Research Journal of Applied Sciences, Engineering and Technology 4(14): 2167-2170, 2012
ISSN: 2040-7467

Submitted: March 05, 2012 Accepted: April 03, 2012 Published: July 15, 2012

Based on Negative Selection Algorithm to Generate Long-Period Pseudo-Random Sequence

Mingyu Yue
Department of Electronic Information, Sichuan University, Chengdu 610065, China
Chongqing City Management College Information Center, Chongqing 400055, China

Abstract: This study presents a method of generating long-period pseudo-random sequence based on a plaintext and negative selection algorithm for real applications. It is a new method that extracts the generated factor from the plaintext. Moreover, the generated factor as a group is expanded by the way of the m-sequence and generates pseudo-random. The autocorrelation and statistical properties of the method are better. If it used for data encrypted, it has a good anti-attack performance. In addition, the generate factor can also be used for immune identification in receiver end, to verify data integrity. So there will be a good prospect.

Key words: Generated factor, long-period, negative selection algorithm, pseudo-random sequence, simulation

INTRODUCTION

The rapid development of the Internet bring information security technology, such as text encryption, image encryption, the encrypted video, they are increasingly being concerned. Therefore, Looking for a high quality random sequence is very urgent. It is an integral part of encryption technology. Random sequence has two kinds: physical method to generate true random sequence and the use of simulation algorithm generating pseudo random sequence. According to some probability distribution generated by random sequence is truly random sequences. However, the real random sequences are usually poor in the speed and efficiency relatively and It requires more storage space, these drawbacks restrict its application. E usually generate pseudo-random sequence through the instructional process or method, it has the similar characteristics with white noise. Because the pseudo random sequences can be easily generated, processing, recycling, which are widely used in communication, such as cryptography, testing and other fields (Zhang *et al*., 2007; Chen and Zhong, 2007).

″Most of the existing method for generating pseudo-random sequence are linear congruence method, linear and nonlinear feedback shift register (Menezes *et al*., 1996). Pseudo random sequences are widely used in information security, it is from the use of short pseudo-random bit pseudo-random generator derived. In fact, safe sequence is derived from the algorithm, therefore, it is not completely random. or safety, pseudo random sequence of keys must be truly random sequences and have the following marked characteristics: the maximum cycle, since the correlation is good, should be a value of two and should have high complexity (Zheng *et al*., 2008).

In this study, on the basis of previous research, proposed one kind method based on the artificial immune negative selection algorithm of growing period of pseudo random sequence. Simulation results show that: this method as long as the parameters are selected properly, it has very good auto correlation characteristic, statistics and attack, improve application effect.

THE BASIC METHODOLOGY

First of all, the plaintext message M by the transmission size grouping M_1, M_2,… M_n. And then through the artificial immune negative selection algorithm to generate r pseudo random sequence generating factor (K_1, K_2, … K_r). The size of the r from the plaintext length and packet number to select. Finally in the r factor through the pseudo random sequence generation algorithm for B (k) corresponding to the generated pseudo-random sequence K (K_1, K_2, … K_r, K_{r+1}, … K_n)

The second definition:

$$B(K) = K_1, K_2,\ldots K_i,\ldots K_r, K_{r+1},\ldots K_n, (i = 1, 2,\ldots n)$$

Encryption required for the pseudo random sequence. and,

$$k_i = \begin{cases} k_i, 1 \le i \le r \\ k_{i-r+1} + k_{i-r}, r+1 \le i \le n \end{cases} \quad (1)$$

At the same time, R is defined by the negative selection algorithm to generate several factors, L as a production factor of length. If the characteristic polynomial $f(x) = 1 + x + x^r$ for primitive polynomial, then B (K) for generating factor as the unit of M sequence. The m sequence is a uniform distribution of the pseudo random sequence:

$$\text{The period} = L^* 2^r\text{-}1 \quad (2)$$

If the pseudo random sequence is used to encrypt the file, then according to the file size, we can choose the right L and r, let B (K) sequence in a cycle of length greater than or equal to the length of the file, "thus it theoretically meet the sequence cipher security conditions (Sudha *et al*., 2007; Chandra *et al*., 2007).

Algorithm and application principle: In 1994, Forrest, Perelson etc proposed egative selection algorithm (Forrest and Perelson, 1994; Forrest *et al*., 1997) successfully simulates the immune tolerance. At present, immune cell self-tolerance mainly by negative selection algorithm to achieve. Here, we use artificial immune negative selection algorithm principle and its corresponding algorithm to generate a fixed size pseudo random sequence generating factor. ue to the adoption of the artificial immune negative selection algorithm to generate factor through tolerance and cannot be the plaintext message packets in a packet with the affinity (Li, 2004) is less than the given threshold, it can also generate the use factor to provide identification information conditions.

The first domain U, U contains the length of L binary arbitrary string and a numerical finite set. U is divided into two subset of M and NM, set M is called plaintext information self set, contains a valid length for L binary string (plaintext information representation) collection, should be as much as possible characterization of plaintext information.

For example: a legal document binary form: {10110111001100011011010 1...}, if L = 8, then obtains the set M $\{A_1, A_2, A_3, ...\}$ = {(10110111), (00110001), (10110101) ...}.

The M Set element should not be too much, so when matching with antigen detection system, saving operation time and save the storage space of system. But also should not be too little, should maintain diversity. The M set and the NM set have the following relationship:

$$\{ M \} \cup \{ NM \} = U \quad (3)$$

$$\{ M \} \cap \{ NM \} = \Phi \quad (4)$$

The plaintext message packet concentration data representative of the normal data characteristics, if directly from the selected packet as a production factor, then it may be mistaken for normal plaintext information packet error or falsified information packet. It must be egative selection process (Li, 2004). Through the negative selection algorithm to remove the plaintext information packet itself to match the production factor, thus realizing the plaintext message packet self tolerance, generate ideal growth factor.

According to the plaintext message set M, for which each element of each bit are nverse operations (Huang and Li, 2008). These data are combined into one set, denoted by NM set. NM = { N1, N2, N_i,} . From the NM concentration selecting one candidate detector elements of N_i, N_i and M in each data set each plaintext information packet data computing. If the N_i is expressed as (N_1 N_2 ... N_L), Ai expressed as (A_1 A_2 ... A_l), the threshold ϵ, ϵ is integer. Calculation between the Hamming distance, get D value:

$$D = \sum_{i=1}^{l} \delta \quad \text{when } n_i = a_i \quad \delta = 1; \text{ then } \delta = 0. \quad (5)$$

when $D \leq \epsilon$, this representation is generated by a tolerance factor, become mature pseudo random sequence generating factor. If it exceeds the threshold, this factor is not qualified, must from NM concentrated delete this factor.

Procedure standard negative selection algorithm
Begin
While the size of a given set of detectors do not DO
Generates a random candidate encryption factor;
Begin
Calculate the affinity between candidate encryption factor with the element;
If the candidate does not recognize any body element
Then the candidate is placed in encryption factor set
End
End.

Using such B (K) pseudo random sequence of sequence cipher encryption method to encrypt plaintext M or information hiding, get the cryptograph C:

$$C = E(M, K) \quad (6)$$

$$C = M \oplus K = M_1 M_2 \ldots M_n \oplus K_1 K_2 \ldots K_n = C_1 C_2 \ldots C_n \quad (7)$$

The receiving end receives the encrypted C, then the corresponding inverse transform M decryption:

$$M = C \oplus K = C_1 C_2 \ldots C_n \oplus K_1 K_2 \ldots K_n \quad (8)$$

SIMULATION AND ANALYSIS

In the MATLAB7 simulation experiment platform, get the" How are you I ' m fine Thank you How are you How are you I ' m fine Thank you How are you How are you I ' m fine Thank you How are you How are you I ' m fine Thank you How are you" 128 English letters (binary the size of the file just for 1K) encrypted experiment. In this experiment, according to L = 8, B (k) cycle for $L(2^r-1)$ is greater than or equal to 1024 can be calculated from the r> = 8, then randomly selected 8 generation factor. In this experiment the encryption factor as follows:

k_1 = 0011010, k_2 = 10111100
k_3 = 11100110, k_4 = 10011110
k_5 = 11011111, k_6 = 11000000

Fig. 1: Comparison of plaintext and ciphertext

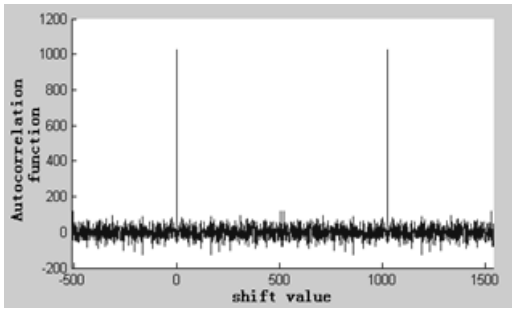

Fig. 2: Computed result of self-correlation function

k_7 = 10001100, k_8 = 10011101

According to the key sequence generation algorithm based on sequence generated key string to the actual length of 1080, more than 1024 bits, but we only take the top 1024 to 128 English letter sequence cipher encryption. The encrypted ciphertext and plaintext ASC II code display contrast as shown in Fig. 1:

From the comparison of plaintext and ciphertext plaintext, despite the obvious cycle and statistical characteristics, but the encryption after its marked period does not exist, masked plaintext cycle characteristics, specifically to prevent information leakage, can effectively prevent he known plaintext or ciphertext only attack (Sudha *et al*., 2007; Chandra *et al*., 2007).

According to the m sequence of the autocorrelation characteristics of theoretical analysis, the m sequence of shift values for the integer multiple of the period, its autocorrelation value maximum value and other value, displacement, the autocorrelation value constant-1. Key sequence generation algorithm to generate the serial key string autocorrelation values the simulation of the MATLAB7 as shown in Fig. 2:

According to Fig. 2, can be clearly seen only in 0 and 1080, the whole cycle is the maximum and elsewhere in the 1 point swing, this fully illustrated in front of the algorithm has good autocorrelation characteristics.

The simulation was found in the pseudo random sequence generation algorithm to generate the serial key string balance test is not ideal (ideal value is 1), this is mainly because the selected growth factors do not have better balance, but on the whole, the growth factor is a seed, while the key sequence is generated according to the m sequence generation method, so it has little effect.

Static test of sequence: We can choose the following measures to test he statistical characteristics of random binary sequence (Zheng *et al*., 2008):

1 frequency test: A binary sequence of length n, 0S to n_0, 1S to n_1 and $n_0 + n_1 = n$:

$$x^2 = \frac{2}{n_0 + n_1}\left\{\left(n_0 - \frac{n_0 + n_1}{2}\right)^2 + \left(n_1 - \frac{n_0 + n_1}{2}\right)^2\right\} \quad (9)$$

Degrees of freedom is 1, when x^2 <3.8415, it passes the frequency test.

2 serial test: $n_{i,j}$(i, j = 0, 1) for counting, when the i with the j and $n_{00} = n_{01} = n_{10} = n_{11} = n/4$:

$$x^2 = \frac{4}{n-1}\sum_{i=0}^{1}\sum_{j=0}^{1}\left(n_{ij} - \frac{n-1}{4}\right)^2 \quad (10)$$

Freedom is 2, when x^2<5.9915, it passes the serial test.

3 poker test: Select m, m<n and, then divided into n/m group sequence. m length binary sequence 2 m. That state is s, $s \le 2^m$, p_i counts i status, i∈[1, s], to look forward to each state is $n/(m2^m)$.

$$x^2 = \frac{m2^m}{n}\sum_{i=0}^{2^m-1}\left(p_i - \frac{n}{m2^m}\right)^2 \quad (11)$$

Its degrees of freedom for the 2^m-1.
For example: if m = 3, 2^3-1 = 7; when x^2<14.0671, it passes the poker test.

4 run test: Suppose that y is running number.
Average value:

$$E = \frac{2n_0n_1}{n_0 + n_1} + 1 \quad (12)$$

Variance is as follows:

$$V = \frac{2n_0n_1\left(2n_0n - n_0n_1\right)}{\left(n_0 + n_1\right)^2\left(n_0 + n_1 - 1\right)} \quad (13)$$

Statistics are as follows:

Table 1: Pseudo random performance

			Poker test			
Types of test	Frequency test	Serial test	M = 2	M = 3	M = 4	Run test
10000b	0.00	2.43	0.27	6.33	15.12	1.55
20000b	0.00	0.86	0.58	7.64	8.89	0.76
30000b	0.00	0.47	1.76	10.13	6.96	0.53
40000b	0.00	1.78	0.57	8.26	5.85	0.46
T	3.82	5.89	7.85	14.12	26.12	1.96

$$Z = \frac{y - E}{\sqrt{V}} \quad (14)$$

when | Z |<1.96, it passes the running test.

5 test results: Table 1 shows 4 types of tests were performed on four bit length test. T is the threshold. According to the Table 1, the sequences have good statistical property.

CONCLUSION

Based on the artificial immune negative selection algorithm to generate long-period pseudo-random sequence and sequence extension algorithm integrated information characteristics and traditional grouping encryption thought and because the growth factor is not predictable, so that it has better reliability and safety for software implementation. The autocorrelation and statistical properties of the method are better. If it used for data encrypted, it has a good anti-attack performance. In addition, the generate factor can also be used for immune identification in receiver end, to verify data integrity. In the pseudo random sequence generation, time and space efficiency is slightly reduced, but the influence cannot be too big. This is the future of the problem to be solved.

REFERENCES

Chandra, S., K.R. Sudha and P.V.G.D. Prasad Reddy, 2007. Data Encryption technique using Random number generator. IEEE International Conference on Granular Computing, pp: 576-579.

Chen, S. and X.X. Zhong, 2007. Chaotic block iterating method for pseudo-random sequence generator. J. China Univ. Posts Telecommun., 14(1): 45-48.

Forrest, S. and A.S. Perelson, 1994. Self-Nonself Discrimination in a Computer [J]. IEEE Symposium on Security and Privacy. 1994. IEEE Computer Society Press, Oakland, CA, pp: 202-213.

Forrest, S., S. Hofmeyr and A. Somayaji, 1997. Computer immunology [J]. CACM, 40(10): 88-96.

Huang, G. and X. Li, 2008. An intrusion detection system based on immune principle [J]. Micro Electr. Comput., 25(8): 192-194, (in Chinese).

Li, T., 2004. Computer Immunology. Publishing House of Electronics Industry, Beijing, (in Chinese).

Menezes, A., P. Oorschot and S. Vanstone, 1996. Handbook of Applied Cryptography. CRC Press, New York, USA, pp: 169-203.

Sudha, K.R., A. Chandra Sekhar and P.V.G.D. Prasad Reddy, 2007. Cryptography protection of digita signals using some Recurrence relations. IJCSNS Int. J. Comput. Sci. Netw. Security, 7(5): 203-207.

Zhang, J.S., X.M. Wang and W.F. Zhang, 2007. Chaotic keyed Hash function based on feedforward-feedback nonlinear digital filter. Phys. Lett. A, 362(5-6): 439-448.

Zheng, F., T. Xiao-jian, S. Jing-yi and L. Xue-yan, 2008. Pseudo-random sequence generator based on the generalized Henon map. J. China Univ. Posts Telecommun., 15(3): 64-68.