DEVELOPMENT OF A NEW WAY TO ENCRYPT THE ARABIC LANGUAGE LETTERS USING THE SYMMETRIC ENCRYPTION SYSTEM [+]

Haifaa Abdul-Zahra Atee [*]

Abstract:

It is not secret that the algorithms and methods used to encrypt messages in Arabic language are few and old, there are no modern encryption algorithms to encrypt the Arabic letters along the lines as the encryption algorithms of the letters in English, though they are few and uncommon use. On this basis, has been developed a way to encrypt the Arabic character letters by using a symmetric encryption system, XOR logical function, and binary & decimal encoding schemes to convert letters to decimal ambiguous numbers does not indicate the content of the message. Been using an encryption key length (15) digits, five of them to encode the character that the letter start with it, and the remaining parts are used with the XOR function. Possible to include or not include the key within the message, according to what was agreed upon between the parties. Encrypt each two symbols together and thus eliminating the problem of the most frequent letters in the Arabic language and by which they can predict the original text of the message. The encrypted text figures twice the size of the original text.

استحداث طريقة جديدة لتشفير الرسائل باللغة العربية باستخدام نظام التشفير المتناظر

هيفاء عبدالزهره عاتي

المستخلص:

لا يخفى ان الخوارزميات والطرق المستخدمة لتشفير الرسائل باللغة العربية قليلة وقديمة ولا توجد خوارزميات تشفير حديثة لتشفير الرسائل باللغة العربية على غرار خوارزميات التشفير للحروف الانكليزية، وان وجدت فهي قليلة وغير شائعة الاستخدام . وعلى هذا الأساس فقد تم استحداث طريقة لتشفير الرسائل باللغة العربية باستخدام نظام التشفير المتناظر Symmetric system ودالة XOR المنطقية وأنظمة الترميز الثنائي والعشري لتحويل الحروف إلى أرقام مبهمة لا تدل على محتوى الرسالة . تم استخدام مفتاح تشفير طوله(١٥) رقم ،خمسة منها لترميز الحرف الذي تبدأ به الرسالة الباقية تستخدم مع دالة XOR .ممكن تضمين أو عدم تضمين المفتاح داخل الرسالة وحسب ما متفق عليه بين الطرفين. تم تشفير كل رمزين معا وبذلك تم القضاء على مشكلة الحروف الأكثر تكرارا في اللغة العربية والتي من خلالها يمكن التكهن بحروف النص الأصلي للرسالة . تم الحصول على نص مشفر على شكل أرقام حجمه ضعف حجم النص الأصلي.

Introduction:

Since times immemorial, security of data to maintain its confidentiality, proper access control integrity and availability has been a major issue in data communication[1].

[+] **Received on 25/10/2010 , Accepted on 24/2/2011 .**

[*] **Assistant Lecturer / Foundation of Technical Institutes**

Cryptography always played an important role in peoples lives. It is believed that basic cryptographic techniques were known and used by Egyptians around 4500 years ago. For many centuries cryptography was used mainly by states to achieve secrecy of communication.[2]

The earliest encryption of messages was found since as early as 3500 BC during the years of The Sumerians (the earliest known civilization in the world). Further occurrence of cryptography was then found in 1900 BC, where unusual hieroglyphic symbols (writing system used by the ancient Egyptian) were used instead of the common ones. This type of cryptography was not to provide confidentiality of the messages, but focuses more on the concept of providing mysterious information[3]. Encryption used in ancient civilizations to conceal information and correspondence, but the encryption science founder entrepreneur owes his birth and upbringing of scientist mathematicians and linguists, the Arabs during the golden age of Arab civilization and the most famous Faraaheedi and Canada. Scientist provided mathematical concepts and progress of the most important combinations and permutations.

Nowadays commonly use the term "encrypt" is indicative of a concealment of information, but the encryption word imported from European languages (Cipher) in turn came originally from the Arabic language.

It is well known that the internet these days are the largest center for the transfer of information (such as financial movement) in an encrypted form , to prevent the integrity and security of the futility of hackers, vandals, and thieves[4].

Julius Caesar developed cryptosystem known as code Caesar (Caesar Cipher)[5], which was coded text (Cipher text); to secure contacts and correspondence with the leaders of armies. Later emerged many of the machines that carry out the encryption .

Some of the methods used in encryption, are [1,6]:

- Caesar cipher.
- Monoalphabetic cipher
- Playfair cipher
- Hill cipher

Most of the methods used to encrypt the Arabic language has adopted the principles of substitution and transposition (rearrangement of letters according to certain algorithm). In both of these principles the try and error method used to decode the small messages, while the large messages using the table of repeated letters and count the number of repetition each letter in the message, in Arabic language the letters that most frequently recurs are أ , ل , م….etc[10].

In this paper we develop a method for encrypting Arabic message characters and convert the text to numbers, each two characters encoded together to eliminate the problem of characters that recur frequently in the texts in Arabic, which used to guess the hidden letters in ciphertext. These operations are simple and easily to implement. Without knowing the data, tables of decryption, and tables of encryption, it is difficult to do cryptanalysis. The encryption system proposed designed in C++ language and tested on some of short messages.

Encryption:

Encryption refers to the translation of data into an encoded format for the purpose of achieving data security. Reading an encrypted file requires access to a secret key, which enables the decryption of this file.[7]

The strength and effectiveness of the encryption depends on two main factors: the algorithm and key length (estimated to take action (bits)). Traditional encryption operations are transposition and substitution.[4,7]

On the other hand, the decryption is the process of converting data into its original form, using the appropriate key to decode the code.

There are two types of technology used in the encryption, these are[7]:

- Symmetric encryption.
- Asymmetric encryption

some of a symmetric encryption methods are:
Blowfish, Digital Encryption Standard (DES), Tiny Encryption Algorithm (TEA), Triple DES, and International Data Encryption[8].
Some of a Asymmetric encryption method are:
Pretty Good Privacy(PGP) and Reivest, Shamir & Aselman(RSA)[8].

Symmetric encryption:

Symmetric-key block ciphers are the most prominent and important elements in many cryptographic systems. Individually, they provide confidentiality [5].

In a symmetric encryption , the sender and receiver uses the same secret key to encrypt and decrypt the message. The parties agree at the outset on a password that will be used. That can contain the pass phrase, large and small letters and other symbols. Then, turning encryption software pass phrase to a binary number, and other symbols are added to increase the length. The resulting binary number key encrypt the message. After receiving the encrypted message, the receiver use same secret key to decoding the cipher text (cipher text or encrypted text), as translated software again secret key to form a dual key (binary key) who re-convert the cipher text to its original form the concept. The large gap in this type of encryption was lies in the exchange of the secret key, but nowadays this is not problem , the parties can exchange the keys by internet or any other device.

Asymmetric encryption (Public key):

In public-key encryption systems each entity (a) has a public key (e) and a corresponding private key (d)[5].

Asymmetric encryption came as a solution to the problem of safe distribution of keys in symmetric encryption, Instead of using a single key, asymmetric encryption uses two keys to connect them to a relationship. It is alleged the two keys public key (public key), and the private key (private key). The private key is known to one side only that is a sender, and is used to encrypt and decrypt the message. The public key used to encrypt messages, only the owner of the private key decode the message.

The first problem is to encrypt the data using the method of encryption key non-symmetrical complex process taxes the computer's processor takes a lot of time, so it is not for the exchange of messages is relatively long, and the solution to this problem lies in the use of the method the fastest in encryption, a method of encryption key is symmetric.

A second problem in this encryption system is that it is still capable of penetrating easily by using the attack allegedly attacked the man in the middle (Man in the middle attack).

There are three main objectives behind the use of cryptography are as follows[7]:

1 - confidentiality or privacy (Confidentiality):

To be maintaining the privacy of the e-mail, you should not be able to access them only by the parties concerned are allowed to do so. To maintain privacy, it is necessary to control the process of entering in internet, more common methods of control are: the use of passwords (passwords), and firewall (firewall), as well as clearance certificates (authorization certificates)

This service is used to save the content of information of all persons except that told them to see it.

2 - Integration of data (Integrity):

The processes of information transfer and storage Must be protected, so as to prevent any change to the content of deliberate or unintentional. The importance in maintaining the content is useful and reliable. Often, the human errors and intentional tampering operations are the cause of damage or distort the data. The result is that the data becomes useless, and unsafe for use. In order to avoid distortion or damage to data, must use techniques such as: electronic fingerprint of the message (message digest) and encryption (encryption), it is useful to also use the software anti-virus (anti-virus software) to protect storage devices from violations of viruses that cause damage or distort the data.

This service is used to save the information of the change (delete or add or modify) by persons unauthorized to do so.

3 - Proof of Identity (Authentication):

Must make sure the identity of the parties involved in the exchange of data, since both parties must know the identity of the other to avoid any form of deception (such as forgery and identity fraud). There are some solutions and procedures to verify the identity of the parties relating to such as: passwords (passwords), and digital signatures (digital signatures), and digital certificates (digital certificates) issued by a third party. It can also enhance security by relying on some perceived advantages such as: fingerprint (finger print), and audio, in addition to the picture.

This service is used to establish the identity of dealing with the data (authorized).

4. Non-repudiation :

Is a service which prevents an entity from denying previous commitments or actions. When disputes arise due to an entity denying that certain actions were taken, a means to resolve the situation is necessary. For example, one entity may authorize the purchase of property by another entity and later deny such authorization was granted. A procedure involving a trusted third party is needed to resolve the dispute[4].

The primary goal of encryption is to provide these services to the people to be maintaining the security of their information.

<u>Encryption & Decryption Algorithms:</u>

Encryption is a process of coding information which could either be a file or mail message into cipher text in a form unreadable without a decoding key in order to prevent anyone except the intended recipient from reading that data. Decryption is the reverse process of converting encoded data to its original un-encoded form, plaintext.

A key in cryptography is a long sequence of bits used by encryption / decryption algorithms[8]. For example, the following represents a hypothetical 40-bit key:

00001010 01101001 10011110 00011100 01010101

A given encryption algorithm takes the original message, and a key, and alters the original message mathematically based on the key's bits to create a new encrypted message. Likewise, a decryption algorithm takes an encrypted message and restores it to its original form using one or more keys.

There are two primary approaches to encryption: symmetric (private key) and Asymmetric(public-key). Symmetric encryption is the most common type of encryption and uses the same key for encoding and decoding data. This key is known as a session key. Asymmetric encryption uses two different keys, a public key and a private key. One key encodes the message and the other decodes it. The public key is widely distributed while the private key is secret.

There are three types of cryptographic algorithms:

Secret Key Cryptography (SKC): Uses a single key for both encryption and decryption .

- Public Key Cryptography (PKC): Uses one key for encryption and another for decryption .
- Hash Functions: Cryptographic hash functions play a fundamental role in modern cryptography [4],it is use a mathematical transformation to irreversibly "encrypt" information.

A) Secret key (symmetric) cryptography. SKC uses a single key for both Encryption & Decryption.


Plaintext ----------► Cipher text ----------►Plaintext

B) Public key (Asymmetric) cryptography. PKC uses to keys, one for encryption and the other for decryption.


Plaintext ----------► Cipher text ----------►Plaintext

C) Hash function (one way cryptography). Hash function have no key , since the plaintext is not recoverable from the cipher text.

Figure (1) illustrate the Three types of cryptography: secret-key, public key, and hash function

New Approach To Cryptograph The Arabic Letter Messages

This system consist of 6 parts:

P : Is a finite symbol represent the plaintext.
C : Is a finite symbol represent the cipher text.
K : Is a finite symbol represent the keys domain.
E : Is a finite symbol represent the encryption text.
D : Is a finite symbol represent the decryption text.
M : is a matrix to encrypt & decrypt the message.
For each $k \in K$ there is encryption transfer such that:

$e_k \in E$ (for encryption)

$d_k \in D$ (for decryption)

$d_k (e_k (X)) = X$ and $d_k : C \rightarrow P$ and $e_k : P \rightarrow C$ for each $X \in P$

$M = (P, 29) \in P$, 29 is a number of Arabic character codes.

New Encryption Method:

In this system each Arabic character corresponding to the number, the number of the first letter in the message represent first part of key, the length of key consist of 15 bits, 5bits represent the number code for the first letter in the plain text, others will be number within the range 2^{10} , (1024) ,now the first letter will take number 1 and so on , each character may take any number from 0 up to 28. table(1) illustrate the representation of Arabic character and corresponding numbers for it.

Table(1) Arabic character and corresponding numbers

number	Letter	number	Letter	number	Letter	number	Letter
24	م	16	ط	8	د	0	ء
25	ن	17	ظ	9	ذ	1	ا، أ،إ،آ ،ى
26	هـ ، ة	18	ع	10	ر	2	ب
27	و ، ؤ	19	غ	11	ز	3	ت
28	ي	20	ف	12	س	4	ث
		21	ق	13	ش	5	ج
		22	ك	14	ص	6	ح
		23	ل	15	ض	7	خ

To encrypt the plain text (طريق السلامة) ,first, the word Letters must be written without spaces then should be know the code number corresponding to the letter ط (first char in the message) this will be the first part of the key, this number is 16 as shown in Table(1). Then must assign numbers to other letters as follow (Table(2)) :

Table(2) cades of letters

28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك	ق	ف	غ	ع	ظ	ط
ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك	ق	ف	غ	ع	ظ
ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك	ق	ف	غ	ع
ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك	ق	ف	غ
غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك	ق	ف
ف	غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك	ق
ق	ف	غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل	ك
ك	ق	ف	غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م	ل
ل	ك	ق	ف	غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن	م
م	ل	ك	ق	ف	غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ	ن
ن	م	ل	ك	ق	ف	غ	ع	ظ	ط	ض	ص	ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	ء	ي	و	هـ

When apply the previous table on the plaintext (طريق السلامة) has been obtained for similar codes to different characters, as shown in table(3). The characters (ط ، ل ، ة) have the same

code (0) , so the characters (ق ، ل) have the code 2 ,this is a good indicator of the strength of key.

Table (3) code numbers of the message

ة	م	١	ل	س	ل	ا	ق	ي	ر	ط
0	28	6	0	19	2	10	2	10	22	0

Must convert this decimal number to binary :

0	28	6	19	2	10	22
00000	11100	00110	10011	00010	01010	10110

To increase the strength of the key, will use the XOR function, Key will be chosen at random, which is part of the key that will send the message recipient, this key is located within the boundaries of the function 2^{10} (1024), in our research we will choose the number 1012.

When convert 1012 to binary will get the number 1111110100, then we use this number to applied the XOR function with each 2 char in the plaintext as shown below:
Encoding for first set of 2 char (ط ر) is 308 as shown :

binary 1st 2char → 1011000000 ⊗ binary of 1012 → 1111110100 → binary result 0100110100 → decimal result 0308

Encoding for second and third set of 2 char (ي ق)(ا ل) is 958 as shown :

binary 2nd and 3rd 2char → 0001001010 ⊗ binary of 1012 → 1111110100 → binary result 1110111110 → decimal result 0958

Second and third set of 2 char has same code because the chars (ي ا) and (ق ل) have the same codes in the table (3).

binary 4th 2char → 0000010011 ⊗ binary of 1012 → 1111110100 → binary result 1111110111 → decimal result 1015

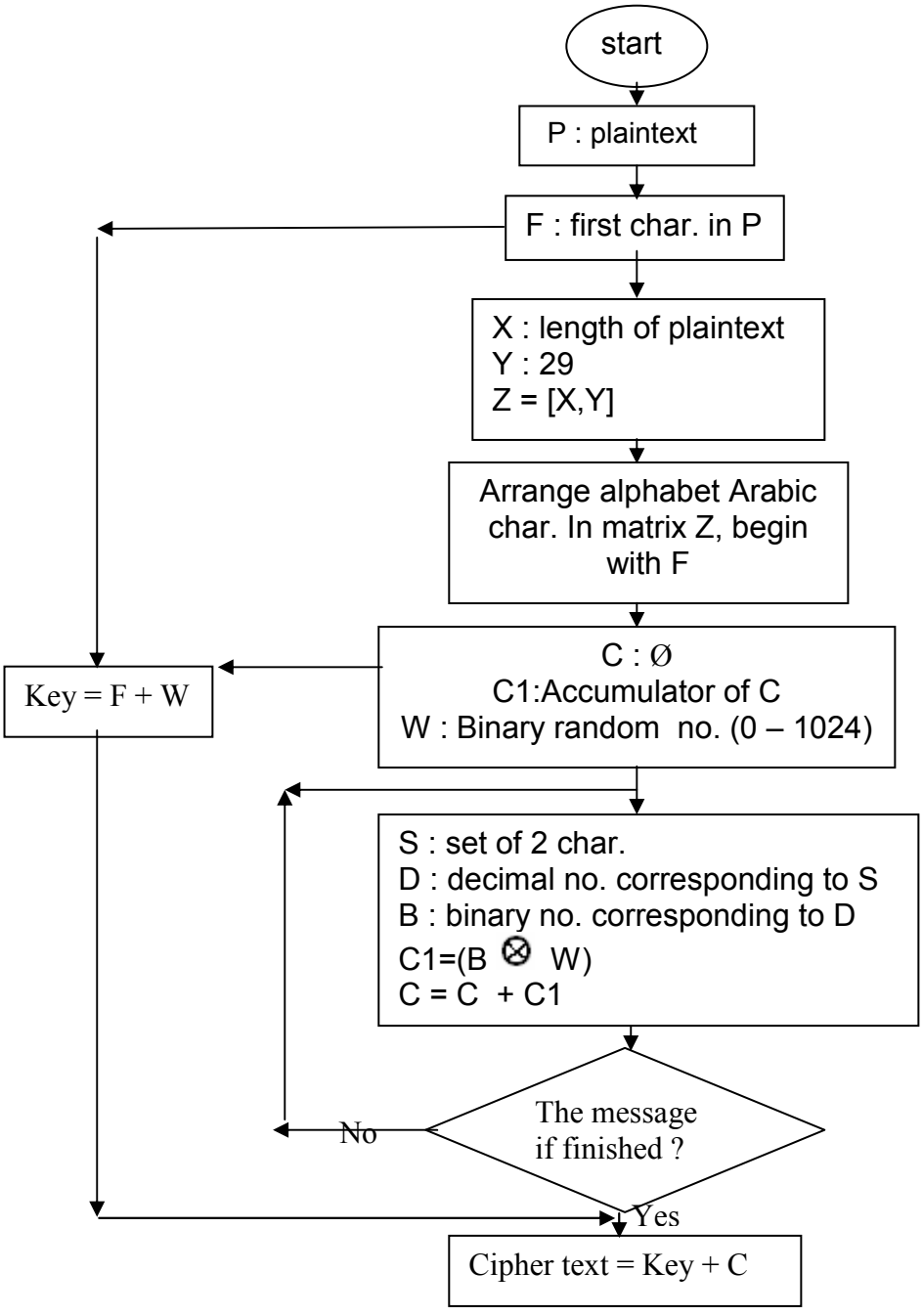
binary 5th 2char → 1110000110 ⊗ binary of 1012 → 1111110100 → 0001110011 → decimal result 0115

Last char (ة) will be as below:
00000 ⊗ 1111110100 → 1111110100 → 1028
Now obtained this series of numeric which represent the cipher text that will be send to the receiver of message .
Plain text = 10280115101509580958059 6
So, we can send the key with the cipher text if there is no way to send it alone.

Send message = key + plaintext
Send message = 6 + 2* P
Send message =16101210280115101509580308

Key Cipher text

Figure (2) illustrate the flowchart of encryption operation.
So, we can duplicate key's size or exchange specific positions in the message to increase the strength of the key.



Figure(2) : Flowchart of Encryption Operation

New Decryption method:

The cipher text will be send to receiver to decrypt it and return the original plain text, As explain in the next steps :

1- If the key send with the message, the receiver must be split to key's part and cipher text's part, as shown:

Key = 161012
Cipher text = 102801151015095809580596

2 - Must getting the first letter in the message and the number that used with the XOR function.
First letter in the message is 16 ⟶ ط , and 1012 is the number that used with the XOR function.

3 - Convert 1012 to binary number:
1012 ⟶ 1111110100

4 – Split the cipher text to 4 char as all:
102801151015095809580308

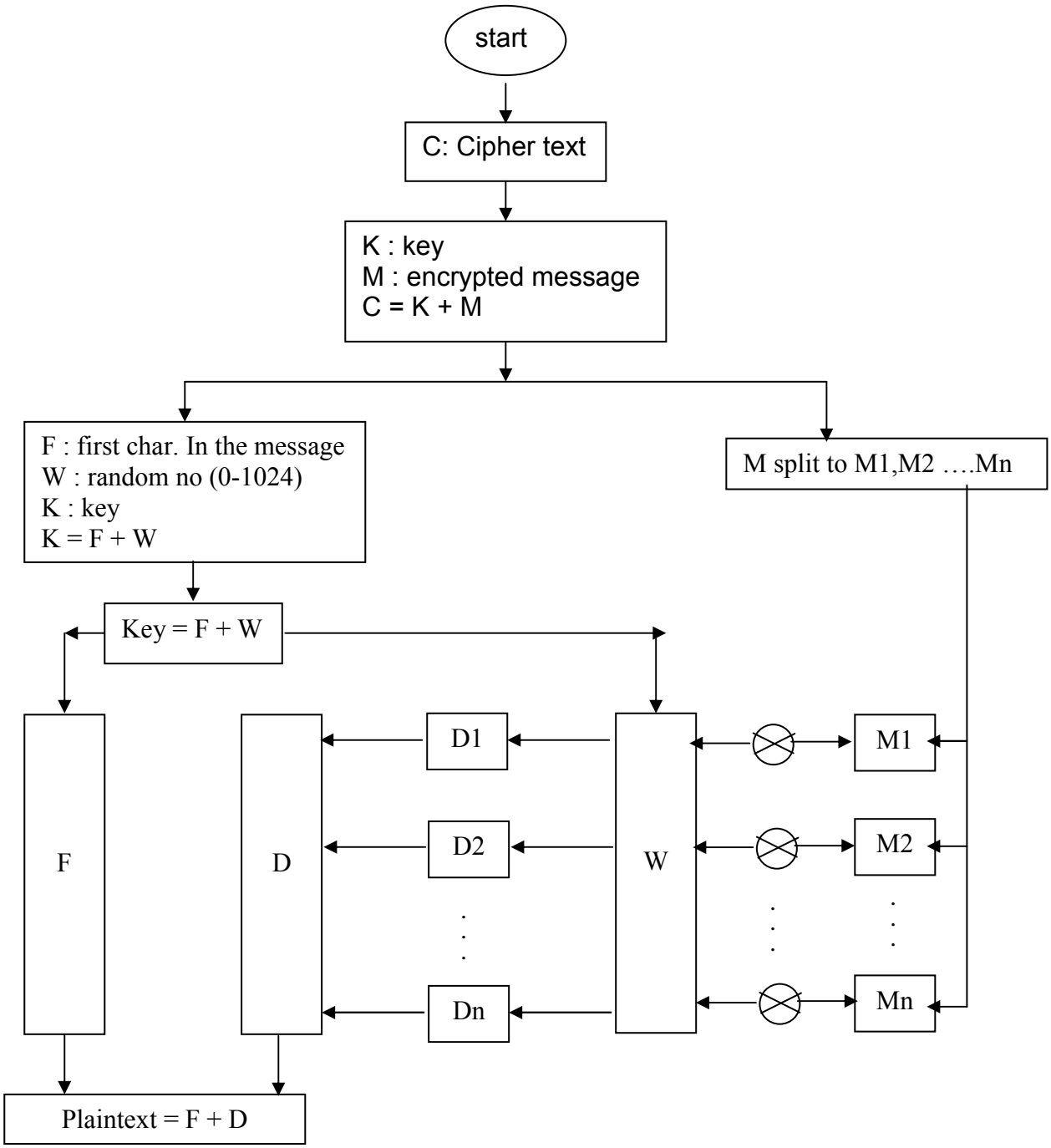
5- take each 4 char and convert it to binary numbers:
0308 ⟶ 0100110100
0958 ⟶ 1110111110
0958 ⟶ 1110111110
1015 ⟶ 1111110111
0115 ⟶ 1111110111
1028 ⟶ 1000000100

6- Use XOR function for each binary number in the above step with the 1111110100 (1012)

0100110100 ⊗ 1111110100 ⟶ 1011000000
308 1012 22 0

By using the table(2):
0 = ط (the first letter)
22 = ر (the second letter)

Then we continue in this operation until retrieve all the plaintext.
Figure (3) illustrate the flowchart of decryption operation.



Figure(3) : Flowchart of Decryption Operation

Conclusions:

1. All characters , numeric, and symbols could be encrypted with the digits from 0 to 9 .
2. Cipher text don't refers to any thing about the plain text or which language used to writing the message. So that , this method could be used to any other language.
3. The size of the encrypted message is twice the size of the original message plus the key length.
   Cipher text = 2*plain text + key length

4. Possible to increase the complexity of encryption using XOR function more than once, as well as encrypt each three or more characters together.
5. Variable matrix was used, the number of columns is as the number of Arabic language characters code (29), as shown in table(1). The number of rows should be variable as the length of characters in the message.
6. Encrypt messages in this way makes the process of encrypting each message is different from the other due to change in the size of the matrix for each message.
7. strength of encryption increase When increase the key length.

References:

1- Mahapatra A. & R. *Dash. Data Encryption and Decryption by Using Hill Cipher Technique and Self Repetitive Matrix*. MSc, computer science Bechelor of Technology, Electronics & instrumentation engineering, Rourkela.2007.
2- Savercheuko .I. , classical cryptography.2005.
3- T. Yeanlee, H. Minglee, J.shiehsu. Data Transmission Encryption & Decryption Algorithm in Network Security,2008.
4- Wayner,Peter. Disappearing Cryptography and Information Hiding, Kaufmann Publishers, Amsterdam.2009.
5- A. Menezes, P. V.Orschot, and S.Vanstone. Applied cryptography .CRC Press.1996.
6- K. Ahmed .*Universal Lossless Compression Technique with built in Encryption*. MSc, computer science , University of Essex .2006
7- J. Gordon .Introduction to cryptography . 1998.
8- J. Buchmann. *Introduction to cryptography*. Technical university of Darmstadt, Germany.2004.

٩- المراياتي،محمد. الطيان،محمد حسان .علم التعمية واستخراج المعمي عند العرب.١٩٩٧.
١٠-ميرعلم، يحيى. اسهامات علماء التعمية في اللسانيات العربية.٢٠٠٢.