P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2018,12 (2) :49-56

A Proposed Method for Achieving the Confidentiality of Arabic Texts

Dr. Yaseen Hikmat Ismaiel

Lecturer: Department of Computer Science
College of Computer and Mathematic Science Mosul University

ARTICLE INFO

Received: 12 / 8 /2018
Accepted: 13 / 9 /2018
Available online: 3/1/2019
DOI: 10.37652/juaps.2022.171807

Keywords:
transposition cipher.
substitution cipher.
Arabic text cryptography.

ABSTRACT

Due to the rapid development of distributed computing systems, the use of local networks, and the huge expansion of the Internet, the process of maintaining the confidentiality of information becomes important and necessary. Encryption is one of the most important ways to keep information confidential and to prevent unauthorized people from disclosing this information. Many cryptographic algorithms have appeared in different ways and methods. But most of these algorithms have been devoted for encrypting texts in the English language. Because of the increasing important and sensitive information exchange in the Arabic language by users on the Internet, an urgent need has appeared to build dedicated encryption systems for the Arabic language. The aim of this research is to provide a proposed encryption method that uses the idea of combining the methods of transposition and substitution cipher for the purpose of Arabic texts confidentiality. This research depends on using diacritics in the Arabic language to perform the process of substitution cipher. The efficiency of the proposed method has been tested on different texts; the method provides high speed execution, the length of the encrypted texts is almost close to the clear text length, and the confidentiality of the resulting encrypted texts is high.

1. Introduction:

Cryptography has been used for a long time. In the past, primitive methods were used to keep the information transmitted confidential. The information in that period was sent by people, birds, etc. Now, with the great development in the information technology and the proliferation of the Internet, there are large quantities of important and sensitive information requiring protection sent over the network. Encryption is defined as the process of converting data from its natural structure to another mysterious incomprehensible entity through complex algorithms to protect or send data to third parties in a secure manner, ensuring that only those authorized people to access such data are able to view their content. In order to be able to access the content, these people must first decrypt the encrypted data. Decryption is the process of retrieving data from its encrypted form to its original one by knowing the encryption method and the used key. Encryption systems are divided into two basic types depending on the type of the used key in encryption and decryption [1][2]:

- ✓ *Secret key systems (symmetric encryption)*: This method relies on the use of one key, that is, both the sender and receiver use the same secret key to encrypt and decrypt the message.
- ✓ *Public Key Systems (asymmetric encryption)*: In this type of encryption, two keys are used to connect a mathematical relationship; one of which is used for encryption and another for decoding.

Encryption systems can also be divided in terms of the nature of their handling of clear text characters to obtain encrypted texts into two types [2][3]:

- ✓ Transposition systems: in which the clear text characters are rearranged according to a particular formula.
- ✓ Substitution Systems: here clear text characters are replaced by letters, numbers or symbols according to the encryption method used.

———* Corresponding author at: Department of Computer Science College of Computer and Mathematic Science Mosul University.E-mail address: Yasino79@yahoo.com

2. Problem statement:

The use of the Internet and network is growing rapidly. In recent years, the growth of texts with Arabic content and number of users on the Internet has greatly increased. Arabic is a widely spoken language with more than 375 million speakers and over 155 million or over forty percent of these Arabic-speaking people use the Internet. The number of Arabic speaking internet users has grwon by a factor of sixty in the last fifteen years (2000-2015) [4].

This growth in Internet use has been accompanied with the growth in the amount of sensitive and important data in Arabic, which are exchanged over the internet. These sensitive data require good coding methods for the purpose of maintaining their confidentiality and preventing unauthorized persons from disclosing them. Most encryption methods are designed to encrypt texts in English, as there are very few coding codes that deal with Arabic.

3. Research Objectives :

The objective of this research is to find a new method for encrypting Arabic texts to provide Arab users with the ability for confidentially exchanging sensitive Arabic texts. The new method makes the Arabic text more secure and renders guessing the correct keys and plaintext more difficult to unauthorized persons. Also, it should provide fast execution speed and maintain the length of the encoded text compared to the length of the clear text.

4. Literature Review :

As previously discussed, there are a few coding methods that are designed primarily to deal with texts in the Arabic language. Some researchers, especially in the recent period, have attempted to build and design algorithms and cryptographic systems which deal with texts in Arabic. This section is devoted for reviewing some of the scholarly work of researchers in this field.

For instance, in 2009 Abdullah [5] investigated a novel algorithm for compressing and encrypting Arabic short text messages (SMS messages). The author has changed Arabic characters' coding from Unicode to base64 coding scheme and has developed a runt version of lossless Huffman coding scheme.

In addition, in 2011, Atee [6] developed a way to encrypt the Arabic characters (letters) by using symmetric encryption XOR logical function, and binary and decimal encoding schemes to convert letters to decimal ambiguous numbers.

Moreover, in 2013, Alqahtani et al. [7] introduced a new approach for encrypting Arabic letters by using vigenere cipher. The authors assigned numbers to Arabic letters (أ-ي) and space and numbers from (0-9), so the addition is carried out via taking the modular of 39.

Furthermore, Aysan and kuppuswamy [8] in 2014 proposed the mixed encryption algorithm based on a simple multiplication and logarithm function with Caesar cipher to encrypt Arabic texts.

In 2016, Rihan and Osma [9] introduced a cryptography technique for the Arabic language using neural network. The authors used hebbian network with genetic concept.

In 2016, Hamid [10] used back propagation Algorithm with artificial neural to build a cipher system for encrypting any Arabic text to prevent any data attack during the transition process.

In 2016, Habeeb [11] used Genetic Algorithm (GA) to attack an Arabic encrypted text by Vigenere cipher. The algorithm was tested to find the key letters for different cipher text sizes and key lengths.

In 2017, Hashim et al [12] used symmetric and asymmetric cryptosystems to encrypt texts that are written in English or Arabic by adopting Vigenere Cipher and RSA Cryptosystem.

In 2017, Abduljabbar [13] proposed a new approach to encrypt text messages based on genetic algorithm operators. The author generated 8 bit chromosome to encrypt plain texts after selecting randomly crossover points.

In 2017, Hadi [14] proposed a system which depends on encryption to solve many problems for SMS in the Arabic language.

Finally, in 2017, Najim al-din and Shaban [15] proposed a new encryption method to encrypt Arabic texts by using the principle of integration to provide better security and increase the complexity of guessing the correct keys and correct plain texts.

5. Proposed Method:

After reviewing the previous studies in this field and studying the methods of encoding the Arabic texts proposed by the researchers to determine their strengths and weaknesses, a set of basic points are developed and defined as determinants when designing the proposed method. These determinants can be summarized as follows:

1. The proposed method should provide faster execution in both encryption and decryption.
2. There should not be a significant increase in the length of the resulting encoded text compared with the length of the plain text.
3. The method should provide a high level of confidentiality for the encrypted text and should suit texts of different lengths.

The proposed method incorporates the concept of transposition encryption methods with substitution as follows:

A- Substitution Encryption:

In this research, we have relied on the diacritic marks in the Arabic language and used them to perform efficient substitution processes as follows:

1. Remove dots from the Arabic characters and use diacritics (ً, ٍ, ٌ) instead of them, as shown in Table 1.
2. Use diacritics (ِ,ُ,َ,ْ,~ ,ّ) to substitute the most frequent Arabic characters [11], as shown in Table 2.
3. The non-use or appearance of the characters (which are compensated in Tables (1, 2)) in the resulting encrypted text might cause suspicion to the intruder or attacker. Also, in order to provide a compression percentage of the length of the resulting encrypted text to achieve the second point of the parameters of the proposed method, the characters that have been substituted in Tables (1, 2) were used to substitute for some question tools (cases of Hamza and space and some common passages in the Arabic language, as shown in Table (3)).

Table 1 dotted characters substitution

ة	ي	ق	ف	غ	ظ	ض	ش	ز	ذ	خ	ج	ث	ت	ب
ةً	يٍ	قّ	فً	غً	ظً	ضً	شّ	رً	دً	حً	حٍ	ثّ	تً	بٍ

Table 2 most frequent Arabic characters substitution

ن	و	ه	م	ل	ا
نّ	~	هْ	مَ	لُ	لِ

Table 3 removed characters substitution

مِن	ب
الى	ت
عن	ث
على	ج
في	خ
ء	ذ
وَ	ز
ئ	ش
ى	ض
هل	ظ
مَن	غ
أين	ف
ما	ق
متى	ي
Space	ا
ال	ل
لا	م
قد	ه
حيث	و
سوف	ن
إن	ة

B- Transposition Encryption:

For the purpose of the transposition process, the column substitution encoding method is used as follows:

1. Enter the resulting text from the substitution processes (step A) to the matrix whose number in columns is five from the lower right corner vertically.
2. The encryption key which is used to replace the matrix columns i calculated according to the following equations:

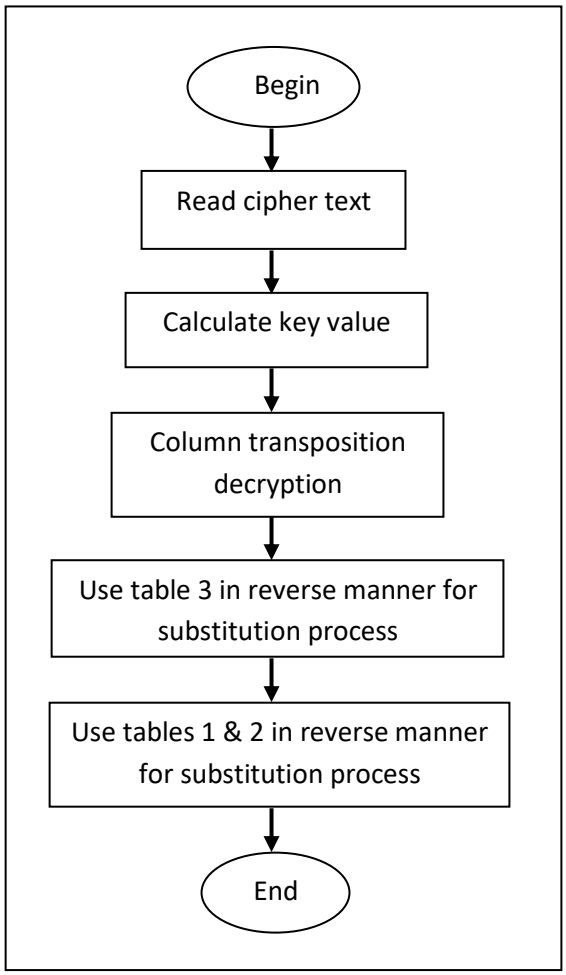


Figure 2 decryption process

A = text length - the most repetitive character
B = the most repetitive character -2
C = the less repetitive character + 3
D = Number of diacritics ً + 5
E = Number of diacritics ً + 7

Therefore, the key formula is ABCDE. After the process of encoding the columns, we get the text encoded in Arabic. The encryption process for the proposed method can be illustrated by the flowchart in Figure (1).

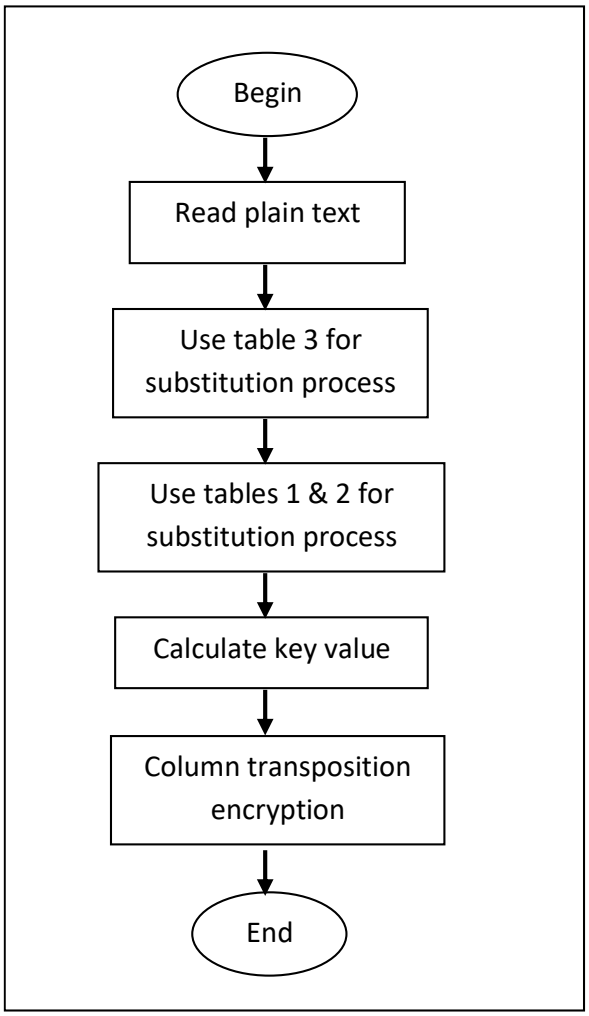


Figure 1 encryption process

When the decryption process is performed for the proposed method, the cryptographic process steps are used reversely to obtain the clear text. The steps of the decryption process for the proposed method can be illustrated by the flow diagram in Figure (2).

6. Result and Discussion:

A- At the beginning, an example of encryption of the proposed method will be clarified:

In this example, we have the plain text " من حسن إسلام المرء تركه ما لا يعنيه". The steps of encryption of the proposed method will be used in detail and according to the flowchart in Figure 1 to get the cipher text as follows:

- At the beginning, we use table 3 to conduct substitution operations for some characters and sections as shown below:

م	ب
ن	
ا	
ح	ح
س	س
ن	ن
ا	
إ	إ
س	س
ل	م
ا	
م	م
ا	
ا	ل
ل	
م	
ر	
ء	ذ
ا	
ت	ت
ر	ر
ك	ك
ه	ه
ا	
م	ق
ا	
ا	
ل	م
ا	
ا	
ي	ي
ع	ع
ن	ن
ي	ي
ه	ه

The characters resulting from the substitution process are shaded to distinguish them from the original characters of the encoded text.

- Tables 1 and 2 are then used to perform substitution operations for certain characters, of course, not the shaded characters.

م	ب	ب
ن		
ا	ا	ا
ح	ح	ح
س	س	س
ن	ن	ً
ا	ا	ا
إ	إ	ِ
س	س	س
ل	م	م
ا		
م	م	َ
ا	ا	ا

ا	ل	ل
ل		
م	م	َ
ر	ر	ر
ء	ذ	ذ
	ا	ا
ت	ت	ً
ر	ر	ر
ك	ك	ِ
ه	ه	ْ
	ا	ا
م	ق	ق
ا		
	ا	ا
ل	م	م
ا		
	ا	ا
ي	ي	ى
ع	ع	ع
ن	ن	ً
ي	ي	ي
ه	ه	ْ

After the substitution processes are completed, the column transposition method is used. At the beginning, calculate the value of the key used to replace the columns as follows:

A = text length - the most repetitive character

A = 30 – (ا) 7 = 23

B = the most repetitive character -2

B= (ا) 7 – 2 = 5

C = the less repetitive character + 3

C = (ب،ت،ث،،) 0 + 3 = 3

D = Number of diacritics ْ + 5

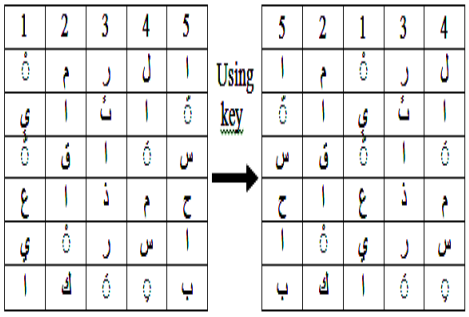
D = 0 + 5 = 5

E = Number of diacritics ّ + 7

E = 2 + 7 = 9

Key = ABCDE = 23 5 3 5 9 = 52134

After calculating the value of the key, the string of characters resulting from the substitution processes is entered into the matrix from the lower right corner and vertically as shown below:



To extract the encrypted text, the values are drawn from the matrix by specifying a particular angle and direction, for example, "upper left corner and horizontally:

Cipher ="ا م ْ ر ل ّ ا ي ا ً ت س ق ّ ا َ ح ا ع ذ م ْ ا ي ر س ب ك ا َ ِ"

B- For the purpose of demonstrating the efficiency of the proposed method, a set of Arabic texts of different lengths and subjects has been encrypted. The lengths of all ciphered texts were less than the lengths of the plain text, as shown in Table 4 and Figure 3.

Plain text length	Cipher text length
25	22
35	30
76	54
97	75
145	121
319	285
892	837

Table 4 comparison between plain and cipher texts length

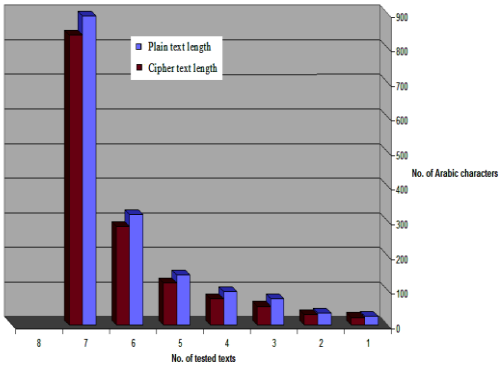


Figure 3 comparisons between plain and cipher texts length

C- Also, to illustrate the efficiency of the proposed encryption method by comparing the statistics of the plain text characters and cipher text, the frequency of letters was calculated for seven different texts. It was found that there is a difference in frequencies of most of the characters, as shown in Table 5 and Figure 4.

Table 5 plain and cipher texts letters frequencies'

Arabic letters	Plain text letters frequency	Cipher text letters frequency
ا	40	54
ب	3	7
ت	7	3
ث	1	0
ج	3	5
ح	3	3
خ	2	3
د	6	6
ذ	3	5
ر	2	2
ز	1	2
س	5	5
ش	2	0
ص	3	3
ض	2	1
ط	3	3
ظ	2	1
ع	9	5
غ	4	1
ف	8	2
ق	4	6
ك	6	6
ل	22	9
م	18	5
ن	15	3
ه	19	4
و	16	3
ي	18	2

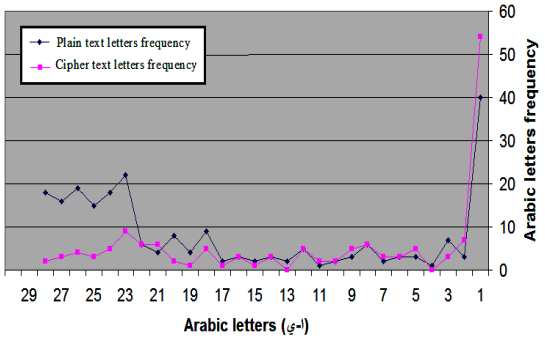


Figure 4 plain and cipher texts letters frequencies'

7. Conclusion:

A- The use of diacritics is a new and effective way of encoding Arabic texts.

B- In most of the encrypting Arabic texts methods, the length of the encrypted text is much larger than the length of the plain text. In the proposed method, the length of the encrypted text is less than the length of the plain text. Increasing the length of the clear text leads to increasing the probability of many substitution processes and, therefore, there will be a significant decrease in the length of the encoded text.

C- In symmetric encryption methods, sending the key to the recipient is required to be used in the decryption process. In the proposed method, the key is calculated at the receiving party based on the encrypted text received and there is no need to send the key. Thus, there will be a new key for each encrypted text, which increases the strength of the encryption method.

D- The use of diacritics to perform substitution operations has contributed to a significant change in the frequencies of cipher text characters compared to the plain text, thus making it difficult to conduct frequency analysis attack.

8. Reference:

1. William Stallings, 2014, "Cryptography and Network Security Principles and Practice, Sixth Edition, Pearson Education, Inc., Printed in the United States of America.

2. Behrouz A. Forouzan , 2007, "Cryptography & network security" , tata Mc Graw-Hill Behrouz A. Forouzan , 2007, "Cryptography & network security" , tata Mc Graw-Hill.

3. C. Margare and M. Steven, 2013, "The Mathematics of Encryption An Elementary Introduction", the American Mathematical Society. Mathematical World Volume 29 .

4. Elzenati hesham , 2016 , "Arabic XML documents : summarizing , managing , and securing", doctoral dissertation , singidunum university , department of postgraduate studies , Belgrade.

5. Abdullah A. , 2009 , "Enhancing cost and security of Arabic SMS messages over mobile phone network" , Rafidain journal of computer sciences and mathematics , vol.6 , no.3 , university of mosul , Iraq.

6. Atee Haifaa , 2011 , "Development of a new way to encrypt the Arabic language letters using the symmetric encryption system" , Al-mustansiriyah Journal of science , Al-mustansiriyah university , Iraq.

7. Alqahtani yahya and et al. , 2013 , "New Approach of Arabic encryption , decryption technique using vigenere cipher on mod 39" , international jouranal of advanced research in it and engineering , VOL. 2 , NO.12.

8. Aysan mohammed and kuppuswamy prakash , 2014 , "Hybrid combination of message encryption techniques on Arabic text : using new symmetric key and simple logarithm function" , International Journal of scientific knowledge , vol. 5 , no. 4.

9. Rihan shaza and osma saif , 2016 , "ARABIC cryptography technique using neural network and genetic algorithm" , international research journal of computer science (IRJCS) , issue 05 , vol.3.

10. Hamdi oday , 2016 , "Arabic text encryption using artificial neural networks" , engineering and technology journal , vol. 34 , no. 5.

11. Habeeb Rokaia , 2016 , "Arabic text cryptanalysis using genetic algorithm" , Iraqi journal electrical and electronic engineering , vol.12 , no. 2.

12. Hashim Hayder and allkufi mohammed , 2017 , "A proposed method for text encryption using symmetric and asymmetric cryptosystem" , International journal of computer trends and technology" , vol.50 , no.2.

13. Abduljabbar Riyadh , 2017 , "Fast approach for Arabic text encryption using genetic algorithm" , European journal of scientific research , vol. 144 no.4 , pp. 342-348.

14. Hadi ameer , 2017 , "toward trust and more characters of Arabic short message service using encryption" , journal of engineering and applied sciences , vol. 12 , no. 21 , med well journals.

15. Najim al-din basim and shaban saad , 2017 , "A new algorithm for encryption Arabic text using the mathematical equation" , Diyala journal of engineering sciences , vol. 10 , no.01 , pp.21-30 , university of diyala , Iraq

طريقة مقترحة لتحقيق سرية النصوص العربية

د. ياسين حكمت إسماعيل

قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل
Yasino79@yahoo.com

المستخلص:

نظرًا للتطور السريع لنظم الحوسبة الموزعة ، استخدام الشبكات المحلية ، والتوسع الهائل في الإنترنت، أصبحت عملية الحفاظ على سرية المعلومات مهمة وضرورية . يعد التشفير أحد أهم الطرائق للحفاظ على سرية المعلومات ومنع الأشخاص غير المصرح لهم من الكشف عنها . ظهرت العديد من خوارزميات التشفير بأساليب وطرائق مختلفة. ولكن تم تخصيص معظم هذه الخوارزميات لتشفير النصوص في اللغة الإنجليزية. بسبب تزايد تبادل المعلومات الهامة والحساسة في اللغة العربية من قبل المستخدمين على الإنترنت، ظهرت حاجة ملحة لبناء أنظمة تشفير مخصصة للغة العربية. الهدف من هذا البحث هو تقديم طريقة تشفير مقترحة تستخدم فكرة دمج طرائق التشفير الإبدالية والتعويضية لتحقيق السرية للنص العربي. يعتمد هذا البحث على استخدام علامات التشكيل في اللغة العربية لأداء عملية التشفير التعويضية. تم اختبار كفاءة الطريقة المقترحة على نصوص مختلفة ؛ وفرت الطريقة سرعة تنفيذ عالية ، عدم زيادة في طول النصوص المشفرة مقارنة بالنصوص الواضحة ، وكذلك سرية عالية للنصوص المشفرة الناتجة.