack on the proposed method the results showed this cryptanalysis couldn't guess the correct cipher text, so it fail to break the cipher text of the proposed method as it shown in the results below of this cryptanalysis technique.

1000100 0000010 0011110 0001111 0010000 0001111 0000010 0001101 0000110 0010101 0010000 0001100 0010101 0000011 0011100 0001100 0000110 0010111 1000100 0000010 0011110 0001111 0010000 0001111 0000010 0001101 0000110 0010101 0010000 0001100 0010101 0000011 0011100 0001100 0000110 0010111

Also by applying m-138 cipher text only attack, the results showed the fail of this type of cryptanalysis to break the cipher text of the proposed method and guessing the correct plain text as shown below:
The plain text obtained from the m-138 cipher text only attack for the cipher text obtained from proposed method is:

EREDTHEANDTHEREASTHEREANDEREDTHEANDTHEREASTHEREANDEREDTHEANDTHEREASTHEREANDEREDTHEANDTHEREASTHEREANDEREDTHEANDTHEREASTHEREANDEREDTHEANDTHEREAST

By using frequency attack the results showed that this cryptanalysis also fail to break the cipher text of proposed method as shown below in the Figure (6):

Total word count :	0
Number of different words :	0
Complexity factor (Lexical Density) :	0%
Readability (Gunning-Fog Index) : *(6-easy 20-hard)*	0
Total number of characters :	143
Number of characters without spaces :	0
Average Syllables per Word :	0
Sentence count :	1
Average sentence length (words) :	0
Max sentence length (words) :	0
()	
Min sentence length (words) :	0
()	
Readability (Alternative) beta : *(100-easy 20-hard, optimal 60-70)*	206.8

Figure 4: frequency attack on the proposed method

By applying the berlekamp Massey cryptanalysis to guess the correct equation used in the proposed method and then guess the correct keys to obtain the correct plain text for the proposed method, the results showed that the cryptanalysis technique couldn't guess the correct equation used in the proposed method, so it's very difficult to break the cipher text as it shown in the Figure (7) below:

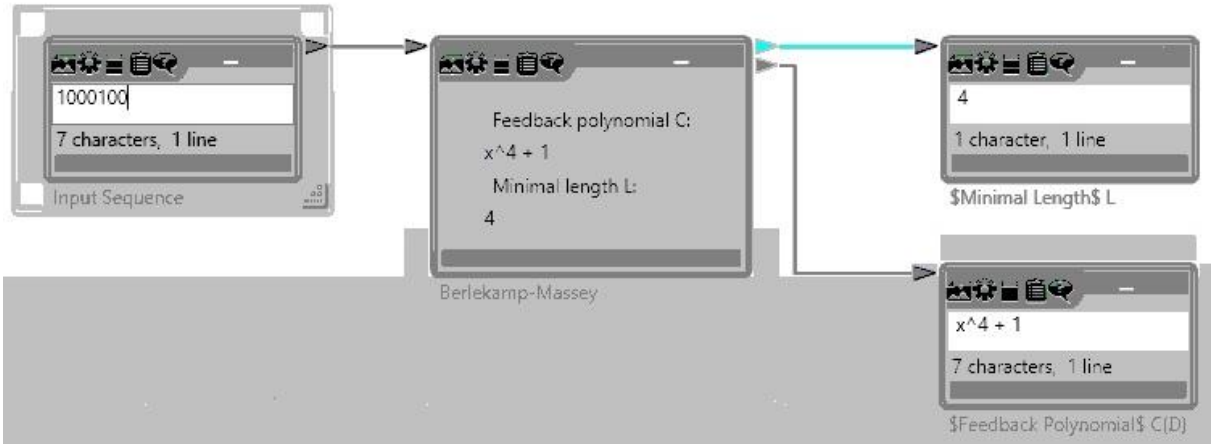


Figure 5: berlekamp Massey cryptanalysis on the proposed method

By applying the linear feedback shift register cryptanalysis to obtain the equation used in the proposed method to obtain the correct plain text for the proposed method, the results showed that the cryptanalysis technique couldn't guess the correct equation used in the proposed method, so it's also very difficult to break the cipher text as it shown in the Figure (8) below:

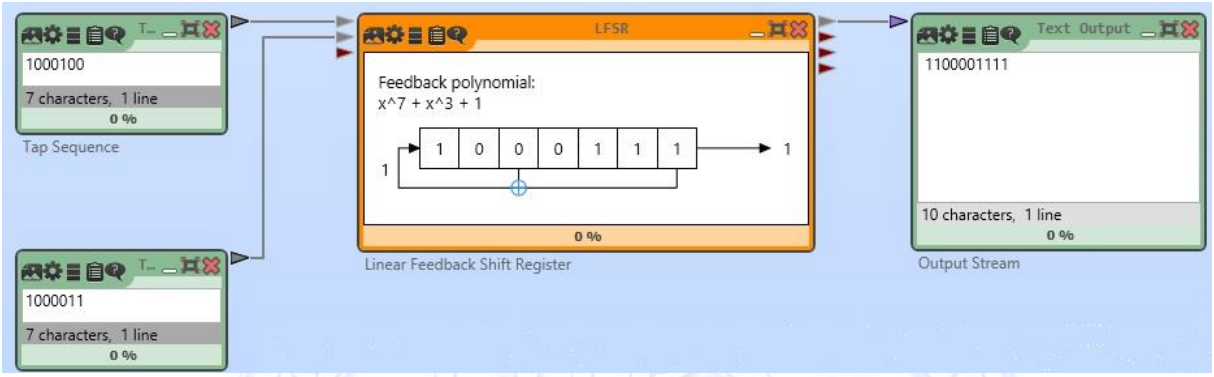
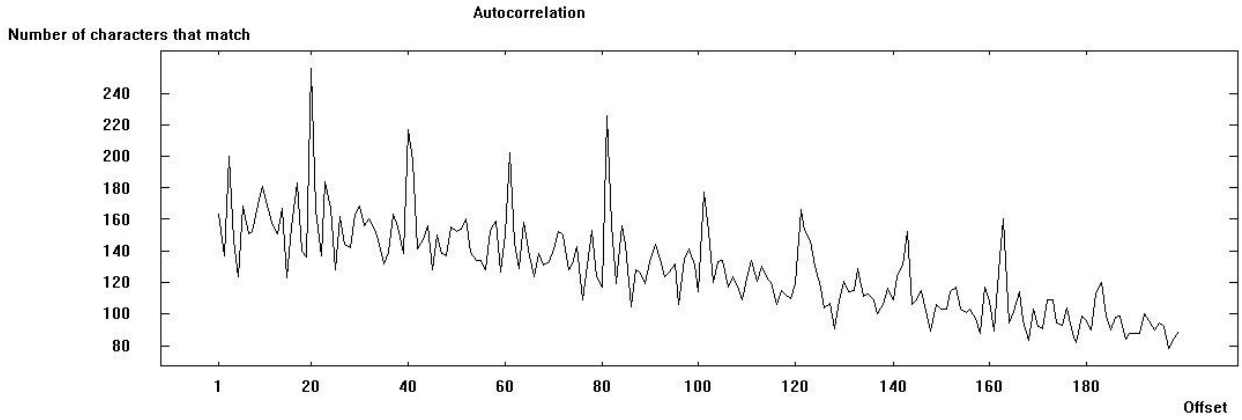


Figure 6: linear feedback shift register cryptanalysis on the proposed method

also by using auto correlation attack in order to trying to break the cipher text of the proposed method, the results showed that the correlation between the characters is very poor so it means that, it's difficult to guess the correct characters of the cipher text as it shown below in the Figure (9):



Finally, by applying the one of the most powerful cryptanalysis technique which is side – channel attack on the proposed method, the results showed incapability to break the cipher text of the proposed method even with this cryptanalysis technique as it shown in the Figure (10) below:

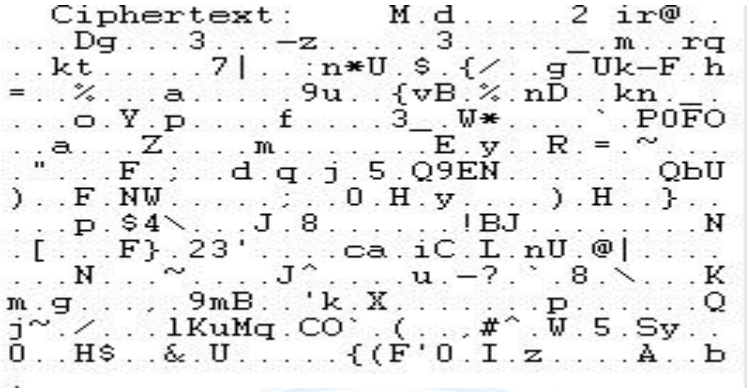


Figure 7: side-channel attack on the proposed method

Conclusion and future work

This research propose a new technique to encrypt and decrypt the message using a first order equation for three variables and two random numbers as a keys and xored the result of the equation with the third shard value as a third key to encrypt the characters depending on the position of the character in the message.

By using different types of cryptanalysis methods such as berlekamp Massey cryptanalysis, linear feedback shift register (LFSR), autocorrelation attack, brute force attack , frequency attack, m-138 cipher text only attack and side – channel attack on the proposed algorithm, the results showed that this method investable against this types of cryptanalysis, and couldn't guess the correct equation and doesn't guess the correct keys, because of guessing the correct equation and three variables requires number of attempts up to n! times which requires many years to break the cipher text, from the result above we concluded 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.

References

1. Carter, B. and T. Magoc, Classical Ciphers and Cryptanalysis. space, 2007. 1000: p. 1.
2. Dhavare, A., R.M. Low, and M. Stamp, Efficient cryptanalysis of homophonic substitution ciphers. Cryptologia, 2013. 37(3): p. 250-281.
3. 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.
4. 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.
5. Luciano, D. and G. Prichett, Cryptology: From Caesar ciphers to public-key cryptosystems. The College Mathematics Journal, 1987. 18(1): p. 2-17.
6. Mishra, A., ENHANCING SECURITY OF CAESAR CIPHER USING DIFFERENT METHODS. International Journal of Research in Engineering and Technology, 2013. 2: p. 332.
7. Schneider, B., Applied Cryptography: Protocols, algorithms, and source code in C. 1996: John Wiley & Sons.
8. Srivastava, V.K., A.K. Srivastava, and M. Khan. A Symmetric Key Cryptographic Algorithm. in International Journal of Engineering Research and Technology. 2012. ESRSA Publications.
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. Vobach, A., Pseudo-random transposition cipher system and method. 1996, Google Patents.
11. Wong, C., Security Metrics, A Beginner's Guide. 2011: McGraw Hill Professional.