A Study of Different Complexity Evaluation Approaches for Some Cryptosystems

Sabiha F. Jawad
Al-Mustansyria University Iraq-
E-mail: sabiha_fj@yahoo.com

Abstract

This paper provides analysis for different approaches used for security evaluation of some cipher systems. Many tests (computational complexity, statistical and Information theoretical)
are performed on the ciphered outputs of a number of classical cipher systems ranging from mono-alphabetic, to poly-alphabetic. The results were encouraging, and can be considered as an approach to enhance the security evaluation tasks (especially for the used skewness , kurtosis, unicity distance and the Key Entropy parameters), that is really considered as a hard problem facing the cryptosystem design.

الخلاصة

يقدم هذا البحث تحليل لمختلف المفاهيم المستخدمه في تقويم امنية بعض انظمة التشفير .هناك العديد من الاختبارات مثل (التعقيد ، الاحصائية ونظرية المعلومات) والتي تكون ناتجه من انظمة التشفير الكلاسيكي منها الانظمه الهجائيه الاحاديه والانظمه الهجائيه المتعدده.
تعتبر المعاملات من اهم المشاكل التي تواجه تصميم انظمة التشفير .
skewness , kurtosis, unicity distance and the Key Entropy

1- INTRODUCTION

Good cryptographic systems should always be designed so that they are as difficult to break as possible. It is possible to build systems that cannot be broken in practice (though this cannot usually be proved).

The designer of a cipher system faces an important question about the validity of his designed system. Hence his main goal is to find rules or parameters by which he can decide about the “importance” of his design. This "importance" includes the complexity aspects. As it is well known in the cryptographic domain, the newly designed cipher system is faced with a great challenges by the cryptanalysis efforts to break the cipher system. To point out the main weak points in the designed cipher system. Therefore, many well known groups working in cryptographic subject establish and intelligent group of specialist who works mainly to find some parameters helping in the security evaluation aspect. The evaluation process has slightly different meaning from the cryptanalysis process; Fig.(1) shows the two different processes.The outcome from the evaluation process is a weighting parameter helping in evaluation of security of tested cipher system, while the cryptanalysis process provides the plaintext, the key or the algorithm.

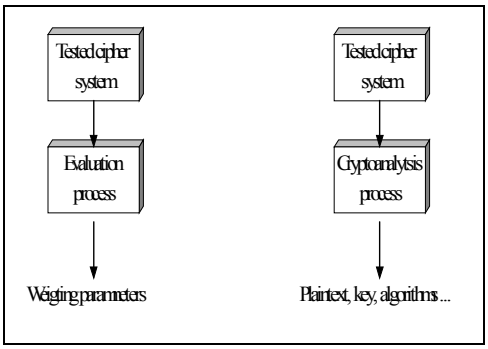

Fig.(1) Evaluation and cryptanalysis process.

So the evaluation process is a necessary process that must be performed by the designer to help in avoiding any weakness which happens during the design stages. There are many evaluation approaches followed for different cipher systems. Piper [F.Piper, 1989] developed two main approaches,which are theoretical and practical approaches.

The statistical approach (frequency counting of letters with different combinations) was applied to the evaluation of monoalphapetic substitution ciphers. While the polyalphabetic substitution cipher resists such statistical evaluation. This fact leads the researchers to a modified statistical evaluation method which is composed of index of coincidence and simple statistical calculations[Douglas,2006]. Fig.(2) shows a block diagram of these concepts.

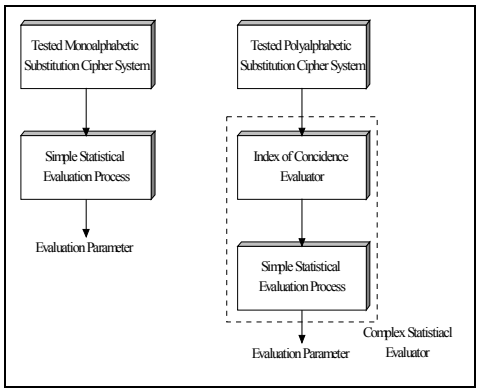

Fig.(2) Different statistical evaluators for different substitution cipher system.

Generally, it is possible to combine many different concepts to adapt the evaluation process to the tested cipher systems.

2- Computational Complexity Approach

Computational complexity theory is a branch of the theory of computation in theoretical computer science and mathematics that focuses on classifying computational problems according to their inherent difficulty. In this context, a computational problem is understood to be a task that is in principle amenable to bring solved by a computer (which basically means that the problem can be stated by a set of mathematical instructions). Informally, a computational problem consists of problem instances and solutions to these problem instances.

The evaluation mainly based on the measure of the security of the ciphering algorithm. The security is related to the complexity of the algorithm. Hence the complexity can be considered as an evaluation parameter of the designed cipher system. The computer specialists use a well

known measure of this complexity which is called a Big "O" measure. Table 1 Big "O" complexity measures for different cipher systems.
Where: n, is the number of alphabetic, m, is the number of the keys and d, is the number of security keys.

Hence, the algorithm is said be secure if it can not be solved, or analyzed (broken) even when we use the computer with a limited time and a limited memory.This security is said to be complete. In the other hand, we say that a system has a measured security if we can analysis this algorithm by a computer with a highly speed and large memory [Th.Bier,2004]. The complexity measure of an algorithm depends mainly on time and memory which are required to analyze the algorithm by the computer. The evaluation of the complexity of the algorithm depends on the type of the algorithm.For example, the complexity of cipher algorithm differs if the algorithm is monolaphabetic type or it is a polyalphabetic type. While the complexity of the transposition cipher algorithm depends on the used secret key. On the other hand the complexity of the stream cipher depends on the randomness of the key generator [D.R.Hankersons ans et al.,2000].

3- STATISTICAL EVALUATION APPROACHES

In this section,some important concepts for the statistical evaluation approach used to identify between substitution cipher algorithms (either monoalphabetic or polyalphabetic), and identification the value of the secret keys used by the substitution ciphers based on some statistical parameters as the index of coincidence and some modified statistical tests depending on the measurements of the expected value, variances, skewness and kurtosis [F.Piper,1989], [J.Seberry,1989].

3.1The index of coincidence (Ic) parameter:

The index of coincidence (Ic) is a numerical parameter used to identify the cipher algorithm whether it is a monoalphabetic or polyalphabetic according to a specified relations and values which are calculated for different key lengths (k) as shown in table (2).

The Ic parameter is given by the form

$$Ic = \frac{\sum_{\lambda=A}^{Z} f_\lambda - (f_\lambda - 1)}{n(n-1)} \quad (1)$$

$$\sum_{\lambda=A}^{Z} f_\lambda = f_A + f_B + \ldots + f_Z \quad (2)$$

where,
f_λ: represents the frequency of the letter λ in the cipher text.
n: number of the letter in the cipher text.

According to the values of k, the cipher algorithm can be identified either as mono-alphabetic (with k = 1) or poly-alphabetic (with k ≥ 2). Generally, as k is large, this means that the cipher algorithm is more complex. The detailed procedure of using such evaluation parameter is shown in Fig.(3) [F.Piper,1989], [D.W.Davies,1989].

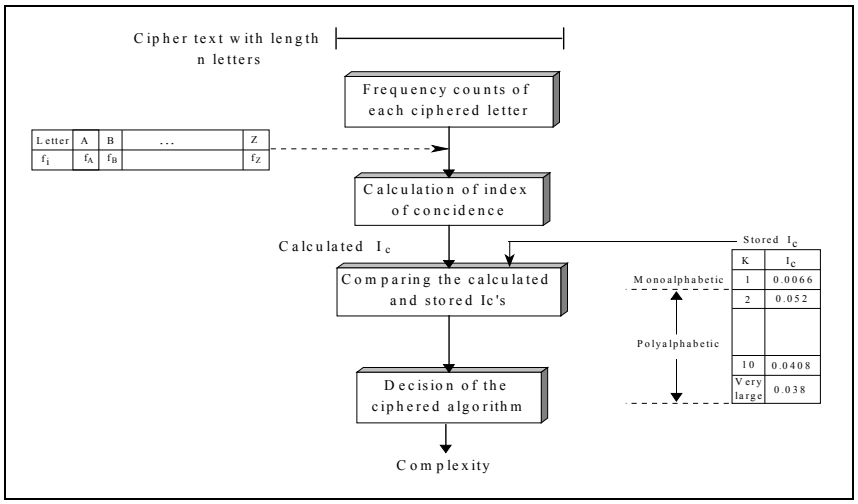

Fig.(3) Index of coincidence parameters used in evaluation procedure.

3.2 Kasisiki Test parameter:

This parameter depends on the calculation of the distances between the repeated similar patterns in the cipher text. This parameter is used as either a separate evaluation parameter for the estimation of the used key length, or it can be used in connection with Ic parameter. The procedure of application of Kasiski parameter in evaluation is as follows:

1- Calculation of the distances (di) between the repeated occurrences of similar patterns of ciphered letters.

2- Factorization of these distances (di), i.e., $d_i = f_{i1} \times f_{i2} \times ... \times f_{in}$

3- Calculation of the common factors which represent the expected keys used by the ciphered algorithm.

Fig.(4) shows the block diagram of this procedure. According to the value of (k) [F.Piper,1989].

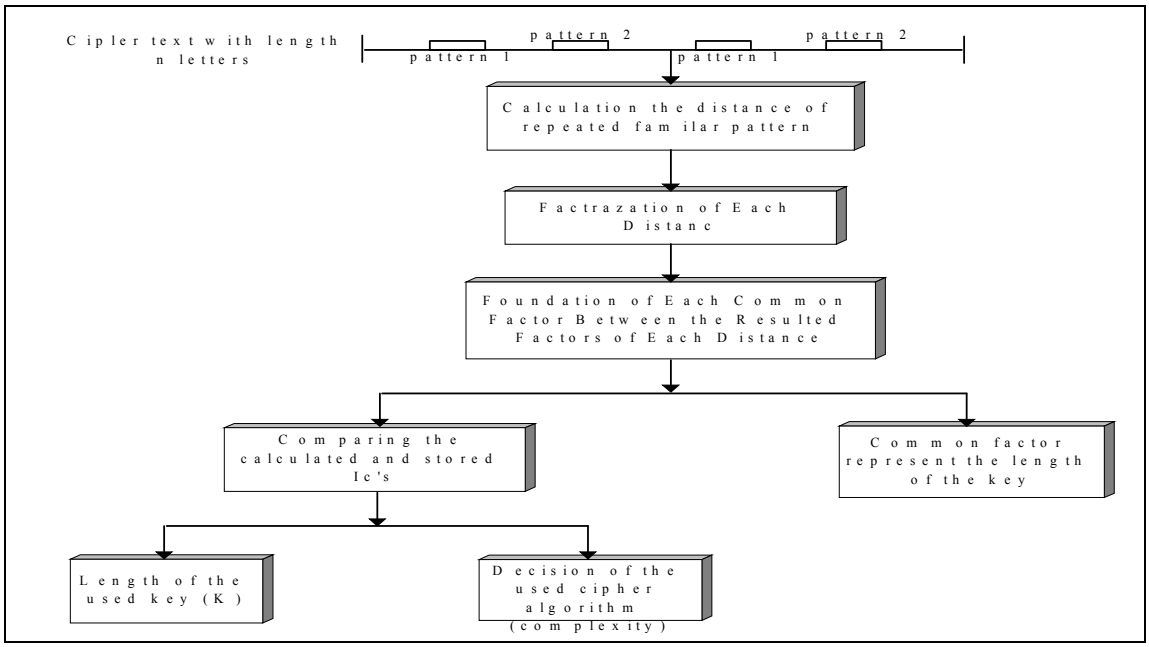

Fig.(4) Block diagram of the application of the Kasiski Test parameter for Complexity Evaluation

3.3 Histogram of the ciphered text:

The statistical English language shows that the letters of a long plain text relative frequencies histogram is shown in Fig.(5) which represents the histogram of the plaintext letters of the English Language [J.Seberry,1989].

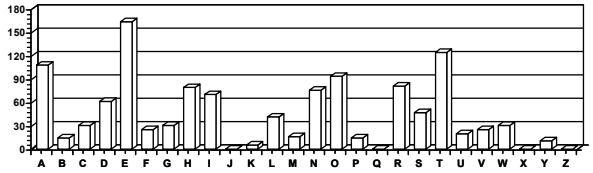

Fig.(5) Histogram for the ASCII values of Gettysburg adress.

To make use such histogram concept as an evaluation parameter, the procedure requires the calculation of the histograms of the ciphertexts (resulting from different cipher algorithms). In [J.Jalbot,2006], it was an efforts made in studying the features of the histogram resulting from a multiplicative cipher algorithm with different keys using a rectangular window of a specified length to identify these factors. And by sorting there features in a data base for any input cipher text.This data base will be a good rule to identify whether the multiplicative ciphered used or not and even identifying the multiplicative key.

3.4 Modified statistical evaluation parameters:

In this subsection, we use some statistical parameters such as the expected value, variance, skewness and kurtosis, as an evaluation parameters. In [D,R.Hankersons ans et al,2000], the parameters were used as evaluation parameters for the monoalphabetic and polyalphabetic ciphers algorithm. The expected value of the plaintext or the cipher text can be defined mathematically by the relation:

$$E(x) = \frac{1}{n} \sum_{i=A}^{Z} (p_i \times F_i) \qquad (3)$$

where; n is the length of the message,
F_i is the relative frequencies for the letter i,
p_i is the probability of the appearance of the letter i.

The cipher text resulting from mono-alphabetic algorithms has a range for the expected value lower than the range of the expected values for the cipher texts results from a poly-alphabetic algorithms. In other words if the range the expected value between 0.061 and 0.069 (depending on the length of the ciphertext), then the cipher algorithm is a mono-alphebetic algorithm. If the expected value greater than (0.069) then cipher algorithm will be a poly-alphabetic algorithm (depending on the length of the ciphertext and the secret key). The variance of the plain text is represented mathematically by the equation:

$$var(x) = \sum_{i=A}^{Z} (p_i)^2 - 0.038 \cdot \qquad (4)$$

As in the expected value behavior, the variance has similar property, i.e., the cipher text resulting from mono-alphabetic algorithm has a range smaller than the range of the variance for the cipher text which results from a poly-alphabetic algorithm. In other words, if the range of the variance between 0.96 and 0.97 (depending on the length of the ciphertext) then the cipher algorithm is a mono-alphabetic algorithm. If the variance greater than (0.97) then

cipher algorithm will be a poly-alphabetic algorithm (depending on the length of the ciphertext and the secret key).

The skewness is the degree of symmetry or departure from symmetry of a distribution. The skewness for the cipher text can be described by the equation.

$$\text{Skewness} = [\text{E(x)}]^3 - [\text{E(x)}]^2 \times 0.038 + \frac{2 \times (0.038)^3}{\left[(\text{var(x)})^2\right]^{\frac{3}{2}}}\text{L} \qquad (5)$$

where, E(x) is the i-th expected value, and Var(x) is the i-th variance.

The cipher text resulting from a mono-alphabetic cipher has a range less than the range of the skewness for the cipher text which results from poly-alphabetic algorithm. In other words if the range of the skewness between 0.00025 and 0.0003 (depending on the length of the ciphertext), then the cipher algorithm is a mono-alphabetic algorithm. If the skewness greater than (0.0003) then cipher algorithm will be a polyalphabetic algorithm (depending on the length of the ciphertext and the secret key).

The kurtosis is the degree of peak ends of a distribution, usually taken relative to a normal distribution.A distribution having a relatively high peak, such as the curve of Fig.(6a), is called leptokurtic while the curve of Fig.(6b) which is flat-topped is called playkurtic. The normal distribution Fig.(6c) which is very peaked are very flat-topped is called mesokurtic [http://www.math.colorado.edu.].

Fig.(6). a) Leptokurtic b) PlatyKurtic c) Mesokurtic

The kurtosis of the cipher text is given by the equation (6)

$$Kurtosis = [E(x)]^4 - 4[E(x)]^3 \times 0.038 + 6[E(x)]^2 \times (0.038)^2 - \frac{3(0.038)^4}{\left[(\text{var}(x))^2\right]^2} \qquad (6)$$

where,

E(x) is the i-th expected value, and var(x) is the i-th variance.

The cipher text resulting from mono-alphabetic algorithm has a range for kurtosis less than the range of kurtosis from a cipher text resulted from a poly-alphabetic algorithm. In other words if the range the kurtosis between 0.00002 and 0.000024 (depending on the length of the ciphertext) then the cipher algorithm is a mono-alphabetic algorithm. If the kurtosis greater than (0.000024) then cipher algorithm will be a poly-alphabetic algorithm (depending on the length of the ciphertext and the secret key).

4- Evaluation Approach Based on Information Theory

The information theory parameters can be used as an evaluation parameters, since they are important parameters. The "unicity distance" can be used to identify the substitution monoalphabetic cipher system from the polyalphabetic cipher system. The unicity distance can be defined as follows; it is minimum numbers of the letters which is required to cryptanlysis a cipher text, and it is given by the equation:

$$U_d = \frac{H(k)}{R} \qquad (7)$$

where

$$H(k) = \sum_{A}^{Z} p_c(k) \log_2 \frac{1}{p_c(k)} \quad (8)$$

is the key entropy, and R is the redundancy [http://www.nelaxs.com.people/nerp/automata/p-and-np]. Where the redundancy is the number of letters which can be deleted from a word or a sentences without producing any change in meaning. When the unicity distance for the cipher text is greater than or equal to 1.35, then the encipher algorithm is poly-alphabetic algorithm, but when the unicity distance is less than (1.35), then enciphering algorithm is additive cipher algorithm [4].

5- Conclusion

-In theory any cryptographic methods with a key can be broken by trying all possible key in sequence. If using brute force to try all possible keys is the only option, the required computation power increases exponentially with the length of the key. However, the key length is not the only relevant issue. Many Ciphers can be broken without trying all possible keys.

-The computational complexity evaluation approach is well known approach, and it depends mainly on calculation of Big O(.) for the cryptosystem under test. As much this Big O(.) will be complex, the cryptosystem will be more secure.Hence this Big O(.) represent a clear parameter for the security aspect of the cryptosystems.

-The Index of coincidence and the Kasiski parameter represent a good measure for identifying the cryptosystem if it belongs to mono-alphabetic or poly-alphabetic. This aspects will evaluate the complexity of that system. Since it well known that the poly-alphabetic cryptosystems (like Vigenere and others) are more secure than mono-alphabetic.

-The histogram calculation of the cipher text can be considered as promising step towards evaluating the security of poly-alphabetic cipher system.

- With the Modified statistical parameters, we used four , mean value, variance, skewness and kurtosis. Each one differs in it's main mathematical equation. We found that using the third order and fourth order statistical equation can provide good evaluation statistical parameters for the complexity of cipher systems.

-Unicity Distance and key Entropy parameters that belong to Information Theory, help in offering good evaluation parameters for the complexity.

Table (1) Degree of complexity measured b y O(.) parameter.

The type of the cipher algorithm	Substitution (Mono-alphabetic)	Affine (Mono-alphabetic)	Vigenere (Poly-alphabetic)	Beaufort (Poly-alphabetic)	Columnar Transposition
Number of operation n	O(n)	O(n + m)	$O(n^d)$	$O(n^d)$	O(s!)

Table (2) The relation between k and Ic.

k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**large**
Ic	0.0066	0.0520	0.0473	0.0450	0.0436	0.0427	0.0420	0.0415	0.0411	0.0408	0.038

References:

Douglas R. Stenson, 2006, " Cryptography Theory and Practice", Chapman & Hall/CRC

D. R. Hankerson ans et al., 2000," Coding Theory and Cryptography – the Essential", CRC Press

D.W. Davies and W. L. Price, 1989, " Security for computer Networks", John Wiely &Sons.

F. Piper, and H. Beker, 1982, " Cipher System: The Security of Communication", Noorthwood Books- London

J. Seberry, and J. Pieprzyk, 1989, " Cryptography- An Itroduction to Computer Security", Printice Hall .

J.Talbot, and D. Welsh, 2006, " Complexity and Cryptography- An Introduction", Cambridge University Press

" NP-Complete Problems", http://www.netaxs.com/people/nerp/automata/p-and-np

"Strength of Cryptographic Algorithms", http://math.colorado.edu/heba/crypto/strength.html

Th. Bier, and M. R. Wahiddin, 2004, " Mathematical Aspect of Classical and quantum Cryptography", Research centre, IIUM