Text Cryptography based on Arabic Words Characters Number

Omar Fitian Rashid*[1], Mohammed Jasim Mohammed[2,3], Mustafa Sabah Mahmood[1]

[1] Department of Computer Technology Engineering, Al-Hikma University College, Baghdad, Iraq

[2] Department of Medical Instrumentation Engineering Techniques, Al-Hikma University College, Baghdad, Iraq

[3] Department of Mathematics, College of Science, University of Anbar, Ramadi, Iraq

*Corresponding author: omaralrawi08@yahoo.com

Abstract

Cryptography is a method used to mask text based on any encryption method, and the authorized user only can decrypt and read this message. An intruder tried to attack in many manners to access the communication channel, like impersonating, non-repudiation, denial of services, modification of data, threatening confidentiality and breaking availability of services. The high electronic communications between people need to ensure that transactions remain confidential. Cryptography methods give the best solution to this problem. This paper proposed a new cryptography method based on Arabic words; this method is done based on two steps. Where the first step is binary encoding generation used to encode messages (characters, numbers, and special characters) to binary numbers. Three encoding ideas are proposed for this reason. At the same time, the second step is used Arabic language words to represent the obtained binary sequences to achieve the final secret text. The obtained results exhibited that sending a text containing secret messages in a different language (Arabic) cannot be suspicious when read.

***Keywords**:* Cryptography, Text cryptography, Arabic characters, Encryption, Decryption.

1. Introduction

With the coronavirus pandemic, internet user numbers are about 4.93 billion in 2021 [1]. Approximately all the communications between companies, universities, and people were being transformed online instead of the normal one. Therefore, a secure communication channel between sender and receiver is very important to protect sensitive data transfer through the Internet [2]. To keep data from intruders and attackers, these data must be transferred in a way that cannot for the intruder to access, read, or understand it. This done by using data cryptography or hiding techniques. Data hiding can used to protect information from the unauthorized user and provided a high data security [3]. Where the main data hiding goals are security, authenticity, and data integrity. Data hiding can classify into cryptography and steganography. Figure 1 shows the classification of protect data.

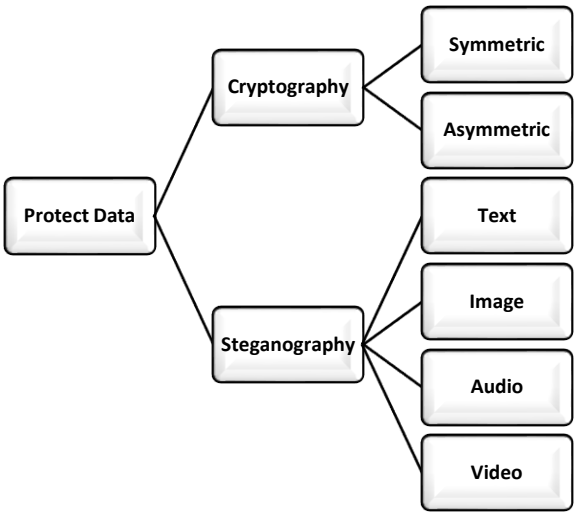

Figure -1 Protect data classification

Cryptography is the method that utilized to convert the message into ciphertext that is unintelligible to unauthorized user [4], this method is done based on some mathematical procedures, or encrypting principles [5]. Cryptography can classify either stream cipher or block cipher based on the way to processed data. Stream cipher is processed data using bit by bit and used a random key to obtained the ciphertext. While, block cipher processes data using unit of bits, then grouped it to blocks. On another hand, Steganography applied to secure communication based on information hiding within cover such as video, audio, image, and text [6]. Where text is the most used to cover communication, because is great used compared to other media. About 63% of users are used Internet only to send email and message text, 20% of them are used social media sites, and 50% used internet to making voice and video calls [7]. Arabic Text Steganography is a modern steganography method, the evaluation of Arabic Text Steganography Methods is listed below [8]:

1. A dot method can boost the invisibility and can be classified as less robust because the hidden information can be lost if changed the font format. The maximum embedded bits numbers per position is just two.

2. Diacritics method is enhanced the robustness of the stego text, but secret text is not encrypted and embedded it in sequence locations. This method has a low invisibility and both cover text and stego text are not identical. The maximum embedded bits numbers per position is ranges from one to four.

3. Kashida method is a wide used method which can resist versus copy-paste operation, on another hand, the drawbacks of this method are the capacity and robustness.

4. Unicode method: this method has a high invisibility. This method can be enhanced by a copy-paste operation, font format, and OCR.

5. Sharp edges method is obtained a high invisibility, capacity, and robustness. With the adding of non-sequence embedding.

6. Integrated method: this method is done by combine more than one method to get better performance and overcome the previous methods' limitations.

Arabic language is unique because this language has (28) characters as shown the Table 1.

Table 1- Arabic characters

No.	Latter	Letter pronunciation	No.	Latter	Letter pronunciation
1	أ	Alef	15	ض	Daa
2	ب	Baa	16	ط	Taa
3	ت	Taa	17	ظ	Dhaa
4	ث	Thaa	18	ع	Ain
5	ج	Jeam	19	غ	Ghain
6	ح	Haa	20	ف	Faa
7	خ	Khaa	21	ق	Qaaf
8	د	Dal	22	ك	Kaaf
9	ذ	Dhal	23	ل	Lam
10	ر	Raa	24	م	Meam
11	ز	Zaa	25	ن	Noon
12	س	Sean	26	ه	Haa
13	ش	Shean	27	و	Waw
14	ص	Sad	28	ي	Yaa

Several researches were proposed in this fields, a DNA symmetric cryptography method was suggested by Vikram et al. [9] which applied to enhance information security, where their obtained results proved that this method has a high security level. A new cryptography method was proposed by Rashid [10] by using DNA Cryptography, RNA, and Amino Acid. The evaluation was done based on six text files have various sizes. Sajisha and Mathew [11] proposed an information security method based on advanced encryption standard (AES) algorithm and DNA steganography. This technique has multilayer security levels. Rismayani and Susanto [12] proposed a new method to secure files transmission by using AES and DES methods.

2. Materials and Methods

The proposed text cryptography method is done based on Arabic language characters, where the proposed method consists of two steps:

The first step: binary encoding generation that used to encode message (characters, numbers, and special characters) to binary numbers. Three encoding idea is proposed for this reason. First encoding is done by calculating all possible message values, where characters values are equal to 52 values, while possible number values is equal to 10 value, and finally special character values consists of 17 values. The total values are equal to 79 values; therefore, seven binary numbers can handle and represent all possible values, a random binary number with seven digits is generating for each possible value. Table 2 show the first binary encoding method that used to encode all possible message values to binary numbers with length of seven digits

Table 2- First binary encoding method

No.	Binary Numbers	Character	No.	Binary Numbers	Character	No.	Binary Numbers	Character
1	0111000	A	28	1100001	b	55	1010000	2
2	1110100	B	29	1010110	c	56	0101101	3
3	1011111	C	30	0001111	d	57	1110101	4
4	1111001	D	31	0101101	e	58	0111000	5
5	0101011	E	32	1010101	f	59	0001000	6
6	0110000	F	33	0001100	g	60	1010011	7
7	0000010	G	34	1110111	h	61	0011111	8
8	1000010	H	35	1010011	i	62	1010000	9
9	1011100	I	36	0110010	j	63	1101000	.
10	0011111	J	37	1111101	k	64	0010000	(
11	1000100	K	38	1100101	l	65	0000001	)
12	1001111	L	39	1000100	m	66	0111011	?
13	0100110	M	40	1000011	n	67	0100000	!
14	1000001	N	41	1110110	o	68	1111000	+
15	0010010	O	42	0110011	p	69	0000000	-
16	1001101	P	43	0100110	q	70	1000000	/
17	1101100	Q	44	0111010	r	71	1001011	*
18	0010100	R	45	1101100	s	72	0011000	=
19	1001110	S	46	0101001	t	73	1010000	@
20	0100101	T	47	1000001	u	74	1100100	#
21	1111001	U	48	0100001	v	75	1010000	$
22	0101100	V	49	0010110	w	76	1110110	%
23	1011001	W	50	0001000	x	77	1001111	^
24	0010100	X	51	1010011	y	78	0010011	&
25	0110100	Y	52	0101010	z	79	1111111	Space
26	1110111	Z	53	0001100	0			
27	0101011	a	54	0010001	1			

Second encoding idea is utilized by reducing binary numbers from seven digits to five digits, this led to reduce the secret text size. This done by divided message values into three groups (sets), the first group consist of upper-case characters, while the second group includes the lower-case characters, and the final group consists of both digits and special characters values. In this encoding idea, five binary digits are randomly generated to encode three possible values, each value is belonged to one group. Example: the binary numbers 10001 is represented: character “A” from first group, character “a” from second group, and “0” from third group. In order to decide which character is represented by this binary number, i.e. which group is chosen, each group has its unique binary number that represent this group and this number is put before characters representation to know the specific character. The binary number that represented the first group is 00001, while the binary number that represented the second

group is 00010, and the binary number that represented the third group is 00011. Table 3 show the second binary encoding method that used to encode all possible message values to binary numbers with length of five digits.

Table 3- Second binary encoding method

No.	Binary Numbers	1st Group	2nd Group	3rd Group	No.	Binary Numbers	1st Group	2nd Group	3rd Group
1	10001	A	a	0	16	01001	P	p	+
2	01010	B	b	1	17	10100	Q	q	-
3	10111	C	c	2	18	01111	R	r	/
4	11001	D	d	3	19	00100	S	s	*
5	00101	E	e	4	20	11100	T	t	=
6	10010	F	f	5	21	10101	U	u	@
7	11000	G	g	6	22	01110	V	v	#
8	10011	H	h	7	23	01000	W	w	$
9	00110	I	i	8	24	10000	X	x	%
10	01011	J	j	9	25	10110	Y	y	^
11	10100	K	k	.	26	01101	Z	z	&
12	11010	L	l	(	27	00001	Set 1	-	-
13	01100	M	m	)	28	00010	-	Set 2	-
14	00111	N	n	?	29	00011	-	-	Set 3
15	11011	O	o	!	30	11111	Space	Space	Space

Third encoding idea is utilized by reducing possible binary numbers values from 30 values to 20 values, which led to reduce comparison time. This done by divided message values into six groups, where the upper-case characters are divided into first and second group, while the lower-case characters are divided into third and fourth group, and finally, the both digits and special characters values are divided into fifth and sixth group. The encoding is utilized by using five binary digits are randomly generated to encode six possible values, each value is belonged to one group. Example: the binary numbers 10010 is represented: character “A” from first group, character “N” from second group, character “a” from third group, character “n” from fourth group, number “0” from fifth group, and special character “?” from the last group. Same as second encoding method, each group has its unique binary number that represent this group and this number is put before characters representation to know the specific character. The binary number that represented the first group, second group, third group, fourth group, fifth group, and sixth group are equal to 00001, 00010, 00011, 00100, 00101, and 00110 respectively. Table 4 show the third binary encoding method that used to encode all possible message values to binary numbers with length of five digits.

Table 4- Third binary encoding method

No.	Binary Numbers	1st Group	2nd Group	3rd Group	4th Group	5th Group	6th Groups
1	10010	A	N	a	n	0	?
2	10011	B	O	b	o	1	!
3	10100	C	P	c	p	2	+
4	11010	D	Q	d	q	3	-
5	11011	E	R	e	r	4	/
6	01001	F	S	f	s	5	*
7	10100	G	T	g	t	6	=
8	01111	H	U	h	u	7	@
9	10110	I	V	i	v	8	#
10	11100	J	W	j	w	9	$
11	10101	K	X	k	x	.	%
12	01110	L	Y	l	y	(	^
13	01000	M	Z	m	z	)	&
14	00001	Set 1	-	-	-	-	-
15	00010	-	Set 2	-	-	-	-
16	00011	-	-	Set 3	-	-	-
17	00100	-	-	-	Set 4	-	-
18	00101	-	-	-	-	Set 5	-
19	00110	-	-	-	-	-	Set 6
20	11111	Space	Space	Space	Space	Space	Space

While the second step in the proposed method is using Arabic language characters to represent the achieved binary sequences from step one to get the final secret text. This done by writing Arabic paragraph based on binary numbers, this mean represented each binary number with Arabic word, if the number is “1” then used a word contains odd characters number, and if the number is “0” then used an Arabic word contains even characters number. Where the used paragraph is containing understandable sentences and the reader cannot suspicious there is any secret message inside it.

3. Results and Discussions

To understand the proposed cryptography method steps to encode message, the following example will explain and highlight these steps and their results. Suppose the plaintext is "This is my new code P6074"

The binary numbers sequence that achieved from applying first encoding method for previous plaintext is equal to “01001011110111101001111011001111111101001111011001111111110001 00101001111111111000011010110100101101111111101011011101100001111010111111111 11001101000100000011001010011111010 1”, Table 5 shows the first encoding method in details.

Table 5- First encoding method in details

T	**h**	**i**	**s**	**space**	**i**	**s**	**space**	**m**
0100101	1110111	1010011	1101100	1111111	1010011	1101100	1111111	1000100
y	**space**	**n**	**e**	**w**	**space**	**c**	**o**	**d**
1010011	1111111	1000011	0101101	0010110	1111111	1010110	1110110	0001111
e	**space**	**P**	**6**	**0**	**7**	**4**		
0101101	1111111	1001101	0001000	0001100	1010011	1110101		

While, when applying second encoding method, the obtained binary numbers sequence for the same plaintext is equal to "00001111000001010011001100010011111001100010011111011001 01101111100111001010100011111101111101111001001011111100001010010001111000100 11001100101", Table 6 shows the second encoding method in details.

Table 6- Second encoding method in details

1st Group	**T**	**2nd group**	**h**	**i**	**s**	**space**	**i**	**s**
00001	11100	00010	10011	00110	00100	11111	00110	00100
space	**m**	**y**	**space**	**n**	**e**	**w**	**space**	**c**
11111	01100	10110	11111	00111	00101	01000	11111	10111
o	**d**	**e**	**space**	**1st Group**	**P**	**3rd Group**	**6**	**0**
11011	11001	00101	11111	00001	01001	00011	11000	10001
7	**4**							
10011	00101							

Finally, when using the third proposing encoding method, the obtained binary numbers sequence for the same plaintext is equal to "000101010000011011111011000100010011111100011 1 0110 00100010011111100011010010010001110111110010010010000111101100100111 0 01111100011101000010010011000110001111011111110001010100001011010010010011 1111011", Table 7 display the third encoding method in details.

Table 7- Third encoding method in details

2nd set	**T**	**3rd set**	**h**	**i**	**4th set**	**s**	**space**	**3rd set**
00010	10100	00011	01111	10110	00100	01001	11111	00011
i	**4th set**	**s**	**space**	**3rd set**	**m**	**4th set**	**y**	**space**
10110	00100	01001	11111	00011	01001	00100	01110	11111
4th set	**n**	**3rd set**	**e**	**4th set**	**w**	**space**	**3rd set**	**c**

00100	10010	00011	11011	00100	11100	11111	00011	10100
4th set	**o**	**3rd set**	**d**	**e**	**space**	**2nd set**	**P**	**5th set**
00100	10011	00011	00011	11011	11111	00010	10100	00101
6	**0**	**7**	**4**					
10100	10010	01111	11011					

After that, Arabic language words used to represent the achieved binary sequences to achieved the final secret text. If the number is "1" then used a word contains odd characters number, and if the number is "0" then used an Arabic word contains even characters number. The secret text in this example using the first encoding method is equal to:

ان عملية نقل البيانات المستخدمة في يومنا هذا كانت ولازالت من اهم الامور في حياتنا اليومية، ومن اول عمليات نقل المعلومات هو البريد والصحف و الجرائد. وفي وقتنا الحالي تمكن الانسان من نقل الرسائل والصور والفيديوات لاسلكياً وايضا ظهرت اجهزة التليفون واصبحت العملية اسرع واسهل وهذه العملية في تطور دائم. حيث يحتاج الانسان ان يتواصل مع الاخرين لغايات متعددة اهمها الاطمئنان على احوال بعضهم البعض ونقل المعلومات والصور فيما بينهم ونقل المعلومات التي تحدث في العالم الخارجي لذلك الاتصالات تعد من اهم الاختراعات التي توصل اليها الانسان في عصرنا هذا لذلك تم تطوير واستمرار عملية التطوير بسبب اهمية الموضوع.

4. Conclusion

A new cryptography method is using Arabic words; where this method is utilized based on two steps. The first step is generated binary encoding methods to encode message into binary numbers. Three encoding methods is proposed for this reason. First encoding method is done by calculating all possible message values, while the second encoding method is utilized by reducing binary numbers from seven digits to five digits, and the third encoding method is utilized by reducing possible binary numbers values from 30 values to 20 values. Then used Arabic language words to represent the obtained binary sequences to get the final secret text. The achieved results showed that the sending text that contain secret message in a different language (Arabic language) and cannot suspicious when read it.

Refrences

[1]. C. Mcclain, E. Vogels, A. Perrin, S. Sechopoulos and L. Rainie, "Key Internet Statistics to Know in 2021 (Including Mobile)", Washington, DC, USA, Dec. 2021.

[2]. J. Boehm, J. Kaplan, M. Sorel, N. Sportsman and T. Steen, "Cybersecurity tactics for the Coronavirus Pandemic", Dec. 2021.

[3]. O A. C. Oluwakemi, A. S. Kayode, and O. J. Ayotunde, "Efficient data hiding system using cryptography and steganography", International Journal of Applied information Systems (IJAIS), 2012.

[4]. M. K. Sharma, A. Upadhyaya, and S. Agarwal, "Adaptive steganographic algorithm using cryptographic encryption RSA algorithms", Journal of Engineering, Computers & Applied Sciences, 2013.

[5]. V. Tyagi, "Data hiding in image using least significant bit with cryptography", International Journal of Advanced Research in Computer Science and Software Engineering, 2012.

[6]. A. H. Mohsin, A. A. Zaidan, B. B. Zaidan, S. A. B. Ariffin, O. S. Albahri, A. S. Albahri, M. A. Alsalem, K. I. Mohammed, and M. Hashim, "Real-time medical systems based on human biometric steganography: A systematic review", Journal of Medical Systems, vol. 42, no. 12, pp. 1-20, Oct. 2018.

[7]. C. Mcclain, E. Vogels, A. Perrin, S. Sechopoulos and L. Rainie, “The Internet and the Pandemic”, Washington, DC, USA, Dec. 2021.

[8]. R. Thabit, N. I. Udzir, S. M. Yasin, A. Asmawi, N. A. Roslan, and R. Din, “A Comparative Analysis of Arabic Text Steganography”, Applied Sciences, vol. 11, no. 15, pp. 6851.

[9]. A. Vikram, S. Kalaivani, and G. Gopinath, “A Novel Encryption Algorithm based on DNA Cryptography”, 2019 International Conference on Communication and Electronics Systems (ICCES), 2019.

[10]. O. F. Rashid, “Text Encryption Based on DNA Cryptography, RNA, and Amino Acid”, E- Proceedings of the 5th International Multi-Conference on Artificial Intelligence Technology (MCAIT 2021) Artificial Intelligence in the 4th Industrial Revolution, 2021.

[11]. K. S. Sajisha, and S. Mathew, “An encryption based on DNA cryptography and steganography”, 2017 International conference of Electronics”, Communication and Aerospace Technology (ICECA), 2017.

[12]. Rismayani and C. Susanto, "Using AES and DES Cryptography for System Development File Submission Security Mobile-Based", *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1-7, 2020.