Artificial Immune Systems for Information Security: Comparative Analysis of Negative and Positive Selections

Sergei Bezobrazov
Brest State Technical University, Moskovskaja str. 267, 224017 Brest, Belarus, bes@bstu.by

Abstract: *The artificial immune system is a new, perspective system for protection of computer systems from viruses. Training and selection of detectors is necessary in artificial immune system, as it prevents break-in of unnecessary detectors. This paper presents comparative analysis of two methods detectors selection: negative and positive selections. The results of experiments are discussed in the paper.*

Keywords: artificial immune system, anomaly detection, negative selection, positive selection.

I. INTRODUCTION

Evolution of new information technologies give not only unique opportunities for active development of economics, politics, state and society, but also stimulate appearance and evolution computer crime. Striking examples of computer crime are creation and distribution of computer viruses.

New viruses are appearing permanently, which use security vulnerability of operating systems. Number of viruses is increase (Figure 1) and damage from it is rise [1].

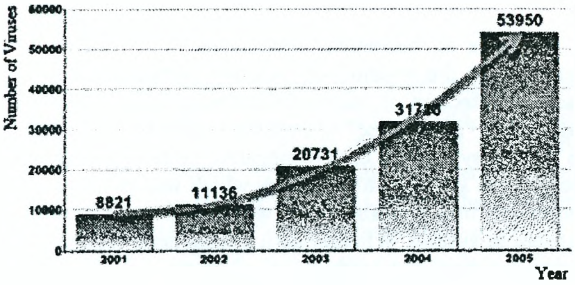

Fig. 1. Increase in the number of new viruses

Modern antivirus software doesn't secure computer systems in full measure. Traditional antivirus software has series of weaknesses. Some of them:

- Developer's mistakes in interpretation and understanding of new unknown virus. As a result antivirus software doesn't work correctly;
- Actual antivirus bases are needed for successful detection viruses, as a rule bases located on website of developer. It takes a long time for lurking and downloading new antivirus bases. The last viruses spread all over the world for several hours and antivirus software with outdated bases were powerless in the face of new security threat;
- Some viruses can infect antivirus software;
- Heuristic algorithms which use antivirus software for detecting unknown viruses wide of the perfection. In practice computer users usually deal with misoperation of these algorithms and finally disable heuristic analyzer.

All this reduce to search of new methods in information security construction. One of such methods is Artificial Immune System's method which based on basic principles of biological immune system [2]. Biological immune system is a unique system of defending body from harmful bacterium's and viruses. Immunity is based on synthesis of special proteins – antibodies, which capable to bind with foreign material – antigens. The immune system is capable to memorize and to keep the information about viruses, which infected body, in immune memory. This ability allows immune system to cope with repeated infection very quickly.

Artificial Immune System (AIS) have some powerful capabilities, such as pattern recognition, future extraction, immune memory, learning, adaptability, distributed structure [3].

This paper presents comparative analysis of two methods detectors (antibodies) selection: negative and positive selections. We have developed two models of artificial immune systems. The first model used a method of negative selection, and the second - a method of positive selection.

The rest of the paper is organized as follows. The section 2 describes algorithm of negative selection. In the section 3 the positive selection algorithm is described. In the section 4 the artificial immune system is described, based on negative selection and positive selection. Section 5 presents experimental results. Conclusion is given in section 6.

II. NEGATIVE SELECTION METHOD

Basic elements of immune system are lymphocytes - white cells. Lymphocytes are formed from stem cell in a bone marrow. Young lymphocytes go to thymus and to lymph nodes where they undergo to multiple-stage training and selection. Detectors which react against 'self' (organism's cells) are destroyed. As a result survive only those detectors which don't react against 'self'. Mature lymphocytes have detectors on the surface. They are capable to detect a specific antigen. Lymphocytes circulate through the body and implement the immune function - detection of harmful bacteria [4].

Selection of detectors is necessary, as it prevents break-in of unnecessary detectors. In computer system generation of antibodies represents random process. Detectors are generate at random, therefore there is a probability of that some from them will react against 'self' files. The mechanism of selection prevents penetration of undesirable antibodies in system. The most used algorithm of selection is the algorithm of negative selection proposed by S. Forest [5]. The essence of negative selection is: antibodies are compared with self files. If the detector reacts against 'self' file it declares as the negative detector and destroyed. Survive only those detectors which are structurally different from test 'self' files. In result mature detectors will ignore clean files and detect viruses. Detection will occur in that case when the detector "to meet" a file structurally similar with the detector (Figure 2).

The algorithm of negative selection can be presented as follows:

- Set *S* of 'self' files is determined;
- Randomly generates set of detectors *R*;
- From set *R* each detector is compared with each 'self' file from *S*;
- If the detector and a 'self' file are similar enough, the detector is destroyed, else the detector "introduced" in system.

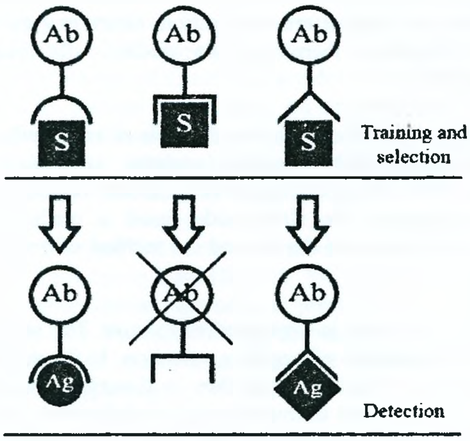

Fig. 2. Negative selection algorithm (Ab – detector, S – test 'self' file, Ag - antigen)

Thus the purpose of negative selection is to provide tolerance for self cells. It deals with the immune system's ability to detect unknown antigens while not reacting to the self cells.

III. POSITIVE SELECTION METHOD

The method of positive selection is exact antithesis of a method of negative selection [5]. In contrast to negative selection, the method of positive selection is guided by development of detectors which are structurally similar with 'self' files. In process of training detectors get the structure maximum similar with structure of 'self' files. And in process of selection those antibodies which are not sim'lar to 'self' files are destroyed. As a result in computer system penetrate and circulate detectors which are structurally similar with 'self' files. If, at comparison with files, there is a detection of structural difference between the detector and a file, artificial immune system gives the signal about anomaly detection (Figure 3).

The algorithm of positive selection can be presented as follows:

- Set *S* of 'self' files is determined;
- Randomly generates set of detectors *R*;
- From set *R* each detector is compared with each 'self' file from *S*;
- If the detector and a 'self' file are not similar structurally the detector is destroyed, else the detector "introduced" in system.

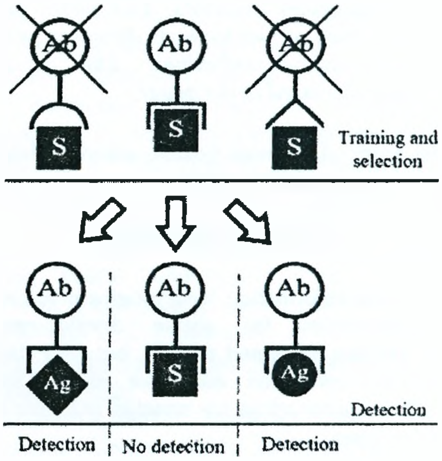

Fig. 3. Positive selection algorithm (Ab – detector, S – test 'self' file, Ag - antigen)

Thus the purpose of positive selection is to acquiring of structure maximum similar to structure 'self' files.

On the basis of theoretical conclusions it is possible to assume, that the method of positive selection is more exact for detection of malicious software as the structure of detectors "is adjusted" to structure 'self' files, i.e. real-life files of system. In case of negative selection we try to receive such structure of detectors which would be similar to structure of a probable virus which at present moment is absent in system. Whether so it in practice? We will try to compare these two methods of selection.

IV. SYSTEM DESCRIPTION

We had been created the elementary artificial immune system composed of several modules: detectors generation, detectors training and selection, the module of detection, cloning and mutation, genetic memory. We will consider these modules in details (figure 4).

A. Detectors generation

Randomly generates 500 detectors, each of which represents a binary string in the size from 32 till 256 bit.

B. Detectors training and selection

Training and selection of detectors is realized using the elementary genetic algorithm [?]. Detectors are compared with predetermined test files. The most able detectors which are a material for the next iterations of genetic algorithm are selected from an initial population. After selection detectors undergo a crossover - the pair parents are selected and randomly determine a break point. After that both half of detectors are crossed, forming new detectors. After crossing there is a process of mutation. The mechanism of mutation consists in entering at random of marginal changes into structure of the detector. The mutation allows detectors to get new, desirable properties which are absent in parents. The next iterations of genetic algorithm occur by analogy of premises: selection - crossover - mutation.

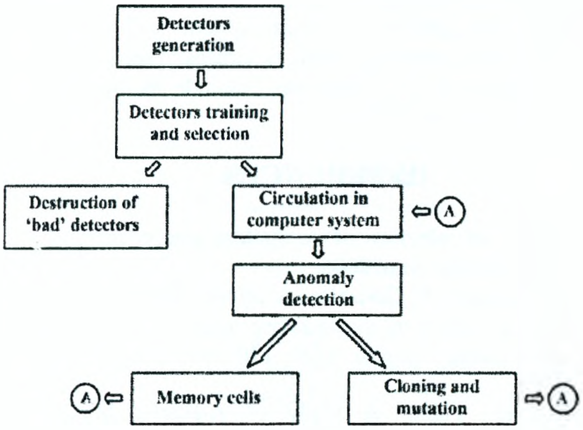
Detectors generation
Detectors training and selection
Destruction of 'bad' detectors
Circulation in computer system
A
Anomaly detection
A
Memory cells
Cloning and mutation
A

Fig. 4. Model of artificial immune system

C. *The module of detection*

"Mature" detectors check by turns a file set consisting of system utilities, various documents and viruses. The rule of comparison consists in search of identical values in corresponding positions in file and in detector. The detectors size is much less than files size. Therefore at comparison the method of "data windowing" was applied, i.e. the detector gradually moved on all length of a file, and total value of coincidence was accepted equal to the maximal value of a window. When there is a detection of enough of concurrences, the artificial immune system give the signal about detection of anomaly.

D. *Cloning and mutation*

As a rule, the viruses infecting computer systems infect a large quantity of files. Process of cloning and mutation is applied to the fast detection of the infected objects. The detector which has detect a virus in system, is exposed to process of cloning, i.e. a large quantity of copies of this detector is created. During cloning, in their structure makes randomly small changes (mutation). The mutation allows clones to get the structure maximum similar with structure of detected virus. Process of cloning and a mutation allows artificial immune system to detect operatively all infected objects of computer system.

E. *Genetic memory*

Genetic memory meant for keeping information about viruses which ever infected computer system. Carriers of this information are memory cells. Memory cells are copies of detectors which detected viruses. Memory cells have the improved properties in comparison with usual detectors. Due to genetic memory the artificial immune system easily deal with repeated infection of computer system.

V. EXPEREMENTAL RESULTS

A. *Negative selection*

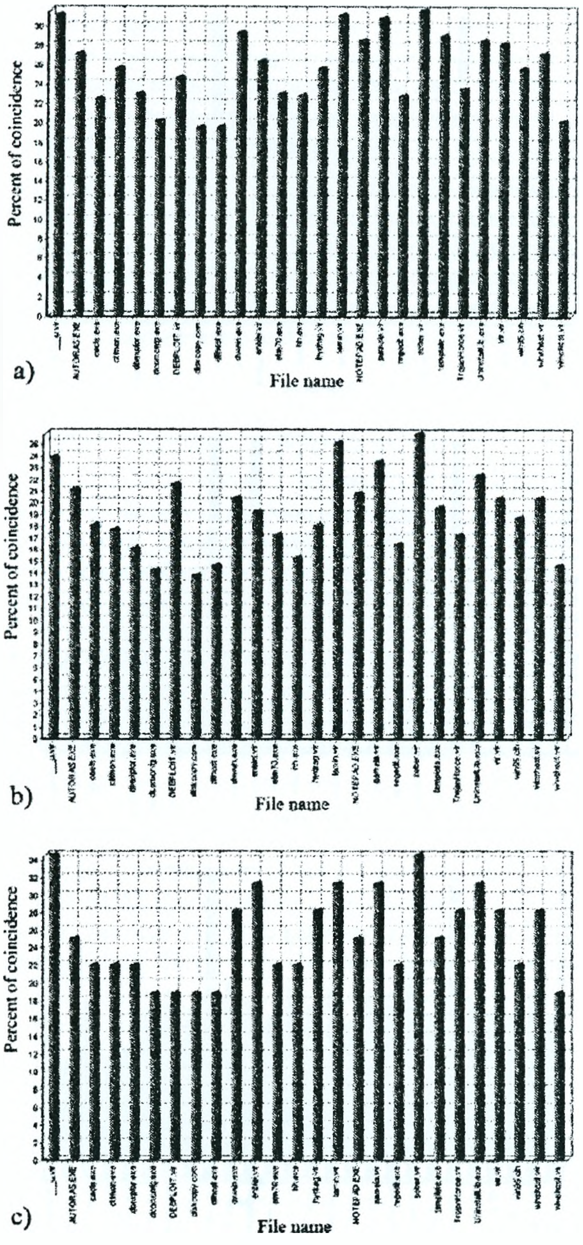

Fig. 5. a), b), c) – results of negative selection

Using negative selection for selection of undesirable antibodies, training took (in our case) an interval approximately equal 2 minutes. However in the detection sometimes there were insignificant errors, i.e. clean files were detected as viruses. In other words misoperation took place.

B. Positive selection

Using positive selection training took (in our case) an interval approximately equal 3 minutes. For training artificial immune system required larger timetable and computational burden. But results were with a high accuracy.

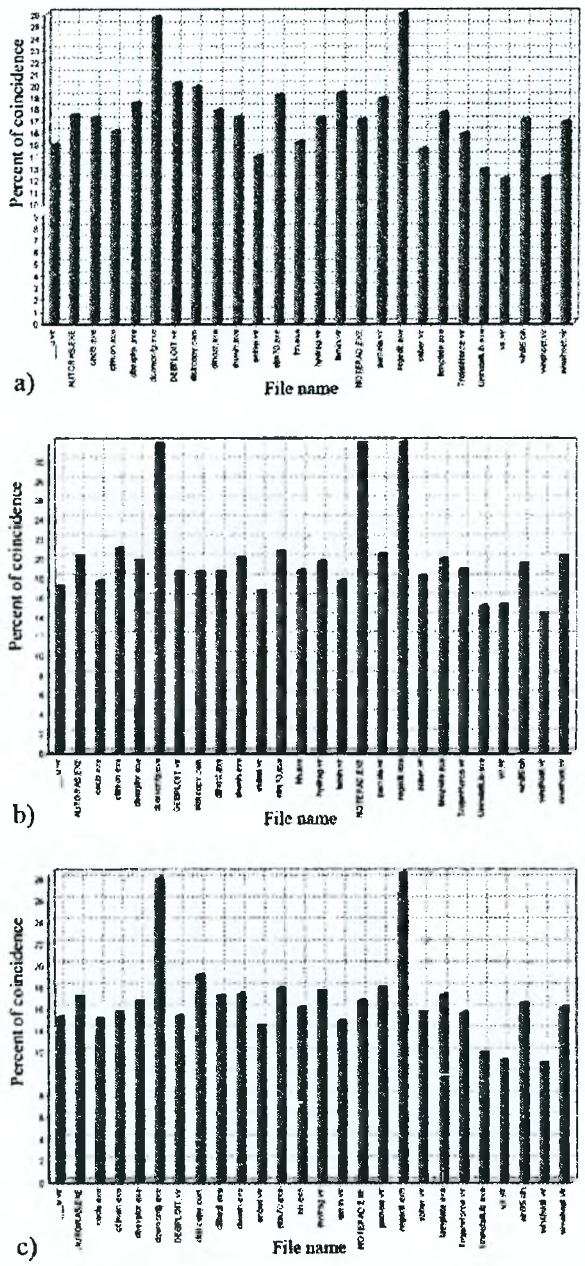

Fig. 6. a), b), c) – results of positive selection

VI. CONCLUSION

Using negative selection for training and selection of 'bad' detectors provides a gain in time and computing. However supposes appearance of misoperations for which elimination additional calculations are necessary.

Method of positive selection is slowly. It needs the bigger time and computing expenses. However the percent of misoperations is very small.

It is significant that training of detectors depend on test files on which there is a training. If files, which mean destructive functions, (for example, format.com or fdisk.com in OS Windows,) do not enter into a set of test files, that the artificial immune system with a high probability will determine them as viruses. Therefore for training detectors it is necessary include as much as possible various 'self' files.

REFERENCES

1. Danger of infection of computers yearly increases. www.viruslist.com, 2005.

2. D. Dasgupta, F. Gonzales. Artificial immune system (AIS) research in the last five years. *Evolutionary Computation*, pages: 123- 130 Vol.1, 8-12 Dec. 2003.

3. Gonzlez, Fabio. A study of Artificial Immune Systems Applied to Anomaly Detection. *PhD. Dissertation*, The University Of Memphis, May 2003.

4. A. Janeway and P. Travers. Immunobiology: the immune system in health and disease. *Current Biology Ltd.*, London, 2nd edition, 1996.

5. In. A formal framework for positive and negative detection. *IEEE Transactions on Systems, Man, and Cybernetics* 34:1 pp. 357-373, 2004.