ISMSIT2017

1st International Symposium on Multidisciplinary Studies and Innovative Technologies Proceedings Book

(November 2-4, 2017, Tokat, Turkey)

Symposium homepage: www.ismsitconf.org

Cryptanalysis using Artificial Bee Colony Algorithm Guided by Frequency based Fitness Value

Arkan Kh Shakr Sabonchi [1*], Bahriye Akay [2]

[1,2] *Computer Engineering Department, Erciyes University, Turkey*

[*] *arkankhaleel@gmail.com*

Abstract – Researchers attention on the optimization methods to solve several applications increases recently by the advances in the technological field. The Artificial Bee Colony algorithm that has been quite popular during the recent years is firstly used in this study to cryptanalyse the substitution cipher that is the fundamental structure of the modern cryptology methods. Artificial Bee Colony algorithm searches for the key used for encryption phase and it is aimed to obtain the decrypted text by the key discovered. Various number of texts and population sizes are examined to obtain the original key by using the character frequency analysis. Consequently, from the results, the Artificial Bee Colony Algorithm displays good performance for cryptanalysis application in substitution cipher.

Keywords – cryptanalysis; substitution cipher; Artificial bee colony algorithm (ABC);

I. INTRODUCTION

The security has a great importance in all information and data systems. Cryptography is a key topic for a secure data transfer. Naturally, the encryption algorithms are the fundamental structures of the crypto system. Cryptology is a combination of the words of Krypto's (hidden) and lo'gos (word) from Greek and called as the privacy science [1]. Cryptology is divided into two categories, cryptography, and cryptanalysis [2,3]. The science that deals with hiding the data, making them trusted and making the data impossible to be read and understood are called as the cryptography. Creating the encrypted data as readable and understandable are referred as cryptanalysis. It is possible to divide the Cryptology methods into two parts; Traditional and Modern methods. The traditional methods are analyzed under two titles such as transposition and substitution [2,3]. Cryptanalysis and Swarm Intelligence applications are used in various cryptology systems to develop new and more reliable cryptology systems. By the food seeking behaviors of ants and acting of bird and fishes as the swarm emerges the particle swarm optimization algorithms (PSO) and Ant Colony Optimization [4,5]. Besides, by the year of 2000, the number of investigations have been conducted on the behaviors of honey bees that act as an intelligent swarm. The algorithm models such as bees algorithm [6], virtual bee algorithm [7], bee system [8], bee colony optimization, artificial bee colony have been emerged as the result of studies conducted. A genetic algorithm that is one of the first examples of these systems was used to cryptanalyze the cryptology systems in 1993, and the results were efficient [9]. Differential Evolution Algorithm [10], Ant Colony Optimization [11], Particle Swarm Optimization [12], and Tabu Search Algorithm are also employed. Moreover, artificial bee colony algorithm hasn't been utilized in cryptanalysis. In this study, the Artificial Bee Colony Algorithm that is a successful optimization algorithm is used in cryptanalysis transaction for the first time. The contribution that is gained from the Bee Colony algorithm will bring a new system to the literature. More secured systems can be obtained by this new approach and can raise the data security to higher levels.

II. SUBSTITUTION CIPHER

A substitution cipher can be categorized as Mono Alphabetic and Poly Alphabetic [15]. It is preferred to use Mono Alphabetic within the scope of this study, and the numerical approach is as follows [16]: Let $P=C=Z_{26}$. K, includes all the possible permutations of 26 symbols as 0, 1, 2, ……., 25. For each permutations $\pi \in K$, $e_\pi(x)=\pi(x)$ and $d\pi(x)=\pi^{-1}(x)$, The random example of permutation π is as follows. This example can also be used in encryption transactions.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
X	N	Y	A	H	P	O	G	Z	Q	W	B	T	S	F	L	R	C	V	M	U	E	K	J	D	I

Fig. 1 Example of substitution cipher

The encrypted text is obtained by replacing the characters in unencrypted plain text based on Figure 1, with the matching characters. For example:

Plain Text: *ERCIYESUNIVERSITESI*
Encrypted Text: *HCYZDHVUSZEHCVZMHVZ*

The decryption processing is the reverse permutation. As is seen in the example below, the second row is written first; then it is continued by the alphabetical order [16].

Encrypted Text: *HCYZDHVUSZEHCVZMHVZ*
Plain Text: *ERCIYESUNIVERSITESI*

It is assumed that the text used during cryptanalyses process is an ordinary text and there is no space or punctuation marks in it.

III. CHARACTER FREQUENCY BASED METHOD

Several techniques of cryptanalyses use the statistical properties of English language. Character frequency based

method is one of the techniques are used in cryptanalysis process. Beker and Piper divided the 26 characters into groups below (unigram, bigram and trigram) [22]. Frequencies of the unigram, bigram and trigrams are shown in Figure 2, 3 and 4.

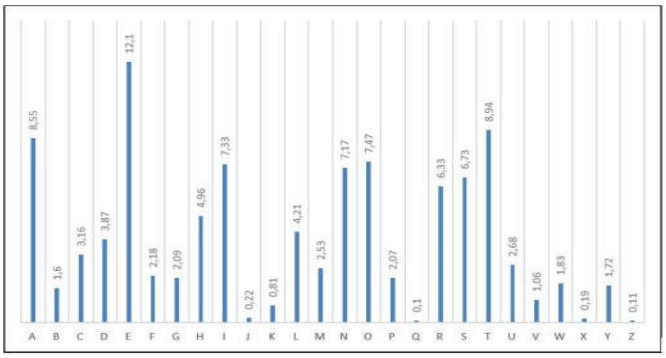

Fig. 2 Unigram statistics

Character frequency based method was firstly used by Matthews in 1993 [9] and by Clark in 1994 [23]. The dictionary database was created in 1995 by Lin and Kao [24]. Then, the fitness value is found by comparing the words in the dictionary and the text decrypted through possible keys before placing the words that are frequently used in the language.

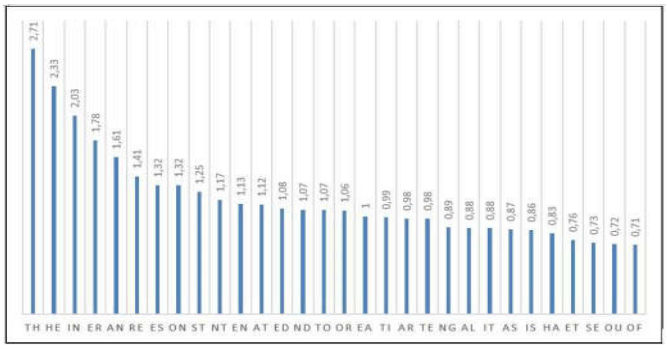

Fig. 3 Bigram statistics

$$fitness(k) = x_1 \cdot \sum_{a\in(A..Z)} |K^{u} - D^{u}| + x_2 \cdot \sum_{a\in(A..Z)} |K^{b} - D^{b}| + x_3 \cdot \sum_{a\in(A..Z)} |K^{t} - D^{t}| \quad (1)$$

fitness(k) = solution value for each key.

x_1, x_2, x_3 = the values between 0 and 1 are received to change the n-gram values.

$a \in (A..Z)$ = shows the English characters.

K = shows the frequency value of standard English characters such as in Table1,2,3.

D = shows the text frequency value for the text decoded.

u,b,t = indicates to unigram, bigram and trigram.

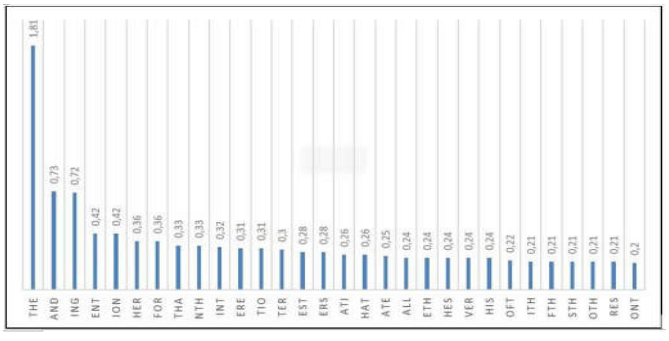

Fig. 4 Trigram statistics

IV. ARTIFICIAL BEE COLONY ALGORITHM

ABC algorithm mimics the food seeking behavior of honey bees. Even though the bees live together, any delay or confusion aren't observed by means of their flawless work distribution and the self-organizing skills. In ABC algorithm, honeybees search for the food sources within a particular structure. It is seen that this algorithm has better performance in comparison with the genetic algorithm and differential Evolution algorithms [17- 21]. There are three different types of bees in ABC algorithm; scout bee, employed bee and the onlooker bee. Each of the employed bees is responsible for a separate food source, and the number of the employed bees equal to the number of onlooker bees. The employed bees search for the food sources and share the relevant information with onlooker bees. Onlooker bees start to move to food sources based on the information received from employed bees. The employed bees turn into scout bees in case of exhaustion of sources, then the scout bees start to find a new source.

A. *Basic steps of ABC Algorithm*

1. Randomly determine initial food sources
2. **do**
3. Send the employed bees to the food sources
4. Calculate the probability values of food sources
5. Send the onlooker bees to the food sources based on the probability values
6. Send scout bees to search for new food sources
7. **while** (Is the termination condition provided?)

The position of a food source is equal to a solution in ABC algorithm. The richness of food source concerning pollen or nectar is defined by the objective function. ABC algorithm starts with generating the food source locations by Eq. (2);

$$x_{ij} = x_j^{\min} + rand(0,1)(x_j^{\max} - x_j^{\min}) \quad (2)$$

In here, x_{ij} represents the position at jthe parameter of the source i; i= {1, 2,…,SN}, j={1,2,…D}; SN shows the size of population and D indicates the total number of position. x_j^{max} and x_j^{min} refer to the lower and upper limits of positions and rand (0,1) is a number randomly produced between 0 and 1.

After the initial population is generated, the seeking phase is started by the employed bees. An employed bee keeps the information (position and quality) of her source in memory and determine a new food source by Eq. (3) within the neighborhood of the relevant food sources. If the quality of the food source is (objective function value) better than the food source in memory, the old food source is deleted from the memory and the new one is stored.

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - k_{kj}) \quad (3)$$

where *vi* is new source produced by changing a single position of the source at i line (the position at j column selected randomly) with the position of the source at k line selected randomly. ϕ_{ij} is a number that is accidentally produced between -1 and 1. Afterwards, the employed bees share the data obtained with onlooker bees after completing the research process in sources they are responsible for. The onlooker bees choose the sources with respect to the probability values

calculated by Eq. (4) based upon the nectar amount of the sources within the data received from the employed bees.

$$pi = \frac{fitness_i}{\sum_{j=1}^{sn} fitness_i} \quad (4)$$

The fitness values in here are calculated by the Eq. (5)

$$fitness_i = \begin{cases} \frac{1}{1+f_i} & f_i \geq 0 \\ 1+abs(f_i), & f_i < 0 \end{cases} \quad (5)$$

f_i shows the objective function value, $fitness_i$ refers to the fitness value of the nectar amount of the source at i line. As the fitness value of a source enhances, the probability of being selected of that source increases as well. The onlooker bees detect new sources by Eq. (3) by using the neighborhood mechanism such as the employed bees after specifying the sources to search in. The nectar amount of the new source is evaluated. If the nectar amount of the new source is adequate, the old source is deleted from the memory, and the new one is stored. In a word, the onlooker bees and the employed bees apply the greedy selection. The exhausted sources are determined by 'limit' parameter and a new random solution is produced by Eq. (2) instead of the depleted source.

V. THE PROPOSED ALGORITHM STEPS

1. Enter iteration number, limit value, key length, food source number and also the text encrypted.
2. Producing of starting population randomly (possible keys).
3. Ensure the text as much as the number of population by using the population randomly produced via Substitution Chiper program.
4. Calculate the solution value for decrypted texts by using the Equation 1 and store the best solution value by comparing the values.
5. Produce new population as much as the number of starting population for the employed bees.
    a. For each solution, generate a solution using Equation 3 and evaluate the solution by Equation 1.
    b. Compare the solution values of new population and starting population, then keep the best value in mind.
6. Compute the probability values for each solution by the selection process by Equation 4.
7. Produce new population by onlooker bees based on the probability values.
    a. For each solution, generate a solution by using Equation 3 and evaluate it by Equation 1 if the probability of the solution is higher than a random value within (0,1).
    b. Compare the solution values of starting a population and new population, then store the best one.
8. If the counter associated with a source exceeds the limit value, replace it with a new and randomly produced solution, then compute the solution value.
9. Repeat until the maximum iteration number is reached.

As is seen in Figure 5, it is possible to turn the original key and the plain text to a ciphered text by using the substitution method. The cipher of decoded text is obtained by reaching optimal keys produced for the Equation 1 based on the fitness values besides being produced of the stochastic keys to solve the cipher text using Artificial Bee Colony Algorithm based on character frequency method. Since the problem analyzed in this study is a discrete optimization problem, the same characters should not be repeated in the solution representation while producing new solutions and initializing the population [26].

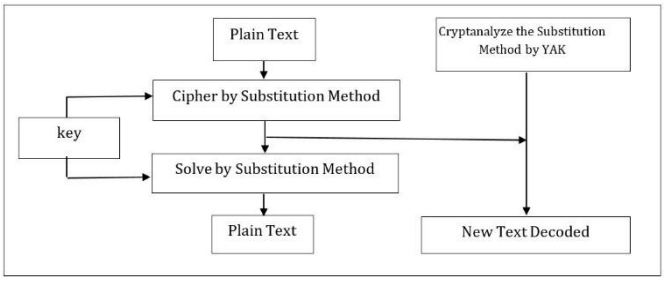

Fig. 5 Cryptanalysis steps of substitution cipher with ABC

The different mutations operators can be used to produce new populations in the stages of the bee in employed and the scout bee [27]. The key and the plain text decoded are received after the comparison between unigram, bigram or trigram in the ciphertext, and the standard language character probabilities used with possibilities.

The effectiveness of this attack is based on bigram and trigram frequencies. Equation 1 helps to solve the fitness value of possible keys in this research [15,23,5]

On the other hand, the plain text decoded is obtained by the key as a result of the comparison between the characters of the standard language used and the characters of encrypted text. Since there is not a frequency value for the numbers and symbols as a result of the literature scan, wrong results may be received if this method to estimate the key if the key contains character except than 26 alphabets.

VI. EXPERIMENTAL STUDY AND FINDINGS

In this study, the algorithm is tested on texts with different lengths such as 1000, 2000, 3000. Populations with 20, 50 and 100 sources are used for each length. Furthermore, the results in Table 1, 2, and 3 are obtained when the maximum iteration value is arranged as 1000.

Table 1. Text length have 1000 characters, and number of ıteratıon ıs 1000

Number of Populations	Exact character number	
	For the key	*For the plain text*
20	13	311
50	19	681
100	14	379

The results for different numbers of populations and trial values are given in Table 1 when the length of the encrypted text is 1000 and maximum iteration number is 1000. The best

result value is found when the number of populations is 50. As it is, 19 characters are found as exact from 26 key lengths.

Table 2. Text length 2000 characters and ıteratıon number 1000

Number of Populations	Exact character number	
	For the key	*For the plain text*
20	14	1123
50	14	1158
100	15	1114

Table 2 gives the results for different numbers of populations when the encoded text length is 2000 and maximum iteration number is 1000. The best result value is obtained when the number of populations is 100. 15 characters of the key with 26 lengths are found as exact.

Table 2. Text length 3000 characters and ıteratıon number 1000

Number of Populations	Exact character number	
	For the key	*For the plain text*
20	15	1934
50	16	1926
100	18	2122

Table 3 shows the results for different numbers of populations when the encoded text length is 3000 and maximum iteration number is 1000. The best result value is found when the number of populations is 100. 18 characters of the key with 26 lengths are considered as exact.

The similarity ratio between the text obtained by deciphering using the obtained key and the plain text (original text) is also supplied. Even though it is parallel with the exact character number in optimal key, there are changes in some of the keys. For example, according to Table 1, the exact character number of the text is 681 when the right character is 19 for the key. According to Table 2, the exact character number for the text is found as 1123 and 1158 if the right character number for two different keys is 14. And about Table 3, when the right character number for the possible key is 16, the exact character number in text after deciphering is found as 1926; the right character number for another key in the same table is 15, the exact character number in text after deciphering is found as 1934. Sometimes these changes may be proportional to the number of correct characters in the key, and sometimes they are based on the right character found, not in the number of specific characters found in possible keys.

VII. CONCLUSION

ABC algorithm which is a swarm intelligence algorithm is firstly used to get the keys of texts decrypted by the substitution method based on character frequency method. From the results of ABC algorithm, it can be said the algorithm produces good results for cryptanalysis, and it can be utilized as an alternative method as well. The purpose of next studies is to use ABC algorithm with modern encryption algorithms and compare it with different classical encryption algorithms.

ACKNOWLEDGMENT

This study is supported by Erciyes University, Scientific Research Projects Unit under contract number FDK-2016-7085.

REFERENCES

[1] B. Scheneier, Applied Cryptography Second Edition: protocols, algorithms, and source code in C: John Wiley and Sons, 1996.

[2] W. Stallings and M. P. Tahiliani, *Cryptography and network security: principles and practice* vol. 6: Pearson London, 2014.

[3] Z. Obaid, A. Sabonchi, and B. Akay, "KLASİK KRİPTOLOJİ YÖNTEMLERİNİN KARŞILAŞTIRILMASI," Engineering Sciences, vol. 11, pp. 100-108, 2016.

[4] M. Dorigo, "Optimization, learning and natural algorithms," *Ph. D. Thesis, Politecnico di Milano, Italy,* 1992.

[5] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on, 1995, pp. 39-43.

[6] D. Pham, "Karaboga. D, Ghanbarzadeh A, Koc E, Otri S, Rahim S and Zaidi M.,"The Bees Algorithm", Technical Note," *Manufacturing Engineering Centre, Cardiff University, UK,* 2005.

[7] X.-S. Yang, "Engineering optimizations via nature-inspired virtual bee algorithms," *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach,* pp. 317-323, 2005.

[8] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Applied soft computing,* vol. 8, pp. 687-697, 2008.

[9] R. A. Matthews, "The use of genetic algorithms in cryptanalysis," Cryptologia, vol. 17, pp. 187-201, 1993.

[10] G. S. Wulandari, W. Rismawan, and S. Saadah, "Differential evolution for the cryptanalysis of transposition cipher," in *Information and Communication Technology (ICoICT), 2015 3rd International Conference on*, 2015, pp. 45-48.

[11] M. F. Uddin and A. M. Youssef, "An artificial life technique for the cryptanalysis of simple substitution ciphers," in *Electrical and Computer Engineering, 2006. CCECE'06. Canadian Conference on*, 2006, pp. 1582-1585.

[12] S. M. Hameed and D. N. Hmood, "Particles swarm optimization for the cryptanalysis of transposition cipher," *Journal of Al-Nahrain University,* vol. 13, pp. 211-215, 2010.

[13] P. Garg, "GENETIC ALGORITHMS, TABU SEARCH AND SIMULATED ANNEALING: A COMPARISON BETWEEN THREE APPROACHES FOR THE CRYPTANALYSIS OF TRANSPOSITION CIPHER," *Journal of Theoretical & Applied Information Technology,* vol. 5, 2009.

[14] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: artificial bee colony (ABC) algorithm and applications," *Artificial Intelligence Review,* vol. 42, pp. 21-57, 2014.

[15] S. Omran, A. Al-Khalid, and D. Al-Saady, "Using Genetic Algorithm to break a mono-alphabetic substitution cipher," in *Open Systems (ICOS), 2010 IEEE Conference on*, 2010, pp. 63-67.

[16] D. R. Stinson, *Cryptography: theory and practice*: CRC press, 2005.

[17] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," Applied soft computing, vol. 8, pp. 687-697, 2008.

[18] D. Karaboga and B. Akay, "Artificial bee colony (ABC) algorithm on training artificial neural networks," in Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th, 2007, pp. 1-4.

[19] B. Akay and D. Karaboga, "Wavelet packets optimization using artificial bee colony algorithm," in Evolutionary Computation (CEC), 2011 IEEE Congress on, 2011, pp. 89-94.

[20] B. Akay and I. Kirmizi, "Structural optimization of wavelet packets using swarm algorithms," in Evolutionary Computation (CEC), 2012 IEEE Congress on, 2012, pp. 1-5.

[21] E. Hancer, B. Xue, M. Zhang, D. Karaboga, and B. Akay, "A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information," in Evolutionary Computation (CEC), 2015 IEEE Congress on, 2015, pp. 2420-2427..

[22] F. Piper, *Cryptography*: Wiley Online Library, 2002.

[23] A. Clark, "Modern optimisation algorithms for cryptanalysis," in Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on, 1994, pp. 258-262.

[24] F.-T. Lin and C.-Y. Kao, "A genetic algorithm for ciphertext-only attack in cryptanalysis," in *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, 1995, pp. 650-654.

[25] M. F. Uddin and A. M. Youssef, "Cryptanalysis of simple substitution ciphers using particle swarm optimization," in *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, 2006, pp. 677-680.

[26] B. Akay, E. Aydogan, and L. Karacan, "2-opt based artificial bee colony algorithm for solving traveling salesman problem," in 2nd World Conference on Information Technology (WCIT-2011), 2012, pp. 666-672.

[27] B. Akay and X. Yao, "Recent advances in evolutionary algorithms for job shop scheduling," in Automated Scheduling and Planning, ed: Springer, 2013, pp. 191-224