ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

A Survey - Appropriateness of Negative Selection and GeneticAlgorithm for Network based Intrusion Detection

Arsha R R[1],Anju J S.[2]

P.G. Student, Department of Information Technology,Govt. Engineering College, Bartonhill, Trivandrum, India[1]

Assistant Professor, Department of Information Technology,Govt. Engineering College, Bartonhill, Trivandrum, India[2]

Abstract:In the area of computer security, purpose of Intrusion Detection (ID) not only attempts to discover abnormalaccess to computers by analyzing various interactions but also reporting all the abnormal behaviors of the system. Due to the growing of internet applications, the needs of security are increasing. The use of artificial immune algorithms in intrusion detection is an appealing concept for two reasons. Firstly, the human immune system provides the human body with a high level of protection from invading pathogens, in a robust, self-organized and distributed way. Secondly, current techniques used in computer security are not able to cope with the dynamic and increasingly complex nature of computer systems and their security. The main goal of HIS is to differentiate self from potentially harmful non-Self . An intrusion detection system (IDS) is a type of security software designed to automatically alert administrators when someone or something is trying to compromise information system through malicious activities or through security policy violations. It serves as an extra wall for protecting critical systems due to the ability of hackers to subvert other protection systems like firewalls. A network-based intrusion detection system (NIDS) is used to check and analyze network traffic to protect a system from network-based threats. A NIDS searches for any suspicious patterns by checking all inbound packets.

Keywords:Artificial immune system, Negative selection algorithm (NSA), Intrusion detection system(IDS), Genetic algorithms (GA).

1. INTRODUCTION

The normal behavior of a computing system can be characterized by observing its properties over time. The problem of detecting intrusions can be viewed as finding non permitted deviations of the characteristic properties in the monitored network system. This assumption is based on the fact that intruders' activities in some way different from the normal users' activities. However, it may be very difficult to realize or detect such differences in real-time before any damage has been done. The existing immunity-based intrusion detection methods emulate one or the other mechanisms of the natural immune system and shown as promising in detecting some type of intrusions.

The first NSA algorithm which was proposed by forest et. al. is an exhaustive approach. The limitation of this approach is the computational difficulty of generating valid detectors, which grows exponentially with the size of the self . So, to solve the problems of the exhaustive approach, we need a technique for implementing the NSA which locates a detector instead of selecting them at random as in the case of the exhaustive approach. For this, we use an evolutionary approach using a genetic algorithm (GA).

An adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics are termed as Genetic Algorithm. As such they represent an intelligent exploitation of a random search used to solve optimization problems. Each generation consists of a population of character strings that are analogous to the chromosome that we see in our DNA. Each individual represents a point in a search space and a possible solution. The individual in the population are then made to go through a process of evolution. GAs are basically used in IDS to generate rules used to detect anomalies[4] . They were inspired by the biological evolution (development), natural selection, and genetic recombination. GAs use data as chromosomes that evolve through: selection (usually random selection), cross-over (recombination to produce new chromosomes), and mutation operators. Finally, a fitness function is applied to select the best (highly fitted) individual. The process is repeated for a several generations until reaches the individual (or group of individuals) that closely meet the desired condition[5] . GAs are very promising in the computer security field, especially in IDS. They have been applied for intrusion detection since the 1990's , and still being used up till the current time. GA is usually used to generate rules for intrusion detection, and they usually take the form if condition then action, where the condition part used to detect the anomalous ones as they test the fields of incoming network connections .

2. THE NEGATIVE SELECTION ALGORITHM

Artificial Immune System (AIS) covers many models inspired by the biological immune system. The first model, negative selection algorithm (NSA), among AIS models was introduced by Forrest et al. Many researches have been performed after the introduction of NSA. These researches proposed various NSA, and they are differentiated in data representation, detector representation, self-definition and matching rule.

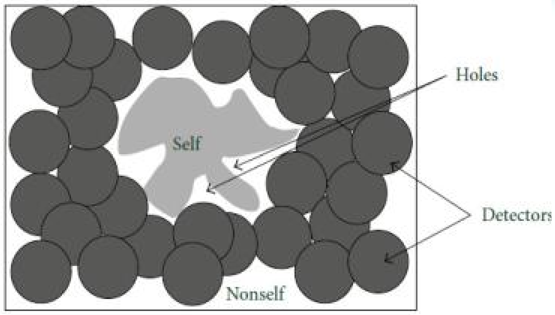


Fig.1 The NSA

Fig.1 shows Randomly generate candidate detectors (represented by dark circle); if they match any self (i.e., if any of the points covered by the detector are in the self-set), they are eliminated and regenerated until getting enough valid detectors

Forrest et al (1994; 1997) proposed and used anegative selection algorithm for various anomaly detectionproblems. This algorithm defines self by buildingthe normal behavior patterns of a monitored system. Itgenerates some random patterns that are compared toeach self pattern defined. If any randomly generatedpattern matches a self-pattern, this pattern fails to becomea detector and thus it is removed. Otherwise, itbecomes a detector pattern and monitors subsequentprofiled patterns of the monitored system. During themonitoring stage, if a detector pattern matches anynewly profiled pattern, it is then considered that newanomaly must have occurred in the monitored system.

Most works in negative selection used the problemin binary representation [6]. There are at leasttwo obvious reasons of this choice: first, binary representationprovides a finite problem space that is easier toanalyze; second, binary presentation is straightforwardto use for categorized data. However, many applicationsare natural to be described in real-valued space.Furthermore, these problems can hardly be processedproperly using negative selection algorithm in binaryrepresentation [8]. On the other hand, this work andsome other works [9][10] demonstrated that despite theintrinsic difficulty of real-valued representation, it canalso provide unique opportunity in dealing with higherdimensionality.

Matching rule is one of the most important componentsin a negative or positive pattern detection algorithm[6][8][11][12]. For binary representation,there exist several matching rules like rcb (r-contiguousbit), r-chunks, and Hamming distance [6][8]. For realvaluedrepresentation, however, the Euclidean distanceis primarily used [8][9][10][13]. Matching is determinedwhen the distance between a data point and somedetector is within a certain threshold. In some cases,variations of Euclidean distance are used, such as, a Euclideandistance defined in a lower dimensional spaceprojected from the original higher dimensional problemspace [13]. Independent of the type of matchingrule, the detectors usually have some basic characteristics,e.g., the number of bits, r, in binary representation,or the distance threshold, to decide a matching in realvaluedrepresentation, that are constant throughout theentire detector set. However, the detector features canreasonably be extended to overcome this limitation.

2.1. Self-nonself (SNS) Model

SNS model focuses on the adaptive nature of the immune system, i.e., it uses the adaptive immune system and its memory or self-learning capability. In this model, the B cells (which are called detectors in AIS) would have antigen specific receptors that can recognize non-self or foreign bodies and in turn initiate an immune system response that is specific to the system where this AIS model is applied. In this technique, the first step according to Forrest et al. [1] involves randomly generating detectors (which is the AIS's equivalent of B cell in HIS). These detectors that are still immature are then exposed to a set of self structures. Any detector that reacts or matches any member of the self set is eliminated. The remaining members of the detector set that were nonreactive with any member of the self set become mature detectors. This detector selection technique is called negative selection and the algorithm used to perform this computation is called a negative selection algorithm (NSA)[1].

When a mature detector encounters a pathogen or non-self entity that it has been exposed to previously during the negative selection phase, it mounts a very rapid and efficient response that is called the secondary response. When a new pathogen that does not bind with any mature detector is encountered, the immune system mounts a primary response during which the immune system tries to learn the pattern of this unseen pathogen so that it can mount a secondary response next time the same pathogen shows up.

2.2. Study on NSA

Hofmeyr and Forrest [19],[14] proposed the artificial immune system (ARTIS) method, and they applied it to intrusion detection. This method represents detectors as bit strings, and uses the r contiguous matching rule. Beside these, the study defines the life-cycle of a detector, so it provides dynamic detector populations and adaptation ability in a continuously changing environment. In this life-cycle, a detector can be in one of the five states: immature, mature, activated, memory or death Gonzalez et al. [16] presented the effects of the low-level representation and its matching rules on the performance of NSA in covering the non-self space. This study indicates that the matching rule for NSA needs to be chosen when it represents data accurately in problem space. They explored and compared the different binary matching rules: r-contiguous matching, r-chunk matching, Hamming distance matching, and Rogers and Tanimoto matching.

Gonzalez et al. [17] proposed a Real-Valued Negative Selection (RNS) algorithm. RNS algorithm uses real numbers to represent self/non-self space RNS algorithm and binary NSA were compared for anomaly.detection problem. Then, merits and demerits of the real-valued representation were presented based on the binary representation. Real-valued representation have some advantages : closer to original problem space, allowing the use of methods from computational geometry to speed-up the algorithms, facilitating the use of other machine learning methods to find useful high level knowledges are the advantages of real valued representation [20]. But it have some Disadvantages like making analysis of the problem space harder, not suitable for the representation of categorical attributes. [7] discussed about a method, Wireless sensor networks utilize large numbers of wireless sensor nodes to collect information from their sensing terrain. Wireless sensor nodes are battery-powered devices. Energy saving is always crucial to the lifetime of a wireless sensor network. Recently, many algorithms are proposed to tackle the energy saving problem in wireless sensor networks.

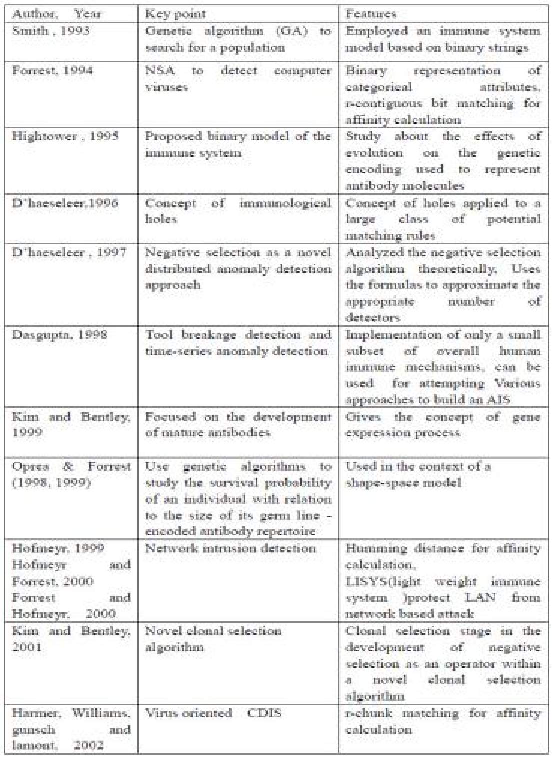
Dasgupta and Gonzalez [15] explored positive selection and negative selection, and they were compared using real-valued representation. Detectors are represented as rectangle with real numbers. Based on this comparison, advantages and disadvantages of these approaches were described. This comparison showed that positive selection is more precise, but it needs more time and space resources. The negative selection is less precise, but it needs fewer time and space resources. Real-valued representation is used in many applications due to the nature of applications domains, i.e. intrusion detection from network traffic.

The non-self coverage gets difficult for the problems with natural real-valued representation. This is because, the real-valued space is continuous and the boundary of self and non-self is ambiguous in this space. Therefore, the non-self coverage is a major issue for real-valued NSA (RNSA) [17], [18], [21], [22], [23], [24]. Detector representation and self-definition are the determinant for the non-self coverage. A part of research have been focused on the detector representation and distribution in the non-self space to maximize the coverage[18],[23],[21]. On the other hand, the recent research is focused on adaptive-self that implicates the variable self radius [25], [22], [24] . The self radius is an important value to control the detection rate and false alarm rate. In real-valued NSA, the detectors are usually represented as circles or rectangles for two dimensional problems. Nevertheless, some NSA use mixture of specific geometrical shapes to represent the detectors. However, some of NSA generate the detectors with different sizes. Based on the data and detector representation, the matching rule is changed, and Euclidean distance matching is usually used in real-valued representation.

Dasgupta and Gonzalez [15], [26] represent the detectors generated by genetic algorithm as rules. Gonzalez et al. proposed [18] a Randomized Real-Valued Negative Selection Algorithm (RRNS). This algorithm takes the detector radius and the self variability threshold (self sample radius) as parameters, so each self sample and detector is represented as circles in twodimensional problem space. These circles have a fixed size specified by the relevant parameter. Based on the self radius parameter, the algorithm uses Monte Carlo method to estimate the volume of self region.

Balachandran et al [23] present a work focused on developing a framework for generating multishaped detectors in real-valued NSA. This new extended realvalued NSA uses multiple shape (sphere, rectangle or ellipse) detectors for covering two-dimensional non- Self space. In this NSA, self space is also specified by the constant self radius parameter. Ji and Dasgupta [21],[27],[28],[29] proposed a new real-valued NSA, which generates variable size detectors. In this NSA, the detectors are represented as circles in twodimensional space and the radii of these circles are variable. On the other hand, the radius for all self samples is taken as the constant parameter and used to check whether a new generated detector is in any self circle or not. If it is, then discarded, otherwise the distance between the center of detector and the nearest self sample is assigned to this detector radius. This is called boundary-aware method [28]. Zeng et al [22] introduce a self-adaptive negative selection algorithm (ANSA). ANSA can adapt the varieties of self/non-Self space by adjusting self radii and detectors radii. Yuel et al [24] worked on optimization of self set for real-valued NSA

The main problem of the negative selection algorithm is a severe scaling problem. The definition of larger self set was essential to cover diverse types of network intrusions[12].



Author, Year	Key point	Features
Smith , 1993	Genetic algorithm (GA) to search for a population	Employed an immune system model based on binary strings
Forrest, 1994	NSA to detect computer viruses	Binary representation of categorical attributes, r-contiguous bit matching for affinity calculation
Hightower , 1995	Proposed binary model of the immune system	Study about the effects of evolution on the genetic encoding used to represent antibody molecules
D'haeseleer,1996	Concept of immunological holes	Concept of holes applied to a large class of potential matching rules
D'haeseleer , 1997	Negative selection as a novel distributed anomaly detection approach	Analyzed the negative selection algorithm theoretically. Uses the formulas to approximate the appropriate number of detectors
Dasgupta, 1998	Tool breakage detection and time-series anomaly detection	Implementation of only a small subset of overall human immune mechanisms, can be used for attempting Various approaches to build an AIS
Kim and Bentley, 1999	Focused on the development of mature antibodies	Gives the concept of gene expression process
Oprea & Forrest (1998, 1999)	Use genetic algorithms to study the survival probability of an individual with relation to the size of its germ line - encoded antibody repertoire	Used in the context of a shape-space model
Hofmeyr, 1999 Hofmeyr and Forrest, 2000 Forrest and Hofmeyr. 2000	Network intrusion detection	Humming distance for affinity calculation, LISYS(light weight immune system)protect LAN from network based attack
Kim and Bentley, 2001	Novel clonal selection algorithm	Clonal selection stage in the development of negative selection as an operator within a novel clonal selection algorithm
Harmer, Williams, gunsch and lamont, 2002	Virus oriented CDIS	r-chunk matching for affinity calculation

TABLE I. COMPARISON OF OPPORTUNISTIC ROUTING PROTOCOLS

3. THE GENETIC ALGORITHM

Genetic algorithms have proven to be an enormously powerful and successful problem-solving strategy[3]. Genetic algorithms have been used in a wide variety of fields to find solutions to problems that are more difficult than those faced by human designers. Thus, the solutions they come up with are often more efficient, more elegant, or more complex than anything comparable a human engineer would produce.

The detector set will increases with the development of the system. but it is impossible to generate the detectors dramatically for a finite system. As the intrusion pattern changes rapidly, we must eliminate old and invalid detectors. Instead of deleting it, Hofmeyr suggested making the detectors dynamic [2]. Ayara et al. and Gonzlez and Dasgupta [10] tried to give detectors a period of time before eliminating them. Kim and Bentley investigated a further extension of DynamiCS [30]; when memory detectors show a poor degree of self-tolerance to new antigens, they will be eliminated. Li proposed a receptor editing inspired real NSA [31].If new detectors are generated by taking some feedback from previous detectors instead of random, then the new detector can be better suited for the non-Self antigens. Hightower et al. [32], Perelson et al. [33], and Oprea and Forrest [34] employed a Genetic Algorithm (GA) to study the effects of evolution in the genetic encoding of the antibody molecules, which can be seemed as a feedback strategy.

The input to the GA is a set of potential solutions to that problem, encoded in some fashion, and a metric called a fitness function that allows each candidate to be quantitatively evaluated. These candidates may be solutions already known to work, with the aim of the GA being to improve them, but more often they are generated at random.

The GA then evaluates each candidate according to the fitness function activity; the candidate with good fitness has high chances to get selected than the one with average fitness. Various functions are basically used to test the fitness of any particular individual. These individual with high fitness can be termed as promising candidates. These promising candidates are kept and allowed to reproduce. From them multiple copies are made, but the copies are not perfect; random changes are introduced during the copying process. These digital offspring then go on to the next generation, forming a new pool of candidate solutions, and are subjected to a second round of fitness evaluation. Those candidate solutions which were worsened, or made no better, by the changes to their code are again deleted; but again, purely by chance, the random variations introduced into the population may have improved some individual, making them into better, more complete or more efficient solutions to the problem at hand. Again these winning individual are selected and copied over into the next generation with random changes, and the process repeats.The expectation is that the average fitness of the population will increase each round, and so by repeating this process for hundreds or thousands of rounds, very good solutions to the problem can be discovered.

4. CONCLUSION

The increased network connectivity and easy access to information and resources through Internet and World Wide Web makes the security issues one of the most important factors in today’s computing. The promise of Electronic Commerce also contributing to the explosive growth of the Internet and the underlying communication networks. Though there are many security-related products and technologies, yet the potential threats and vulnerabilities are intractable. Intrusion detection is an important part of computer security. It provides an additional layer of defense against computer misuse (abuse) after physical, authentication and access control. Different models of intrusion detection have been developed, and many IDS software available for use. Commercial IDS products such as NetRanger (www.cisco.com), Real Secure (www.iss.net), Omniguard Intruder Alert (www.axent.com) work on attack signatures. These signatures needed to be updated by the vendors in order to protect from new types of attacks. There is a working group established to design a common intrusion detection framework (CIDF) for providing a common intrusion specification language. However, no detection system can catch-all types of intrusions and each model has its strength and weaknesses in detecting different violations in networked computer systems. An influx of new approaches is needed to enhance security measures.

Researchers have been exploring various

artificial intelligence based approaches for intrusion/misuse detection. Recent works on immune-based computer security emulated one or the other functional components of the natural immune system. In particular, Forrest et al. used a negative selection algorithm to detect changes in the protected data and program files. In another work, they applied the algorithm to monitor UNIX processes where the purpose is to detect harmful intrusions in a computer system. Kephart suggested another immunologically inspired approach for virus detection . In this approach, known viruses aredetected by their computer-code sequences (signatures)and unknown viruses by their unusual behavior withinthe computer system.From the survey, NSA nd GA canbe viewed as a good approach to solve the intrusions inthe real time world.

REFERENCES

[1] S.A. Hofmeyr and S. Forrest, Architecture for an artificial immune system, Evolutionary Computation, vol. 8, no. 4, pp. 443473, 2000

[2] S. A. Hofmeyr and S. Forrest, An Immunological Model of Distributed Detection and Its Application to Computer Security, The University of New Mexico, Albuquerque, NM, USA, 1999.

[3] “Adaptive Learning: Fly the Brainy Skies.” Wired, vol.10, no.3 (March 2002).

[4] Wei Li, Using Genetic Algorithm for Network Intrusion Detection, Proceedings of the United States Department of Energy Cyber Security Grou, Training Conference, Vol. 8, pp. 24-27,2004.

[5] Fabio A. Gonzalez and Dipankar Dasgupta, An Immunity-based Technique to Characterize Intrusions in Computer Networks, Evolutionary Computation, IEEE Trancactions, Vol. 6(3), pp.281-291, 2002.

[6] Esponda, F., S. Forrest, P. Helman, A Formal Framework for Positive and Negative Detection Scheme, IEEE Transaction on Systems, Man, and Cybernetics, 2003

[7] Christo Ananth, T.Rashmi Anns, R.K.Shunmuga Priya, K.Mala, “Delay-Aware Data Collection Network Structure For WSN”, International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Volume 1,Special Issue 2 - November 2015, pp.17-21

[8] Gonzalez, F., D. Dasgupta, J. Gomez, The Effect of Binary Matching Rules in Negative Selection, Genetic and Evolutionary Computation Conference (GECCO), Chicago, 2003

[9] Gonzalez, F., D. Dasgupta, L. F. Nino, A Randomized Rea-Valued Negative Selection Algorithm, 2nd International Conference on Artificial Immune System (ICARIS), UK. 2003

[10] Gonzalez, F., D. Dasgupta, Anomaly Detection Using Real-Valued Negative Selection, Genetic Programming and Evolvable Machine, vol. 4. pp. 383-403, 2003

[11] Ceong, H. T., et al, Complementary Dual Detectors for Effective Classification, 2nd International Conference on Artificial Immune System (ICARIS), UK, 2003

[12] Kim, J., et al, An evaluation of negative selection in an artificial immune system for network intrusion detection, in Proceedings Genetic and Evolutionary Computation Conference (GECCO), San Francisco, 2001

[13] Dasgupta, D., et al, MILA Multilevel Immune Learning Algorithm, Genetic and Evolutionary Computation Conference (GECCO), Chicago, 2003

[14] S. A. Hofmeyr and S. Forrest.Architecture for an artificial immune system.volume 8, pages 443473. Evolutionary Computation Journal, 2000.

[15] D. Dasgupta and F. Gonzalez. An immunity-based technique to characterize intrusions in computer networks.pages 10811088. IEEE Transactions on Evolutionary Computation, 2002.

[16] F. Gonzalez, D. Dasgupta, and J. Gomez.The effect of binary matching rules in negative selection.Pages 195206.In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2003), 2003.

[17] F. Gonzalez, D. Dasgupta, and R. Kozma.Combining negative selection and classification techniques for anomaly detection.volume 1, pages 705710. Proceedings of the 2002 Congress on Evolutionary Computation, 2002.

[18] F. Gonzalez, D. Dasgupta, and L. F. Nino.A randomized real-valued negative selection algorithm.pages 261272, 2003.

[19] S. A. Hofmeyr and S. Forrest. Immunity by design: An artificial immune system. volume 2, pages 12891296. In Proceedings of the Genetic and Evolutionary Computation Conference, 1999.

[20] F. Gonzalez and D. Dasgupta.Anomaly detection using real-valued negative selection.volume 4, pages 383403. Genetic Programming and Evolvable Machines, 2003.

[21] Z. Ji and D. Dasgupta. V-detector: An efficient negative selection algorithm with ”probably adequate” detector coverage. volume 179, pages 13901406. Information Sciences, 2009.

[22] J. Zeng, X. Liu, T. Li, C. Liu, L. Peng, and F. Sun. A selfadaptive negative selection algorithm used for anomaly
detection. volume 19, pages 261266. Progress in Natural Science, 2009.

[23] S. Balachandran, D. Dasgupta, F. Nino, and D. Garrett.A framework for evolving multi-shaped detectors in negative selection.pages 401408. In Proceedings of the 2007 IEEE symposium on foundations of computational intelligence, 2007.

[24] L. Xii X. Yuel, F. Zhang and D. Wangl.Optimization of self set and detector generation base on real-value

negative selection algorithm.2010 International Conference on Computer and Communication Technologies in Agriculture Engineering, 2010.

[25] G. B. Bezerra, T. V. Barra, L. Nunes de Castro, and F. J. Von Zuben. Adaptive radius immune algorithm for data clustering.pages 29030. In: Artificial Immune Systems: 4th International Conference, 2005.

[26] F. Gonzalez and D. Dasgupta.An immunogenetic technique to detect anomalies in network traffic.Pages 10811088. In Proceedings of the genetic and evolutionary compuation conference,2002.61

[27] Z. Ji and D. Dasgupta. Applicability issues of the realvalued negative selection algorithms. pages 111118. In Proceedings of the genetic and evolutionary computation conference, 2006.

[28] Z. Ji and D. Dasgupta.A boundary-aware negative selection algorithm.In Proceedings of the international conference on artificial intelligence and soft computing, 2005.

[29] Z. Ji and D. Dasgupta. Real-valued negative selection algorithm with variable-sized detectors. In proceeding of: Genetic and Evolutionary Computation, 2004.

[30] Kim J, Bentley PJ. Immune memory in the dynamic clonal selection algorithm. Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS ’02); 2002; pp. 5967.

[31] Li GY, Guo T. Receptor editing-inspired real negative selection algorithm. Computer Science. 2012;39:246251

[32] Hightower R, Forrest S, Perelson AS. The evolution of secondary organization in immune system gene libraries.Proceedings of the 2nd European Conference on Artificial Life; 1994; Brussels, Belgium.pp. 458470.

[33] Evolution and somatic learning in V-region genes. Perelson AS, Hightower R, Forrest S Res Immunol. 1996 May; 147(4):202-8.

[34] Oprea M, Forrest S. How the immune system generates diversity: Pathogen space coverage with random andevolved antibody libraries. 1999;(99-02-014)