Genetic Regenerator of Pseudo-Random Sequences R.A NMJ

Abderrahim Sabour[1], **Ahmed Asimi**[2] **and Aboubakr Lbekkouri**[1]

[1]*Department of Mathematics and Computer Science University Mohammed V, Rabat, B.P 14106, Morocco*
[2]*Department of Mathematics and Computer Science University Ibn Zohr Agadir, B.P 8106, Morocco*

Summary
In this article, we propose a new regenerator of pseudo-random sequences R.A NMJ, based on evolutionist algorithms, in that case the genetic algorithms, in order to regenerate pseudo-random sequences cryptographically reliable. R.A NMJ is an algorithm that simulates the evolution, via dissipative functions, of a population of individuals created from a password. Each individual evolves since its creation in a dynamic manner, thus generating a chaotic behaviour. The algorithm R.A NMJ will create at the time of process I the initial population from a password, which will evolve at the time of process II. Process III, that is iterative will regenerate at each iteration a binary sequence where all individuals of the population will contribute; a sequence is combined with the clear message, thus generating the cipher. The addressee, equipped with a similar generator deciphers the received message while applying the same process. For the evaluation of the statistical characteristics of regenerated sequences by R.A NMJ, we have used the software NIST Statistical Test Suite, for many passwords and which lead to nice results.

Key words: *evolutionist algorithms, pseudo-random sequences, chaotic behavior.*

1. INTRODUCTION

The flood encryption systems [1], [2], [3] are cryptosystems with secret key [4]. Their running process is relatively simple: a pseudo random generator [5], [6], [7] creates a binary sequence is combined with the plaintext, thus generating the cryptogram. In this article, we propose a new regenerator of pseudo-random sequences R.A NMJ, based on evolutionist algorithms, in that case the genetic algorithms [8], [9] which are generally used to obtain in a right time approximated solutions to difficult optimization problems, for which we do not know more efficient classical methods. However R.A NMJ [11], [12], [13], [14], [15] does not have anything to do with these optimization problems, since it does not have any function to optimize (evaluation function of individuals), which makes all individuals having the same fitness. This forced the definition of a selection method without fitness that we call order function whose role is to; produce a new distribution of individuals of the population. In this approach, the individual is composed of two blocs: Data block and Control block. The reproduction of block data of the individuals of the population is made via mating functions I and II, the first is used at the time of process II and at each iteration, allows to double the size of block data of all individuals, the mating function II operates at the time of process II, on less than 5% of the individuals of the population and does not change the size of block data. The basic idea of this approach is to exploit the complexity of evolutionary systems, by defining new dissipative functions. The first obstacle was to assign an initial population to an arbitrary password while keeping a high avalanche effect, task that forced the introduction of a class of two or three states functions called observer. This latter allows non injective binary transforms (Lemma) that assigns to each sequence another sequence, the arrival sequence is an irreversible projection of the starting sequence. This transform is used to create the initial population, at the time of the mating and at the time of the contribution of individuals. The study of the statistical characteristics of the regenerated sequences from passwords was done by software NIST Statistical Test Suite. R.A NMJ is composed of three main processes:

The first creates the initial population of N1 individuals of size T1. The second process, iterative, supplied with the initial population will evolve this latter while doubling at each iteration the size of the block data of all individuals.
The third initialized by mature population, regenerates a pseudo-random sequence.

2- NOTATIONS AND DEFINITIONS

From now on, we designate by:
PW : Password regarded as binary sequence, its size may vary from 1 to 100 characters.
$|n:m|$: the set of integers between n and m with $n < m$.

Manuscript received January 5, 2007
Manuscript revised January 25, 2007

N : the set of integers.

$Nd(P)$: the number of individuals of the population P.

$L(s)$: the length of the state sequence s. If s is a binary sequence (i.e. two sates sequence 0 and 1). $L(s)$ is the number of 0 and 1 that constitutes s. If s is three states sequence α, β and λ, $L(s)$ is the number of α, β and λ that appears in s. The function defined from N to $\{0,1\}$, respectively in $\{\alpha,\beta,\lambda\}$ is called two states function, respectively three states. And the set of these functions denoted F_2, respectively F_3. For each two states or three states function we assign a unique sequence f defined by : f=F(0)F(1)F(2)F(3)F(4)……..F(n)…. And if there exists an integer k such that f=F(0)F(1)F(2)F(3)F(4)…F(k)

F(0)F(1)F(2)F(3)F(4)…F(k)…, we say that F is periodic with period F(0)F(1)F(2)F(3)F(4)…F(k), called primitive signal f, that we denote $Sp(\mathrm{F})$, therefore for every integer $n \in N$ F(n)=F(n mod $Sp(\mathrm{F})$). And if f is a finite sequence, we extend it to a unique infinite periodic sequence whose size is a multiple of that of f. This shows that we have a bijection between the set of two or three states periodic functions and the set of finite sequences and the periodic infinite sequences.

Γ_2 : the set of two states functions 0 and 1 whose assigned sequences are periodic.

Γ_3 : the set of three states functions whose assigned sequences are periodic. We define a subset Ω of Γ_3 by: G is an element of Ω if and only if $Sp(G) = \alpha...\alpha\beta...\beta\lambda...\lambda\beta...\beta$.

If we denote by :

X : the frequency of α in $Sp(G)$.

Y : the frequency of β between α and λ in $Sp(G)$.

Z : the frequency of λ in $Sp(G)$.

T : the frequency of β in $Sp(G)$ following λ.

Then G will be denoted $G = [X, Y, Z, T]$.

Definition 1. An observer G is a three states periodic function such that $G(0) = \alpha$ and $G(n) = G(n \bmod L(Sp(G)))$ for all $n \in N$. In other words Ω is the set of observers.

Definition 2. An individual is composed by two blocks: Data block and Control block.

The Control block is composed by four sub blocks:

1- *Mating sub-block*: its value indicates the parameter values X1, Y1, X2, Y2, SigX2, SsX, BI and BC, used by the mating function, and whose domains are :

$D_{X1} = D_{X2} = |1:4|$ and $D_{Y1} = D_{Y2} = |1:4|$;

$D_{SigX2} = |0:1|$; $D_{SsX} = |0:1|$;

$D_{BI} = |0:1|$ lay out bit;

$D_{BC} = |0:1|$ caliber bit.

2- *Priority sub-block*: its value permits to set a priority between the individuals at the time of the order function or the mating function. Its value changes after each competition thus allowing a dynamic behavior for the individuals.

3- *Contribution sub-block*: This sub block determines the values of the parameters X, Y, Z, T and the direction of the circuit of the Data block of the individual I in question (Ss), used by the contribution function and whose allowed values are : $D_X = |5:8|$; $D_Y = |1:4|$; $D_Z = |5:8|$; $D_T = |1:4|$ and $D_{Ss} = |0:1|$.

4- *Age sub-block*: enumerates the frequency of iterations of process III that an individual has carried out without a change in its Data block, its value intervene at the time of the mating since it is the older individual who will be reproduced. At the time of process III the transition from the current population to the following population is a generation.

3. THE ALGORITHM R.ANMJ

The principle of the algorithm is described by the following diagram:

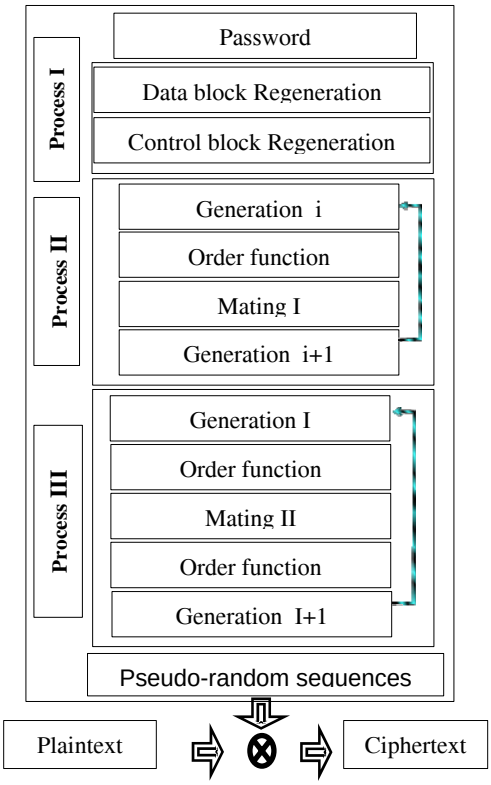

Fig. 1 Tabular general of the R.A NMJ

R.ANMJ is composed by three processes, the first will create the initial population, the second will make her evolve, and the third will allow the regeneration of pseudo-random sequences that will be combined with the messages to be ciphered or deciphered.

3-1 Process I

The objective of this process is the regeneration of the individuals of the initial population starting from a given password, which is done in two steps:

A. Regeneration of Data blocks of the individuals.

B. Regeneration of Control blocks of the individuals.

A. Data block Regeneration of initial population individuals

Data block Regeneration of initial population individuals is associated to a given password is done while applying to this latter the notion of observer. To handle long passwords, we must add to these observers the position (Pos) which allows them to point differently at the password, and the direction of the circuit (Sm), hence we handle passwords of more than 100 characters. This allows to redefine the observer G in the following way:

$$G = [X, Y, Z, T, Sm, Pos].$$

Where:

X : the frequency number of α in $Sp(G)$

Y : the frequency number of β between α and λ in $Sp(G)$.

Z : the frequency number of λ in $Sp(G)$.

T : the frequency number of β in $Sp(G)$ following λ.

Sm : denotes the direction of the circuit of the password.

$Pos(k)$: denotes the bit index of the starting of the processing to the kth individual. It takes values from, $|0 : L(PW)-1|$ initialized to 0. And that we define in the following manner.

$$Pos(0) = 0$$

$$Pos(k+1) = (Pos(k) + (L(PW)*17) \bmod (X*Y + Z*T)) \bmod L(PW)$$

for all $k \geq 1$.

Remark 1. The range of values, denoted Di, that X, Y, Z and T is variable and it is left to the designers to choose their suitable values. The original version of R.ANMJ uses a class of three states functions defined by :

$D_X = |2 : 7|$, $D_Y = |1 : 6|$, $D_Z = |2 : 7|$ and $D_T = |1 : 5|$

With 16 characters as the size of the Data block of the individuals. The previous choice will regenerate a population of 2160 individuals, since the values of D_X, D_Y, D_Z and D_T determine the number of the individuals of the population valued at 1080, and since this version runs at every state the password in both direction, this implies that the population will be doubled, which will give 2160 individuals.

Definition 3. Let v be the transform defined by:

$$v : \{\alpha, \lambda\} \times \{0,1\} \to \{0,1\}$$
$$(\alpha, k) \mapsto k$$
$$(\lambda, k) \mapsto \bar{k}$$

with $\bar{0} = 1$ and $\bar{1} = 0$.

Definition 4. The function W :

$$\Omega \times \Gamma_2 \to \Gamma_2$$
$$(G, S) \to W(G, S) = S',$$

Allows the transit, for a given observer G , from a password (binary sequence associated to S) to the corresponding Data block, by picking only the first T1 bits of the sequence associated to S'.

To determine the value of S' , we construct two sequences $(t_i)_i$ and $(k_i)_i$ through:

$$t_0 = \inf\{n \in N / G(n) = \beta\}$$
$$k_0 = \inf\{n \in N / G(n) \neq \beta\}$$

We have two cases: $t_0 > k_0$ or $t_0 < k_0$, in other words $G(0) = \beta$ or $G(0) \neq \beta$. And since $G \in \Omega$, then $G(0) = \alpha$, therefore $t_0 > k_0$ and $k_0 = 0$.

And we define

$$t_{i+1} = \inf\{n \in N / n > k_i; G(n) = \beta\}$$
$$k_{i+1} = \inf\{n \in N / n > t_i; G(n) \neq \beta\}$$

The function S' is defined by:

$$S'(n) = v(G(m), S(m))$$

with $m = k_r + d$, $n = d + \sum_{i=0}^{r}(t_i - k_i)$ and $0 \le d \le t_{r+1} - k_{r+1} - 1$.

Lemma. The function W is not injective.

Proof . Let S_1 and S_2 be two elements of Γ_2 such that:
$Sp(S_1)$ =11000110011000111101010110110111011011000011001100110010001111,
$Sp(S_2)$ =11010110101001111111010000101011011011010001110011001100101010

And let G be an observer defined by $G = [3,2,2,3,0,0]$.
And since

$Sp(W(G,S_1)) = Sp(W(G,S_2))$=11000100000101011 0100000100110;

therefore $W(G,S_2) = W(G,S_2)$. This shows that the function W is not injective. And we check that there exists at least $2^{E\left[\left(\frac{Y+T}{X+Y+Z+T}\right)*L(S_1)\right]}$ sequences S_i such that $W(G,S_1) = W(G,S_i)$, which will give in that case about 2^{30} sequences S_i . With $E[A]$ denoted the integer part of A and $G = [X,Y,Z,T]$.

B. Control block Regeneration of initial population individuals

The Control block of an individual is determined from its Data block, while starting the reading from the first octet. To guarantee a dynamic behavior to the individuals, we recount the Control value block, except the age sub-block that remains unchanged. Following each competition we determine the octet to be pointed from the current state in order to determine the new Control block.

3.2 Process II

This process evolves the population of individuals created at the time of process I. This process is iterative and allows to obtain after k iterations, of a stronger population (ideal distribution for the possible states) based on two functions (the order function and the mating function).

A. Order function

The mechanism of this function is as follows:
We consider a population I_1, I_2…I_N de N individuals ranked in this order; and we compare the priority block of I_1 and I_2. The winner (the person having more priority otherwise the first) dictate the direction of the movement and the number of paces over the looser who will take the new released position after the shift of intermediary persons. And to make fuzzy the current state of the population, we redefine the Control block of two persons in competition. Next we move to the following couple. This goes on N/2 times.

B. Mating Function 1

The mating function I allows at each iteration of process Il, doubling the size of Data blocks of all individuals of the population. Each person owns a mating sub block in its control block, which permits to determine the way of make the mating : Each one of the two persons contributes to the evolution of his conjoint while applying his mating function, and accepting the transformation imposed by the other.

At this step of their evolution, the contribution gauging of both persons is primordial, one recalculate the values X1, Y1, X2 and Y2 in the following way:

- T = X1+Y1+X2+Y2
- X1= min (X1, X2)
- Y1= min (Y1, Y2)
- X2= T/2 – X1
- Y2= T/2 – Y1
-

Let's apply this rule on the following operators:

1. $[3, 4,5, 8] \rightarrow [3,4,7,6]$
2. $[4,5,2,4] \rightarrow [2,4,5,3]$
3. $[2,3,4,5] \rightarrow [2,3,5,4]$
4.

These values will be the new parameters of the mating function. The mating function allows doubling the size of the Data block, at each iteration of process Il (the number of iterations is fixed by the designer).

3.3 Process III

The task of the first two processes was to choose N2 (N2= 2160) persons having Data block of T2 bits (T2=1024 bits). In other words, it concerns choosing N2 individuals among $2^{T2}\left(2^{1024} = 1.79 \cdot 10^{308}\right)$, knowing that the two processes I and II have allowed:

- Widening the spectre of Data blocks of the individuals of the population.
- An acceptable non correlation rate between the persons.

The real challenge that process III must face starting from a given population is:

- Regenerate a binary sequence while making fuzzy the contribution of each person of a given generation, which makes fuzzy the contribution of a given generation, via the contribution function.
- Keep the non correlation of the persons forming the populations who succeeded, via the mating function II.
- Regenerate a pseudo random sequence cryptographically reliable.

A. Mating function II

The mating functions II, at each iteration of process III, evolve Data blocks of some individuals, it operates on less than 5% of the individuals of the population and does not change the size of Data block. The process changes a little bit regarded to the one displayed in process II, since the two persons in competition obey to the following rules:

R1: Two persons in competition, the younger always wins.
R2: If two persons in competition have the same age, we apply the same rules as in the case of the order function.
R3: The age of the person is coded on 4 octets. At the time of he mating function, the loser initialize his age at 0.
R4: The age of the person is incremented by 1 at the time of each contribution to the binary sequence.

Rules **R1**, **R2**, **R3**, and **R4** allow to make more difficult the mating function introduced in process II. The function remains the same for the mating process.

B. Contribution function

At this point, ail individuals of the population will contribute, one by one, while referring to their contribution sub block which will allow them to determine the parameters of the contribution function; function that calls on the notion of observer to remedy for the problem of the fixed size of the individuals (the number and the size of the persons are finite, and to make it useful these two information call on the observer.

Example:

Take the following operator: $G = [X, Y, Z, T, Ss, Posi]$

Domains: $D_X = D_Z = |5:8|$ and $D_Y = D_T = |1:4|$

Ss : Circuit direction of the Data block of the individual.

$Posi$ **:** Denotes the bit index of the starting determined from the priority sub block.

We have $4*4*4*4*2*2^{L(Data)} = 512*2^{L(Data)}$ ways of contributions for an individual at the time of the elaboration of the mask $(L(Data) \geq 1024)$, with about 60 different sizes almost likely equally. The size with what contributes a given person varies, in the case of this operator, from 55.55% to 88.88% in a likely equally manner. This allows to make fuzzy the contribution of our individuals. The same probability as well as the remaining 10% in the worst of the hidden cases, make quasi impossible the deduction of the generation in question. We repeat process III as many times as necessary.

4. R.ANMJ STATISICALTESTS

For these tests we have regenerated each of the passwords, in this case 'a', 'R.A NMJ', 'AhmedASIMI', 'ALBEKKOURI', 'Abd.SABOUR' and 'Université IBN ZOHR Faculté des Sciences Département de Mathématiques et d'Informatique Agadir Maroc', a sequence from which we removed the first 300 blocks of 1000000 bits each. Which makes more than 150 iterations of process III for a population of 2160 persons with a size of 1024 bits, and we obtained interesting results.

As example, in the three following graphics, we present respectively the results of the three principal tests at the time of study of pseudo-random sequences which are:

- *Linear Complexity Test*
- *Maurer's Universal Statistical Test*
- *Discrete Fourier Transform (Spectral) Test.*

And we have used for these tests the password 'R.A NMJ'. In practice, the reason that statistical hypothesis testing works is that the reference distribution and the critical value are dependent and generated under a tentative assumption of randomness. If the randomness assumption is, in fact, true for the data at hand, then the resulting calculated test statistic value on the data will have a very low probability (e.g., 0.0001) [10] of exceeding the critical value .This value will be presented in figures 2,3 and 4 by the line (C.V).

Linear Complexity Test

The focus of this test is the length of a generating feedback register. The purpose of this test is to determine whether or not the sequence is complex enough to be considered random. Random sequences are characterized by a longer feedback register. A short feedback register implies non-randomness.

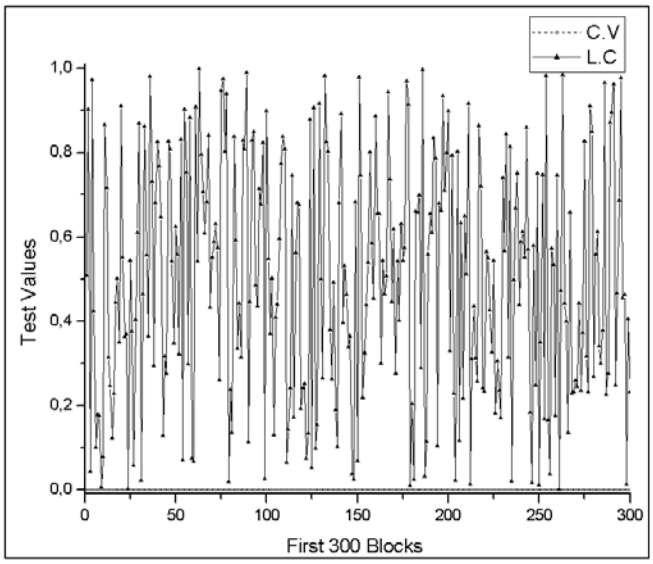

Fig. 2 Result of Linear Complexity Test

Maurer's Universal Statistical Test

The focus of this test is the number of bits between matching patterns. The purpose of the test is to detect whether or not the sequence can be significantly compressed without loss of information. An overly compressible sequence is considered to be non-random.

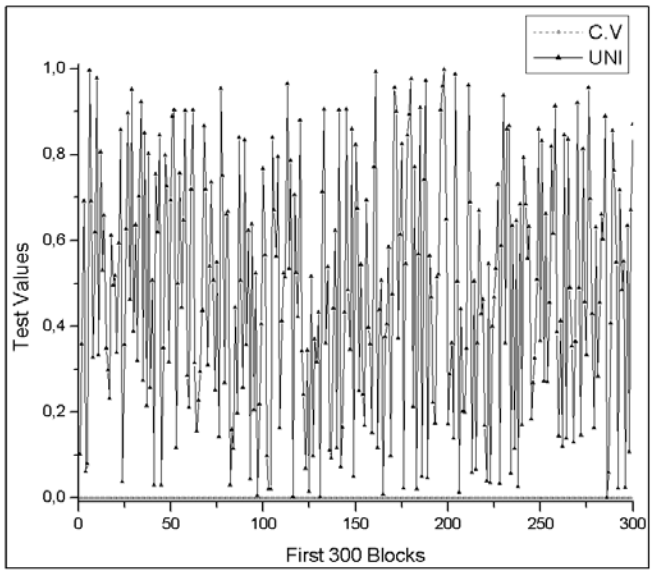

Fig. 3 Result of Maurer's Universal Statistical Test

Discrete Fourier Transform (Spectral) Test

The focus of this test is the peak heights in the discrete Fast Fourier Transform. The purpose of this test is to detect periodic features (i.e., repetitive patterns that are near each other) in the tested sequence that

would indicate a deviation from the assumption of randomness.

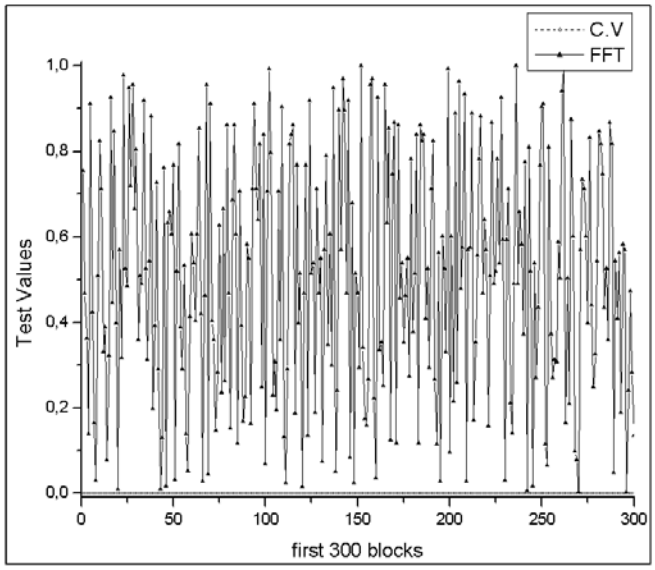

Fig. 4 Result of Discrete Fourier Transform Test

In conclusion, for this part we note that the response of the system towards the different tests are positive, moreover this response follows an evolution which requires a self study.

5. SOME R.ANMJ CHARACTERISTICS

- Only R.A NMJ initialization, which is the password, is secret and governs completely the regenerator output: it constitutes then the secret key of the cipher and decipher system.
- With regard to the speed of this generator, it is true that it will require initialization timing, process I&II, but this time could be quickly caught, since the number of calls of the mating functions diminish without damaging the quality of the generated sequences. The algorithm structure is adapted to the parallel implementation, which makes possible to reach the highest debits.
- R.A NMJ adopts a local safety solution, having as unique objective to make from the authentication of each individual a new problem.
- Even though when it concerns an iterative process, we cannot talk about a block, since the individuals contribution at a given iteration is non uniform, because every person decides the process to be followed while referring to its Control block. Moreover, the contribution function exploits the observer notion in order to increase the contribution size of the individuals while making out of their authentication a problem. This problem gets complicated since lifespan of a person depends on his position in every generation of the mating.
- The objective is to regenerate arbitrary segments, while making the liaisons between segments difficult to detect.
- Finally, we should point out the flexibility of the technique which leaves the choice to the designers as regard to the number of persons and the limits of Control block parameters. ln other words, modifying one of its parameters regenerate a different sequence, while keeping the chaotic behaviour of the system.

REFERENCES

[1] B. Schneir, Applied Cryptography. Wiley 1996, Second Edition.

[2] C. Shannon, A mathematical theory of communication, Bell system Journal, vol. 27 pp. 379-423 part I, 1948.

[3] C. Shannon, A mathematical theory of communication, Bell system Journal, vol. 27 pp. 623-656 part II, 1948.

[4] G. Zémor, Cours de Cryptographie. Cassini 01/2001.
landscapes. Foundations of Genetic Algorithms 3, Scan Francisco, Morgan Kaufmann, 243-269 (1995).

[5] P. Rogaway et D. Coppersmith, A Software-Orentid Encryption Algorithm. Fast Software Encryption, Cambridge Security Workshop Proceeding, p.56-63. Springer-Verlag, Berlin, 1994.

[6] R.A. Rueppel Analysis and Design of Stream Ciphers. Springer-Verlag, Berlin, 1986.

[7] R.C. Merkle. Method and apparatus for Data Encryption. U.S Patent #5,003,597, 26 mars 1991.

[8] J.H. Holland, Genetic Algorithms, Scientific American, 44-50 (1992).

[9] J. Horn, D.E. Goldberg, Genetic Algorithms difficulty and the modality of fitness

[10] A. Rukhin and all. A Statistical Test Suite For Random and Pseudorandom Number Generators For Cryptographic Applications, NIST Special Publication 800-22 (with revisions dated May 15, 2001)

[11]. A. Sabour, A. Asimi and A. Lbekkouri, Genetic Regenerator of pseudo-random sequences R.A NMJ, International Symposium on AI 50 Years' Achievements, Future Directions and Social Impacts (ISAI'06), Aug. 1-3, 2006, Beijing, P.R.China.

[12]. A. Sabour, A. Asimi and A. Lbekkouri, Etude théorique de l'observateur : Fonctions dissipatives du R.A NMJ, Les journées d'Optique et du traitement de l'information « OPTIQUE'06 », Rabat 2006.

[13]. A. Sabour, A. Asimi and A. Lbekkouri, Etude du Comportement Chaotique du R.A NMJ via la Sensibilité aux conditions Initiales. 3rd International Symposium on Image/Video Communications over fixed and mobile networks, Tunisia 2006

[14]. A. Sabour, A. Asimi and A. Lbekkouri, Régénérateur génétique des suites binaires cryptographiquement sûres R.A NMJ, Colloque International Sur l'Informatique et ses Applications IA'2006 Oujda, 31 Octobre, 1 et 2 Novembre 2006

[15]. A. Sabour, A. Asimi , A. Lbekkouri and H. Bouyakhf, Etude de la Complexité des Classes de Fonctions à Trois Etats, The Maghrebian Conference on Sofware Engineering and Artificial Intelligence (MCSEAI'06), Agadir Morocco, December 07 to 09, 2006.



Abderrahim Sabour received the DESA in Engineering Computer Science from the University Mohammed V-Agdal in 2004. His research interest includes artificial intelligence and the application of evolutionist algorithms to computer security.



Ahmed Asimi received his PhD degree in Number Theory from the University Mohammed V-Agdal in 2001. His research interest includes Number theory and cryptology. He is an assistant professor at the faculty of Science at Agadir since 2004.



Aboubakr Lbekkouri received his PhD degree in Algebra from Syracuse University (NY) in 1986. His research interest includes Number theory and cryptography. He is a full professor at the faculty of Science at Rabat since 1990.