ARAE CIPHER SYSTEM

OSAMA MAHMOUD ABU ABBAS, KHALID MOHAMMAD NAHAR, AND MOHAMMAD AHMAD TUBISHAT
Computer Science Department, IT Faculty, Yarmouk University, Jordan

ABSTRACT

Data encryption is the translation of data into a form that unintelligible without a deciphering mechanism. This paper is intended to introduce a new approach for encryption, ARAE cipher system. It uses Arabic letters and their diacritics for encrypting English messages and vice versa. A pseudo random generator is used to generate integer numbers to represent each character in Arabic language. The same numbers are used again after sorting them to represent the English characters. The conclusions that are extracted indicate the efficiency of ARAE system according to security and time performance.

Keywords: *Encryption, Decryption, Pseudo Random Generator, Arabic Characters, RSA, Sorting Algorithm.*

1. INTRODUCTION

Cryptographic algorithms exist to protect information first, by transforming data into a form that is meaningless to humans, such as a string of zeroes and ones, and, second, by performing certain manipulations to these transformed data so that even a specially designed machine cannot recover the original text unless provided with a secret key. The data that has been transformed can be transmitted electronically, so that even if an eavesdropper manages to read the message, its true content will remain hidden. The receiving party, however, is equipped with a secret key, so it can read the original data. This process of hiding data is referred to as encryption. The process of reversing the transformation is called decryption. Sometimes the entire process that includes both the encryption and decryption of data is called cipher. Plaintext refers to data in plain or unencrypted form. Ciphertext refers to data in encrypted or enciphered form [5].

ARAE cipher system encrypts any English message using Arabic letters and their diacritics and vice versa. Each Arabic character is associated with an integer number produced by pseudo random generator which has four factors: a, b, m, and x_0. Each English character is assigned one of these numbers after sorting them using any efficient sorting algorithm. In the encryption process for an English message, for example, each English character in the message is encrypted to one of the Arabic characters according to the integer number associated with it. The decryption process is performed in the same way.

Following are some of the previous studies that introduced new approaches in encryption or improved old ones. However, ARAE system uses different approach and new ideas:

- K. Verma, Mayank Dave and R. C. Joshi presented a cryptanalysis method based on Genetic Algorithm and Tabu Search to break a Mono-Alphabetic Substitution Cipher in Adhoc networks. They have also compared and analyzed the performance of these algorithms in automated attacks on Mono-alphabetic Substitution Cipher. As a result they concluded that the use of Tabu search is largely an unexplored area in the field of Cryptanalysis and generalized version of these algorithms can be used for attacking other ciphers as well [9].
- Mark G. Simkin discusses five encryption techniques: transposition ciphers, cyclic substitution ciphers, Vigenere ciphers, exclusive OR ciphers, and permutation ciphers. Accompanying these discussions are explanations of how instructors can demonstrate these techniques with spreadsheet models [10].
- Kwok-Wo Wong, Sun-Wah Ho, and Ching-Ki Yung modified the chaotic cryptographic scheme so as to reduce the length of the ciphertext to a size slightly longer than that of the original message. Moreover, they introduced a session key in the cryptographic scheme so that the length of the ciphertext for a given message is not fixed [4].
- Jun Wei, Xiaofeng Liao, Kwok-wo Wong, and Tao Xiang proposed a new chaotic cryptosystem. Instead of simply mixing the chaotic signal of the proposed chaotic cryptosystem with the ciphertext, a noise-like variable is utilized to govern the encryption and decryption processes. This adds statistical sense to the new cryptosystem [3].
- Chien-Yuan Chen, Cheng-Yuan Ku, and David C.Yen found ways to utilize the LLL algorithm to break the RSA system even when the value of d is large. According to their proposed cryptanalysis, if d satisfes $|\lambda - d| < N0.25$, the RSA system will be possible to be resolved computationally [2].
- Chang-Doo Lee, Bong-Jun Choi, and Kyoo-Seok Park proposed a block encryption algorithm which is designed for each encryption key value to be applied to each round block with a different value. This algorithm needs a short processing time in encryption and decryption, has high intensity, and can be applied to electronic commerce and various applications of data protection [1].

2. ARAE CIPHER SYSTEM

ARAE cipher system holds new methodology in encryption and decryption. This methodology is summarized in the following steps:

1. A set of all Arabic characters and their diacritics are encoded alphabetically in array using pseudo random generator. In this step, each Arabic character is assigned a unique number.

2. Each English character is assigned one of the numbers generated in step 1 after sorting them using any efficient sorting algorithm. The English characters and their corresponding numbers are stored alphabetically in array.
3. Any message written in English language can be encrypted using Arabic characters by dividing the message into characters then match and substitute each character in the message with its corresponding character in Arabic array, or vice versa.
4. The encrypted message and the encrypted factors of the pseudo random generator are sent to the receiver.
5. The Arabic and English arrays are built again by the receiver using the same pseudo random generator factors (after decrypting them). According to these two ordered arrays the receiver decrypts the message in the same way was encrypted.

3. PSEUDO RANDOM GENERATOR

Pseudo random generator has four factors that determine the set of numbers produced by it. These factors are a, b, m, and x_0. The most popular recursion formula for the pseudo random generator is:

$$\boxed{\mathbf{X_{n+1} = aX_n + b\ (mod\ m),\ for\ n>=0}}$$

where a, b, and m are fixed integer constants and x_0, called the seed, represents the first integer number in the series. So, starting from x_0, the formula gives rise to sequence of integers between 0 and m-1. The maximum number of different integers could be produced by pseudo random generator is m. It would be sensible for m to be some positive integral power of 10. For example, suppose $x_0 = 89$, $a = 1573$, $b = 19$, and $m = 10^3$, then:

$X_1 = 1573 * 89 + 19 \pmod{10^3} = 16$, and

$X_2 = 1573 * 16 + 19 \pmod{10^3} = 187$, and so on.

It is very important to note that using the same factors (a, b, m, and x_0) generates the same set (X_1, X_2, ….X_n). The operation of division by m is most efficiently done if $m = r^k$ for some positive integer k. For most computers, this entails setting $m = 2^k$ where k is selected so that m is large and the numbers involved are within the accuracy of the machine. The pseudo random generator can produce no more than m different numbers before the cycle repeats itself again and again [8].

4. BUILDING ARABIC ARRAY

In this phase, the pseudo random generator is used to generate a set of random numbers and then assign each random number produced to one of the Arabic characters.

In Arabic array all shapes of Arabic letters and diacritics are used and stored in alphabetical order. Storing Arabic characters in alphabetical order is very important in the decryption process as will be explained in section 8. For each Arabic letter, three shapes are used. The shape of the letter when it appears at the beginning of the word, the shape of the letter when it appears in the middle of the word, and the shape of the letter when it appears at the end of the word. For example, the Arabic letter ك, has three shapes: كـ at the beginning of the word, ـكـ in the middle of the word, and ـك at the end of the word. All kinds of diacritics are used also (ٌ ُ ْ ً َ ٍ).

Each Arabic character is assigned a unique random number produced by pseudo random generator. Figure 1 shows an example on the Arabic array

Figure1: Arabic Array

ا	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	…
68	17	18	11	71	14	56	22	59	29	36	35	44	…

5. BUILDING ENGLISH ARRAY

In this phase, the numbers that are produced by the pseudo random generator for building the Arabic array are sorted in ascending order using any efficient sorting algorithm. Then, they are assigned to the ordered English characters. The English characters and the numbers that are assigned to them are stored alphabetically in an array. Figure 2 shows an example on the English array.

Figure2: English Array

A	B	C	D	E	F	G	H	I	J	K	L	M	…
11	14	17	18	22	29	35	36	44	56	59	68	71	…

6. ENCRYPTION PROCESS

Now, after construction of Arabic and English arrays, any English text could be encrypted to Arabic characters by matching each English character and Arabic character holding the same number. Vice versa is also possible; any Arabic text could be encrypted to English characters by matching each Arabic character and English character holding the same number.

For example, according to figures 1 and 2, the Arabic message ساذج is encrypted by ARAE system to GLKM, because the Arabic letter س is associated with the random number 35 in the Arabic array which is, in its turn, associated with the letter G in the English array. The Arabic letter ا is associated with the random number 68 in the Arabic array which is, in its turn, associated with the letter L in the English array. Also the Arabic letter ذ is associated with the random number 59 in the Arabic array which is, in its turn, associated with the letter K in the English array. Finally, the Arabic letter ج is associated with the random number 71 in the Arabic array which is, in its turn, associated with the letter M in the English array. However, as it was previously explained, vice versa is possible. For example, the English message CAFE is encrypted by ARAE system to د ر ث د.

7. SENDING PROCESS

In this phase, the sender sends the encrypted message, being an English or Arabic one, to the receiver. He also sends the pseudo random generator factors: a, b, and m to the receiver after encrypting them using RSA algorithm. The last factor of pseudo random generator, the seed x_0, could be arranged by an agreement between the parties to increase the security of ARAE system.

Since the three factors a, b, and m could be attacked by brute force method, the seed x_0 must remain a secret between the parties. As such it is very difficult to be attacked. Some examples on agreement between the parties are: x_0 is the birthday of sender's mother, x_0 the time of landing the nuclear bomb on Hiroshima and Nagasaki, or any other agreement.

Encrypting only three pseudo random generator factors by RSA takes very little time compared with encrypting and decryption all the message by RSA. This increases the performance of ARAE system and reduces the encryption and decryption time.

RSA algorithm is summarized in the following steps [6, 7]:

1. Choose two large primes, p and q.
2. Compute n = p*q and z = (p-1)*(q-1).
3. Choose a number relatively prime to z and call it d.
4. Find e such that e*d=1 mod z.
5. To encrypt a message (plaintext), P, compute C = P^e (mod n).
6. To decrypt a message (ciphertext), C, compute P = C^d (mod n).

Note that to perform the encryption, you need e and n and to perform the decryption, you also need d and n. Therefore, the public key consists of the pair(e, n), and the private key consists of (d, n).

8. DECRYPTION PROCESS

In this phase, the receiver receives only the encrypted message and the encrypted three pseudo random generator factors: a, b, and m. The seed x_0 is known to the receiver by an agreement between him and the sender.

The three factors: a, b, and m are decrypted by RSA algorithm and then used with the seed x_0 to rebuild the Arabic array by using the same generator to produce the same numbers that were produced in the encryption process. The Arabic characters must be stored in alphabetical order in the Arabic array before assigning them random numbers. This is very important to ensure building the same Arabic array in the encryption process.

Then, the receiver sorts the numbers that are produced by pseudo random generator and assigns them to the ordered English characters.

Finally, the receiver finds the random numbers associated with each character in the encrypted message in one of the arrays (according to the language of the original message) and matches it with the same number in the other array. The character associated with that number in the other array is the decryption of the encrypted message character.

There is, still, only one problem, how can the receiver know the language of the encrypted message that has been received? Especially, if the message is a mix of Arabic and English characters. To solve this problem, the receiver must check the language for every character, if it is Arabic then the original character is English and he must match this character with a character in the English array. But, if it is English then the original character is Arabic and he must match this character with another character in the Arabic array. However, special characters that are common between English and Arabic languages are considered English if they lies between English characters and considered Arabic if they lies between Arabic characters.

The receiver can check the language of any character by checking its ASCII code.

9. CONCLUSIONS

ARAE system is an encryption system for English and Arabic texts. After analyzing the results of ARAE system, the following conclusions are obtained.

- ARAE system is very efficient since it needs fixed time for building the Arabic and the English arrays. Also it needs fixed time for matching the characters of the message since the number of elements for the two arrays are fixed. In addition, RSA algorithm is applied only to three factors: a, b, and m which needs certain and fixed time. The results shows that the average time for encrypting text of size 1000 characters is 0.06 milliseconds, the average time for encrypting text of size 10000 characters is 0.8 milliseconds and the average time for encrypting text of size 100000 characters is 50 seconds.
- ARAE system is also very secure since it uses a hybrid of RSA algorithm (for encrypting the factors: a, b, and m) and a secret between the parties (for hiding the seed x_0). So the attackers need two levels of attack to break ARAE system. One level is to attack RSA for obtaining the factors and another level is to attack the secret for obtaining the seed x_0.
- The pseudo random generator is very effective because it can produce m different numbers without duplications and because it depends on more than one factor which gives the system more security.

REFERENCES

[1] Chang-Doo Lee, Bong-Jun Choi and Kyoo-Seok Park, "Design and evaluation of a block encryption algorithm using dynamic-key mechanism". *Future Generation Computer Systems*, Volume: 20, Issue: 2, Pages: 327 – 338, 2004.

[2] Chien-Yuan Chen, Cheng-Yuan Ku b and David C.Yen, “Cryptanalysis of large RSA exponent by using the LLL algorithm”, *in Proceedings of The Tenth National Conference on Information Security,* Taiwan, Pages:45-50, 2000.

[3] Jun Wei, Xiaofeng Liao, Kwok-wo Wong, and Tao Xiang, “A new chaotic cryptosystem”, *Chaos Solitons & Fractals* 30 (5): 1143-1152 Dec 2006.

[4] Kwok-Wo Wong, Sun-Wah Ho, and Ching-Ki Yung, “A chaotic cryptography scheme for generating short ciphertext”, *Physics Letters A*, Volume 310, Number 1, Pages: 67-73, 2003.

[5] Peyravian Mohammad, Roginsky Allen and Zunic Nev. “Hash-Based Encryption System”. *Computers & Security,* Volume 18, Issue 4, Pages: 345-350, 1999.

[6] Rivest, R., Shamir A., and Adleman L. "A Method for Obtaining Digital Signatures and Public Key Cryptosystems", *Communications of the ACM,* Volume 21, Number 2, Pages: 120-126, 1978.

[7] Tanenbaum, Andrew S., *Computer Networks,* Fourth Edition, Prentice Hall PTR, 2003.

[8] Byron J. T. Morgan, *Elements of Simulation*, Chapman and Hall, 1984.

[9] A. K. Verma, Mayank Dave and R. C. Joshi, "Genetic Algorithm and Tabu Search Attack on the Mono-Alphabetic Subsitution Cipher in Adhoc Networks", *Journal of Computer Science* Volume 3, Number 3, pages: 134 -137, 2007.

[10] Mark G. Simkin, "Using Spreadsheets to Teach Data Encryption Techniques", *AIS Educator Association,* Volume 1, Number 1, pages 27 – 37, 2006.