ISSN No:-2456-2165

Performance Evaluation of Non-Latin Characters Based (Arabic) Symmetric Encryption Algorithm

[1]Adebayo Ademola Riliwan; [2]Oluokun Samuel Gbenga, [3]Ogunode Rotimi Samuel, [4]Seyi Osunade

Abstract:- Cryptography is the art of keeping information secure by transforming it into form that unintended recipients cannot understand. Encryption is a form of data security in which information is converted to cipher text to ensure confidentiality, authentication and integrity of user data. Issue relating to the choice of encryption algorithm among the Arabian IT community is another challenge. This research evaluated the performance of three symmetric non-Latin character-based encryption algorithms for Arabic text: First-Order-Equation-of-Three-Variables, Modulo-37-cipher, and Atbash-Substitution. The algorithms were assessed based on the following performance metrics namely: encryption time, decryption time, execution time, throughput, memory usage, and avalanche effect. The results revealed that the Atbash-Substitution algorithm had the highest throughput at 35.63 B/ms and the lowest memory usage at 197.90 MB. It also exhibited a weak avalanche effect. In contrast, the Modulo-37-cipher algorithm showed poor performance in terms of throughput. Therefore, the Atbash-Substitution algorithm demonstrated superior performance for Arabic text, being the fastest and most memory-efficient among the evaluated algorithms.

Keyword:- *Cipher, non-latin, throughput, encryption, algorithms, modulo-37-cipher.*

I. INTRODUCTION

As the Internet and other forms of electronic communication become more prevalent in the non-Latin speaking countries/ groups, electronic security is becoming increasingly important (Al-Omari,2018). Thus, it becomes essential to protect e-mail messages, credit card information, and corporate data, by means of encryption that conform with the Latin alphabets and numbers.Various cryptographic algorithms have been proposed and implemented to achieve the security requirements such as Authentication, Confidentiality, and Integrity of the Arabic language. There are basically two types of encryption techniques; symmetric and asymmetric. Symmetric cryptography is the one which uses a single key for encryption and decryption. The Symmetric encryption techniques provide cost-effective and efficient methods of securing data without compromising security however; sharing the secret key is a problem. On the other hand, asymmetric techniques solve the problem of distributing the key for encryption, but; they are slow compared to symmetric encryption and consume more computer resources.

In networking, security depends solely on cryptography (meaning "secret writing"), which is the science and art of transforming messages to make them safe and immune to attack (Hamouda, 2020).Cryptography is said to be an Arab-born science that is improved by western scientists (Al-Omari, 2018). A great historian in cryptology, David Khan stated that "cryptology was born in Arabic world"and the fact was later confirmed in some Arabic treatise found in Istanbul's Suleymanye library in addition to the work of other scholars who wrote about cryptography and cryptanalysis in the Arab world (Al-Omari, 2018).

Data that can be perused and perceived with no difficulty or special measures is called plain text. The method of changing plaintext and making it meaningless is called encryption. The encryption process produces unreadable and meaningless output called cipher-text. The process of retrieving plain text from cipher text is known as decryption (Kuppuswamy and Alqahtani, 2014).

Therefore, this paper aimed at comparing the three aforementioned algorithms which were evaluated based on the encryption and decryption time, throughput, memory used and avalanche effects.

II. ALGORITHMS UNDER EVALUATION

Modulo 37 Cipher Encryption Algorithm: Kuppuswamy and Alqahtani, (2014) developed an encryption algorithm cipher that uses modulo 37 in its mathematical model. The algorithm provided an effective use of key algorithm on Arabic characters.

The algorithm for key generation is as follows: Step 1 – any natural number say n was selected Step 2 – the inverse of n was calculated with modulo 37 (key 1) say k Step 3 – select another random negative number say n1 Step 4 – the inverse of the n1 was calculated with modulo 37(key 2) say k1
The algorithm process for the encryption phase is as follows: Step 1 – user ID in the synthetic table was assigned a value Step 2 – the synthetic number was multiplied by any random selected natural number Step 3 – calculation was done using modulo 37 Step 4 – random negative number was selected and multiplied with the result in step 3 Step 5 – calculation was done using modulo 37 {CT =(PT* n*n1)mod 37}

The algorithm process for the decryption phase is as follows;
Step 1 – encrypted text was multiplied by key1 and key2
Step 2 – calculation was done using modulo 37
Step 3 – Result of the calculation produce 'R'

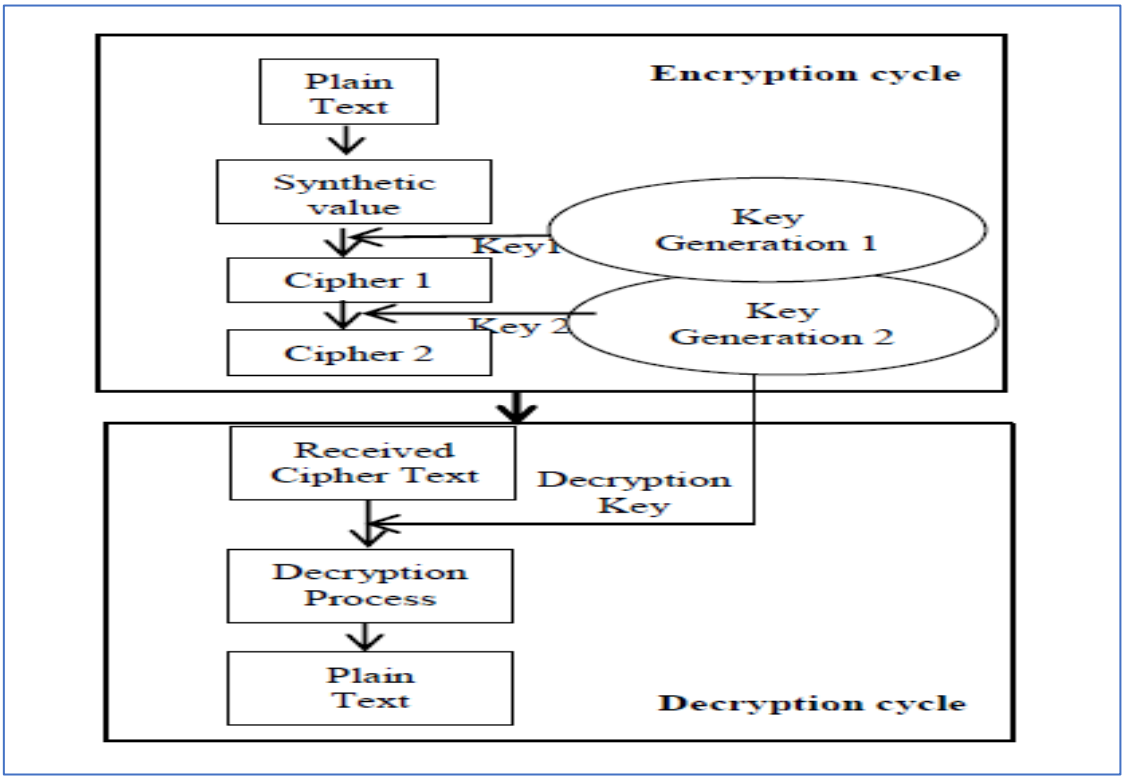

Fig. 1: Modulo 37 Cipher Encryption and Decryption Architecture (Kuppuswamy and Alqahtani, 2014)

- **First Order Equation of Three variables Encryption Algorithm:** Shaban (2017) in his work proposed the algorithm where two random numbers are used to generate key. The algorithm technique makes use of different equation in the key generation mechanism and encryption mechanism.

The algorithm process for the key generation phase is as follows:
Step 1 – a first order equation of three variables is chosen say 2x+y-3z, where x represents the character of the message and y, z are two random numbers
Step 2 – a random number say key1 is chosen and it is assigned to y
Step 3 – another random number say key2 is chosen and it is assigned to z
Step 4 – the chosen equation and the two random numbers selected are only known by the sender and receiver

The algorithm process for the encryption phase is as follows:
Step 1 – the equation values is computed for each Arabic character in the plain text
Step 2 – the character obtained is converted to the binary format

Step 3 – the XOR of key1 and the odd position characters is computed; same as the XOR of key2 and the character in the even position computed
Step 5 – All messages are later converted to the binary format transmitted to the receiver over the internet

ISSN No:-2456-2165

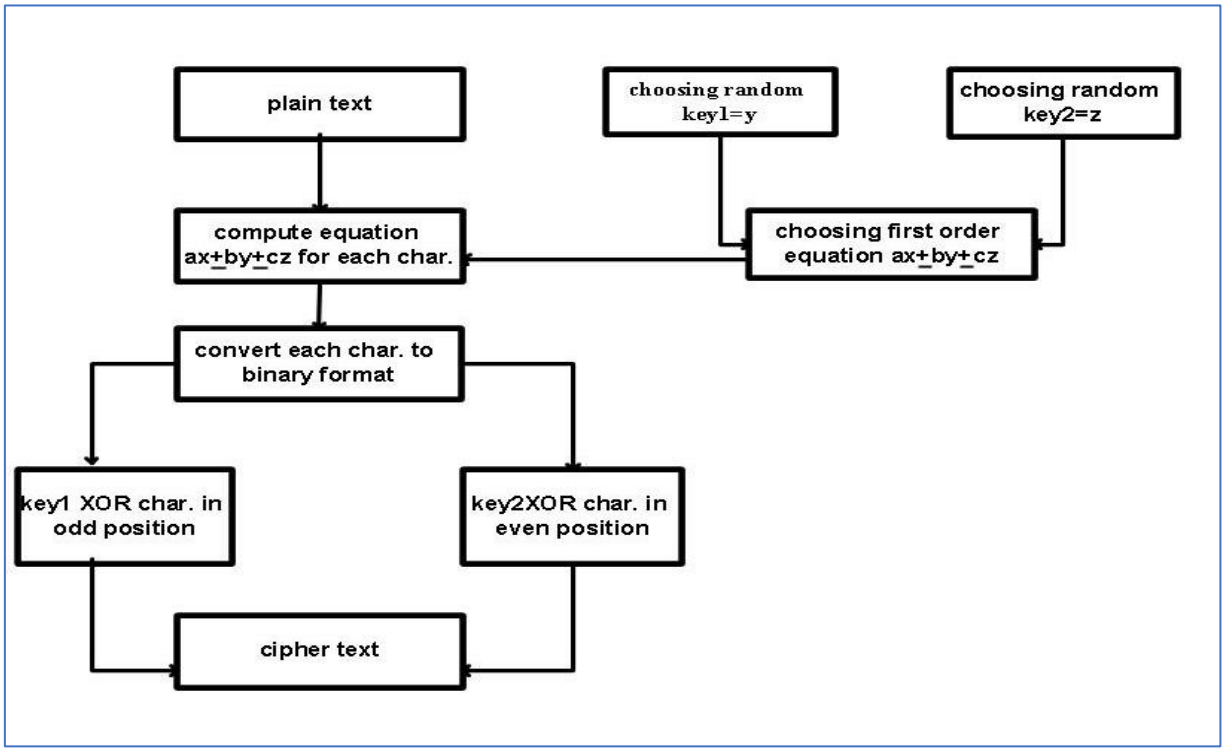

Fig. 2: First Order Equation of Three variables Encryption Phase (Basim, 2017)

The algorithm process for the decryption phase is as follows:

Step 1 – the same key1 and key2 are used to decrypt the message

Step 2 – XOR between key1 and the character in the odd position and key2 with the character in the even position is computed

Step 3 – the binary format is converted to the numeric value for each message characters

Step 4 – the inverse of the first order equation is calculated to find the value of each character in the message such as x=(3z-y)/2; where the value of the x represent the character that we want to decrypt and y represent key1 and z represent key2

Step 5 – lastly, we converted the binary format of the message to character format

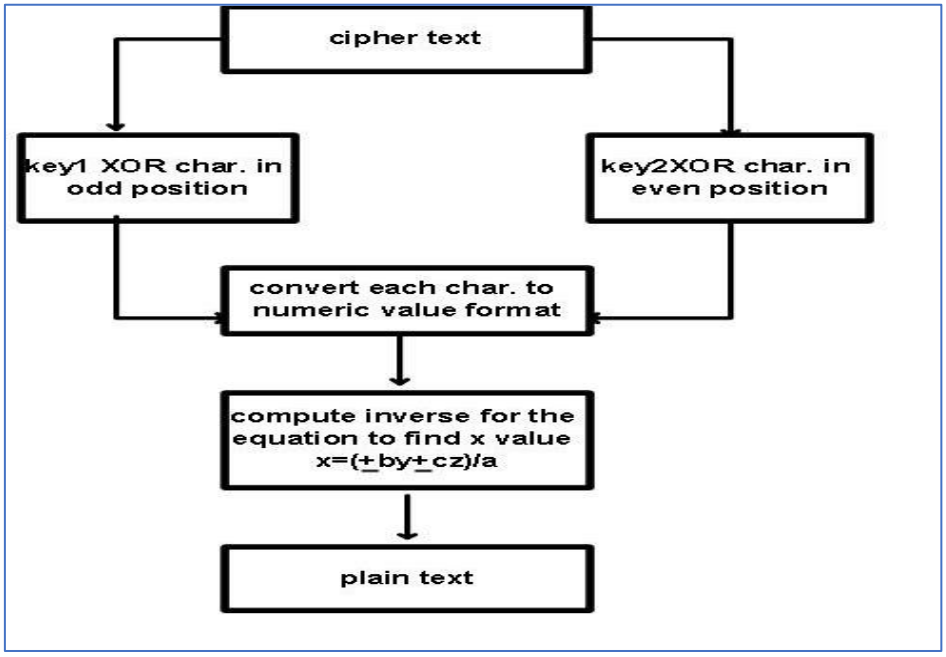

Fig. 3: First Order Equation of Three variables Decryption Phase (Basim, 2016)

III. IMPLEMENTATION

This study implemented the algorithms in phpprograming language. The authors built on existing php extension for the implementation of the algorithms under evaluation. The text files used were between 94B to 980B, which consisting of only Arabic text as input for encryption. The encrypted output of each text is copied, which in turn is input for decryption. For the sake of comparison the study used the same input textof the same size for all algorithms throughout the experiment. Also the study used the same system for all implementations and analysis work, so that memory and processor conditions remain same for all algorithms for comparison. All block cipher algorithms are set in a same mode which is MCRYPT_MODE_CBC. CBC stands for Cipher Block Chaining.It works by XORing each plaintext block of text against the plaintext block preceding it, then encrypting it, this makes duplicate plaintext blocks different when encrypted.

IV. SYSTEM PARAMETERS

The experiment are conducted using Intel(R) Core(TM) i7-3687U CPU @ 2.10GHz, 2601 Mhz, 2 Core(s), 4 Logical Processor(s), 12GB RAM. The simulation program is compiled using the default settings in .NET 2013 visual studio for C# windows applications. The experiments was performed couple of times to ensure that the results are consistent and are valid to compare the different algorithms. The value of the information is proportional to the risk of information which means when high valued information there will be a great need for the information to be protected and secured.

Encryption algorithms consume a significant amount of computing resources such as CPU time, memory and computation time (Mandal, 2012). Determining the appropriate algorithm suited for a particular data type, and scenario there is need to answer the following questions and possibly problems;

- To determine time taken by an algorithm to processes a file during encryption and decryption.
- To estimate the amount of CPU time consumed in the process.
- To Calculate the Memory Utilized.

This study is based on the design of a benchmark application using php program for testing the selected encryption algorithms (Modulo 37 Cipher Encryption Algorithm, First Order Equation of Three variables Encryption Algorithm and Atbash Substitution) in order to evaluate the resource utilization and time consumed by each algorithm with different Arabic input data size to determine the most appropriate algorithm for Arabic text. Various encryption algorithms have been developed in time past for various purposes. They all have their strengths for encryption and decryption but they also have their weaknesses in times of attack (i.e unauthorized person(s) attempting to decipher encrypted data in a forceful manner without the appropriate decryption keys). Encryption algorithms are said to have improved from time to time but not all of them are reliable for every kind of data to be ciphered.

That is why this study embarks on testing three recent encryption algorithms for Arabic text to determine which one is best suitable and at what time. Due to the fact that three of these encryption algorithms will be implemented to achieve the aim of the paper, it is important to note that the work does not primarily focus on the development of the encryption algorithms, but on the resource utilization of this algorithm to determine if they are suitable for the task at which they are implemented on.

Each of the three algorithms will be implemented solely for the testing of their encryption and decryption strengths, weaknesses as well as the resources they make use of during the process.

Each of the encryption techniques has its own strong and weak points. In order to apply a suitable cryptography algorithm to an application, we should have knowledge regarding performance, strength and weakness of the algorithms. Therefore, these algorithms must be analysed based on several features. In this paper, analysis is done with following metrics under which the cryptosystems can be compared are described below:

- Encryption time (E_t):This is the time taken to change original text (plain text) to scramble text (cipher text).It is important that encryption processes of algorithms are fast enough to meet real time requirements. Therefore, the speed at which encryption takes place in each of the case study algorithms will be determined. Encryption time is dependent on the key size, plaintext block size and mode of encryption process. For this research, encryption time was measured in milliseconds.
- Decryption time (D_t): this is the time expended in the process of retrieving original text (plain text) from scrambled text (ciphertext). it is desired for the decryption time to be lesser so as to improve system responsiveness. Decryption time impacts performance of system. Forthis research, decryption time was measured in milliseconds.
- Throughput (T_p): The throughput of the encryption scheme is the size ofthe plain text in bytes divided by the encryption time in millisecond. It was defined as the measure of the data-transfer rate through a networking scheme. Throughput is considered an indication of the overall performance of the system. Throughput can be calculated using equation 1.

$$T_p = \frac{P_t}{E_t} \quad (1)$$

Where; $PT_s: Data\ Size;$

$T_p: Throughput;\ E_t: Encryption\ Time$

- **Execution Time (EX_t):** This refers to the total time expended in converting the original text (plain text) to scrambled text (cipher text) i.e. encryption time and the time to retrieve original text (plain text) from scrambled text (cipher text) i.e.decryption time. Execution time can be calculated using equation 2.

$$EX_t = E_t + D_t \quad (2)$$

Where; EX_t: $Execution\ Time$;
E_t: $Encryption\ Time$;
D_t: $Decryption\ Time$

- **Memory Used:** The memory requirement is dependent on the number of operations to be performed by the algorithm, key size used, initialization vectors and the type of operations. It is desirable that the memory required should be as small as possible because it greatly has cost implication on the system to be used.
- **Avalanche Effect (A_E):** This is the diffusion reflects of the cryptographic strength of an algorithm. It shows the significant of changes made to an input (plain text) to produce an output (cipher text). This wasachieved by measuring the level of dissimilarity (Hamming distance). We calculate the Hamming distance as a sum of bit by bitxor considering the ascii value, as it becomes easy to implement programmatically. Avalanche effect reflects performance of cryptographic algorithm.

$$A_E = (\text{hamming distance} \div \text{file size}) \quad (3)$$

V. SIMULATION PROCEDURE MODEL DESIGN

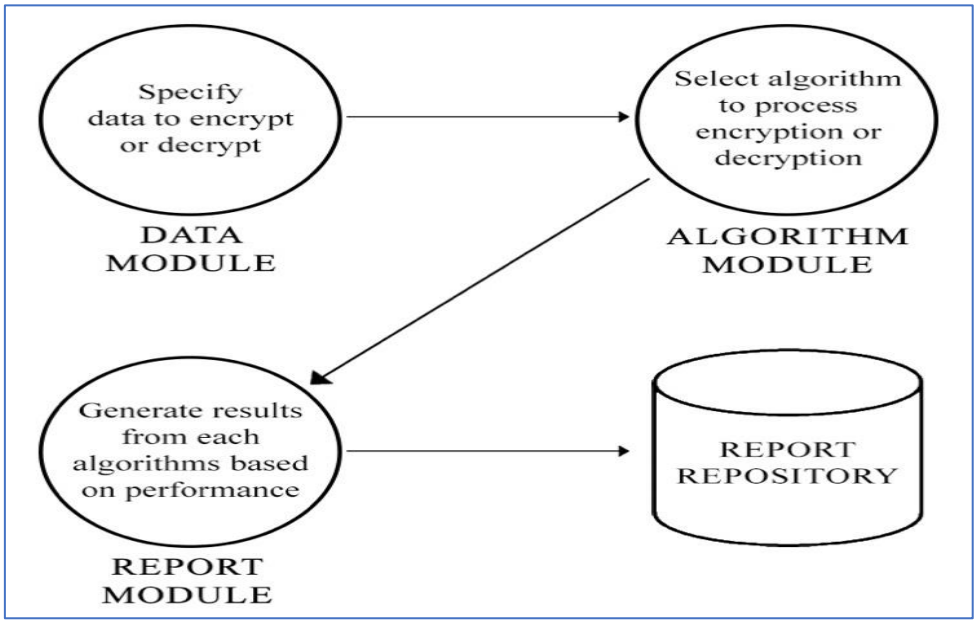

Fig. 4: Simulation Procedure Model Diagram

For clarification, simplicity and a better understanding of the application,figure 4 shows the modules involved the evaluation processes.

- The first module (data module), specified the data to be encrypted or decrypted using different algorithms. The data accepted by this module is the Arabic text.
- The second module (algorithm module), involved the selection of any of the implemented algorithm to be used for encryption and decryption of the data specified in the data module.
- The third module (report module), displayed the resulting report of encryption and decryption process based on the selected performance metrics. The report will include result of the encryption, the data size, the encryption parameters, and memory usage. All these parameters will be displayed for each algorithm that was executed independently of the others (i.e. every algorithm will be executed one after the other so that each can have full system resources at its disposal). The result of this module is the key to determining the most secure and appropriate encryption algorithm for each data type, size and scenario.

The model would be able to perform encryption and decryption of non-Latin character based (Arabic). It is assumed the computer system involved in the model runs Microsoft Windows operating system because this is a platform where the developed system will perform optimally.The evaluation system was developed using PHP because it is a general-purpose language endowed with numerous implementations. The choice of the language is made due to its excellent flexibility, combinability and the availability of specialist writing the language.

VI. RESULT AND DISCUSSION

From the data recorded for ten (10) different Arabic texts that are of different size, the evaluation of the following metrics was deduced and represented with tables and charts.

- **Encryption Time:** As shown in table 1 and represented in figure 5, On average, Atbash substitution has 12.66ms, Modulo-37-cipher 19.07ms while first order equation of three variables 21.79ms. Also, the encryption time increases with increase in the input data size. Encryption time is directly proportional to the input data size.

ISSN No:-2456-2165

Table 1: Encryption Time Comparison for Selected Arabic Encryption Algorithms

ENRYPTION TIME (ms)											
	980 bytes	**742 bytes**	**726 bytes**	**456 bytes**	**430 bytes**	**426 bytes**	**316 bytes**	**260 bytes**	**116 bytes**	**94 bytes**	**Avearge Time**
First Order Equation of Three Variables	47.29	38.39	34.29	21.42	20.31	20.02	14.74	11.78	5.22	4.43	21.79
Modulo 37 Cipher	40.48	31.94	29.74	19.3	18.20	18.03	13.19	11.00	4.91	3.87	19.07
Atbash Substitution	27.75	21.01	18.56	12.62	12.18	12.07	8.92	7.39	3.29	2.77	12.66

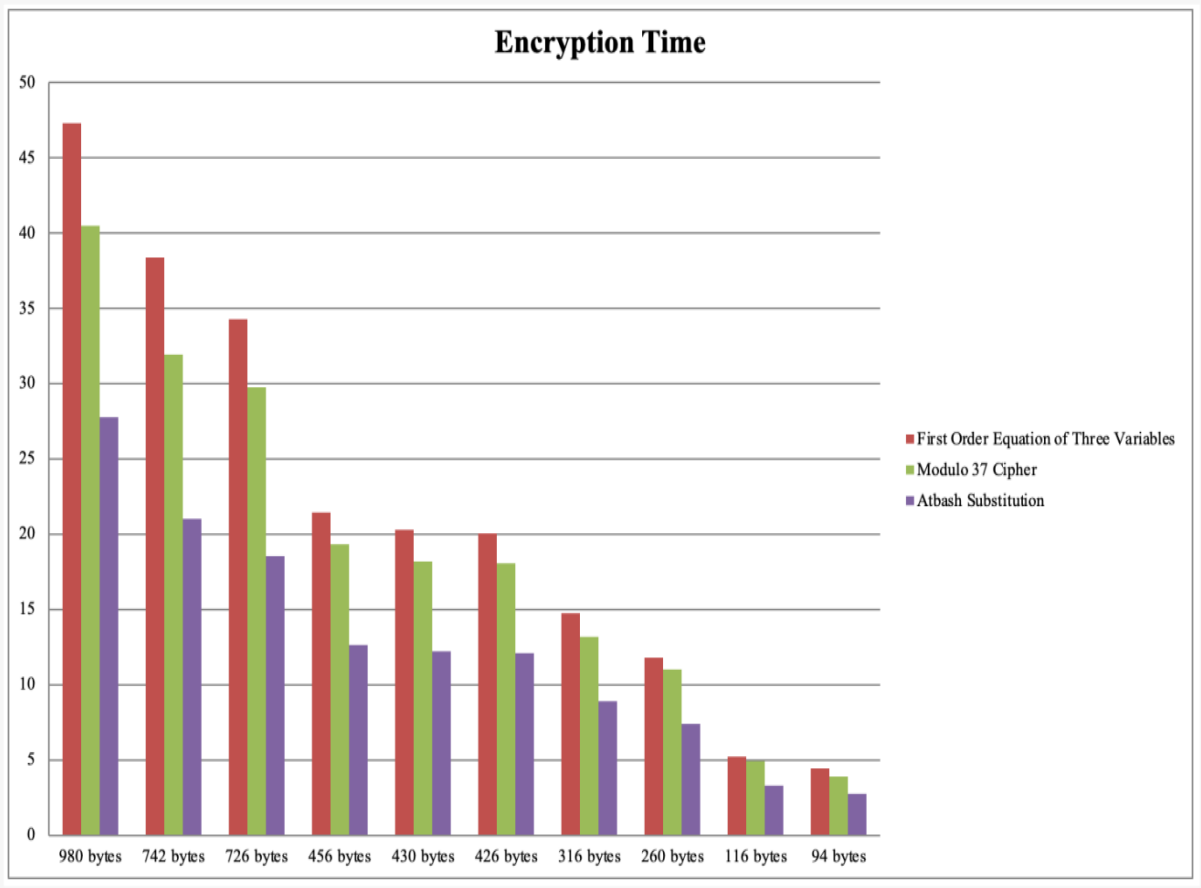

Fig. 5: Encryption Time Comparison for Selected Arabic Encryption Algorithms

- **Decryption Time:** As shown in table 2 and illustrated in figure 6,modulo-37-cipher has 12.79ms, first order equation of three variables 13.01ms and Atbash substitution has 12.66ms. Also, the decryption time increases with increase in the input data size. Encryption time is directly proportional to the input data size.

Table 2: Decryption Time Comparison for Selected Arabic Encryption Algorithms

DECRYPTION TIME (ms)											
	980 bytes	**742 bytes**	**726 bytes**	**456 bytes**	**430 bytes**	**426 bytes**	**316 bytes**	**260 bytes**	**116 bytes**	**94 bytes**	**Average**
First Order Equation of Three Variables	28.00	21.82	20.59	13.49	12.29	12.14	9.01	7.52	3.37	2.78	13.10
Modulo 37 Cipher	27.69	20.97	20.51	12.93	12.15	12.04	8.86	7.01	3.31	2.46	12.79
Atbash Substitution	46.32	35.04	32.16	20.81	20.26	20.12	14.97	12.15	5.56	4.29	21.17

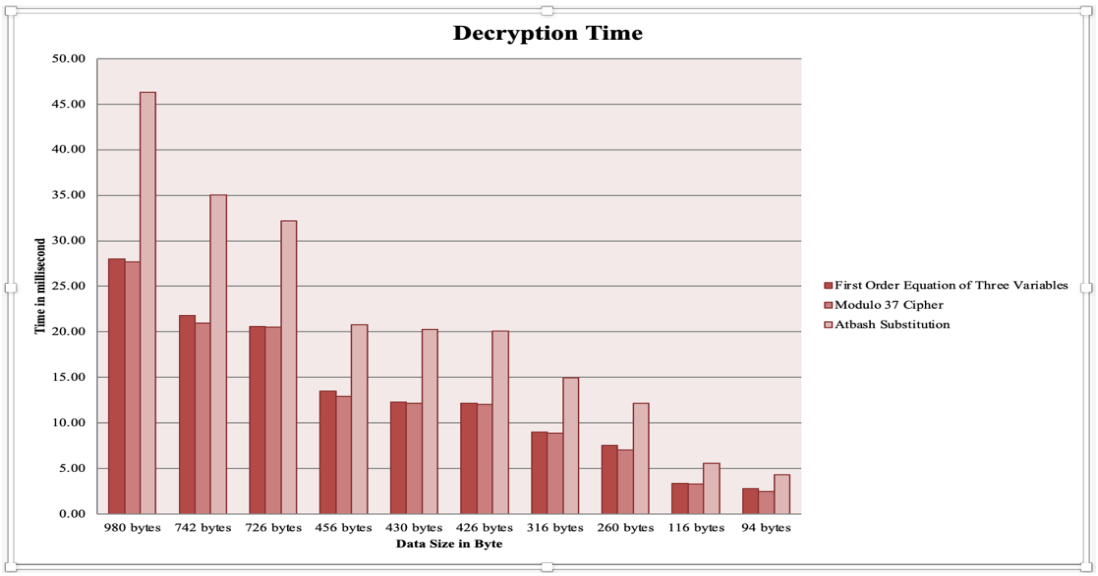

Fig. 6: Decryption Time Comparison for Selected Arabic Encryption Algorithms

- **Execution Time (EX_t):** As shown in table 3 and illustrated in figure 7, modulo-37-cipher has 31.86ms, Atbash substitution 33.82ms and first order equation of three variables 34.89ms.

Table 3: Execution Time Comparison for Selected Arabic Encryption Algorithms

EXECUTION TIME (ms)											
	980 bytes	**742 bytes**	**726 bytes**	**456 bytes**	**430 bytes**	**426 bytes**	**316 bytes**	**260 bytes**	**116 bytes**	**94 bytes**	**Average**
First Order Equation of Three Variables	75.29	60.21	54.88	34.91	32.60	32.17	23.75	19.30	8.59	7.21	34.89
Modulo 37 Cipher	68.17	52.91	50.25	32.23	30.35	30.07	22.05	18.01	8.22	6.33	31.86
Atbash Substitution	74.07	56.05	50.72	33.43	32.44	32.18	23.89	19.54	8.85	7.06	33.82

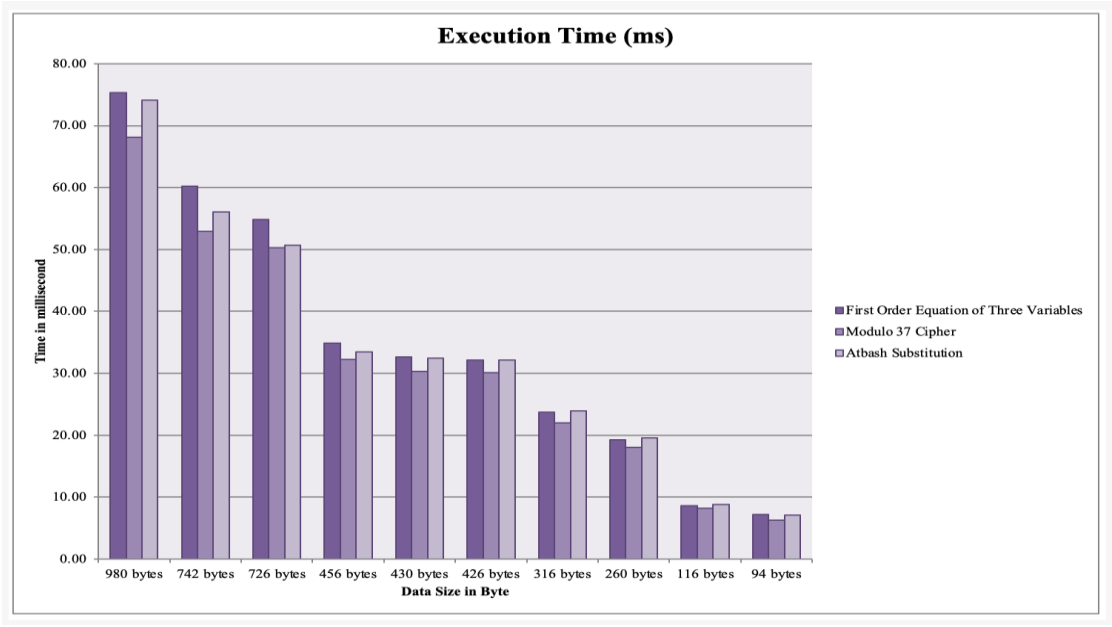

Fig. 7: Execution Time Comparison for Selected Arabic Encryption Algorithms

- **Throughput (T_p):**Table 4 and figure 8 shows the throughput for the selected encryption algorithm for non-Latin character (Arabic), on average, Atbash substitution has 35.63b/ms, modulo-37-cipher 23.83b/ms, and first order equation of three variables 21.19b/ms.

ISSN No:-2456-2165

Table 4: Throughput Comparison for Selected Arabic Encryption Algorithms

THROUGHPUT (B/ms)											
	980 bytes	742 bytes	726 bytes	456 bytes	430 bytes	426 bytes	316 bytes	260 bytes	116 bytes	94 bytes	Average
First Order Equation of Three Variables	20.72	19.32	21.17	21.29	21.17	21.29	21.44	22.07	22.22	21.22	21.19
Modulo 37 Cipher	24.21	23.23	24.41	23.62	23.63	23.63	23.96	23.63	23.63	24.29	23.82
Atbash Substitution	35.31	35.31	39.11	36.13	35.30	35.29	35.43	35.18	35.26	33.94	35.63

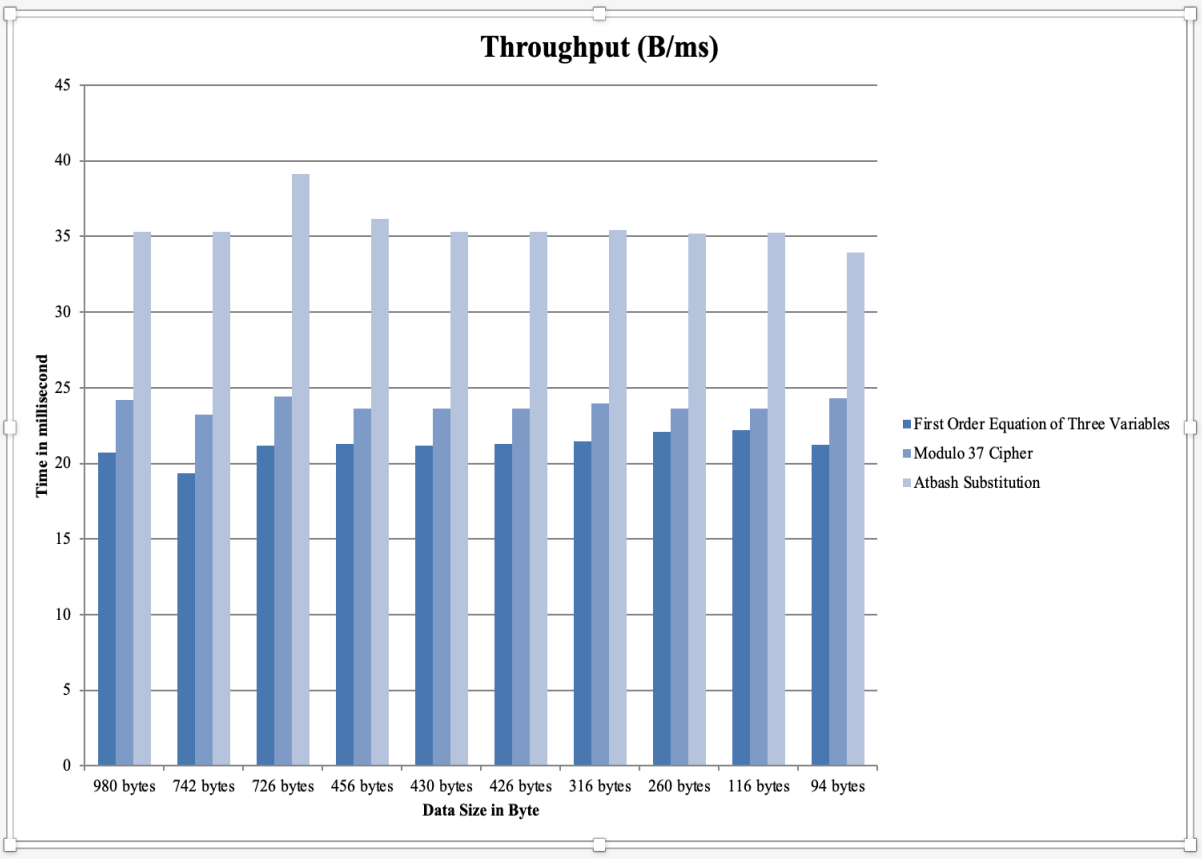

Fig. 8: Throughput Comparison for Selected Arabic Encryption Algorithms

- **Memory Used:** As shown in Table 5 and represented with the chart in the figure 9, on average, it was deduced that Atbash Substitution used 197.04 MB, modulo-37-cipher 196.31 MB, and first order equation of three variables 197.04 MB.

Table 5: Memory Usage Comparison for Selected Arabic Encryption Algorithms

MEMORY USAGE (MB)											
	980 bytes	**742 bytes**	**726 bytes**	**456 bytes**	**430 bytes**	**426 bytes**	**316 bytes**	**260 bytes**	**116 bytes**	**94 bytes**	**Average Memory**
First Order Equation of Three Variables	196.65	198.65	197.75	197.30	196.20	196.10	196.95	197.35	196.90	196.55	197.04
Modulo 37 Cipher	195.75	196.65	196.85	196.25	195.15	196.60	196.10	197.20	195.45	197.10	196.31
Atbash Substitution	198.90	197.70	198.35	196.70	198.90	198.20	197.10	197.25	197.55	198.35	197.90

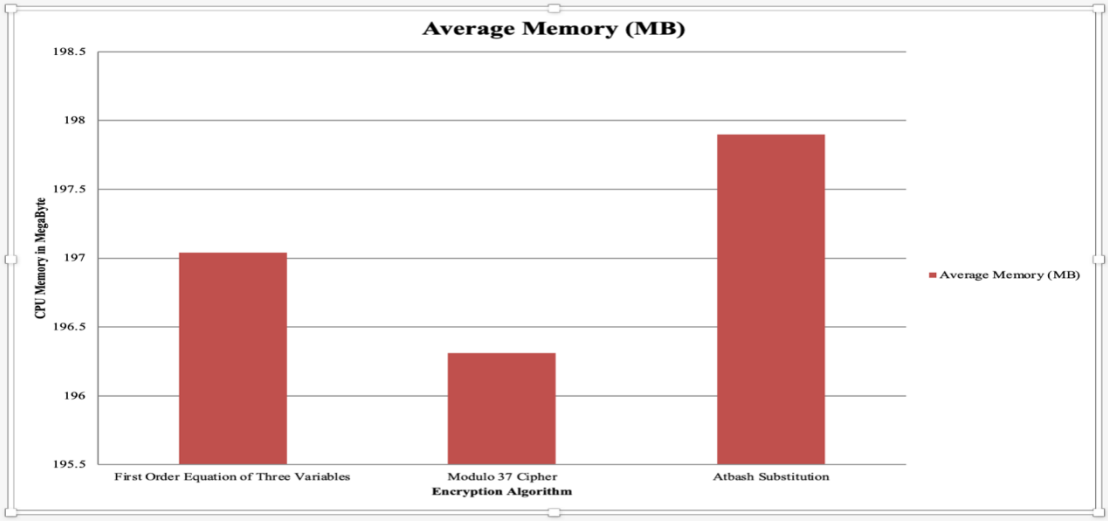

Fig. 9: Memory Usage Comparison for Selected Arabic Encryption Algorithms

- **Avalanche Effect**: This explains the diffusion reflects of the cryptographic strength of each encryption algorithm. It provides the significant changes made to an input (plain text) to produce an output (cipher text). As shown in the decryption GUI of each of the encryption algorithm, it is deduced that Atbash substitution has weak avalanche effect in that the number of bits of the input plain text produce the same number of bits of the output cipher text while others have strong avalanche effect.

VII. CONCLUSION

Cryptography is the art of keeping information secure by transforming it into form that unintended recipients cannot understand. Encryption is a form of data security in which information is converted to cipher text to ensure confidentiality, authentication, integrity, availability and identification of user data. Issues relating to the choice of encryption algorithm take into consideration the block size, CPU resources usage and time for encryption/decryption process are another challenge in the IT community among the Arabian text users. Hence, this research was a performance evaluation of non-Latin character-based encryption algorithm.

The result obtained by the application is exported to Microsoft excel for further analysis. The result support that the throughput of any encryption algorithm is inversely proportional to the period of the encryption process. In terms of throughput, Atbash substitution is the fastest, followed by First Order Equation of Three Variables and Modulo 37 Cipher. The encryption time is low in Atbash, followed by Modulo 37 Cipher and First Order Equation of Three Variables. On memory usage, Atbash substitution required the least memory followed by Modulo 37 Cipher and First Order Equation of Three Variables. Lastly, the avalanche effect is strong in Modulo 37 Cipher and First Order Equation of Three Variables while weak in Atbash substitution.

It was concluded that Atbash substitution has a better performance for Arabic text than the other two algorithms compared since it is the fastest in terms of throughput and uses low memory space. First Order Equation of Three variables showed poor performance results compared to other algorithms since it requires more memory and has poor throughput.

REFERENCES

[1]. Ahmad, S., Alam, K. M. R., Rahman, H., and Tamura, S. (2015, January). A comparison between symmetric and asymmetric key encryption algorithm based decryption mixnets. *International Conference on Networking Systems and Security (NSysS)* (pp. 1-5). IEEE.

[2]. Al-Omari, A. H. (2018). ABJAD Arabic-Based Encryption. *International Journal of Advanced Computer Science and Applications*, 9(10).

[3]. Altamimi, A. S. H., &Kaittan, A. M. (2021). A Proposed Arabic Text Encryption Method Using Multiple Ciphers. Management.

[4]. Elminaam, D. S. A., Kader, H. M. A., &Hadhoud, M. M. (2008). Performance evaluation of symmetric encryption algorithms. *IJCSNS International Journal of Computer Science and Network Security*, 8(12), 280-286.

[5]. Hamouda, B. E. H. H. (2020). Comparative study of different cryptographic algorithms. *Journal of Information Security*, 11(3), 138-148.

[6]. Kuppuswamy, P., &Alqahtani, Y. (2014). New innovation of Arabic language encryption technique using new symmetric key algorithm. *International Journal of Advances in Engineering & Technology*, 7(1), 30.

[7]. Masram, R., Shahare, V., Abraham, J., &Moona, R. (2014). Analysis and comparison of symmetric key cryptographic algorithms based on various file features. *International Journal of Network Security & Its Applications*, 6(4), 43.

[8]. Shaban, S. A. (2017). A new algorithm for encrypting Arabic text using the mathematical equation. *Diyala journal of engineering sciences, 10(1)*, 21-30.

[9]. Stallings, W. (2006). Cryptography and network security principles and practices 4th edition.