Cryptanalysis Knapsack Cipher Using Artificial Immune System

Eman Th. Al-Obaidy
Veterinary Medicine College
University of Mosul, Iraq

Received on: 16/07/2008 **Accepted on: 14/10/2008**

ABSTRACT

In this work, the use of an artificial immune system (AIS) in cryptanalysis is explored. This AIS uses the clonal selection principle for the cryptanalysis of knapsack cipher. Results showed that the proposed approach is good especially when the effect of the control parameters on the performance of the clonal selection is well taken into consideration. The program is written in Turbo C.

Keywords: Artificial immune system (AIS), cryptanalysis, knapsack cipher, Turbo C.

تحليل شفرة **Knapsack** باستخدام النظام المناعي الاصطناعي

ايمان العبيدي
كلية الطب البيطري، جامعة الموصل

تاريخ الاستلام: 2008/07/16 **تاريخ القبول: 2008/10/14**

الملخص

تم في هذا البحث التحري عن إمكانية استخدام نظام مناعي صناعي في تحليل الشفرة. يستخدم هذا النظام المناعي الصناعي مبدأ (انتخاب مجموعات من فصائل واحدة) لتحليل شفرة Knapsack. أظهرت النتائج ان الأسلوب المقترح كان جيدا" خصوصا اذا أُخذ تأثير معاملات السيطرة في الأداء لمبدأ (انتخاب مجموعات من فصائل واحدة) بنظر الاعتبار. تمت كتابة البرنامج بلغة (Turbo C).

الكلمات المفتاحية: نظام مناعي صناعي، تحليل الشفرة، شفرة Knapsack، لغة (Turbo C).

1. Introduction

The verbtrate immune system is a rich source of theories and acts as an inspiration for computer–based solutions. Over the last few years there has been an increasing interest in the area of artificial immune system [1].

Most AIS aim at solving complex computational or engineering problems, such as pattern recognition, elimination and optimization [2]. This research investigates using AIS in the cryptanalysis of the knapsack cipher problem.

2. The Immune system

The human immune system is a complex system of cells, the basic component of the immune system is the lymphocytes or the white blood cells. Lymphocytes exit in two forms, B cells and T cells. These two types of cells are rather similar, but differ with relation to how they recognize antigens and by their functional roles, B-cells are receptors of antigen=BCR, T-cells are receptors of antigen =TCR B-cells are capable of recognizing antigens free in solution, while T cells require antigens to be presented by other accessory cells. Each of this has distinct chemical structures and produces many Y shaped antibodies from its surfaces to kill antigens. Ab's are molecules attached primarily to the surfaces of B cells whose aim is to recognize and bind to Ag's as shown in figure 1 [2].

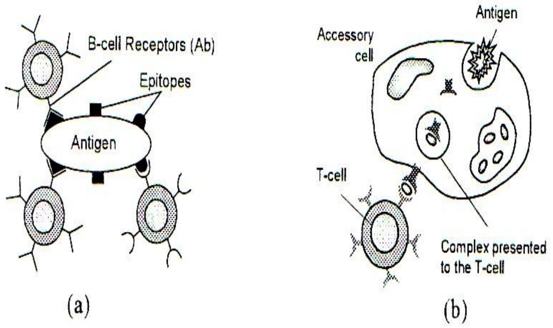


Figure 1: pattern recognition in the immune system. (a) B-cell recognizing an antigen (Ag) free in solution. (b) T-cell recognizing an antigen by an accessory cell.

The cells that originally belong to our body and harmless to its functioning are termed self (or self antigens), while the disease causing elements are named nonself (or nonself antigens). The immune system, thus, has to be capable of distinguishing between what is self from what is nonself. Binding is highly specific, so each detector recognizes only a limited set of structurally related antigen. A striking feature of the immune system is that the processes by which it generates detectors, identifies and eliminates foreign material, and remembers the patterns of previous infections are all highly parallel and distributed. This is one reason why immune system mechanisms are so complicated, but it also makes them highly robust to failure of individual components and to attack the immune system itself. Antigenic recognition is the first prerequisite for the immune system to be activated and to mount an immune response. The recognition has to satisfy some criteria. First the cell receptor recognizes an antigen with a certain affinity, and a binding between the receptor and the antigen occurs with strength proportional to this affinity. If the affinity is greater than a given threshold, named affinity threshold, then the immune system is activated.

The human immune system contains an organ called thymus that is located behind breastbone, which performs a crucial role in the maturation of T cells. After T cells are generated, they migrate into the thymus where they mature. During maturation all T cells that recognize self antigens are excluded from the population of T cells, a process termed *negative selection*. If a B- cells encounters of nonself antigen with a sufficient affinity it proliferates into memory and effector cells a process named *clonal selection.* In contrast if a B- cell recognizes a self – antigen it might result in suppression, as proposed by the *immune network theory* [1] [2][3] [4] [5][6][7]

3. Artificial Immune System

An artificial immune system is a computational system based upon metaphors of the natural immune system or artificial immune systems are intelligent methodologies inspired by the immune system toward real – world problem solving [8]. The most common principles used by AIS are negative selection, clonal selection and immune network theory.

3.1 Clonal Selection

Ab's are molecules attached primarily to the surface of B cells whose aim is to recognize and bind to Ag's. Each B cell secretes a single type of Ab, which is relatively specific for the Ag. By binding to these Ab's and with a second signal from accessory cells such as the T- helper cell the Ag stimulated the B cell to proliferate (divide) and mature into terminal (no dividing) Ab secreting cells, called plasma cells. The process of cell division (mitosis) generates a colne i.e. a cell or set of cells that are the progenies of a single cell.

B cells in addition to proliferating and differentiating into plasma cells, can differentiate into long – lived B memory cells. Memory cells circulate through the blood, lymph, and tissues and when exposed to a second antigenic stimulus, commence to differentiate into plasma cells capable of producing high – affinity Ab's, preselected for the specific Ag that had stimulated the primary response. In figure 2 the clonal selection process is shown[9].

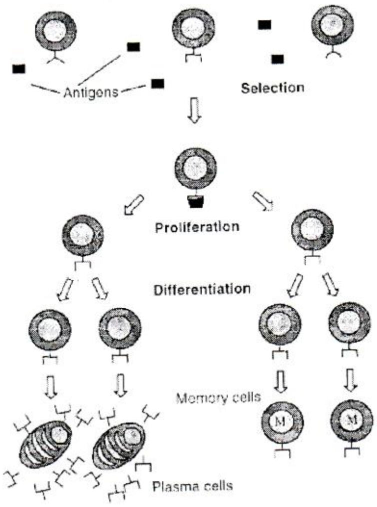


Figure 2: Clonal Selection Process.

De Castro [9] presented an algorithm called CLONALG, which is based on the natural clonal selection, the following list contains the notation which is used to describe the algorithm:

Ab: available antibody repertoire ;
$Ab_{\{m\}}$: memory antibody repertoire ;
$Ab_{\{r\}}$: remaining anti body repertoire;
$Ag_{\{m\}}$: population of antigens to be recognized ;
f_i : vector containing the affinity of all antibodies with relation to the antigen Ag_j ;
$Ab^j_{\{n\}}$: n antibodies from Ab with the highest affinities to Ag_j ;
C^j :population of clones generated from $Ab^j_{\{n\}}$;
C^{j*} : population C^j after the affinity maturation process;
$Ab_{\{d\}}$: set of d new molecules that will replace d low – affinity antibodies from $Ab_{\{r\}}$;
Ab_j* : candidate from C^{j*}, to enter the pool of memory antibodies ;

Using the termination above, the CLONALG algorithm can be described as follows :

1) Choose an antigen randomly from $Ag_{\{m\}}$ and present it to all antibodies in the repertoire Ab;
2) Determine the vector f_j which contains the affinity of the chosen antigen to all the antibodies in Ab;
3) The antibodies with the highest affinity to the chosen antigen are selected from Ab , to compose a new set $Ab^j_{\{n\}}$ of high affinity antibodies ;
4) These selected antibodies are now cloned independently and proportionally to their affinities, to generate another repertoire C^j of clones. The higher their affinity, the more clones are produced ;
5) The repertoire C^j is submitted to an affinity maturation process inversely proportional to the antigenic affinity to generate another repertoire C^{j*} of colnes. But here is the rule, the higher the affinity, the smaller the maturation rate ;
6) Determine the vector f_j* which contains the affinity of the matured clones C^{j*} in relation to the antigen (which was chosen in 1)
7) From C^{j*} another re- selection is done to select the one with highest affinity in relation to the antigen (which was chosen in 1) to be a candidate to enter the set memory antibodies $Ab_{\{m\}}$.If there already exists an antibody (to the antigen chosen in 1)in $Ab_{\{m\}}$ which affinity is lower, then it is replaced by new one;
8) The d lowest affinity antibodies (corresponding to the antigen chosen in 1) from $Ab_{\{r\}}$ are replaced by new individuals. [9][10]

4. The Knapsack Cipher:

The knapsack problem is formulated as follows. Let us assume that the values M_1M_2…..M_n and the sum S are given. Let it be necessary to compute b_1 b_2 ….b_n values, so that $S = M_1b_1 + M_1b_1$……+ M_nb_n . The values of coefficient b_i can be equal to 0 or 1. The 1 value shows that object will fit into the knapsack, 0 values will not into the knapsack.

The Markle-Hellman knapsack cipher encrypts a message as a knapsack problem. The plaintext block transforms into binary string(the length of block is equal to the number of elements in knapsack sequence). One value determines that an element will be in target sum. This sum is a ciphered message. Table 1 shows an example of solving the knapsack problem for the entry numbers sequence:1 3 6 13 27 and 52.

Table 1: Example of Knapsack Encryption [11]

Plaintext	Knapsack sequence	Ciphertext
1 1 1 0 0 1	1 3 6 13 27 52	1+3+6+52=62
0 1 0 1 1 0	1 3 6 13 27 52	3+13+27= 43
0 0 0 0 0 1	1 3 6 13 27 52	52

Easy knapsacks have a sequence of numbers that are superincreasing that is, each number greater than the sum of previous numbers: $a_i >= \sum_{j=1}^{i-1} aj,$

for i =2,……., n (where a_i is the i-th element of the sequence). For example {1,3,6,13,27,52} is a superincreasing sequence but {1,3,4,9,15,25} is not. The superincreasing knapsack is easy to decode, which means that it does not protect the data. Anyone can recover the bit pattern from the target sum for a superincreasing knapsack if the elements of the superincreasing knapsack are known.

Markle and Hellman suggested that such a simple knapsack be converted into a trapdoor knapsack which is difficult to break. The algorithm work as follows:

1. select a simple knapsack sequence. Elements make a superincreasing sequence $A'=(a'_1+a'_2+$……$a'_n)$
2. select an integer value m greater than sum of all elements of superincreasing sequence.
3. select another integer w that the gcd(m,w)=1, that is number m and w are reciprocally prime.
4. find the inverse of the w mod m-w^{-1}
5. construct the hard knapsack sequence A=wA' mod m i.e. $a_i = wa'_I$ mod m

The trapdoor sequence A could be published as a public key (encryption key). The private (secret) key for this cipher consists of a simple knapsack sequence A' so-called trapdoor values m,w, w^{-1}.

The encoding is done as follows: The message is divided into n-bits block (each block contains as many elements as simple knapsack sequence). Values in the message block show that the element will be in the target sum. The target sum of each block is cipher text.

The decoding consists of the following: Each number of the ciphered message is multiplied through w^{-1} mod m and the result of this operation is plaintext.[11][12][13].

5. Related work:

Many of the reserchers use genetic algorithms for attacking the knapsack cipher: Spillman [14] used the knapsack of size 15, A= { 21031, 63093, 16371, 11711, 23422, 58555, 16615, 54322, 1098, 46588, 6722, 34475, 47919, 51446,1 6438} and the fitness function shown below which penalizes solutions which have a sum greater than the target sum.

$$\text{Fitness (M)} = \begin{cases} 1 - (|\text{Target} - \text{Sum}| / \text{Target})^{1/2} & \text{if Sum < Target} \\ 1 - (|\text{Target} - \text{Sum}| / \text{MaxDiff})^{1/6} & \text{if Sum >=Target} \end{cases} \quad \ldots(1)$$

Where (let M = { $m_1, m_2, \ldots m_n$ }, m_i Є { 0 , 1 } be an arbitrary solution and the public Key A = { $a_1, a_2, \ . \ . \ . \ a_n$ }).

$$\text{Sum} = \sum_{j=1}^{n} ajmj \qquad \ldots(2)$$

$$\text{Target} = \sum_{j=1}^{n} a'\,j \qquad \ldots(3)$$

$$\text{FullSum} = \sum_{j=1}^{n} aj \qquad \ldots(4)$$

MaxDiff = max { Target , FullSum – Target }. …(5)

Clarck,J.,A [15] used a knapsack with the same size of Spillman,R algorithm but with a modified fitness function.

$$\text{Fitness} = 1 - (|\text{Sum} - \text{Target}| \ / \text{MaxDiff})^{1/2}$$

He found that this modified fitness function finds the solution more quickly since solutions with their sum greater than the target are not being penalized.

Garg,P.,Shastri,A [12] used a knapsack of size 8 A ={ 21031, 63093, 16371, 11711, 23422, 58555, 16615, 54322} and an improved genetic algorithm with varation of initial entry parameters, also Kolodziejczyk,J

[13] used the same knapsack size in [12] and also an algorithm with different initial parameters(population size, selection, crossover, mutation, termination) and the results are compared with Spillman's results.

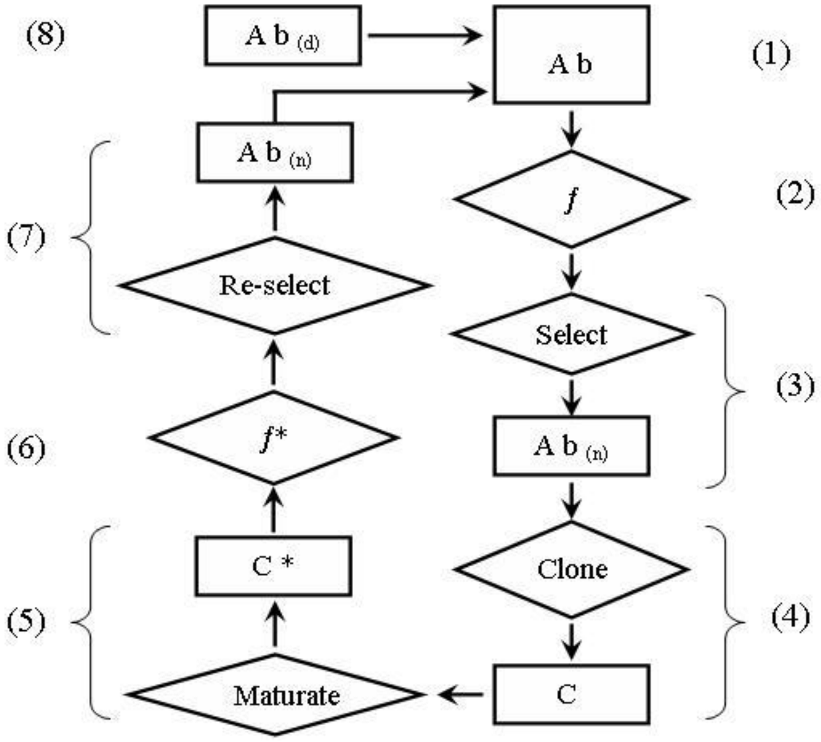
Ahmad, Badrisham,A., [16] used clonal selection algoritm in attacing simple substitution cipher[16] .

In this research the clonal selection algorithm is applied in attacking the knapsack cipher.

6. The Proposed Algorithm:

A clonal selection principle is used in the proposed algorithm for attacking the Knapsack cipher.

CLONALG in [9] is used in this research(the optimization version), which is clarified in figure [3].



Figure[3]: CLONALG : the optimization version.

CLONALG is composed basically of two repertoires (populations of strings): A set of antigens **Ag** and a set of antibodies **Ab.** The set **Ab** can be decomposed into several subsets according to the application under study.
Here is some notes about the algorithm in the optimization version:

- In step 1 there is no need for explicit Ag population just an objective function g(.) to be optimized (maximized or minimized), here an Ab affinity corresponds to the evaluation of the objective function for a given Ab.
- In step 4 the number of clones generated for all these n selected antibodies was given by : Nc= $\sum_{j=1}^{n} round(\frac{B.N}{j})$ where Nc is the total number of clones generated for each of the Ag's, B is a multiplying factor, N is the total number of Ab's
- In step 7 n of the Ab's are reselected to compose the set of Ab instead of selecting the single best individual Ab*$_{j.}$

In the proposed algorithm the 8-elements(Spillman in [14] used 15 elements) sequence of hard knapsack problem (21031 63093 16371 11711 23422 58555 16615 54322) is used to encode 8 bits ASCII code. This sequence has been created from superincreasing sequence (1 3 7 13 26 65 119 267), u equal to 65423 and w integer equal 21031 (w^{-1} = 5363). The MACRO word has been encrypted as shown in table 2.

Table 2: Encryption by Knapsack.

Character	ASCII Code	Target sum (ciphertext)
M	10110010	65728
A	10000010	37646
C	11000010	100739
R	01001010	103130
O	11110010	128821

The following restrictions have been made for incoding :

1. only the ASCII code will be encrypted.
2. The superincreasing sequence will have 8 elements this number of elements guarantee that each character has a unique encoding (there are 256 ASCII codes and 8 elements length will allow to encrypt 2^8 character

A random population of antibodies (binary string 0's and 1's) is used for initialization, the equations from 1 to 5 are used in the algorithm.

7. Results:

In the proposed algorithm each target sum(ciphertext) in table 2 for each character is attacked 7 times with the following entry parameters: N=100, n={10,20,30,40,50,60,70}, (d=N minus n), the mutation rate P_m =0.12 then the results are averaged as shown in table 3:

Table 3: the results of the proposed algorithm.

Character	# antibody	% of search space
M	5	1.9
A	49	19.1
C	25	9.7
R	26	10.1
O	53	20.7
Average	32	12.3

The experiments indicated that the proposed algorithm showed a good performance in finding the correct solution when it's results are compared with the Spillman's algorithm in [14] as shown in table 4.

Table 4: Spillman's results.

Character	# chromosome	% of search space
M	810	2.0
A	80	0.2
C	1860	6.0
R	460	1.0
O	650	0.1
Average	650	1.9

The Spillman's algorithm searches on average less than 2% of the space. This diveregence of the results because of the`area of the possible results in Spillman's work the space is 2^{15} i.e. 32678 and in this work is 2^8 also the algorithms in [12] and [13] search on average 47.9% and 48.4% of the space (2^8), respectively so that the proposed algorithm is good because it searches on average l2.3% of the space (2^8) .

8. Conclusion :

An algorithm that uses a clonal selection principle is used in attacking knapsack cipher. This paper indicates that the clonal selection offers a powerfull tool for cryptanalysis of knapsack cipher especially if the parameters of this algorithm are carefully set, and these parameters are: the size of antibodies repertoire (N), the size of best selected antibodies for cloning (n), the set of new antibodies that will replace low-affinity antibodies (d) and mutation probability (P_m).

By performing a number of trial runs it can be concluded that starting the clonal selection with (N=100, n=10,20,….,70, d=N-n, P_m =0.12) can increase the effeciency and performance of the clonal selection.

The algorithm gives the correct result by searching 12.3% of search space (2^8) while genetic algorithm in [12][13] searches on average 47.9% and 48.4% respectivly of the same space to find the correct result.

REFERENCES

[1] Li, P., Zhong, Y., Zhang, L., ”Applications of Artificial Immune System In Remote Sensing Image Classfication”, State Key Laboratory of information Engineering in Surveying Mapping &Remote Sensing, Wuhan University 129 Luoyu Road, Wuhan, Hubei, 430079,China-zlp62@public.

[2] De Castro,L.N.,Timmis,J., 2002, “Artificial Immune Systems: A Novel Paradigm to Pattern Recognition”, Computing Laboratory, University of Kent at Canterbury,Kent, Canterbury, CT2 7nf, United Kingdom.email {L.N.deCastro,J.Timms}@ukc.ac.uk.

[3] Forrest,S.,Hofmeyr,S.,A., “ John Holland’s Invisible Hand: An Artificial Immune System”, Dept. of Computer Scince, University of New Mexico, Lbuquerque, NM 87131-1386, {steveah,forrest}@cs.unm.edu.

[4] Forrest,S.,Perelson,A.,S.,Allen,L.,Cherukuri,R.,1994,”Self-Nonself Discrimination in a Computer, Dept. of Computer Scince, University of New Mexico, Lbuquerque, NM 87131-1386, forrest@cs.unm.edu.

[5] Dasgupta,D., Forrest,S.,” Artificial Immune Systems in Industrial Application”, Dept. of Mathmatical Sciences, The University of Memphis, Memphis, TN 38119.

[6] Forrest,S.,Hofmeyr,S.,A.,2001,” Engineering an Immune System”, Computer, Dept. of Computer Scince, University of New Mexico, Lbuquerque, NM 87131-1386, {steveah,forrest}@cs.unm.edu.

[7] Somayagi,A., Forrest,S.,Hofmeyr,S.,A.,” Principles of a Computer Immune System”, Computer Scince, University of New Mexico, Lbuquerque, NM 87131-1386, {soma, steveah,forrest}@cs.unm.edu.

[8] Ehret,C.,2006,” Introducing the Artificial Immune System Paradigm”, Telecommunications, Networks and Security, Department of Computer Science, University of Fribourg.

[9] De Castro,L.N.,Von Zuben,F.,J, 2002, “ Learning and Optimization Using the Clonal Selection Principle”,Member, IEEE.

[10] Bachmayer,S., 2007,” Artificial Immune Systems”, Department of Computer Science, Gustaf Haellstroemin katu 2b, Finland, 00014 University of Helisinki, sabine.bachmayer@ helisinki.fi.

[11] MAS 335, 2008,”Cryptography”, Queen Mary , University of London.

[12] Garg,P.,Shastri,A.," An Improved Cryptanalytic Attack on Knapsack Cipher using Genetic Algorithm",International Journal of Information Technology, Volume 3 Number 3.

[13] Kolodziejczyk,J.," The Application of Genetic Algorithm In Cryptoanalysis of Knapsack cipher", Institute of Compute Science and Information Systems, Technical University of Szczecin, ul. Zolnierska 49, 71-210 Szczecin, Poland, Fax:(**4891) 4876439, e-mail : joanna_ kolodziejczyk@ii.tuniv.szczecin.pl.

[14] Spillman,R.,1993,"Cryptanalysis of Knapsack Ciphers using Genetic Algorithms", Cryptologia, volume xvii, no 4, pp.367-377.

[15] Clarck,J.,A.,1998," Optimization Heuristics for Cryptology", PhD thesis, Information Security Research Centre, Faculty of Information Technology, Queensland University of Technology.

[16] Ahmad, Badrisham,A., Mohd Aizaini,M., Subariah,I., Mammi,K., 2006," Clonal selection Algorithn for the Cryptanalysis of a Simple Substitution Cipher", Simposium Kebangsaan Sains Matematik, Langkawi, Malaysia.