INTRA CORRELATION DESTRUCTION TECHNIQUE *FOR ENCRYPTION DATA*

Alaa Noori Mazher
Department of Computer Science and Information System. University of Technology. Baghdad- Iraq.

Abstract

The need for efficient technique for data encryption is clear. Our attention will be devoted to design hybrid technique which, consists of simple and efficient methods such as RLE, vector substitution, and matrix operations. The adopted technique consists of three stages. Each stage contains a specific tool that is suitable for varied type of data such as text, image, and video .In first stage, different methods could be applied depending on the source data format: applying Run Length Encoding (RLE) and Addition Neighboring Element (ANE) to the image, audio and video, while pattern matching is applied to text file format . Stage two involves matrix manipulation and rotation. Element substitution is implemented to the matrix in third stage; a key substitution matrix is used for the process of substituting elements of the matrix. The key matrix will be sent to the destination receiver in order to enable him or her to reconstruct the original data. Finally it must be mentioned that a good result have been obtained using arbitrary plain text and standard image (Leena), where some investigations proved that it is so hard to break the encrypted data.

تقنية تحطيم الترابط الداخلي لتشفير البيانات

علاء نوري مزهر
قسم علوم الحاسبات ونظم المعلومات، الجامعة التكنولوجية. بغداد– العراق.

الخلاصة

إن الحاجة إلى طرائق كفوءة لتشفير البيانات هي واضحة.إذ سيتوجه اهتمامنا إلى تصميم تقنية هجينه والتي تتكون من طرائق وأدوات بسيطة وكفوءة ومتوفرة, مثل تقنية RLE وعمليات المصفوفات وطرق أخرى . تتكون التقنية المستعملة من ثلاث مراحل كل مرحلة تتكون من أدوات محددة وملائمة للاستخدام مع مختلف البيانات كملفات الصورة والصوت والفيديو والنص .

المرحلة الأولى يمكن أن تطبق عدد من الطرائق معتمدة على نوع الملف المعالج حيث تطبق تقنية RLE وكذلك تقنية ANE على ملفات الصور والصوت والفيديو بينما تطبق تقنية مطابقة الأنماط على ملفات النصوص ,أما المرحلة الثانية فتتضمن استخدام عمليات المصفوفات وتدويرها. حيث تقنية تبديل العناصر تطبق على المرحلة الثالثة معتمدة على مصفوفة مفتاح وهذه المصفوفة ترسل إلى وحدة فك التشفير وذلك لتمكينه من إعادة تشكيل البيانات الأساسية .

1. **Introduction**

Encryption plays a key role in most modern application technologies. Suppose we have a message to communicate, but our message might fall into the wrong hands, this will cause a disaster especially if the message contains sensitive information about accounts numbers, or military detail or many other important aspects. By using encryption, we disguise the message so that even if the transmission is diverted, the message will not be revealed. Encryption is a mean of maintaining secure data in an insecure environment. There are many methods and strategies which can be used to encrypt data and it is characterized by its complexity, efficiency, key type such as Rivest-Shamir-Adelman (RSA), Data Encryption Standard (DES), and shifting methods.

Our strategy consists of three stages: the first is to read the plain text or any data then put them into byte stream, in order to implement Run Length Encoding (RLE) methods or pattern matching. The second stage is to use the matrix operation for the byte stream, while the third is to use the multiplication process for the resulted matrix. The General diagram that demonstrates the steps of the encryption process is shown in Figure (1).

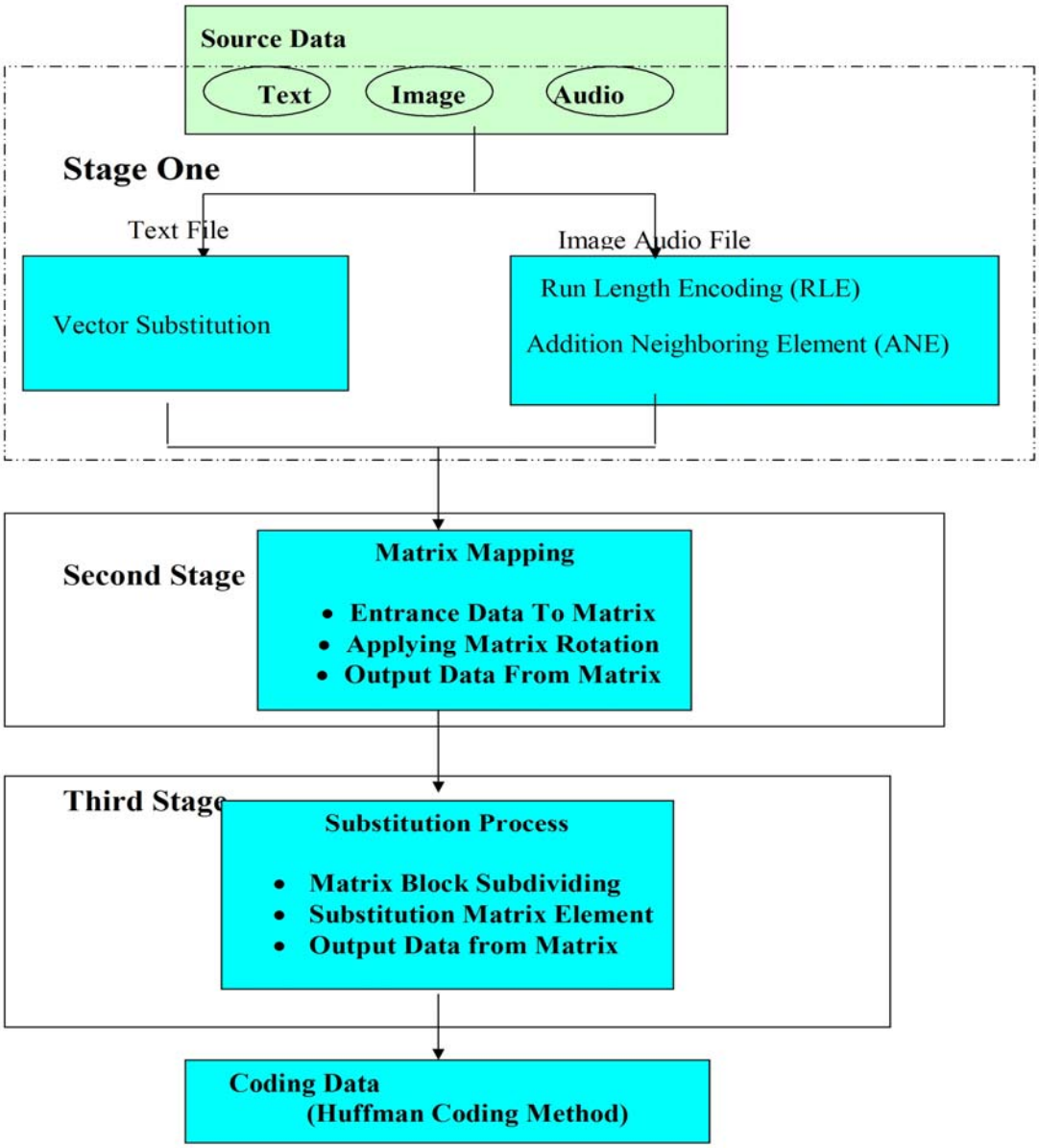

Figure 1: General Encryption Diagram

2. First Stage

The used tools in this stage depend on the type of source data. The suited method for audio, image, and video data is to apply Run Length Encoding (RLE) and the Addition Neighboring Element (ANE), while the suited method for text data is pattern matching technique.

2.1 Run-Length Encoding Technique (RLE)

In this mapping method, the sequence of data bytes are mapped into sequences of integer pairs (g, l), where g denotes the ASCII number of the symbol, and l refers to the length of the adopted symbol, as illustrated in Figure(2)

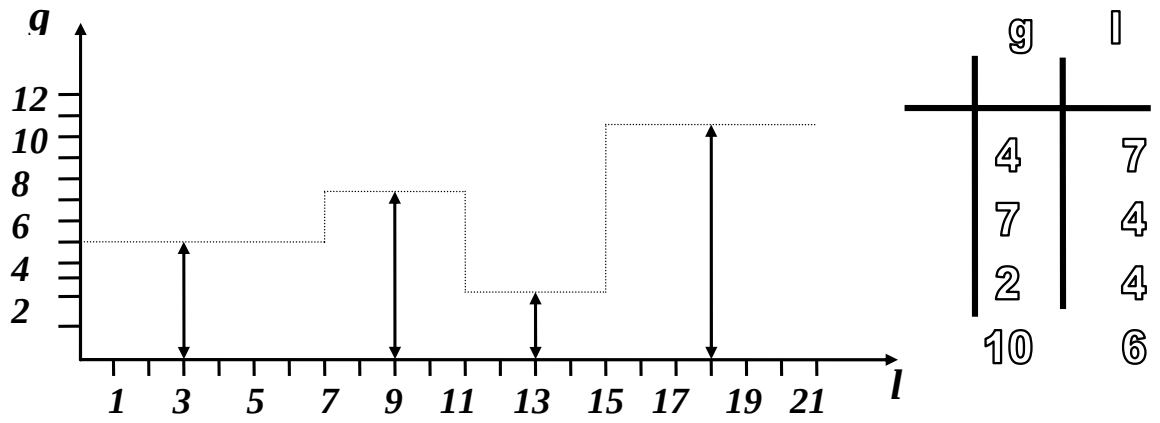

Figure 2: Run-Length Mapping.

This mapping method can be used, efficiently, for data composed of similar regions such as image and video. These two types are characterized by its smooth neighbor pixels. The RLE will destroy the correlation between the pixels. [1,2]

2.2 Addition Neighboring Element (ANE)

The output from the previously stage will be distorted in order to make the interpretability of the image very hard. Each element will be added with the next element in the vector. This process makes the information involving the byte stream spread over a wide range of data or bytes. The reverse process will be implemented in the descriptor unit.

2.3 Vector Substitution

The text data consists of tokens (token is word). The tokens involve high dependence between them and mainly between the neighbor tokens and less with much far tokens. At first, we build a table which consists of the most used terms such as the commonly used verbs, subjects, question words, and others. Each row in the table consists of the term and its key. The table is used later in the substitution in the plain text, where the matching process is applied through the scanning step for the input stream of text. The output of this step is the keys of each token in the stream, if there is no match; it consists of the original tokens as shown in Figure (3). [3]

The table is send to the decryption side in order to re construct the original text. High compression ratio could be obtained using this technique because the new resulted stream is smaller than the source stream.

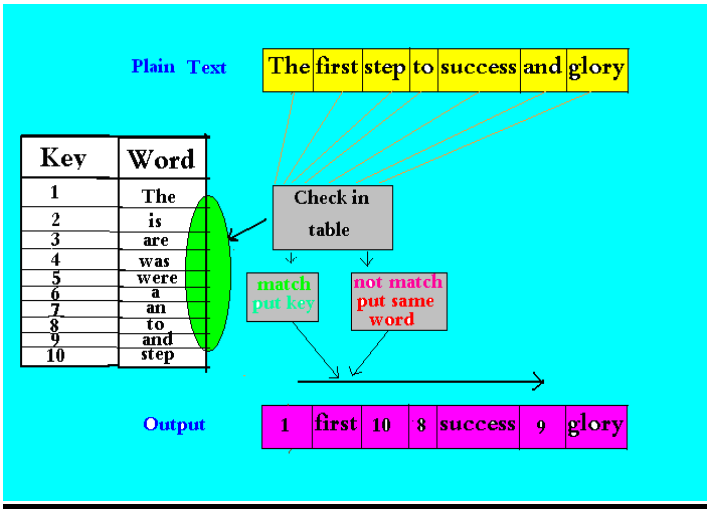

Figure 3: The Matching Process

3. Second Stage

The output data of the first stage is fed into the second stage, where simple matrix operation is applied

3.1 Matrix Mapping

Matrix representation and operation give us high security level. The basic motivation behind the mapping is to transfer a set of medium correlated data (the data lost some of its correlation in the first stage) into matrix elements and to implement specific tool suitable for increasing the variance of the data by implementing some operations that will randomize the result from the previous stage. Our technique will use different mechanism to distribute the data output from the previous stage into matrix element. This configuration is either diagonally filled or filling the odd rows first then evens rows or other. Some of filling types are shown in Table (1). Also this configuration must be available to the decrypted in order to enable him or her in reconstructing the original data. Rotation operation is applied to the matrix but we must be careful in choosing the degree of rotation because some degrees will lead us to loss data. The degrees that could be used is (90,180,270). The output stream is composed using many configurations as in the entrance of data to this process. See Figure (4). [4, 5]

Table 1: the Filling Configuration Types

Type	Code	Description
Row-Column	1	Element of rows
Column-Row	2	Elements of columns
Even- Odd Row	3	Element of odd then even rows
Odd- Even Column	4	Element of even then odd column

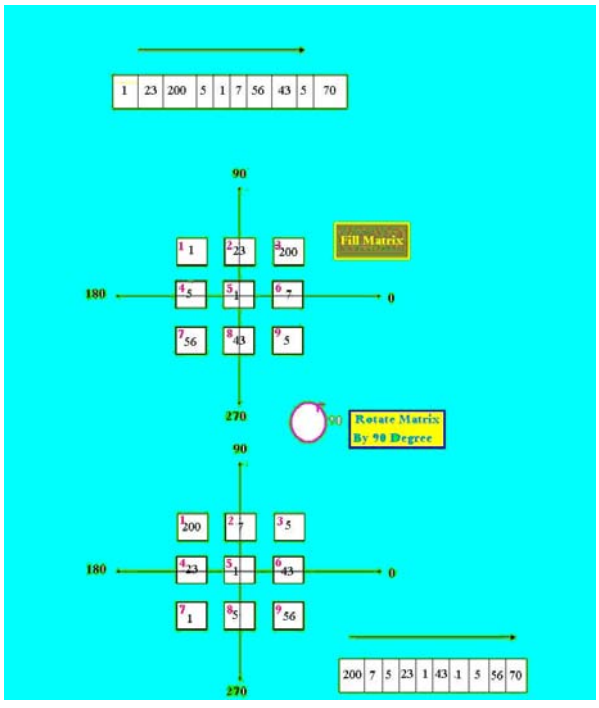

Figure 4: Filling and Rotating the Matrix.

4. Third Stage

In the Final stage an important step is applied, redistributing the element of the matrix.

4.1 Matrix Block Subdividing.

At the entry to the third stage, source matrix elements are grouped into similar size blocks, each is considered as to be an individual matrix. The block subdividing step is an important process for the following reasons;

Data encryption and decryption processes can be executed more efficiently and faster, especially on PC computers, than performing the process on the whole matrix. This is because of the limited memory size of the available PC computers.

Matrix blocks involves more stationary information than the whole matrix data; i.e. small block size may be regarded as to belong to the same informational region, while big block's size contains so many different regions.

Accordingly, less transformed Coefficients may retain to represent, efficiently, the transformed elements.

Therefore, choosing the correct blocks size will lead to improve the overall technique. The small block size leads to simple and easy implementing technique, while choosing large block size is more secure but slow implementation. In our adaptation the most suitable block size is 8X8. [6]

4.2 The Substitution Process.

For each block of the matrix, a substitution operation is applied. The substitution matrix consists of values that identify the position of the coefficient in the matrix to be exchanged with each other. The substitution matrix is sent to the receiver decrypted in order to reconstruct the original matrix. After the process has been applied, the output matrix will be distributed over a stream (one dimensional matrix) as illustrated in Figure (5).

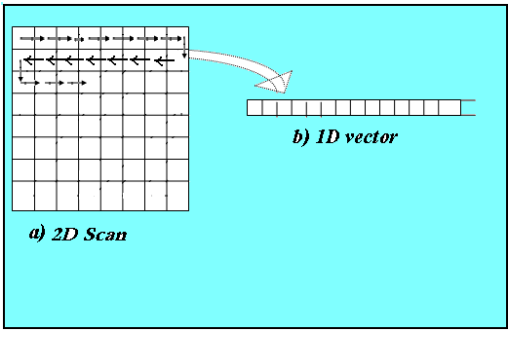

Figure 5: The Output of the Third Stage

5. The Encoding Step:

Up to this point, the cipher stream is presented in the form of a sequence of integer numbers ready to be stored or transmitted through a channel from place to another. In order to compress the data, the storage or the transmission of the coded values is preferable as to be in variable lengths of coded word than in fixed length. Huffman encoding, is one of the most efficient coding method that is used to achieve this requirement. By this error-less encoding method, the shorter binary code is, usually, assigned to higher probability coded word, while longer binary coded is gone to coded word of lowest probability. [7]

The keys that are sending with the encrypted data are shown in Table (2).

Table 2: the Encrypting Keys

Key	Description
RLE Keys	Keys for RLE positions
ANE	Keys for ANE starting
V Dictionary	Vector Substitution Dictionary
MES1C Keys	Matrix Entrance Stage One Configuration
MOS1C Keys	Matrix Output Stage One Configuration
R-Angle	Matrix Rotation Angle
Block Size Keys	Block Size
S-Keys	Substitution Keys
H-Tree Keys	Huffman Tree

6. Decryption and Decoding

In the decryption side which could be a decoder and a decryptor, decoding the compressed information, then the output will be decrypted to get the original data. As a first step, there are many keys to be used in order to help in reconstructing data. These keys are received from the encrypting unit. At first Huffman coding tree is constructed in order to covnert the received binary numbers back into their original quantized values. These coded values should then be transferred from its sequential into matrix representation and as inverse then it constructed (using row column or other) process for filling the matrix, a reconfiguration by using the key matrix is used to resubstitute the matrix element. The next step is the inverse rotation for the matrix by inverse degree, and also by applying the inverse process of the run-length-encoding algorithm and Addition Neighboring Element (ANE) if the original data is image or video for reconstructing the original data. If the

original data is text based, we use the key table that was stored in order to rematch the stream and reconstruct the original data.

7. Discussion and Conclusion

The objective of our paper was to adopt many simple and efficient methods which are used for encryption to get benefit from the facilities of each method in order to design an efficient model for encryption. The RLE and pattern matching were adopted in order to break the dependence that exist in the neighbor elements, and could be important for compression to make the size of the new stream less than original length but not in all cases or data format. Also Addition Neighboring Element (ANE) method will spread the information in the stream into much large range of data. The matrix operation is also used to make the dependents as less as in order to make the process of breaking the cipher text impossible. It must be mentioned that a good result have been obtained using arbitrary plain text and standard image (Leena), where some investigations proved that it is so hard to break the encrypted data and the technique could be applied into varied type of data.

8. Practical Result:

The encryption system was programmed using Visual BASIC. There is no need for Operating System type because this does not affect the work performance or implementation. The practical results are shown in the following images.

Figure 6: Lenna Original Image

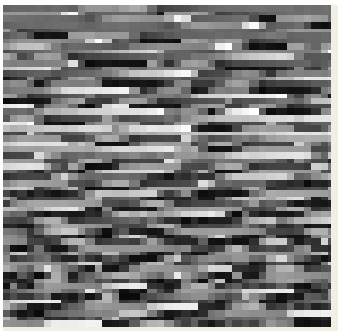

Figure 7: RLE and ANE

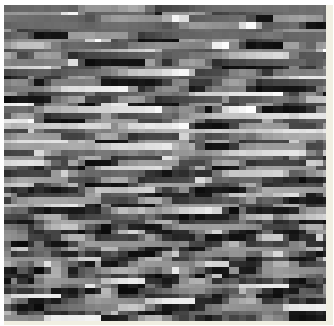

Figure 8: Rotating Matrix

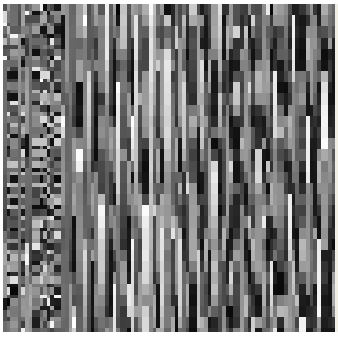

Figure 9: the Substitution Process

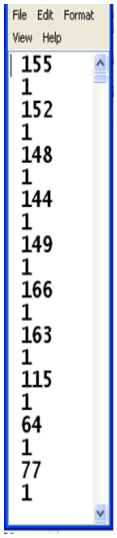

Figure 10: RLE for line of data

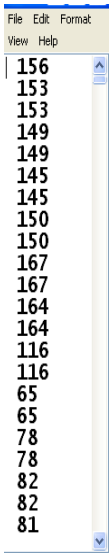

Figure 11: ANE for line of data

Figure 12: the Encryption System

References

1. Bruce,S. N. **2003.** *Practical Cryptography*. First Edition, John Wiley & Sons Inc.pp. 66-67.
2. Adam, Y. **2004.** *Malicious Cryptography*. First Edition, John Wiley & Sons Inc. p 22**,**pp.35-46.
3. Osborne, **2001.** *IPSec Security VPNs*. First Edition, McGraw Hill. pp 46-47,p100.
4. Charles P. Fleeger, **1989.** Security in Computing. First Edition, Prentice-Hall Inc.P83, p116,p123.
5. Gonzalez, R.G., and Wintz,P. **1987**. *Digital image processing*. Second Edition , Addison-Wesley Publishing Comp. pp81-110.
6. Huffman, D.A. **1952**.A method for the construction of minimum-redundancy codes, Proc. IRE. **40** (9):1098-1101.
7. Krawczyk, H. **2001**. The order of encryption and authentication for protecting Communications. CRYPTO LNCS, Volume 2139: 310–331.