مجلة تنمية الرافدين

Tanmyat AL-Rafidain 87(29) 2007
P.P [19- 35]

A Proposed Method of Combining Text Files: Authentication and Encryption

Yasin H. Ismaeil
Assistant Lecturer
Department of Computer Sciences
University of Mosul
e-mail: yaseen_info_2005@yahoo.com

Abstract

This research concerns with proposing a method of combining the encryption and authentication in text files, So the research depended Symmetric encryption methods, transpositions methods, and substitution methods to provide text files security. To provide authentication, a new method for calculating Hash functions was suggested to produce a message digest. The goal of this work is to detect any intrusion activities which includes disclosure, insertion, deletion, and rearranging of the text file while sending through computer's networks.

طريقة مقترحة لدمج الوثوقية والتشفير في الملفات النصية

ياسين حكمت إسماعيل
مدرس مساعد
قسم علوم الحاسبات
كلية علوم الحاسبات والرياضيات
جامعة الموصل

المستخلص

يهتم هذا البحث بتقديم طريقة مقترحة للدمج بين الوثوقية والتشفير في الملفات النصية، حيث أعتمد البحث طرق التشفير المتماثل وطرق التشفير الإبدالية وطرق التشفير التعويضية لتوفير السرية للملفات النصية. لتحقيق الوثوقية، تم اقتراح طريقة جديدة لحساب دوال التمويه أحادية الإتجاه لغرض توليد ملخص الرسالة. يهدف هذا العمل إلى الكشف عن فعاليات المتطفل والمتضمنة الكشف والإضافة والحذف وإعادة ترتيب الملف النصي أثناء إرساله خلال شبكة الحاسبات.

1. Introduction:

Verifying the integrity and authenticity of information is a prime necessity in computer systems and networks. In particular, two parties communicating over an insecure channel require a method by which

تاريخ التسلم ٢٠٠٦/٦/٦_____ تاريخ قبول النشر ٢٠٠٦/١٢/١٩

information sent by one party can be validated as authentic (or unmodified) by the other [Bellare M. ,1998].

Encryption is usually used to provide privacy of data; that is to keep the data secret from people other than the intended recipients. Message Authentication, also often called Message Integrity, provides assurance to the recipient of the data that it came from the expected sender and has not been altered in transit. Encryption itself is not habitually sufficient to ensure privacy, either. Bellovin gives a number of failure modes in IPsec when encryption alone is applied. Most of these are "cut and paste" attacks, where encryption blocks are resent, sent back to the originator, or insert into other messages. A recent field of study is mechanisms for combining message authentication and encryption into one efficient primitive. Even if there are no efficiency gains to be made [Belloivn S. , 1996] [Bellare M. , 1998], [Rose G.].

Nyberg and Rueppel [K. Nyberg ,1994s] presented a digital signature scheme , which is the first signature scheme based on the discrete logarithm problem and provides message recovery. Later, Horster *et al*. [P. Horster, 1994] presented an authenticated encryption scheme modified from Nyberg - Rueppel‘s scheme. Authenticated encryption scheme can be regarded as the combination of message encryption scheme and digital signature scheme. To reduce the communication costs and computational complexity, Hwang *et al*. [S.J. Hwang, 1996] proposed an authenticated encryption scheme with message linkages based on that of Horster *et al.‘s* scheme. Later, Lee and Chang [W.B. Lee, 1997] also proposed another scheme with message linkages based on Lee-Chang‘s scheme. In 2002, Tseng and Jan [Y.M. Tseng ,2002] proposed an efficient authenticated encryption scheme with message linkages and low communication costs based on Horster *et al‘s* scheme. Their proposed method has lower communication costs and less computational complexity than those methods mentioned above. However, there‘s no real authenticated property in Tseng and Jan‘s scheme. Though the intruder couldn't encrypt the message sent from user A to user B, yet he could easily pretend to be A to deliver other messages to B and wouldn't be denied by A. Since the intruder can fool B into accepting an invalid signature (message) as valid, therefore, these schemes do not really achieve the authenticated property.

In this paper a new, simple, and practical constructions of combining message (text file) authentication and encryption was presented. The proposed method can achieve message authentication and security efficiently, and then an intruder won't have the chance to forge to be other's signature to send out the information.

2. Terminology and Definitions:

A. Cipher System:

Encryption is a process of disguising confidential information in such a way that its meaning is unintelligible to an unauthorized person. Decryption is the reverse process means transforming an encrypted message back into its normal form.

A system for encryption and decryption is called a Cryptosystem. Cryptosystems are divided into cipher systems and code systems, as shown in fig. (1). The encryption system concentrates on each character of the word, whereas the code system transforms the word in the plain text to a different word in the code words. Cipher methods are used to implement the cryptosystem. Cipher methods subdivided into transposition, substitution, Stream, and Block ciphers. Transposition encryption encodes a message by reordering the message characters according to some well - defined schemes to produce the cipher text. Whereas, substitution encryption replaces each plain text character by another character in order to produce the cipher text.

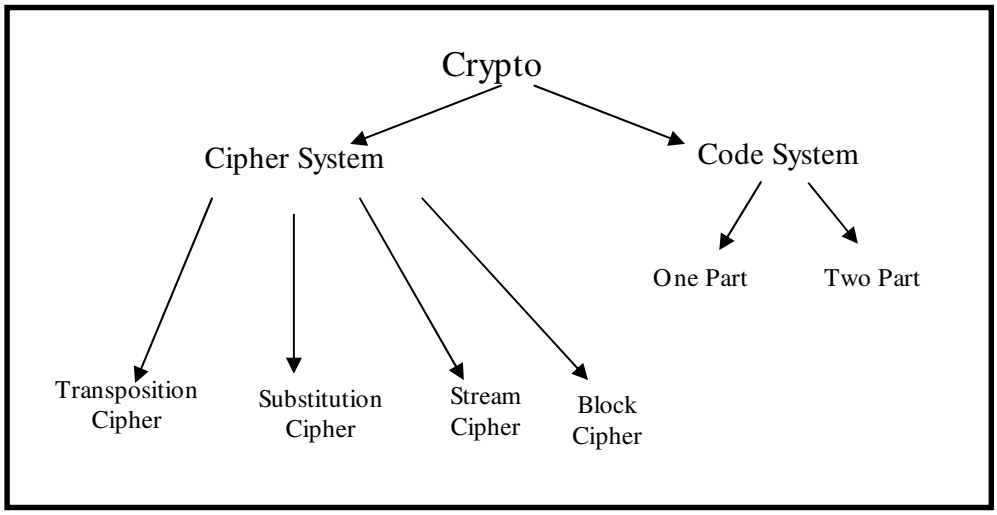


Figure 1
Cryptosystem

Stream encryption concerns with methods that convert the plain text to a cipher text character by character (one character at a time). While block encryption concerns with converting the plain text to a cipher text using a block of characters (more than one character at a time). Finally, cryptosystem are generally divided into symmetric key encryption and asymmetric key encryption. Symmetric key encryption also known as secret key encryption uses a single key to encrypt and decrypt data. The security

of symmetric key algorithms is often directly related to how well the secret key is protected and distributed. Whereas the asymmetric key encryption also known as public key encryption uses two keys. One key, a public key is published widely, whereas the other key must be guarded and kept secret. Given the public key it is computationally infeasible to derive the private key. Asymmetric key cryptography is used either to establish secrecy or authentication or both of them at the same time [Pfleeger C. ,1987] [Seberry J. , 1989] [Hoglund G., 2000].

B. Authentication:

Authentication allows the receiver of a digital message to be confident of both the identity of the sender and the integrity of the message. The authentication service is concerned of assuring that a communication is authentic. In case if a single message such as a warning or alarm signal, the function of the authentication service assure to the recipient that the message is authentic and delivered from authorized source [Seberry J., 1989] [Jan C., 1998].

C. Hash Function:

Hash function H is a transformation that takes a Variable length Message as input m and returns a fixed - size string, which is called the hash value h (that is, h=H(m)). Hash value is small compared to the document itself. In general, hash function is a way of creating a small digital "fingerprint" from any kind of data. The function chops and mixes the data to create the hash value (fingerprint) [Schneier B., 1996] [Stallings W. , 1999].

D. One-way Hash Function:

Hash function H is said to be one - way if it is hard to invert, where "*hard to invert*" means that given a hash value h , it is computationally infeasible to find some input x such that H(x)=h. If, given a message x, it is computationally infeasible to find a message y not equal to x such that H(x) = H(y), then H is said to be collision-free hash function. [Schneier B., 1996] [Stallings W., 1999] [Hoglund G., 2000]

E. Message Digest (MD):

The representation of text in the form of a single string of digits, created using a formula called a one - way hash function. A digest or a hash is a unique set of bytes that is smaller than the input message and represent the input in a unique way. Simply, encrypting a message digest with a private key creates a *digital signature.* It is used to prove the identity of the sender of a message and the message itself [Schneier B., 1996], [Stallings W., 1999].

3. The Proposed Method:

To provide authentication and security in the suggested method, the following scope was dependent:

1. Symmetric encryption methods are used, that is the proposed method used one key (secret key) in encryption and decryption. The usage of one key in both encryption and decryption came from the speed of the encryption process and simplicity of its use.
2. Transposition methods are used frequently either depending on the key or not to provide more Secrecy to the resulted cipher text.
3. Substitution method also used to provide more security and authentication to the resulted cipher text. The substitution cannot be applied to all plaintext characters, but on some of them and the substitution characters same the original characters in its frequency usage in the natural English language. So the above features make substitution method more efficient.
4. To provide authentication services, two mechanisms to calculate message digest was suggested. The first provides a message digest by applying XOR operation to all characters that produces a one byte message digest length. Second mechanism depending on four suggested one - way hash functions and some transposition methods to produce a second message digest with one byte length.
5. Insertion of two message digest values in the resulted cipher text to provide authentication service. Also, some unprinted characters (Meaningless characters) have been inserted in the resulted cipher text to provide more secrecy and randomness.

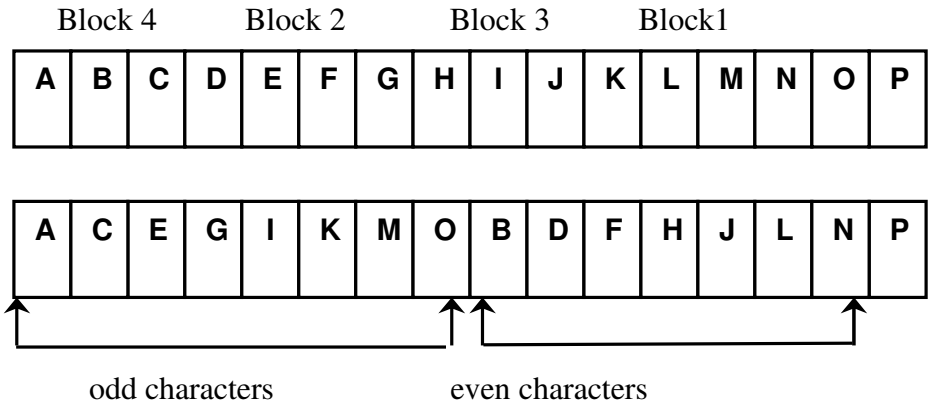
4. Encryption Algorithm:

1. Input plaintext (text file) of any length.
2. Check if plaintext length mod 16 = 0.
   a- If answer was no, append characters (Unprinted characters) until plaintext length mod 16 = 0.
   b- If answer was yes, rearrange plaintext characters in the form of odd characters followed by even characters, i.e.:

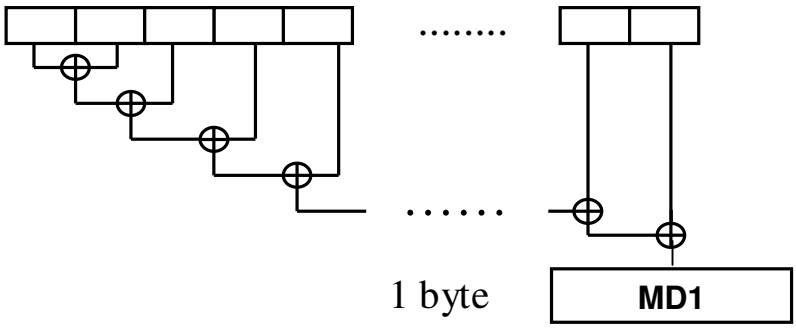
odd characters	even characters

3. Read 4 characters key.
4. Rearrange the key according to the ASCII value of the characters. For example: the keyword send will be arranged to dens and represented as 4231.
5. Now take each 16 characters sequence and repeat the following steps 4 times on them :-
   a- Split the 16 characters sequence into 4 characters blocks.

b- Take the 4 blocks and replace its locations depending on the arranged key. Rearrange the resulting characters sequence in the form of odd characters followed by even characters; i.e. suppose the 16 characters sequence is [abcdefghijklmnop]



6. Perform the logic operation XOR on the characters sequence resulting from step 5, to produce a value called MD1.



7. Split the characters sequence resulting from step 5, into 4 characters blocks and take each 4 blocks arranged them depending on the arranged key , then each block will enter to a specific function depending on the block sequence .

i.e. for each 4 blocks do the following:-

block1	block2	block3	block4
A B C D	E F G H	I J K L	M N O P

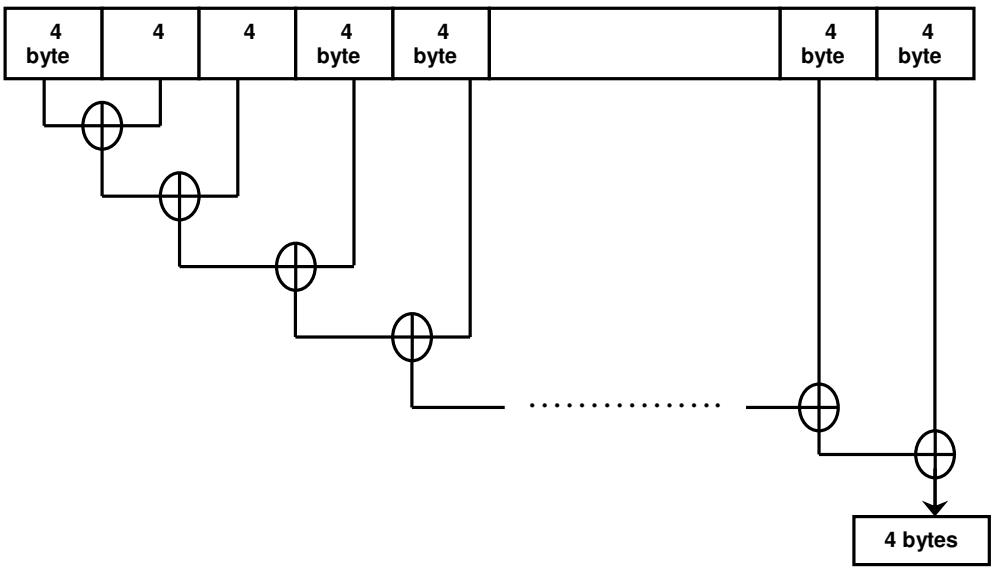
a. Arranged the 4 blocks depending on the arranged key (suppose key = send after arranging would be dens=4231)

block4	block2	block3	block1
M N O P	E F G H	I J K L	A B C D

b. The MD functions sequence is constant :-
i.e.
F1- F (a,b,c,d) = (a^- ^ c) **v** (b ^ d) .
F2- G (a,b,c,d) = (b c) $\oplus$ (a d^-) . $\oplus$
F3- H (a,b,c,d) = (b d^-) $\oplus$ (a^- $\oplus$ c) . $\oplus$
F4- I (a,b,c,d) = (a ^ b) **v** (c ^ d) .

In this process block 4 will be the entry to F4, block 2 will be the entry to F2, block 3 will be the entry to F3 and block **1** will enter F1. The MD functions receive 16 bytes and produce 4 bytes, this is similar to compression function .

c- Do XOR operation on the resulting 4 bytes blocks sequence and produce 4 bytes as a result.

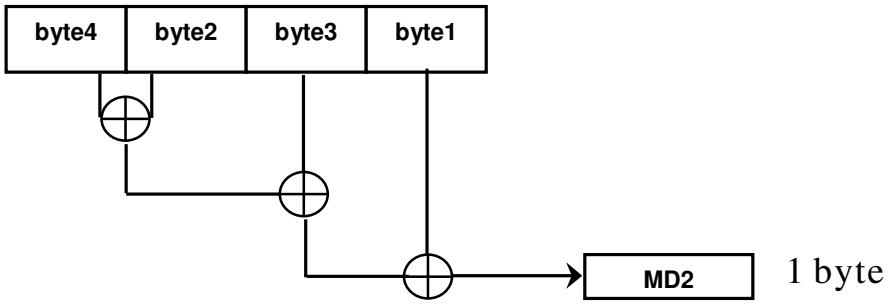


d- Perform XOR operation on the produced 4 bytes after rearranging them depending on the key:-

i.e.

byte1	byte2	byte3	byte4

If the key is send, after arranging would be dens = 4231, so the block sequence would be :-



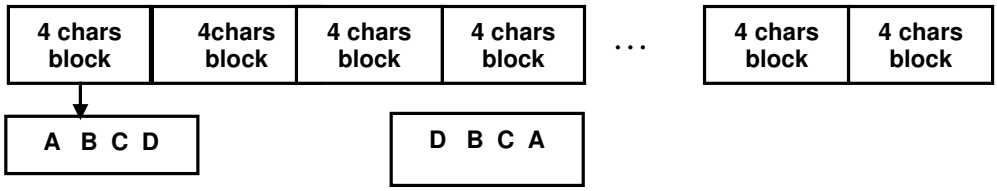
8- Insert MD1, MD2 and some constants in the resulted characters sequence. The constants here are some unprinted characters in order to ensure the plain text randomness, so the constant template of the resulting sequence will be:-

8 bytes	£	md1	£	.....	‰	‰	¤	md2	¤	4 bytes

9- Perform substitution operations on some characters in the resulted sequence using the following table :

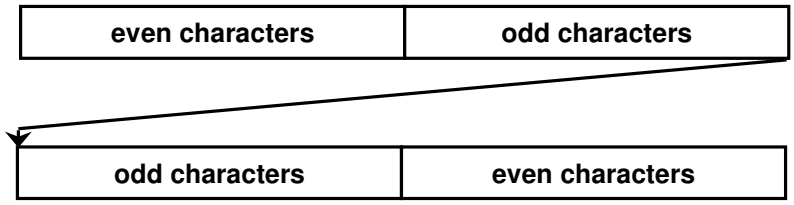
original	Substituted
A	E
E	I
I	O
O	C
C	S
S	T
T	¶

10- Split the result into 4 characters blocks and rearrange the characters in each block depending on the arranged key.



If you know that the arranged key is (4231) , so the arranged characters will be :

11- Rearrange the resulted characters sequence in the form of even characters followed by odd characters and reverse the resulting characters sequence to produce the cipher text.



Cipher text

The encryption algorithm functioning is clarified in figure (2), over - view figure (3) for more details on operations biographies.

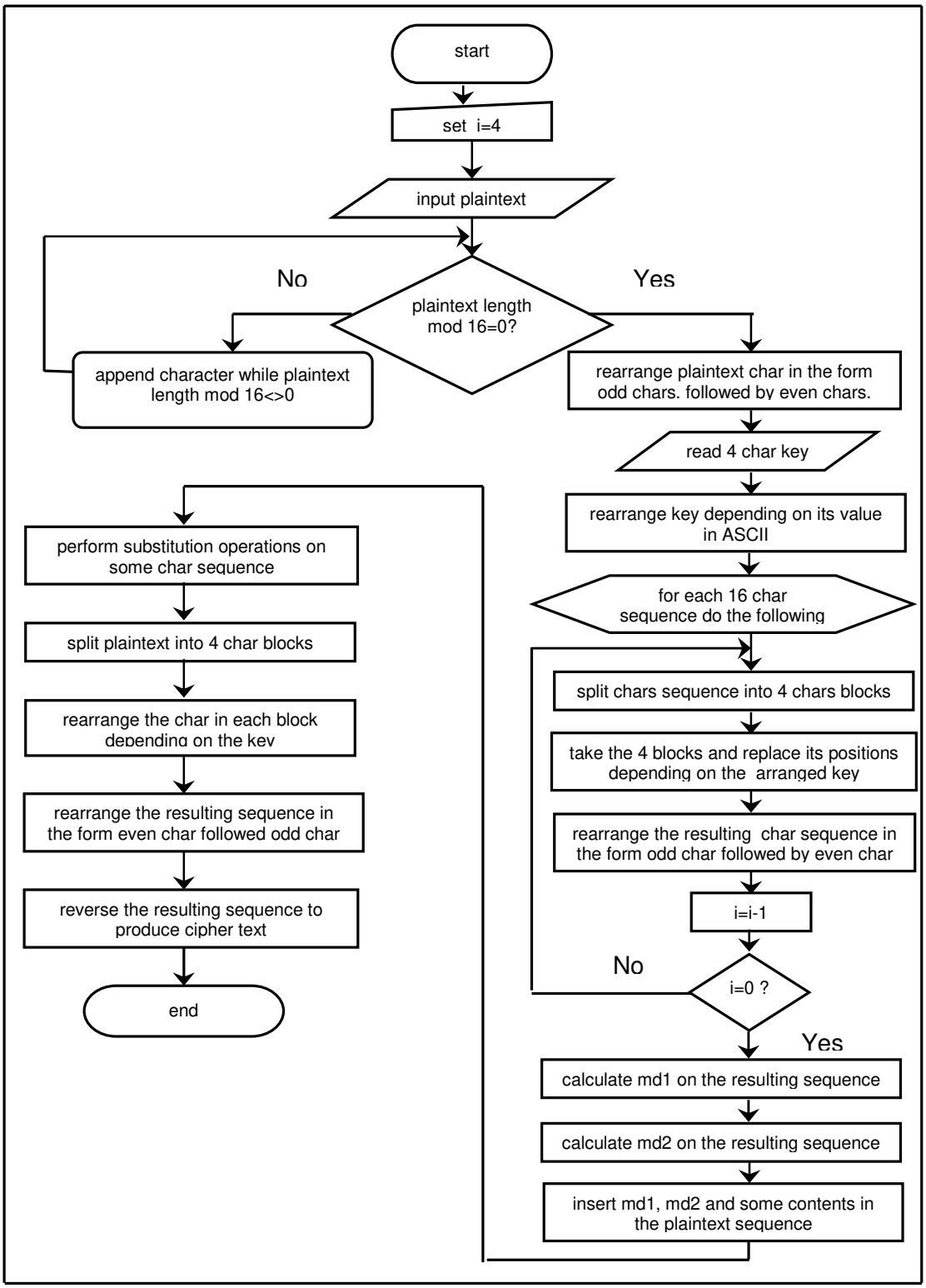


Figure 2 Encryption flowchart

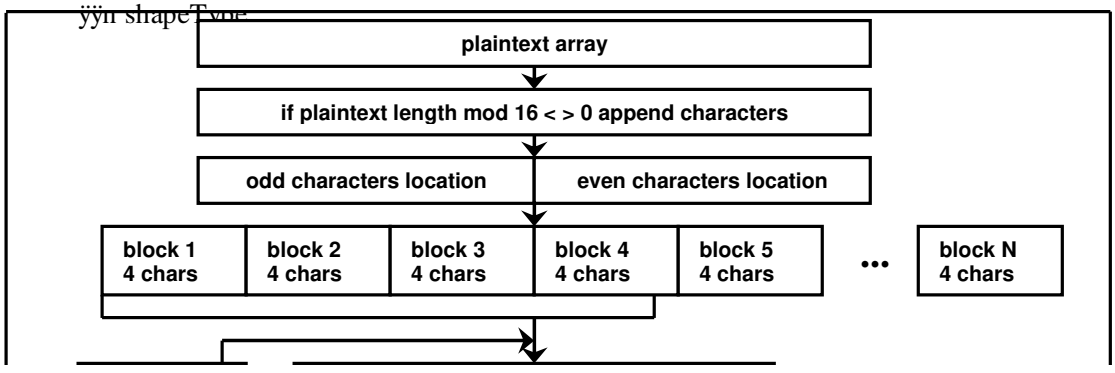


Figure 3 encryption block diagram

5. Decryption Algorithm:

1- Input cipher text (text file).
2- Read 4 characters key.
3- Rearrange the key depending on its value in ASCII.
4- Reverse cipher text sequence and then rearrange characters sequence in the form of even characters followed by the odd characters.
5- Split characters sequence from its mid position, the first sequence represents the even characters and the second sequence represents the odd characters, and construct a new characters sequence in the form of odd character followed by the even character, odd character followed by even character … etc.

odd char	even char	odd char	even char	.....	odd char	even char

6- Split the resulting sequence into 4 char blocks.
7- Rearrange each block characters depending on the arranged key.
8- Perform substitution operations in the opposite way on some characters sequence using the following table:

original	substituted to
A	E
E	I
I	O
O	C
C	S
S	T
T	¶

9- Remove the inserted constants and save the MD1, MD2 values for the time being.

8 bytes	£	md1	£	.....	‰	‰	¤	md2	¤	4 bytes

10- From the resulting sequence calculate the values of new MD1 , MD2 .
11- Compare between the values of calculated MD1, MD2 and the saved MD1 and MD2 .
a- If the two values didn't match, then print "the received text is not authenticated" and the text file can't be decrypted .

b- If the two values matched, then print "the received text is authenticated" and the decryption process will continue .

12-Split the characters sequence from its mid location, the first sequence represent the odd characters and the second sequence represent even characters.

13- Take each 16 characters sequence and repeat the following steps 4 times:

a- Split characters sequence into 4 characters blocks.

b- Take each 4 blocks and replace its locations depending on the arranged key.

c- Split the characters sequence from its mid location, the first sequence represents odd characters while the second sequence represents even characters.

d- The result would be constructing a new characters sequence in the form of odd character, even character, odd character, even character…..odd, even.

odd char	**even char**	**odd char**	**even char**	**.....**	**odd char**	**even char**

14- Finally remove appended characters to produce the original plaintext sequence.

Decryption algorithm functioning is clarified in flowchart (4) .

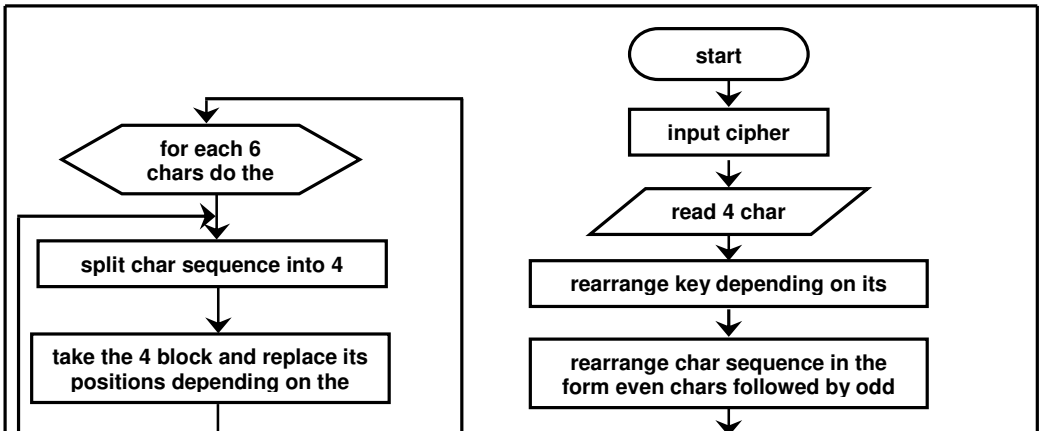


Figure 4 decryption flowchart

6. Samples of Practical Results:

• Example 1 :- If the plaintext is "i catch the spy" and the key is "free", after arranging eefr = 3412, so the resulted cipher text will be :- "e ¶Fh£‰‰ h‡pso¤¤s£° i¶yt"
• Example 2 :- If the plaintext is "meet me at 9 pm" and the key is "help", after arranging ehlp = 2134, so the resulted cipher text will be :- " i P¤£¶‰§i¶m ‡p¤m£9i‰e m"

7. Features of the Proposed Method :

There are some criteria and features used for measuring the performance of the cipher system , So in this section we explain these criteria and how it was implemented in our proposed method . [Beker H. , 1982] [Pfleeger C. ,1987] [Seberry J. , 1989]

A. Diffusion:

The idea behind diffusion is to 'spread' the statistics of the message space into a statistical structure which involves long combinations of the letters in the cryptogram. The effect of it is that the cryptanalyst needs to intercept a much longer cryptogram before he can attempt statistical decipherment. In the proposed method diffusion is implemented by first, applying transposition methods frequently either by applying the key or not to provide more randomness to the resulted cipher text. Second, substitution operations depend on the statistical features of the natural English language. So, the substitution was applied between the most repeated letters in the natural English language.

B. Confusion:

The idea of confusion is to make the relation between a cryptogram and the corresponding key a complex one. It tries to ensure that the majority of the key is needed to obtain even very short cryptograms. In the proposed system, the relation between the secret key and cipher text is more complex, this came from many transposition operations depending on the arranged key, So the cryptanalyst must known the whole key to decrypt the cipher text .

C. Shannon's Five Criteria:

We are in a position to discuss the earnings of five criteria which Shannon suggested should be applied to the cipher system. His suggested important criteria were:

a- The amount of secrecy offered:

In this point, Shannon said that the key must be kept secret and, on occasion, may need to be memorized. In our proposed method, also the key

secrecy is more important and our system provides the authentication service to detect any intrusion activities on the cryptogram.

b- The size of the key:

Consequently the key should be as small or simple as possible. In our method the key is consists of four characters, this means that the key is as simple as possible.

c- The simplicity of the enciphering and deciphering operations:

Enciphering and deciphering should be as simple as possible. If they are done manually, complexity leads to loss of time, errors, etc. The proposed method includes some complex operations, this could be shown in overlapping transposition and substitution methods with some iterations. The complexity of our method came because today we have the advantages of electronics and do not need to bother with mechanical machines.

d- The propagation of errors:

With some cipher systems one error occurred on a transmission can mean that, when the cryptogram is deciphered, whole portions, or even the complete message are garbled , to most communication systems this error propagation should clearly be minimized . To prevent error propagation, we used the idea of block cipher in our system, also the provision of authentication service can detect any error may occur during transmission.

e- Extension of the message:

Finally in some cipher systems the size of the message is increased by the enciphering process, for instance the use of nulls (i.e. adding meaningless characters to swamp the message statistics) causes a larger cryptogram than message. Our proposed method has extension of the message that came from insertion message digest values to provide authentication service, also insertion of small number of meaningless characters to provide more randomness and secrecy of the resulted cryptogram.

ÿÿÿÿpar

8. Conclusions and Discussions:

1. Cryptography is hard to get right, and it has recently become clear that combining encryption and authentication to get practical security is downright difficult. These difficulties are avoidable in our method through additional constraints on the quality of the key, as well as through the simplicity of transposition and substitution operations that have been used in our system, to provide text files security and authentication.
2. The proposed method could be working properly on any plaintext length as entry; the system can be more efficient especially with longer plaintext length.
3. The methods for calculating hash function values to produce the message digest is more efficient, so any intruder activities (insertion, deletion, rearranging, and disclosure) even in one character will be discovered.

References

1. Beker H., Piper F., "Cipher systems the protection" Published by Northwood Publication of communication , London, 1982.
2. Bellare M., Cantti R., Krawczyk H., "Keying Hash Functions for Message Authentication", Springer-Verlag, 1998 .
3. Hoglund G. , Rauch J. , Georgi G., "Hack Proofing Your Network Internet Trade Craft", Published by Syng ., America , 2000.
4. Jan C. A., "Basic Methods Of Cryptography", Published by Cambridge University press, 1998 .
5. K. Nyberg and R. A. Rueppel "Message recover for Signature Schemes Based on the Discrete Logarithm," Advance in Cryptology-EUROCRYPT'94, Springe-Verlag, 1994.
6. Pfleeger C. P., "Security In Computing", The University of Tennessee, Published by Prentice-hall international , inc Printed in the united, 1989
7. P. Horster, M. Michels, and H. Petersen, "Authenticated Encryption Schemes with low Communication Costs," Electronics Letters, Vol. 30,1994.
8. Rose G. , "Combining Message Authentication and Encryption", Qualcomn Australia , ggr@qualcomm.com .
9. S.Bellovin,"Problem Areas for the IP Security Protocol", in Proceedings of the Sixth Usenix Unix Security Symposium, pp.1-16, San Jose, CA, ,1996 .
10. Schneier B., "Applied Cryptography", Published by Katherine Schowalter, Printed in United State of America , 1996
11. S. J. Hwang, C. C. Chang, and W. P. Yang "Authenticated Encryption Schemes with Message linkages," Information Processing Letters, Vol. 58, 1996.
12. Seberry J., Pieprzyk J., "Cryptography An Introduction To Computer Security", Published by Prentice hall of Australia Pty ltd, 1989. .
13. Stallings W., "Cryptography And Network Security Principles And Practice", Second Edition , Published by Prentice-hall , inc , The United State of America, 1999.
14. W. B. Lee and C. C. Chang "Authenticated Encryption Scheme With linkage Between Message Blocks," Information Processing Letters, Vol. 63, 1997.
15. Y. M. Tseng and J. K. Jan "An Efficient Authenticated Encryption Scheme with Message linkages and low Communication Costs," Journal of Information Science and Engineering, Vol. 18, No.1, January 2002.