BALANCING PROCESS TO THE CIPHERING SYSTEM SEC

[1]ZAKARIA KADDOURI, [2]FOUZIA OMARY, [3]ABDOLLAH ABOUCHOUAR AND [4]MOHSSIN DAARI.

[1,3,4] PhD.Std., Department of Computer Sciences. Faculty of Sciences, Mohammed V University-Agdal, Rabat. Morocco.

[2] Professor, Department of Computer Sciences. Faculty of Sciences, Mohammed V University-Agdal, Rabat. Morocco.

LRI Laboratory (Ex: Networks and Data Mining Laboratory)

Department of Computer Science, Faculty of Sciences, Mohammed V University-Agdal, Rabat . Morocco.

E-mail : [1]kaddouri.zakaria@gmail.com, [2]omary@fsr.ac.ma, [3]abdollah.abouchouar@gmail.com, [4]daari.mohssin@gmail.com

ABSTRACT

In this article, we present a new symmetrical encryption system based on symmetric encryption SEC, it formalizes the problem of encryption to a combinatorial optimization problem and uses evolutionary algorithms to solve it. The main objective of this work is to strengthen the resistance of the encryption system "SEC" against cryptanalysis by the characters frequencies study. The Balancing process (BP) is introduced to change the appearance frequencies of characters, and to reach equilibrium. Through the key generated by our algorithm, we illustrate the process of encryption and decryption. Examples of applications will be given at the end of this article, the results show that our new system is robust and resists to any attack by frequency analysis.

Keywords: *Symmetric Encryption, Evolutionary Algorithms, Combinatorial Optimization, SEC, Frequency Analysis.*

1. INTRODUCTION

Cryptography allows to take legible, readable data, and to transform it into unreadable data for the purpose of secure transmission. A key is used to transform it back into readable data when it reaches its destination. Only the use of a secret key can convert the cipher text back into a readable clear text.

There are two main types of cryptography:

Symmetric encryption which is the oldest cryptographic technique used to transmit or store digital data securely. Data that has been encrypted with symmetric encryption is considered to be confidential, in the sense that only entities (persons or systems) who have the key can understand what the data means.

Asymmetric encryption is also known as public-key cryptography. Asymmetric encryption differs from symmetric encryption primarily in the fact that two keys are used: one for encryption and one for decryption. The most common asymmetric encryption algorithm is RSA.

Compared to symmetric encryption, asymmetric encryption imposes a high computational burden, and tends to be much slower. Thus, it isn't typically employed to protect payload data. Instead, its major strength is its ability to establish a secure channel over a non secure medium (for example, the Internet).

This is accomplished by the exchange of public keys, which can only be used to encrypt data. The complementary private key, which is never shared, is used to decrypt.

Given that the evolutionary algorithms are stochastic and use random processes, we notice that they are useful in cryptography [6]. The symmetric encryption system SEC is based on Evolutionary algorithm whose main purpose is to establish an exchange of the appearance frequencies that belong to the different characters of the message to be encrypted as well as their own positions [7].

The main objective of this work is to strengthen the resistance of the encryption system SEC. That is why, a new concept called "balancing" was introduced to change the appearance frequencies of the text characters to be encrypted and to reach an equilibrium. Our system will generate another secret key that will increase the size of the symmetric key, and will strengthen the resistance of our new system to attack by frequency analysis and exhaustive search.

2. EVOLUTIONARY ALGORITHMS

2.1 Definition

Basically, the metaheuristics consist of creating the evolution of a baseline configuration by replacing it repeatedly by a new configuration chosen in its neighborhood.

The evolutionary algorithm is a kind of the metaheuristics [5]. It is useful for optimization when other techniques such as gradient descent or direct, analytical discovery are not possible. It incorporates aspects of natural selection or survival of the fittest and maintains a population of structures (usually randomly generated initially), that evolves according to rules of selection, recombination, mutation and survival, referred to as genetic operators. A shared "environment" determines the fitness or performance of each individual in the population [7]. The fittest individuals are more likely to be selected for reproduction (retention or duplication), while recombination and mutation modify those individuals, yielding potentially superior ones [5].

We present the mechanisms involved in evolutionary algorithms in the flowchart in Figure1.

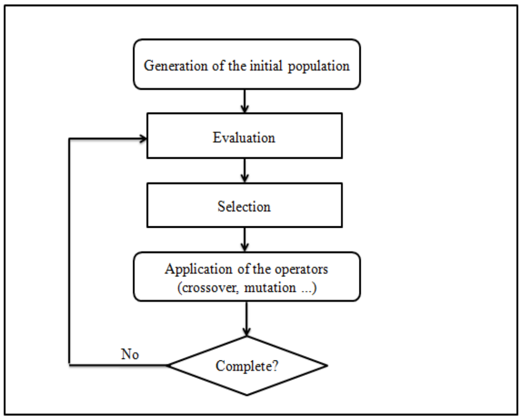

Generation of the initial population
Evaluation
Selection
Application of the operators (crossover, mutation ...)
No
Complete?

Figure 1: Flowchart Of An Evolutionary Algorithm

In precisely, the contribution of our work consists of the design and implementation of symmetric encryption systems with EA as basic tools.

2.2 Algorithm

An evolutionary algorithm generates an initial population P of μ individuals, and then makes the population P (generations) evolves following a repeated pattern.

- <u>For every generation:</u>

1) Select for reproduction: is chosen in the current population P, λ individuals which become parents (mating pool P').

2) Vary operators: by applying the crossover and mutation on the individuals of P, a population P" of λ children is obtained.

3) Evaluate the performance of elements of P" (children).

4) Select for survival: the elements of P (current population) and P" (children) the μ individuals are chosen among these elements which will generate the population P of the next generation.

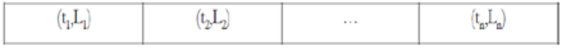
3. DESCRIPTION OF OUR SYSTEM

The Balancing Process (BP) consists of creating equilibrium between the lists of the characters positions that represent the genes of the system. The resistance of the system against attack by frequency analysis will be stronger and the system will become more robust.

We apply the BP on the characters of the plaintext, this process will generate an additional secret key which we call the balancing key, then we apply the SEC.

3.1 Formalisation of the Problem

Let T be a continuation of k characters and $t_1,t_2,…, t_n$ its different characters. Denote by $L_i(1< i< n)$ the list of the different positions of t_i in T before the ciphering. $L_1,L_2,…L_n$ is a partition of the set {1,2,…,n}. T can be represented by vector:



(t_1,L_1)	(t_2,L_2)	…	(t_n,L_n)

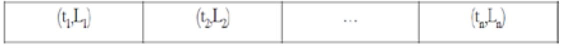
The goal of our works is to alter to the maximum the frequencies of apparition of the characters in the text T and establish more mess in their positions.

3.2 Our Ciphering System

This system is composed of two principal parts: the balancing and the application of evolutionary algorithm as descript in figure 2.

3.2.1 First part: Balancing

The Message M Before The Balance Can Be Represented By The Vector Below:



(t_1,L_1)	(t_2,L_2)	…	(t_n,L_n)

We sort the set of the lists L_1, L_2, …, L_n according to their sizes in the decreasing order then we divide it in three subsets of sizes near to [n/3] (floor of n/3), named respectively : E_L, E_m, E_p.

Let us indicate respectively by N_L, N_m and N_p the cardinals of E_L, E_m and E_p, and by S_k the desirable key size.

The objective of this algorithm is to balance the lists so that all resulting lists have almost the same cardinal.

While the key size is not exceeded:

- The balancing will be applied to E_L U E_p as follows:

o Let us indicate initially the lists which we need to balance. These lists are composed of $L_{P1}, L_{L1},\dots, L_{Pf}, L_{Lf}$ such as f is taken randomly in $\{1,2,\dots, \min(N_m,N_p)\}$, $L_{p1}, L_{p2},\dots, L_{pf}$ are selected in E_p in increasing order of their size, in alternation with $L_{L1}, L_{L2},\dots, L_{Lf}$ which are taken in E_L in the same order, while taking account of the following iterative processing:

- $r \leftarrow 1$; $E \leftarrow \emptyset$
- repeat

$E \leftarrow E \cup L_{pr} \cup L_{Lr}$
If size of $E \leq S_k$ the $r \leftarrow r+1$
Until r=f or size of E is bigger than S_k

- $f \leftarrow r$

We next equilibrate the smaller list of E_p with the largest list of E_l. After the balancing process, these two lists will be deleted from E_p and E_l. The balancing process is again applied to the smaller list of E_p with the largest list of E_l, which will be deleted too, and so on…

o Applying balancing to the lists above as follows:

- Let us take randomly a list l_m in E_m and compute its cardinality N_m=card (L_m)

- Let L_p the list to balance with L_l

Suppose that $L_l = \{p_1, p_2, \dots, p_m\}$,

Let us take randomly an index r in {1,2, ..., m}

Let N_p=card (L_p);
While card(L_p) < N_m .
Put p_k in the list L_p:
➔ $L_p \leftarrow p_r$, r+1 ;
Add the quadruplet:
($[L_l]$; $[L_p]$;N_p ; r) to the balancing key.

o Repeat the process of balancing on:
-$E_p \leftarrow E_p$-{ L_p}.
-$E_l \leftarrow E_l$-{ L_l}.

Until reaching the size of the desired key S_k.

- If we indicate by FO the number of balancing applied then the generated key representative of these operations is a set of FO quadruplet of the from:
- $([L^i l], [L^i p], N^i p, r^i)$

The new ciphered text, T_f, will be denoted by the following:

(c_{f_1}, L_{f_1})	(c_{f_2}, L_{f_2})	……	(c_{f_m}, L_{f_m})

We present the mechanisms involved in balancing process in the flowchart in Figure 3.

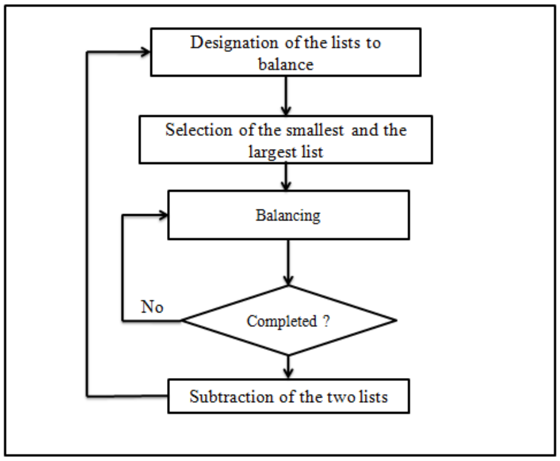

Designation of the lists to balance
Selection of the smallest and the largest list
Balancing
No
Completed ?
Subtraction of the two lists

Figure 3: Flowchart of the Balancing Process

3.2.2 Second part: Application of Our Evolutionary Algorithm SEC

Let M be the message to be encrypted. M is a sequence of n characters. This message can be formed only of numbers, such as bank codes or a mixture of numbers and literary phrases including punctuation.

Let $c_1, c_2,\dots, c_m$ be the different characters of M. Denote by $L_i (1 \leq i \leq m)$ the list of different positions of the character ci in M before the ciphering and card(L_i) the number of occurrences of ci in M.

We notice that: $L_i \cap L_j = ø$ si $i \neq j$, $\forall i, j \in \{1, 2, \dots, m\}$.

Then $L_1, L_2,\dots, L_m$ is a partition of the set {1, 2, …, n}.

The message M can be represented by the vector below:

(c_1,L_1)	(c_2,L_2)	……	(c_m,L_m)

The goal of SEC is to change the appearance frequencies of the characters at a maximum level in the message M and to establish the maximum disorders over their positions. In order to make it happen, we must repeatedly change the distribution lists over the different characters of M. In other words, we must find a permutation σ of {1, 2, …, m} such that the difference between the cardinal of the new list $L_{\sigma(i)}$ assigned to the character ci and the cardinal of the initial list L_i reaches the maximum level. In this case, we are confronted with a combinatorial optimization problem. Nevertheless, the evolutionary algorithms are very effective in this kind of problem. So, they will be applied to permutations' problems [1] [10].

SEC Algorithm :

*1)*Coding
Use an individual (or chromosome) as a vector of size m.

Genes are the lists $L_{pi}(1 \leq i \leq m)$.

L_{pi} is the i^{th} gene which contains the new positions that will take the character c_i.

*2)*Initialization
Creation of the initial population P_0 consists of q individuals: $X_1, X_2,…,X_q$.

Let Original-Ch be the chromosome which genes are $L_1,L_2,…,$ and L_m lists (placed in this order).These lists represent the message before the application of the algorithm. We apply permutations on Original-Ch in order t get an initial population formed by q potential solutions.

Set i :=0;

*3)*Evaluation of individuals
Let X_j is an individual of P_i whose genes are: $L_{j1},L_{j2},…,L_{jm}$.

The evaluation function F is defined on the set of individuals X_j by:

$$F(X_j) = \sum_{i=1}^{n} \text{card}(Lji) - \text{card}(Li)$$

*4)*Selection of the best individuals
The conventional method of the roulette wheel retains the strongest individuals. A Control function is introduced to eliminate the individuals in whom the value of only a minority of genes have changed in comparison with the initial chromosome: Original-Ch [6].

Since this problem is narrowed to a permutation problem with constraints, the genetic operators have been adjusted to this kind of problems

*5)*Crossover MPX(Maximal Preservative X)
This cross is applied to selected individuals with a very precise rate. The best rate is about 60% to 100% [7].

*6)*Transposition Mutation
Choosing the mutation consists of randomly swaping two genes of a chromosome. This operator is applied to individuals from crossing with an appropriate rate, preferably from 0.1% to 5% [7].

Place new offspring in a new population P_{i+1}.

Repeat steps 2, 3 and 4 until a stopping condition.

7) Stopping condition
The function F is bounded because $0\leq F(X) \leq 2*m$, for each individual X.

In fact, the function F has a maximum since it is bounded. According to some researches, the convergence result of fitness function is made but it can be a value close to Max, which can be experimentally determined. Final-Ch denotes the final solution given by our evolutionary algorithm. Final-Ch denotes the final solution given by our evolutionary algorithm. From Original-Ch and Final-Ch, the symmetric key is constructed. This key is called a genetic key [1].

3.3 Deciphering

Decryption is performed in two basic steps:

Step One: decryption by using the genetic key, can find the original message T' from the message T''.

Second step: decryption by using the balancing key allows to find the message T from the encrypted message T'.

- First decrypion

We represent the encoded text T' by a vector of list. Let's by C'_1, C'_2, …, C'_m the different characters of T' and by $L'_1,L'_2,…, L'_m$ their respective lists of positions. Thanks to the genetic key the characters are going to recover their lists of corresponding positions in the text T_f obtained after the first part of ciphering [7].
Indeed, the key can be represented by a vector, that we denote Key, of size m such that:
$key(1)=p_1,key(2)=p_2,…, key(i)=1, …, key (m)=p_m$ where:

The character C'_{p1} is going to be associated to the list L'_{1}.

The character C'_{p2} is going to be associated to the list L'_{2}.

…

The character C'_{m1} is going to be associated to the list L'_{m}.

Thus we get the text T_f.

- Second decryption

Thanks to the balancing key which is clear and direct, For each element ([L^il], [L^ip], N^ip, r^i) of balancing key, we can immediately reconstruct the true positions of lists L_{fi} of each character. Thus we obtain initial text T.

4. EXPERIMENTS RESULTS

4.1. Convergence

To illustrate the performance of the new approach of SEC with the balancing process compared to the system SEC, we tested our program on multiple messages with different sizes. For each message we compared the differences of the evaluations, which determine the convergence of the system.

The following table lists the results:

Table 1: Deviation Values Of Individuals From Each Population In The Both Systems.

Iteration	Deviation of Sec	Deviation of SEC with Balancing
1	39322	24919
2	40541	26778
3	44585	20928
4	40126	24814
5	40334	21064
6	42950	21948
7	41633	20052
8	39173	22390
9	38332	19320
10	34669	18761
11	33218	17408
12	37030	12238
13	32479	13034
14	29521	14137
15	26832	10732
16	28648	13827
17	26171	11670
18	22274	14442
19	24766	11638
20	24938	13009
21	23869	10220
22	24262	12950
23	22470	10414
24	21023	11624
25	18126	9856
26	21026	13080
27	18890	12552
28	19158	11320
29	18087	11322
30	16430	10083
31	18424	10784
32	13216	11088
33	14018	10322
34	13060	11019
35	12556	9552
36	15688	8088
37	16400	6160
38	15994	7392
39	8997	6160
40	8838	4928
41	8370	5060
42	6310	4428
43	5388	1232
44	6874	1206
45	7761	1193
46	5502	1110
47	6866	1051
48	5640	1002
49	5532	0
50	3980	-
51	3377	-
52	2850	-
53	2119	-
54	3360	-
	…	
74	2216	-
75	1192	-
76	2384	-
77	1360	-
78	2284	-
79	1240	-
80	1240	-
81	1788	-
82	1668	-
83	1348	-
84	784	-
85	860	-
86	384	-
87	540	-
88	640	-
89	188	-
90	268	-
91	36	-

The figure bellow represents a graph of the different values of the differences between the individuals of every generation.

We can notice that the system SEC with balancing reaches total convergence after only 49 iterations. Otherwise, without balancing process, the system SEC stops at the 91st repetition with an optimum value of 36

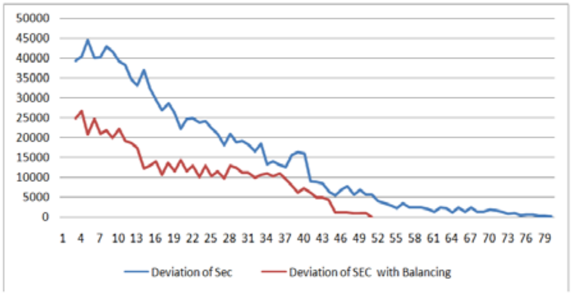


Figure 4: Comparison Of The Deviation Of The System SEC With And Without Balancing Process

4.2. Comparison Of The Frequencies Analysis

The comparison of the frequencies analysis is the main indicator of new system's performance. The table and the figure below compare the analysis of frequencies that belong to an unencrypted and encrypted text by the new system and the encrypted text by SEC.

Table 2 : The Frequency Analysis In The New And The Old System SEC

Frequency analysis in the plain text	Frequency analysis in the text Encoding without BP	Frequency analysis in the text Encoding with BP
102	102	56
97	97	55
91	91	54
90	90	52
89	89	50
80	80	49
79	79	43
77	77	42
65	65	42
62	62	40
44	44	39
36	36	38
32	32	37
29	29	37
27	27	35
26	26	35
24	24	33
21	21	31
20	20	31
20	20	30
20	20	30
20	20	29
15	15	29
15	15	28
15	15	28
14	14	28
14	14	28
13	13	27
11	11	27
11	11	26
11	11	26
11	11	25
10	10	25
10	10	24
8	8	21
8	8	21
7	7	21
4	4	20
3	3	20
1	1	20

According to this comparison, the changes of frequencies appearing in the system using BP are now obvious.

Figure 5. shows a graphical representation of the apparition frequencies in the plaintext, the ciphertext using the BP and the ciphertext without BP.

5. CONCLUSION

Through this work, we were able to achieve an encryption system which consists of defining a new algorithm based on the balancing of the initial lists of the SEC system.

Our system generates another secret key that we call "key balancing" which reinforces the genetic key.

From the results of the occurrences frequencies of the characters obtained, we have proved the robustness of the new system against any attack by the frequencies analysis.

REFERENCES

[1] F.Omary, A.Tragha, A.Lbekkouri, A.Bellaachia, A.Mouloudi « An Evolutionist Algorithm to Cryptography». Brill Academic Publishers – Lecture SeriesAnd Computational Sciences Volume 4, 2005, pp.1749-1752

[2] F.Omary, A.Tragha, A.Lbekkouri, A.Bellaachia, A.Mouloudi « A New Ciphering Method Associated with Evolutionary Algorithm». Lecture Notes in Computer Science – Publisher : Springer Berlin / Heidelberg –ISSN: 0302-9743 –Subject :Computer Science-Volume 3984/ 2006.

[3] F.omary, A.Tragha, A.Bellaachia, A.Mouloudi. « Design and Evaluation of Two Symmetrical Evolutionist-Based Ciphering Algorithms ». International Journal of Computer Science and Network Security (IJCSNS) February 28, 2007 pp 181-190.

[4] Hans DelfsetHelmut Knebl, « Introduction to Cryptography :Principles and Applications».

[5] Gareth Jones , « Genetic and Evolutionary Algorithms ». University of Sheffield, UK

[6] Menezes A.J., Oorschot, P.C. van et Vanstone S.A., « Handbook of Applied Cryptography».(CRC Press, 1997).

[7] Thesis F.Omary, « Applications des algorithmes évolutionnistes à la cryptographie». University of science- Rabat 2006.

[8] Goldberg D.E, Genetic Algorithms in Search, Optimisation& Machine Learning. Addison-Wesley Publishing Company,Inc,1989.

[9] Florin G. et Natkin S.les techniques de la cryptographie.CNAM 2002.

[10] A.Mouloudi, F.Omary, A.Tragha, A.Bellaachia « An Extension of evolutionary Ciphering System». 2006 International Conference on Hibrid Information Tchnology,Novembre 9th – 11th,2006.

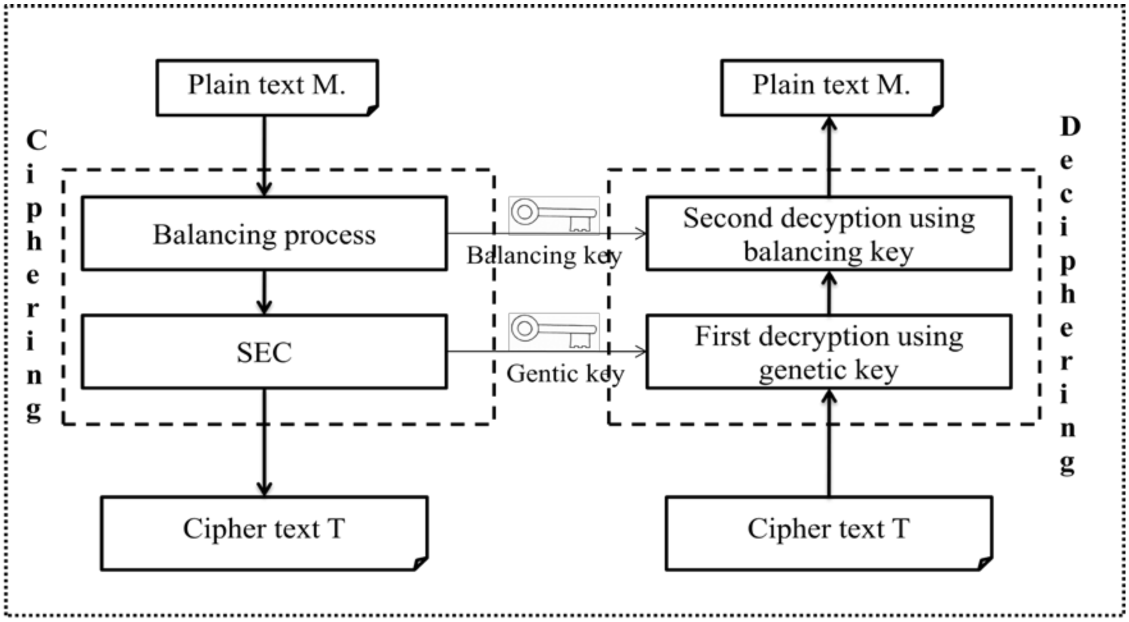


Figure 2: Diagram Of Our Encryption System

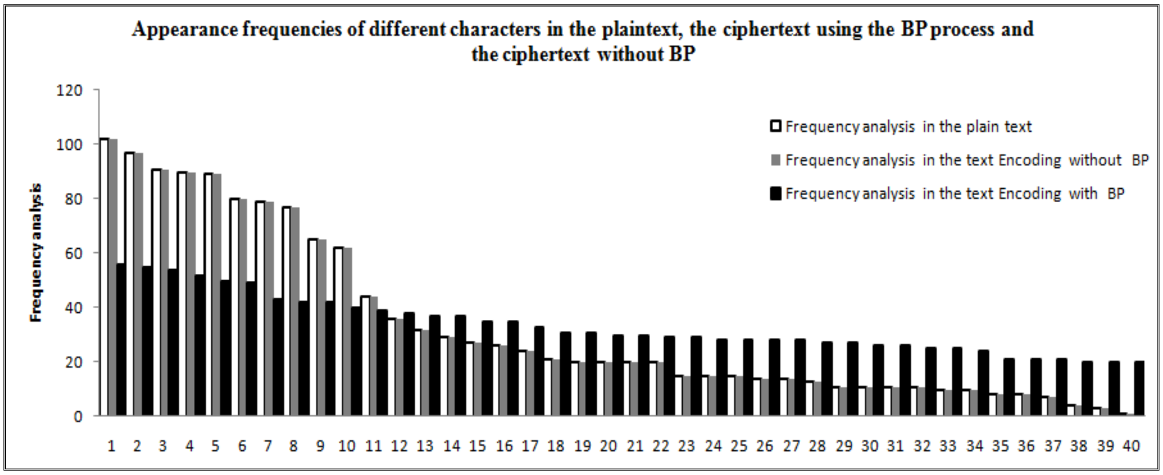


Figure 5: Graphical Representation Of The Appearance Frequencies Of Different Characters In The Plaintext, The Ciphertext Using The BP Process And The Ciphertext Without BP.