Developing an Algorithm for the Application of Bayesian Method to Software Using Artificial Immune Systems

Shafagat Mahmudova (shafagat@gmail.com)
Institute of Information Technology of ANAS https://orcid.org/0000-0003-1817-0756

Research Article

Keywords: Immune system, Software Protection systems, Bayesian method Malware,

Posted Date: June 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-538798/v1

Developing an algorithm for the application of Bayesian method to software using artificial immune systems

Shafagat Mahmudova

Institute of Information Technology of ANAS, Baku, Azerbaijan

e-mail: shafagat_57@mail.ru

ORCID ID: 0000-0003-1817-0756

Abstract

This paper develops a new algorithm by applying the Bayesian method to software using artificial immune systems. An artificial immune system is an adaptive computing system that uses models, principles, mechanisms, And functions used to solve problems in theoretical immunology. Its application to various fields of science is studied. The role that artificial immune systems play in software is invaluable. Methods for detecting malware are explored. Some works in the field of artificial immune system are analyzed and issues to be addressed are identified. The Bayesian method accurately calculates the probability of occurrence of any event under certain conditions. Therefore, the Bayesian method is applied to software using artificial immune systems. By applying this method, fast software performance can be achieved. For this, a new algorithm is developed and experiments are conducted. The developed algorithm is one of the new ones. The results of the experiments provide good performance.

Keywords

Immune system Software Protection systems Bayesian method Malware

Communicated by Harini Balasubramanian.

1. Introduction

The expansion of information technology has led to the development of many fields of science. The technologies based on the working principles of the human body have gained wide popularity: artificial intelligence, visual processing tools, artificial retina, all types of genetic algorithms and so on. The information systems based on the working principles of immune systems have great potential in many areas (Gavrilyuk et al. 2010).

The immune system in medicine is a complex protection mechanism, the main function of which is to protect the body against harmful external substances, toxins, microorganisms and harmful cells.

Protecting the living organism from the persistent impact of dangerous external and internal factors promotes the development of the immune system. The immune system has a destructive response to endogenous substances and does not adversely affect the tissues of the body (De Castro 2002).

Most immunological reactions are short-term and controlled by regulatory mechanisms that impede very strong reactions.

Tolerance is a set of mechanisms in which the immune system prevents destructive reactions to its own body.

Two mechanisms of tolerance are provided in Figure 1.

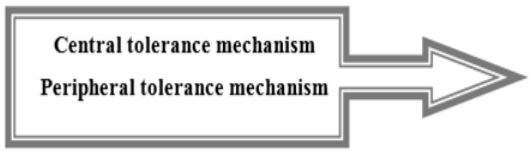

Fig. 1. Two mechanisms of tolerance

Most lymphocytes located in all primary lymphoid organs and acting against their own antigens of the body are destroyed by central tolerance mechanisms.

Immunology is a field of science about the structure and working rues of the immune system, diseases and immunotherapy. Are studies the biological mechanisms of self-protection of the body against any foreign substance.

Artificial Immune System (AIS) is an adaptive computing system that uses the models, principles, mechanisms and functions to describe and solve the problems in theoretical immunology.

Although the natural immune systems have not been completely studied so far, today, there are at least three theories explaining the functioning of the immune system and its interactions (Figure 2).

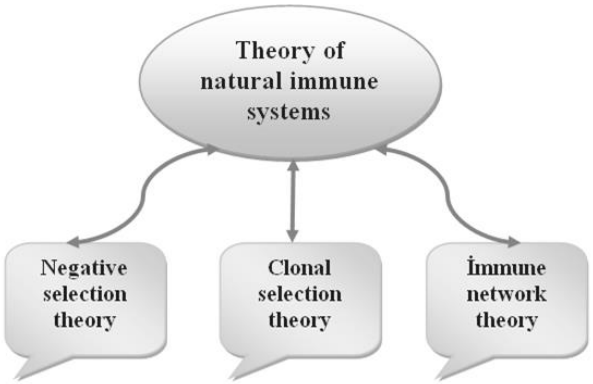

Fig. 2. Theory of natural immune systems

AIS appeared in the works on immune networks in the mid-1980s in the articles by Fermer, Packard and Perelson (1986) and Bersini, Varela (1990). However, AIS was founded only in the mid-1990 s. Forrest and Kephart (Kephart 1994) published the first article on AIS in 1994, and Dasgupta conducted extensive research on the theory of negative selection. Hunt and Cooke started working on the theory of the immune network in 1995. Timmis and Neal continued this work and improved it. The first book on artificial immune systems was edited by Dasgupta in 1999 (Dasgupta 1999).

The technology based on the same principles as a human body is now widely used: the neural networks based on artificial intelligence, visual image processing, artificial retina, and various genetic algorithms. The information systems based on the principles of immunity are very promising for many areas. Currently, artificial immune systems mainly use on type of artificial intelligence to protect against computer viruses, to detect network interventions, and so on.

Here another process also arises where the natural selection of antibodies occurs during clonal selection: survival of only those who live under the identified external body. At the same time, the information about the emerging antibodies is "entered" to the gene library mentioned above. Thus, the gene database contains only the information about the highest threats. The vital features of the immune system needed are highlighted below (Figure 3).

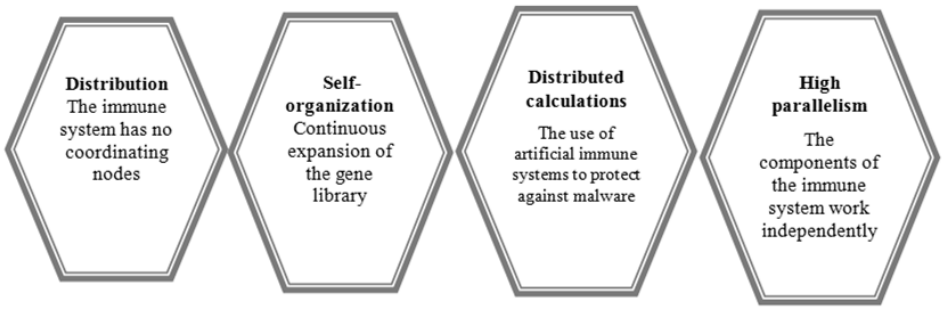

Fig. 3. The key features of the immune system

Unlike the available protection systems, the above-mentioned system does not have a central control system; it is a centralized high-level parallel distributed data processing and analysis system and is particularly beneficial for the protection of distributed computing environments.

AIS comprise the following algorithms:

• **Clonal Selection Algorithm** - A class of algorithms based on the clone selection theory of the obtained tolerance, explaining that the lymphocytes B and T improve their response over time. These algorithms are based on the attributes of Darwin's theory, where the selection is based on the change based on the convergence of the interaction of the antigens and antibodies, and on the cell's division principles and on the somatic hypermutations;

- **Negative Selection Algorithm**;
- **Immune Network Algorithm;**
- **Dendrite Algorithm**.

AIS are a class of automated computing systems based on the principles and processes of the immune system. Typically, such algorithms use the memory and learning capability of the immune system to solve the problem.

2. Related Works

Some of the available studies in the field of AIS are reviewed below.

1. Many studies have been targeted at solving the complex technological problems inspired by immune systems. These problems may include abnormal detection, pattern recognition, system security, and data collection, etc.

2. Large-scale software systems are often difficult to manage and control. In many cases, unexpected events occur in these systems, especially after the upgrades or changes in their environment (e.g. when updating the operating system, etc.). Therefore, in order to avoid this in a changing environment, there is a need for self-adaptive methods to detect errors and monitor performance (Ligeiro 2014). There are similarities between the detection of program errors and the problem of detecting pathogenic microorganisms found in natural immune systems. Inspired by the vaccine and negative clonally selections observed in these systems, an effective adaptive model for software monitoring is developed by analyzing the system resource indicators.

3. Artificial Intelligence (AI) represents the intelligence of the person presented in the machine and software. It is a highly demanded academic field in many modern researches. Leading researchers in AI define this area as "learning and designing the intelligent agents." The term was first invented by McCarthy in 1955 and designated as "the science and technology of creating intelligent machines." The main objectives of AI research are comprehension, knowledge, planning and training, natural language processing (communication), perception and ability to move and manipulate the objects. In fact, the interdisciplinary field of AI is quite broad and includes many sciences and professions, including computer science, psychology, linguistics, philosophy, neurology, and so on. This area was built on the idea of Homo Sapiens as a central intellectual property of people "that an intelligent can describe the mind so precisely that a machine can be created to model it." AI has been a subject of great optimism; however, it has surprisingly failed (Sniecinski et al. 2018).

4. Several traffic signal control systems are developed the intersection traffic control based on optimization techniques and artificial intelligence. The proposed system is applied to the modeled intersection using modern traffic modeling software VISSIM (a visual programming language designed for dynamic systems modeling). The obtained results suggest that the offered system is recommended for use in extreme conditions associated with blocked approaches and high traffic, and that it is competitive and capable to control various traffic scenarios (Louati et al. 2018).

5. The immune system of the human body has a great potential for protecting it against many harmful viruses and external objects. Throughout history, people have been infected with microorganisms. To limit the nature, dimensions and intensity of these microbial invasions, humans have the ability to cope with them. The human immune system is capable to protect the body, skin, cells and tissues from external effects. [8] presents an example of the study of the human immune system through the mathematical models of adaptive immune systems. Extensive simulations are performed to study the effects of external particles on the recovery mechanism of the body. The results confirm the validity of the immunological mathematical model of a human. A strong security and confidentiality in the human body can help to build a strong networking system (Rathore et al. 2018).

6. Dynamic risk identification is used to predict the future risks based on possible risk data. A training model of risk identification technology based on Fuzzy Support Vector Machine (FSVM) is able to fully and automatically identify potential risks, which becomes a key method for the dynamic risk identification. Selection of FSVM parameters is crucial for improving the recognition efficiency and accuracy. Artificial immune algorithm (AIA) is an effective technology for stochastic global optimization and has the benefits of high precision, convergence. Therefore, a new dynamic risk identification model based on the integration of FSVM and the immune optimization algorithm (IOA) is proposed in the work (Bo Yang 2019).

7. (Fanelli Robert 2008) presents a hybrid model for detecting the network intrusions, thus, it incorporates the artificial immune system techniques and the traditional data protection methods. Using the abnormalities detection function, and based on the network monitoring, it comprises the programs to detect intrusions to the congenital immune function based on the network threats detection, misuse, and immunological hazard modeling. The trials

determine the improved detection accuracy as compared to the detection of interventions based on misuse. Further researches and the areas that are important for model improvement are discussed.

8. Proposes an Arithmetic Optimization Algorithm (AOA) (Laith et al. 2021), a new meta-heuristic method which uses the distribution behavior of the main arithmetic operators in mathematics including (Multiplication (M), Division (D), Subtraction (S), and Addition (A)). In a wide range of search spaces, AOA is mathematically modeled and implemented for optimization processes.

9. A population-based optimization algorithm Sine Cosine Algorithm (SCA) was introduced by Mirjalili in 2016. He was motivated by the trigonometric sine and cosine functions. Several SCA variants and applications are available in the literature. A series of computational experiments to validate the performance of the SCA against similar algorithms are conducted and provide good results (Laith et al. 2021).

10. The basic concept of triangular neutrosophic cubic hesitant fuzzy number and their properties are defined (Amin et al 2019). A triangular neutrosophic cubic hesitant fuzzy ordered weighted arithmetic averaging (TNCHFOWAA) operator and a triangular neutrosophic cubic hesitant fuzzy ordered weighted geometric averaging (TNCHFOWGA) operator are developed to aggregate triangular neutrosophic cubic hesitant fuzzy number (TNCHFN) information, and their properties are explored.

As the review of related studies shows, the use of artificial immune systems in various areas is very important and it provides effective performance.

The advantages and disadvantages of protection systems based on artificial immune systems are shown in Figure 4.

Advantages	Disadvantages
A large number of detectors lead to permanency and reliability of the system.	Autoimmune reaction is possible.
No single point of rejection (denial).	Possible autoimmune reaction
Increasing number of nodes in the computing environment distributed in the proposed system also contributes to increased security.	Immunodeficiency is especially possible at a small number of nodes in a distributed computing environment.
All collisions of detectors with malicious objects are stored in memory. This allows for training of detectors	

Fig. 4. Advantages and disadvantages of protection systems based on artificial immune system

3. Methods for malware detection

One of the most serious threats to data security of automated systems is malware. Malware can steal confidential information from computer, infect files, spread throughout computer or entire local network, cause loss of confidential data, encrypt all stored data for malicious purposes and damage the data owners economically, and so on. Malware should be detected and removed as soon as possible to prevent malicious activity (Tokarev et al. 2017).

Malware may conditionally include viruses, worms, Trojans, spyware, adware, etc. (Yevdokimov 2012).

1. Hiding methods. These methods alter the syntax of the program without changing its semantics, making it more difficult to analyze and comprehend malicious codes (Venkatesan 2008). This can be achieved by changing the registers where the variables are placed, but the program code remains unchanged; it changes the rules of execution of processes dealing with other harmful instructions with exchange instructions, etc.

2. Encryption code. Malware encryption includes password decryption/encryption operations, encryption keys, and specific encrypted malicious code itself.

3. Encrypted Malware may include the following types:

- Oligomorphic programs make small changes such as encryption, descriptions that make it difficult to detect some of the features of the code through malware signaling systems;
- Polymorphic programs that encrypt with different encryption algorithms while using a large number of encryption algorithms and keys, and is based on each new infection, because their detection is very complex and time consuming;
- Metamorphic programs that can completely change themselves being unlike the original version, which is the most difficult to detect.

Malware detection systems are often the first defense of a protected computer system. Most of these methods can be classified into the following three classes.

1. **Signature methods** are widely used to detect malware. Although they are very effective, if there is a very small change in the code, it, in turn, changes the signature and becomes ineffective. In addition, this method requires regular updating of the signature database to detect new malware (Biryukov 2012);
2. **Behavioral methods** provide continuous monitoring of the behavior of the program to determine if it is malicious or not. Experiments show that these methods have high levels of false positives.
3. **Heuristic method** primarily uses machine learning and data mining to determine software behavior (Kris 2010).

4. Development of an algorithm using Bayesian method

The significance of the artificial immune systems for software is already reported in previous sections. With the use of Bayesian method, it is possible to determine how protective the immune systems are for software. Bayesian method is briefly explained below.

The Bayesian method accurately calculates the probability of occurrence of any event under certain conditions. Therefore, the Bayesian method is applied to software using artificial immune systems.

Bayesian theorem (or Bayesian formula) is one of the basic theorems of elementary probability theory that enables a precise calculation of the probability of an event under a certain condition (Gmurman 2005). Bayesian formula may be derived from the main axioms of probability theory, particularly from the conditional probability. Bayesian theorem results in a large number of calculations required for its application in practice. Therefore, Bayesian evaluations began to be actively used only after the revolution in computer and network technologies (Daniel 2005).

Bayesian formula is shown below:

$$P(A \mid \mathrm{B}) = \frac{P(B \mid \mathrm{A})\, \mathrm{P(A)}}{P(B)} \tag{1}$$

Here

$\mathrm{P(A)}$–priori probability of a hypothesis of A;

$\mathrm{P(A \mid B)}$ –probability of hypothesis A at the time of event B (a posterior probability);

$\mathrm{P(B \mid A)}$ – probability of event B in case of hypothesis A – is true;

$\mathrm{P(B)}$ – complete probability of event B.

$\mathrm{P\ (A \mid B)\ P\ (A \mid B)}$–conditional probability of AA if event BB occurs.

Experiment. Examples may include temperature maintenance, bridge building scenarios or as virtual entities for use purely in software applications. The frequency of virus spreading within software is 0.001, and the method for immune system protection is 0.9. In this case, the probability of the failure of the positive result is 0.01. Here, the probability of the program being protected by the immune system has to be found, that is, to prove that the virus is false. During the verification, it is necessary to find the probability of uninfected software which is assumed to be infected with the virus.

I– software is infected;

«I» – software is verified to be infected;

H - indicates software is uninfected.

Then the condition given is written as follows:

$P(«I» \mid I) = 0{,}8;$

$P(«I» \mid H) = 0{,}02;$

$P(I) = 0{,}002;$

$P(H) = 0{,}888.$

It is assumed that the software is uninfected, if it is considered to be infected, then it is equal to the conditional probability:

$P(H \mid «I»).$

To find it, let's calculate the probability of being fully infected:

$P(«I») = 0{,}888 \times 0{,}02 + 0{,}002 \times 0{,}8 = 0{,}01936.$

It is assumed that the software is uninfected, and if as a result it is "infected":

$P(H \mid «I») = 0{,}888 \times 0{,}02 / (0{,}888 \times 0{,}02 + 0{,}002 \times 0{,}8) \approx 0{,}917.$

Thus, software assumed to be infected with virus, in fact, is estimated to be uninfected 91.7% during the examination.

If the results of the examination can be considered as a random error, then the re-examination of same software will be independent of the first result. In this case, the software should be re-examined in order to reduce the false positive probability of the results if it is estimated to be infected with virus. It is assumed that the software is uninfected after obtaining the re-examination result as infected with virus, moreover, with the Bayesian formula; it can be calculated as follows:

$P(İ \mid «H», «H») = 0{,}888 \times 0{,}02 \times 0{,}02 / (0{,}888 \times 0{,}02 \times 0{,}02 + 0{,}002 \times 0{,}8 \times 0{,}8) \approx 0{,}2172.$

Here, three software are used for the experiment. The presented rule is applied to each software and the results are summarized in Table 1. Note that the software for these systems is developed by the author of the article (Table 1) and (Figure 5) (Alguliyev 2005).

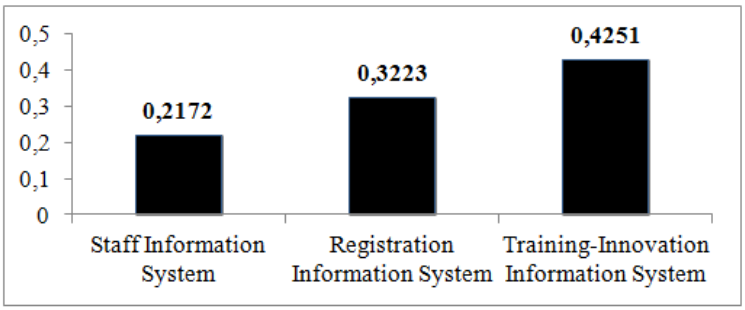

Fig. 5. The probability of 3 systems being uninfected according to the Bayesian formula

Table 1. The probability of 3 systems being uninfected according to the Bayesian formula

System software	Values obtained with Bayesian formula
Staff Information System	0,2172
Registration-Journal Information System	0,3223
Training-Innovation Information System	0,4251

5. Conclusion

The research highlighted important role of AIS in all fields of science. AIS use different algorithms and can be applied to solve various problems. The use of AIS in programming is of particular importance. AIS are used to "clean" malware (viruses, worms, Trojans, spyware, etc.). In this article, a new algorithm was developed and experimented for the application of Bayesian method to software using AIS. The advantages and disadvantages of AIS were shown. To eliminate the disadvantages, perfect AISs should be developed to enable the software more efficient and effective. The application of the Bayesian method results in the fact that, its practical application does not need huge calculations. By applying this method, fast software performance can be achieved. Mention that the proposed method could be optimized using any optimizer.

Notes

Acknowledgements

The author thanks the editors and anonymous reviewers for their helpful comments and suggestions that have led to this improved version of the paper.

Compliance with ethical standards

Conflict of interest

The author declares no conflict of interest.

Compliance with ethical standards

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the author.

Publisher's Note

References

Amin F., Fahmi A (2019) Human immunodeficiency virus (HIV) infection model based on triangular neutrosophic cubic hesitant fuzzy number. International Journal of Biomathematics. 12(5).

https://doi.org/10.1142/S1793524519500554

Google Scholar

Alguliyev R.M, Kazimov T.H., Mahmudova Sh.J., Mahmudova R.S (2005) Corporate Information System « Educational Center». Educational Technology. In: The Inter. Conf., ICTE', (ENG & TECH, Canakkale), 8, p. 294.

Google Scholar

Biryukov A.A (2012) Informat sionnaya bezopasnost': zashchita i napadeniye. DMK Press, Moskva.

Google Scholar

Bo Yang (2019) Mathematical Evaluation of Human Immune Systems For Securing Software Defined Networks Safety Science. In: 6th International Conference on Wireless Networks and Mobile Communications (WINCOM).118: 205.

https://doi.org/10.1109/WINCOM.2018.8629728

Google Scholar

Daniel K (2005) Judgment under Uncertainty: Heuristics and Biases. University Press, Cambridge.

Google Scholar

Dasgupta D (1999) Artificial Immune Systems and Their Applications Artificial Im. Sys. and Their Appl. Springer, Berlin.

Google Sc holar

De Castro, Leandro N (2002) Artificial Im. Sys. A Artificial immune systems: a new computational intelligence approach, New Computational Intel. Approach. Springer, Berlin.

Google Scholar

Fanelli Robert L (2008) A Hybrid Model for Immune Inspired Network Intrusion Detection Artificial Immune Syst. In: International Conference on Artificial Immune Systems (Springer-Verlag, Berlin Heidelberg. 5132: 107.
https://doi.org/10.1007/978-3-540-85072-4_10

Google Sholar

Gavrilyuk S. V, Onyky B. N., Stankevichus A. A (2010) Primeneniye iskusstvennykh immunnykh sistem dlya zashchity sred raspredelennykh vychisleniy ot vredonosnogo programmnogo obespecheniya Bezopasnost. Informatsionnyye Tekhnologii. 17 (3): 32

Google Scholar

Gmurman V. Ye (2005) Teoriya veroyatnostey i matematicheskaya statistika. Vyssheye obrazovaniye, Moskva. Google Scholar

Kephart J (1994) A biologically inspired immune system for computers . In: Fourth Int. Workshop on the Synthesis and Simulation of Living Systems. MA: MIT Press, Cambridge, p. 130.

Google Scholar

Kris K (2010) Komp'yuternyye virusy iznutri i snaruzhi .SbP, Piter. p. 528.

Google Scholar

Laith A., Ali D., Seyedali M., Mohamed Abd E., Amir H.Gandomi (2021) The Arithmetic Optimization Algorithm. Computer Methods in Applied Mechanics and Engineering. 376: 113.

https://doi.org/10.1016/j.cma.2020.113609

Google Scholar

Ligeiro R (2014) Monitoring applications: An immune inspired algorithm for software-fault detection. Applied soft computing. 24: 1095.

http://dx.doi.org/10.1016/j.asoc.2014.08.021

Google Scholar

Laith A. & Ali D (2021) Advances in Sine Cosine Algorithm: A comprehensive survey. Artificial Intelligence Review. 54: 2567-2608.

https://link.springer.com/article/10.1007%2Fs10462-020-09909-3

Google Scholar

Louati S., Darmoul S., Elkosantini, S., Ben Said L (2018) An artificial immune network to control interrupted flow at a signalized intersection an arti. Information Sciences. 433: 70.

https://doi.org/10.1016/j.ins.2017.12.033

Google Scholar

Rathore H., Guizani M, Mohamed A (2018) Mathematical Evaluation of Human Immune Systems For Securing Software Defined NetworksWireless Net.and Mob. Comm. In: 6TH In. Conf. (IEEE, Morocco), p. 187. https://doi.org/10.1109/WINCOM.2018.8629728

Google Scholar

Sniecinski I, Seghatchian J (2018) Artificial intelligence: A joint narrative on potential use in pediatric stem and immune cell therapies and regenerative medicine. Trans. and Aphaeresis Science. 57(3): 422. https://doi.org/10.1016/j.transci.2018.05.004

Google Scholar

Tokarev V.L. Sychugov A.A (2017) Obnaruzheniye vredonosnogo programmnogo obespecheniya s ispol'zovaniyem immunnykh detektorov. Voprosy Zashchity Informatsii. 216.

Google Scholar

Venkatesan A (2008) Code Obfuscation and Virus Detection Master's Projects. 116.San Jose State University. DOI: https://doi.org/10.31979/etd.ez5v-x8jc

Google Scholar

Yevdokimov KN (2012) K voprosu o sovershenstvovanii ugolovnoy otvetstvennosti za sozdaniye, ispol'zovaniye i rasprostraneniye vredonosnykh programm dlya evm (ST. 273 UK RF) Pravo i Zakonodatel'stvo. 136.

Google Scholar

Figures

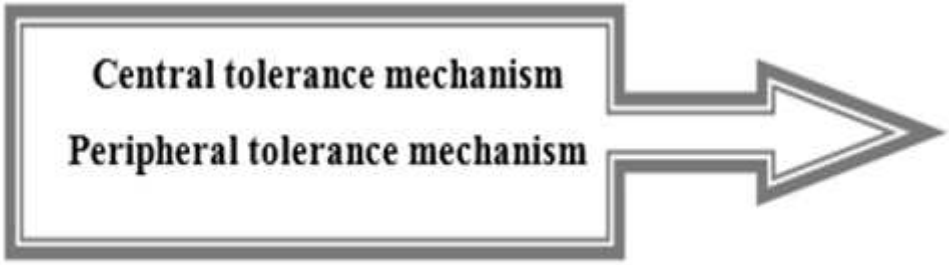

Figure 1

Two mechanisms of tolerance

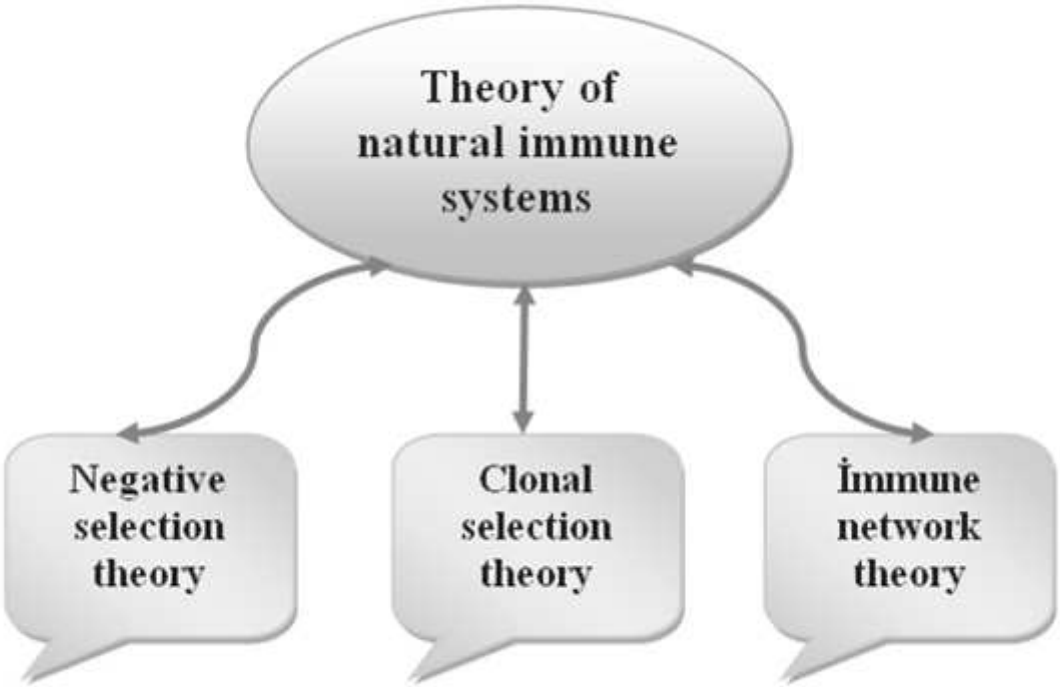

Figure 2

Theory of natural immune systems

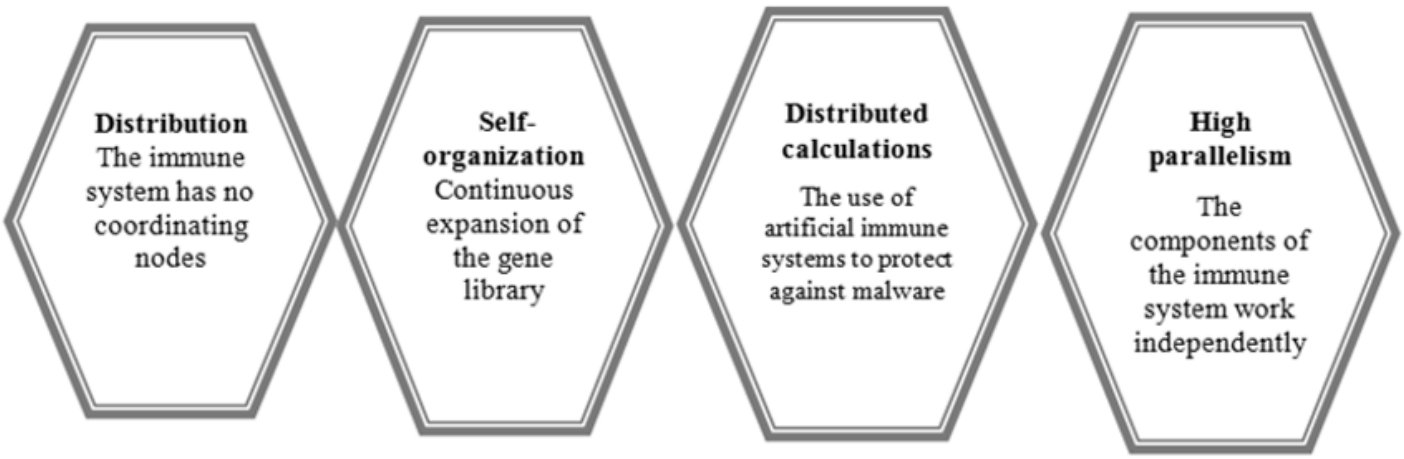

Figure 3

The key features of the immune system

Advantages	Disadvantages
A large number of detectors lead to permanency and reliability of the system.	Autoimmune reaction is possible.
No single point of rejection (denial).	Possible autoimmune reaction
Increasing number of nodes in the computing environment distributed in the proposed system also contributes to increased security.	Immunodeficiency is especially possible at a small number of nodes in a distributed computing environment.
All collisions of detectors with malicious objects are stored in memory. This allows for training of detectors	

Figure 4

Advantages and disadvantages of protection systems based on artificial immune system

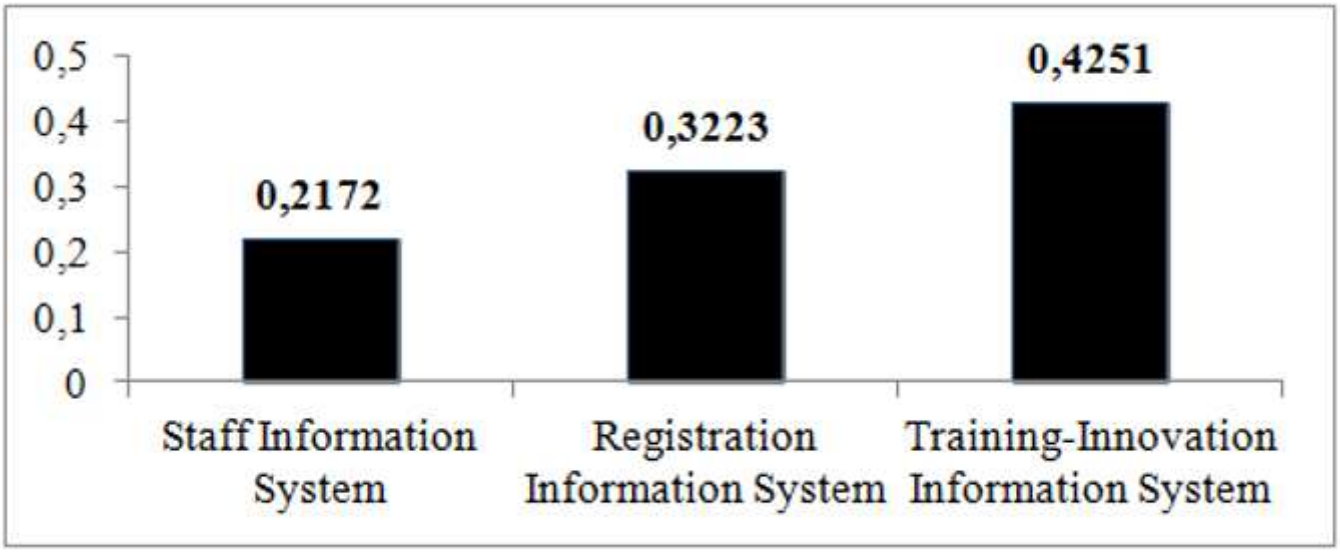

Figure 5

The probability of 3 systems being uninfected according to the Bayesian formula