ISSN:0975-9646

Negative Selection Method for Virus Detection in a Cloud

Agnika Sahu[#1], Prabhat Ranjan Maharana[*2]

[#1]*Wipro Technologies*
Bangalore, India

[*2]*KPIT Info Systems Inc.*
Michigan, USA

***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 Negative Selection Algorithm (NSA) is one of the inspired method to identify self bodies and non self bodies. The whole process is of self and non self body recognition and identifying them. In our proposed work only those cell which are self bodies are rejected as it harmless and non self bodies are send to detector which is harmful.**

***Keywords*— Artificial immune system, Negative selection process, Pattern Matching, 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. For a human body, various detector cells, called antibodies, are continuously generated and distributed to a whole body. The distributed antibodies monitor all living cells and detect non-self cells, called antigens, invading into a human body. 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 negative selection method is used to explain the basic characteristics of a self bodies and non-self bodies. The main goal of intrusion detection is to detect unauthorized use, misuse and abuse of computer systems by both system insiders and external intruders. It monitors any number of hosts on a network by scrutinizing the audit trails of multiple hosts and network traffic.

The new antibody can bind not only to harmful antigens but also to essential self cells. To prevent such serious damage, the human immune system employs negative selection. This process eliminates immature antibodies, which bind to self cells passing by the thymus and the bone marrow. From newly generated antibodies, only those which do not bind to any self cell are released from the thymus and the bone marrow and distribute throughout the whole human body to monitor other living cells. Therefore, the negative selection stage of the human immune system is important to assure that the generated antibodies do not to attack self cells.

The platform represents computer security technology that determines whether the other platform is venomous or not and to distinguish self bodies or non self bodies. 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]. When a new antibody is generated, the gene segments of different gene libraries are randomly selected and concatenated in a random order. 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 Negative 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 wsith 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. [3]

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.
5)

III. NEGATIVE SELECTION MECHANISM

The Negative selection algorithm defines 'self' by building the normal behaviour patterns of a monitored system. It generates a number of random patterns that are compared to each self pattern defined. If any randomly generated pattern matches a self pattern, this pattern fails to become a detector and thus it is removed. Otherwise, it becomes a 'detector' pattern and monitors subsequent profiled patterns of the monitored system. During the monitoring stage, if a 'detector' pattern matches any newly profiled pattern, it is then considered that new anomaly must have occurred in the monitored system. [4]

The purpose of negative selection is to provide allowance for self cells. It deals with the ability to detect unknown antigens while not reacting to the self cells. During the generation of T-cells, receptors are made through a pseudo-random genetic re-arrangement process. Then, they undergo a censoring process in the thymus, called the negative selection. There, T-cells that react against self-proteins are destroyed and only those that do not bind to self-proteins are allowed to leave the thymus.

These matured T-cells then circulate throughout the body to perform immunological functions and protect the body against foreign antigens. Detectors are randomly created and then undergo a maturation phase where they are matched with self, connections. If the detectors match any of these they are eliminated otherwise they become mature. These mature detectors start to monitor new connections during their lifetime. If these mature detectors match anything else, exceeding a certain threshold value, they become activated. This is then reported to a human operator who decides whether there is a true anomaly. If so, the detectors are promoted to memory detectors with an indefinite life span. It is known as negative selection as only those detectors (antibodies) that do not match live on.

This algorithm defines self by building the normal behaviour patterns of a monitored system. It generates a number of random patterns that are compared to each self pattern defined. If any randomly generated pattern matches a self pattern, this pattern fails to become a detector and thus it is removed. Otherwise, it becomes a detector pattern and monitors subsequent profiled patterns of the monitored system. During the monitoring stage, if a detector pattern matches any newly profiled pattern, it is then considered that new anomaly must have occurred in the monitored system.

The basic principles of the negative selection theory are [5] [6]:

1. Define self as a collection S of elements in a feature space U, a collection that needs to be monitored. For instance, if U corresponds to the space of states of a system represented by a list of features, S can represent the subset of states that are considered as normal for the system.
2. Generate a set R of detectors, each of which fails to match any string in S. Discard those that match any element in the self set.
3. Monitor S for changes by continually matching the detectors in against S, if any detector ever matches, then a change is known to have occurred, as the detectors are designed not to match any of the original strings in S.

Inspired by the positive and negative selection processes that occur during the maturation of T cells in the thymus called T cell tolerance. Negative selection refers to the identification and deletion of self-reacting cells that is T cells that may select for and attack self tissues. It is typically used for classification and pattern recognition problem domains where the problem space is modelled in the complement of available knowledge. For example in the case of anomaly detection it prepares a set of example pattern detectors trained on normal (non-anomalous) patterns that model and detect unseen or abnormal pattern.

Negative selection mechanism is inspired by the main mechanism that produces a set of mature T-cells capable of binding only non-self antigens. The first negative selection algorithm was proposed to detect data manipulation caused by a virus in a computer system [7].

The negative 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 key.
3) Partial matching of string takes place, between the randomly generated sting and input sting.
4) If there is partial matching between of the string then those cells are called self bodies and those cells which don't match are called as non self bodies.
5) The self bodies are rejected and the non self bodies are send to detector.

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.

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. [3]

IV. PROPOSED WORK

The objective of this algorithm is to distinguish self bodies and non-self bodies. Here pattern matching and r shift continuous method is used in order to get more set of non self bodies and their samples of non self bodies are kept in the detector ,which can be future used in detecting them.

A. Pattern Matching

Pattern matching is defined as a checking sequence of tokens for the presence of the constituents of same pattern. In Pattern recognition, the match usually has to be exact. The patterns generally have the form of either sequences or structures. Use of pattern matching include outputting the locations (if any) of a pattern within a token sequence, to output some component of the matched pattern, and to substitute the matching pattern.

This technique is usually performed on a computer, by which a group of characteristic properties of an unknown object is compared with the comparable groups of characteristics of a set of known objects, to detect the identity or proper classification of the unknown object. Each string should continuously match with substring of other string

Let S is a string and s1; s2;sn are the substring.S 2 s1; s2;sn

These substrings are match with the other string in order to check whether which substring is matching and which one is not matching.

B. Shift continuous bit distance

Shift continuous bit distance [8] [9] 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.

The algorithm starts with dividing of the input pattern into a number of substrings and then these substrings are matched with the randomly generated key. If the substring is matched with the randomly generated key, it is known as self set and which is not affected by virus where as if the substring don't matches with the randomly generated key it is known as non-self set and which can be affected by virus. Then these non-self set are send to detector to check the danger level of the affected files.

As the string is broken down into a number of substring so virus can be present in any combination of substring so after checking the whole string once we have used a shift operation, so that the randomly generated string could match with all combination of bits.

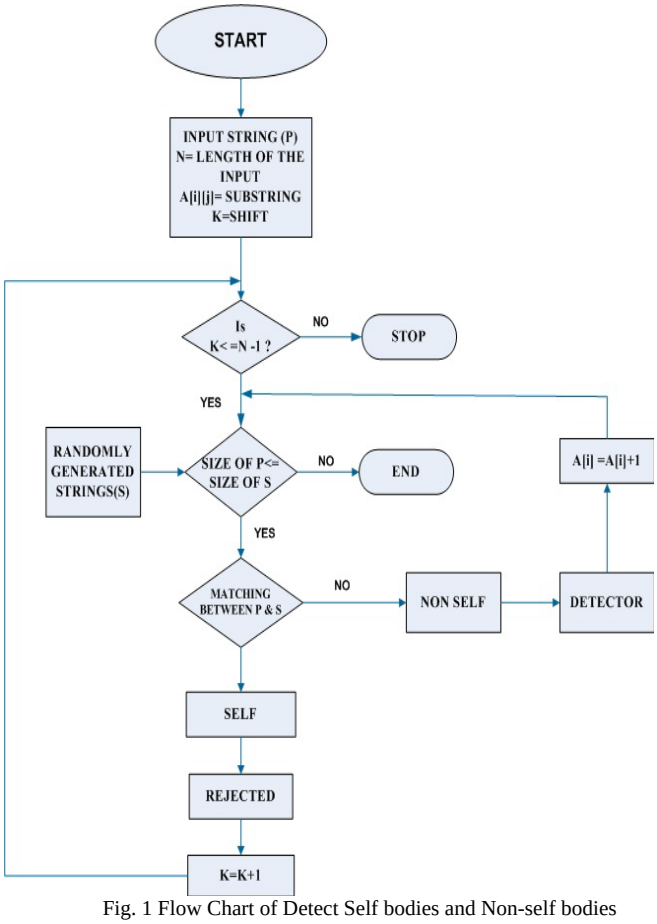


Fig. 1 Flow Chart of Detect Self bodies and Non-self bodies

Self set are a pattern of stable behaviour of a system/process:

1. A collection of logically split segment (equal-size) of pattern sequence.
2. Represent the collection of substring of length S of a string.

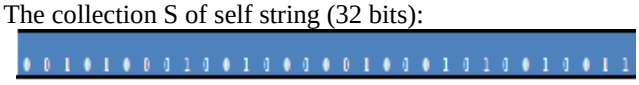
Generators are set of detector R, each of which fails to match with the randomly generated string.

C. Proposed Algorithm

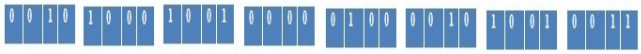
Step 1: Get the input pattern from the file
Step 2: Get the antibody values
Step 3:

```
Begin
    Repeat
      Randomly generate string and place them in S
        For every Input pattern (P)
        While (K<= N- 1)
            While (Psize <=Ssize)
                If (P=S)
                    then self bodies
                else
                    non self bodies
                end if
             Self bodies are rejected where as non self
             bodies are send to detector
            end while
    end while
end
```

Step 4: After once pattern matching takes place shifting of string(S) takes place.
Step 5: Now a new 32 bit string i.e.S1 is generated after shifting.
Step 6: Now again pattern matching process takes place between a new randomly string (R) and Sting S1 by repeating the step 3.
Step 7: Final all the non self bodies are send to detector and self bodies are rejected.
For Example: Let S and R are two array of binary string.
The collection S of self string (32 bits):



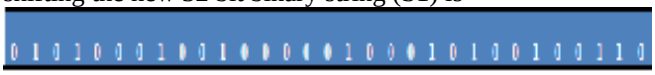
The self string S is broken down into eight substrings, each of length four:



The Randomly(R) generated string (16bits):

The collection S of self (sub) strings (S contains all of the substrings) is to compared with the Randomly generate strings (R0), and then match the strings of R0 against the strings in S. Strings from R0 that match self are eliminated. Strings that do not match any of the strings in S are send to detector collection (R), which is called the non-self.
The R contains the following four substrings:

After pattern matching two substrings from R i.e. the strings 1000 and 1001 being eliminated because they each match a string in S and those strings are called as self. The remaining substrings i.e. 0111 and 0101 are send to detector and those string are called as non self.
Now r –shift continuous process takes place and after shifting the new 32 bit binary string (S1) is



Now pattern matching process again takes places between a new randomly generated string and S1.

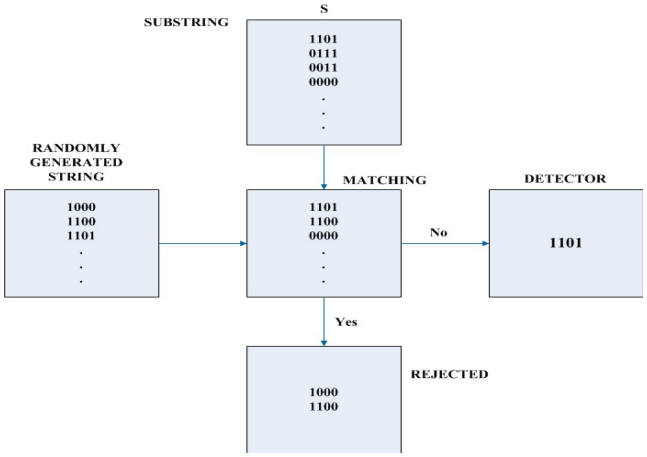


Fig. 2 Matching the Substring with Randomly generated string

V. Conclusions

Thus inspired by Negative Selection Process of AIS matching of input substring with the randomly generated string takes place, by which self bodies can be separated from non self bodies. If the input substring matches with the randomly generated string the self bodies are rejected and which doesn't matches are send to detector set, which may be affected virus. After each and every substring is being matched then shifting operation takes place and a new string is formed. Again the process continues in order to find more non self bodies.

In artificial immune system (AIS), 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] Agnika Sahu and Tanmay Swain, "Clonal Selection Method for Virus detection in a Cloud", International Journal of Computer Science and Information Technologies, Vol. 2 (3), 2011, 1149-1153.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] P.Dhaeseleer, "A Change Detection Algorithm Inspired by the Immune System: Theory, Algorithms and Techniques," Technical Report CS956. The university of New Mexico, Al-buquerque, NM, 1995.
[9] 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

.