ISSN:0975-9646
Agnika Sahu et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 1149-1153

Clonal Selection Method for Virus Detection in a Cloud

Agnika Sahu[#1], Tanmaya Swain[*2], Tapaswini Samant[*3]

[#] *School of Computer Engineering, KIIT University*
Bhubaneswar, India

***Abstract*— The biological immune system (BIS) is to protect the human body against attack from the antigen such as virus, bacteria, fungi and other parasites and eliminates it from the infected cells. Artificial Immune System (AIS) are the machine acquiring algorithm that be some of the principles and effort to accept some advantages of the biological immune system in order to solve the problem field. Among which Clonal Selection Algorithm (CSA) is one of the inspired method to acquired immunity. The whole process is of antigen recognition, cell proliferation and distinction. In our proposed work only those cell which can recognize the antigen are selected to grow. Then these cells are subjected to affinity maturation process which increases the affinity value of selected antigen.**

***Keywords*— Artificial immune system, Clonal selection process, Hamming distance, Shift r continuous.**

I. INTRODUCTION

Artificial Immune systems are the new technique based on the metaphoric concept of the biological inspired computation based on the experimental knowledge of the vertebrate immune system. It is one of the biological processes to destroy or prevent the disease in the body, the immune system is known to be adaptive in terms of function and all the feature are used for solving problems faced in the field of artificial intelligence.

The clonal selection method is used to explain the basic characteristics of an adaptative immune reaction to an antigen input or stimulus of the antigen. It demonstrates the idea that only those cells that can distinguish the antigens are selected to proliferate or multiply. Then these selected cells are then subjected to an affinity maturation or development process, which improves their affinity to the selective antigens.

The Groundwork of conventional computer security technology that determines whether the other platform is venomous or not. As antivirus programs are from where they have only changed a little bit on how they are working. Cloud computing is one of the new models for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1].So, enhancing the cloud security is the key factor to promote the growth of cloud computing. Data record detect virus before the file is executed, they apply some binary information extracted from the program file. The convention method extract signatures [2] from the virus file and then the scanner compare it with the unclassified files to check whether it is a virus or not. As polymorphic virus change their signature while spreading, so it is quite difficult to extract the signature and detect it. The main principle of cloud security technology is to analyse the suspicious files which are collected from end systems and automatically upload to the cloud servers, and after the analysis to determine whether the suspicious file is malware or not.

The rest of the paper is organized as follows. In section II, the detail of a survey was given on Artificial Immune System. The details of the Clonal Selection Process are described in section III. The details of the proposed work are given in section IV. Section V concludes the paper.

II. ARTIFICIAL IMMUNE SYSTEM

Artificial immune systems (AIS) are computational systems inspired by the principles and processes of the vertebrate immune system. The field of Artificial Immune Systems (AIS) is mainly concerned with the structure and functions of the immune system to computational systems, and investigate the application of these systems towards solving computational problems from mathematics, engineering, and information technology. Artificial Immune Systems (AIS) are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving. Basically an immune system has some properties i.e. detection, diversity, learning and tolerance.

1) Detection: Identification takes place in an immune system when the infective fragment and sensory receptor on lymph cell surface is bonded chemically.
2) Diversity: Identification in an immune system is related to non-self bodies of the organism, thus the immune system has a number of sensory receptor, out of which some of the lymph cells will react with the foreign organism.
3) Learning: An immune system has the capability of detecting and eliminating the foreign organism as soon as possible from the human body. This principle allows the lymphocytes to find out and adjust themselves to specific foreign protein structure. It is done by the B-cells.

4) Tolerance: The particles which are mark themselves as self bodies are contain in the chromosomal section.

III. CLONAL SELECTION MECHANISM

The Clonal selection algorithm is used by AIS to define the basic features of an immune response to an antigenic stimulus [3][4]. Clonal selection mechanism describes the basic features of an immune response to an antigenic stimulus. It establishes the idea that only those cells that recognize the antigen proliferate, thus being selected against those that do not.

The main features of the clonal selection theory are [5][6]:

1) The newly cells are replica of their parents (clone) which are submitted to a chromosomal mutation chemical mechanism.
2) Evacuation of newly distinguished lymph cell carrying self - reactive sensory receptor.
3) Development and differentiation on contact of mature cells with antigens.

When an antibody strongly matches an antigen the corresponding B-cell is stimulated to produce clone of itself that then produce more antibodies. This (hyper) mutation, is quite rapid, often as much as one mutation per cell division (de Castro and Von Zuben's 1999).The hyper mutation process enables the new cells to match the antigen more closely. The B-cells with high affinity are selected to differentiate into memory cells which do not secret antibodies but instead remember the antigen pattern. The B-cells that are not simulated as they do not match any antigens in the human body will eventually die. Once the body has successfully defended against an antigen memory cells remain and circulate in the blood, lymph and tissues for the very long period of time. When the same or similar antigen is encountered in the future, memory cells are simulated and more abundant production of antibodies take places. It allows a very quick response to the antigens.

Clonal selection mechanism is inspired by acquired immunity which explains how B and T lymphocytes improve their response to antigens over time called affinity maturation. It is basically focused on the Darwinian [7] attributes of the theory where selection is inspired by the affinity of antigen-antibody interactions, reproduction is inspired by cell division, and variation is inspired by somatic hyper mutation. Clonal selection algorithms are mostly commonly applied to optimization and pattern recognition domains.

The intense role of immune system is to protect the body from the foreign beings. The immune system has the capable for differentiate among the own constituents of our beings and foreign stuff which can damage us. This foreign stuff is known as antigen. The important role played by the immune system is the antibodies. When an antigen is noticed in our body then those antibodies which can distinguish the antigen will multiply by cloning. This procedure is called is Clonal Selection Method. The clonal selection algorithm can be described as follows:

1) Generating a set of candidate keys (K).
2) Determining the n best keys (K*) among the set of the candidate from based on their affinity measures.
3) Cloning or reproducing the n best individuals(C). The clone size is the increasing function of the affinity with the antigen.
4) Submit the population of clone to hyper mutation, where hyper mutation is directly proportional to the affinity of the antibody with the antigen. A matured antibody set is generated (C*).
5) Reselecting the improved individuals from the matured clone set (C*). Some of the member of candidate key (K) is replaced by the improved members of cloned set (C*).
6) The lower affinity cells have a higher probability of being replaced.

The immune system is to run over how to distinguish the self bodies from the non self bodies. The immune system is used to protect the human bodies from the extraneous stuff which are injurious to the organism. The extraneous stuff may be the bacteria, virus, pollen grains, incompatible blood cells and manmade particles.

The clonal selection theory is a theory postulated by Burnet, Jerne, Talmadge, used to describe the functioning of acquired immunity, specifically a theory to describe the diversity of antibodies used defends the organism from invasion [8]. Antibodies are the particles which are produced by the B- lymphocyte cells that are used to neutralize single antigen. B-lymphocytes or white blood cells produce a single or customized antibody of a particular type. Nowadays clonal selection theory is one of the overtaking measures of empiric demonstration. The mechanism of clonal selection process [9] is shown in fig 1.

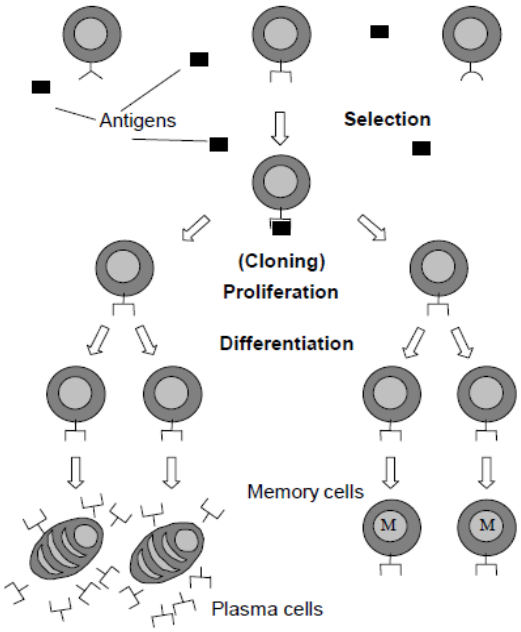


Fig. 1 Clonal Selection Principle

This hypothesis determines that the self bodies have a pre-existing pool of individually specific antibodies which can be recognized with all the antigens with some particularity. When the antigen is matched with a specific antibody, a chemical bonding takes place and replication takes place i.e. more cells are generated with same sense organ or sensory receptor. During the development stage mutation in the clone of the cells take place in order to increase the affinity of the antigen.

IV. PROPOSED WORK

The objective of this algorithm is to increase the affinity values of the selected cells. When these selected cells are subjected to the affinity growth process, which improves the affinity of the selective antigens. The clonal selection algorithm is used by AIS to define the basic features of an immune response to an antigenic stimulus [10] [11]. It sets up the estimate that only those cells they can recognize the antigens are chosen to multiply (proliferate).in order to calculate the affinity values between the antigen and the antibody a distance operator is used.

A. Hamming distance

Hamming distance [12] is the used to find out the best matching position between two binary strings. Hamming distance along with cyclic-shift operation between the strings in the matching process. The equation used in [13] to calculate the maximum hamming distance. It is calculates as follows:

$$HM(x_i^t, x_j^f) = max\{HD(x_{i'}^t, x_j^f)\}$$

where HM(x_i^t, x_j^f) and HD(x_i^t, x_j^f) are the max hamming and hamming distance between x_i^t and x_j^f respectively.

B. Shift continuous bit distance

Shift continuos bit distance [14] [15] is used for binary matching process. It is often necessity to adjust it in a certain position. The shift operation allows the movement of bit both in left and right direction. A shift operation is an operation that requires the operand to be represented in a binary format, viewed as a bit string, and then shift all bit values to the left or right.

C. Affinity

An antibody has a limited detection space where as antigen has a closer distance to it than the antibody which is based on immune theory. The affinity values determine the danger level of the file. Higher the value of affinity more likely virus are present in the file.

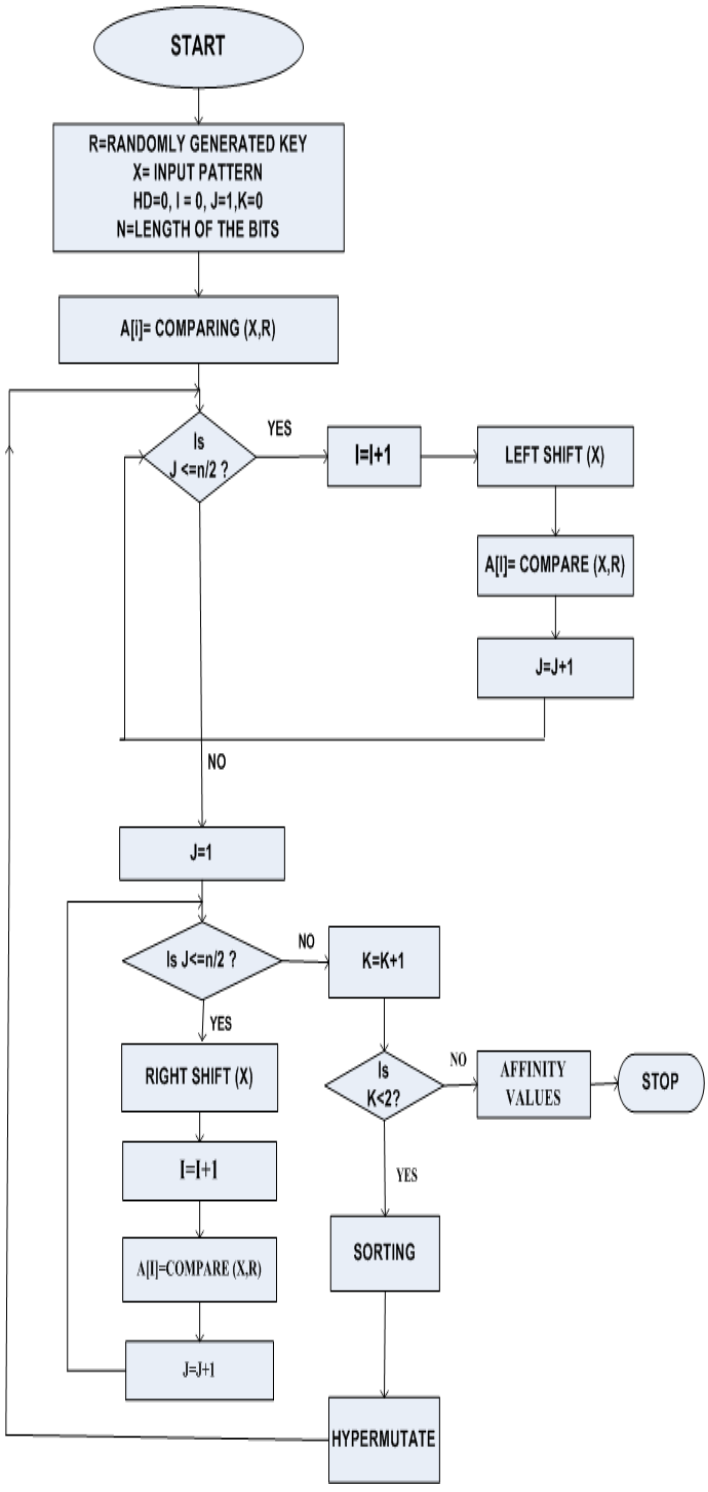


Fig. 2 Flow Chart of Calculating Affinity

The algorithm involves choosing a selected set of antibodies for cloning and development, then selecting a single best grown antibody for placement in the memory population.

In order to calculate the affinity values between the input antigen values and the randomly generated antibodies we have used the hamming distance and shifting operation. Hamming-max distance can avoid the influence of bits mismatching to enhance the ability of matching. Hamming distance along with shifting operation can get best hamming distance values i.e. max hamming distance. At first the anti-clones values are derived from the antibody values which is randomly generated.

A hamming distance value is found out by comparing the antigen values and the randomly generated antibody, then shifting of the bits takes place in the antigen values. At first left shift operation takes place, shifting a single bit of binary bit in the antigen value and comparing it with the randomly generated antibody, by which we will be getting another hamming distance values.

By repeating the left shift operation n/2 times, then right shift operation of the antigen values takes place in the similar way n/2 times. After which a number of hamming distance value is found then sorting the distance found by hamming function into descending order. It orders the anti-clone values. Then hyper mutating the anti-clones values i.e. 1's compliment of the anti clones values 0 converted to 1 and 1 converted to 0.

Again applying the shift operation on the hyper mutates values and finding the hamming distance. Lastly we get the affinity values if the selected antigen.

D. Proposed Algorithm

Step 1: Get the input pattern from the file

Step 2: Get the antibody values

Step 3: Anti-clone values are derived from the antibody values

```
Begin
Distance _H=0
Copy the value of a[i] to d[i]
For (i = 0; i <L; i++)
        If (a[i] =b[i])
                Increment Distance _ H
        End if
End for
Store the value of Distance _ H value in distance _ final [0]
Loop
        Until shift < n
                If shift < =n/2
                        For (i = 0; i <L; i++)
                Left shift a[i] and store in c[i]
                        End for
                End if
        Else
                        For(i=L;i>0;i--)
                Right shift d[i] and store in c[i]
                        End for
                For (i=0;i<L;i++)
                        If (c[i]==b[i])
                        Distance _H= Distance_H+1
                        End if
                End for
                Distance _ final[shift]=Distance _H;
        End loop
Find max value distance _ final
Return distance _max
end
```

Step 4: Sorting the distance found by hamming function into descending order.

It orders the anti-clone values.

Step 5: Hyper mutate: function compliments the anti-clone value: 1→ 0 or 0 → 1.

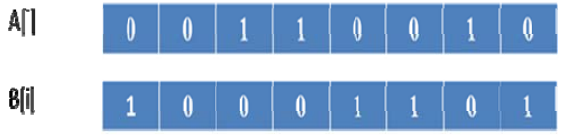
Step 6: Now the antibody values are changed based on the comparison of the distance of the antibody value and anti clone values by repeating the step 3.

Step 7: Final affinity value is calculated.

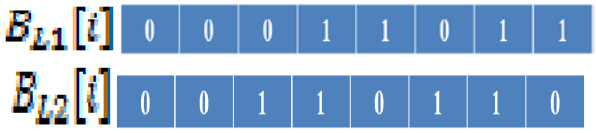
For eg: Let A[i] and B[i] be two binary bits. The number of shifting n=5, so the left shifting is done 2 times and right shifting is done 2 times.

A[i] = Antibody

B[i] = Antigen

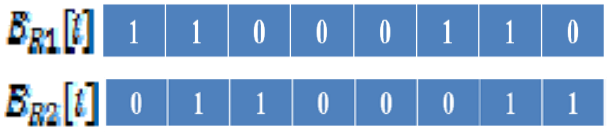


Without shifting hamming distance between A[i] and B[i] is 1. Now shifting each bit and then the hamming distance values is calculated. Left shift B[i] and the hamming distance values are:



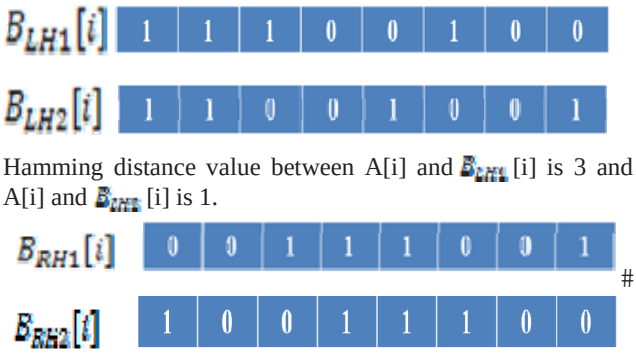
The hamming distance value between A[i] and B_{L1} [i] is 5 and A[i] and B_{L2} [i] is 7.

Right shifting B[i] and the hamming distance values are:



Hamming Distance value between A[i] and B_{R1} [i] is 3 and A[i] and B_{R2} [i] is 4.

Now hyper mutating the shifting value and then finding the affinity value.



Hamming distance value between A[i] and $B_{LH1}[i]$ is 3 and A[i] and $B_{LH2}[i]$ is 1.

Hamming distance value between A[i] and $B_{RH1}[i]$ is 5 and A[i]and $B_{RH2}[i]$ is 3.
The max hamming distance after hyper mutation is 5.

V. Conclusions

Thus inspired by Clonal Selection Process of AIS we have proposed an affinity maturation process in immune response we have presented an algorithm on it. The value provided by the function being optimized for a given input may be related with the affinity to the antigen. As mentioned earlier the algorithm involves stepwise shifting of the antigen values. Hyper mutation is applied to the anti clone values derived before sorting of the distance found by shifting. Shift operation is again applied to the hyper mutate values and then final affinity values are found out The purpose of these distance algorithms is to determine an affinity vector to a detector. This affinity vector is a measure of how alike the examined code is to the detector. A large affinity to the detector means that the file is likely a virus, whereas a small affinity vector value means the file is likely not a virus. The average affinity for a file is calculated to be its danger level. Thus improvement in the affinity values increases the probability for the selective antigens.

In artificial immune system, virus detection program shows a raise in the number of detected threats. If the technology grows plenty to make it into the conventional market, it could be a powerful tool to fighting malware which is currently dying the networks with spam and ruining lives by slipping sensitive data from vulnerable or inadequately protected users across the world.

VI. References

[1] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, 2009.

[2] Kephart, J.O. and Arnold, W.C.Automatic Extraction of Computer Virus Signatures. 4th Virus Bulletin International Conference, pp. 178–184, 1994

[3] Jungwon Kim and Peter J.Bentley,"Negative Selection and Niching by an Artificial Immune System for Network Intrusion Detection", pp.19–25. A late-breaking paper, Genetic and Evolutionary Computation Conference (GECCO '99), Orlando, Florida, USA, 1999.

[4] Jungwon Kim and Peter J.Bentley,"An Evaluation of Negative Selection in an Artificial Immune System for Network Intrusion Detection," Proceedings of the Genetic and Evolutionary Computation Conference (GECCO–2001), pp.1330–1337, 2001.

[5] F. M. Burnet, "Clonal Selection and After", in Theoretical Immunology, G. I. Bell, A. S. Perelson and G.H. Pimbley Jr., Eds., Marcel Dekker Inc., 1978, pp. 63-85.

[6] F.M.Burnet, "The Clonal Selection Theory of Acquired Immunity",Australian Journal of Science, Cambridge University Press, 1959,pp.67-69.

[7] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, 5th ed. 1998.

[8] William E. Paul. Immunology - Recognition and Response : Readings from Scientific America magazine / edited by William E. Paul., USA: Scientific American Inc., W. H. Freeman and Company, 1991.

[9] De Castro,L.N>and Von Zuben,F.J:The Clonal Selection Algorithm with Engineering Application,Proc.of GECCO'00,Workshop on Artificial Immune Systes and their Application,(2000),36-37.

[10] Aickelin, U., Greensmith, J., and Twycross, J.(2004) "Immune System Approaches to Intrusion Detection—a Review". In the Proceeding of the Third International Conference on Artificial Immune Systems (ICARIS–04), pp.316–329.

[11] S.Forrest, A.S.Perelson, L.Allen and R.Cherukuri, "Self-nonself discrimination in a computer", in Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, pp.16–17, Los Alamitos, CA: IEEE Computer Society Press, 1994.

[12] J.O.Kephart,"A biologically inspired immune system for computers", in Proceedings of Artificial Life IV, pp.6–9, MIT Press, Cambridge, MA, 1994

[13] Rui Chao,Ying Tan ,:"A Virus Detection System Based On Artificial Immune System," Computational Intelligence and Security,2009, pp.6-10

[14] P.D'haeseleer, "A change-detection algorithm inspired by the immune system: Theory, algorithms and techniques", Technical Report CS95–6. The university of New Mexico, Albuquerque, NM, 1995

[15] P.Helman and S.Forrest," An efficient algorithm for generating random antibody strings," Technical Report CS-94-07, The University of New Mexico,Albuquerque,NM,1994