International Journal of Security and Its Applications
Vol. 9, No. 4 (2015), pp. 279-288
http://dx.doi.org/10.14257/ijsia.2015.9.4.26

Real-valued Dual Negative Selection Technique for Intrusion Detection

Niu Ling[1] and Feng Gao Feng[2]

[1]*Zhou Kou Normal University, Zhoukou 466001, China;*
[2]*JiYuan Vocational And Technical College, JiYuan Henan 454650, China*
Niuling@zknu.edu.cn, fengjyjava@126.com

Abstract

A novel technique for intrusion detection based on real-valued dual negative selection scheme is proposed in this paper. In traditional real-valued negative selection algorithms, whether the candidate detectors can detect self-set or not totally relies on the affinity extent and the constant-sized mechanism is unfavorable to eliminating the black holes with irregular sizes. The proposed technique introduces the mechanism of variable-sized dual negative selection, in which each mutual detector has to pass three tests. Firstly, the new mutual detector should not be detected by the current existing ones. In other words, the existence of the new detector is necessary. Secondly, those detectors which can detect self-set will be eliminated. Thirdly, the detectors distribution has to be optimized aiming at enhancing the detecting efficiency. Experimental results demonstrate that the proposed technique has much less black holes, fewer detectors and higher detecting rates.

Keywords: *intrusion detection, real-valued, negative selection, detector, variable-sized*

1. Introduction

An increasing number of intrusion cases towards the computer system have attracted more and more people's attention, especially the experts and scholars both at home and abroad. Meanwhile, a lot of technical methods of intrusion detection have been developed. Inspired by the biological immune system, the theory of artificial immune system is established and used to deal with the issue of intrusion detection.

As known to all of us, the immune system is a highly complex, self-organizing, self-adaptive, parallel and distributed system, the function of which is to discriminate self from non-self and defend the organism against external invasions. So far, the main branches of an artificial immune system [1, 2] include negative selection algorithms (NSA) [3, 4] and clone selection (CS). Owing to the inherent special characteristics, NSA has been developed to be the most promising method in the whole biological immune field. The NSA theory stems from the mechanism of immune T cells. T cells are in charge of detecting the potential threatening ones, and those cells, unlike self-cells will be recognized and regarded as the threats by the mutual T cell. As a result, if a T cell is able to recognize a self-cell, it must be eliminated from the immune system. Based on the above principle, we can conclude that a qualified intrusion detection system should recognize all non-self behaviors, namely the "negative" mechanism. Consequently, NSA has been widely used for anomaly detection only requiring normal data to train [5].

String and binary presentation [6] were used to encode the samples and detectors which are very convenient for computer processing, but their intrinsic drawbacks are still obvious. First, the detector efficiency is low. Second, the false-alarm rate is high. Third, in some situations, it is more reasonable to use real values rather than strings or binary presentations. Fourth, constant-sized detector cannot adaptively adjust the search radius

ISSN: 1738-9976 IJSIA

so that a number of detectors are needed to cover the black holes. Fifth, the computational costs is very high. Research results demonstrate that the problems in many applications can hardly be settled properly by using binary representation.

In order to deal with the defects mentioned above, Gonzalez proposed real-valued negative selection algorithm (RNSA) [7]. Unlike string or binary presentations, RNSA not only is much closer to the classic problem space, but also enhances the running speed of the algorithm by computing several index values such as Euclidean distance. However, the initial number of detectors has to be required setting in advance and the radius is a constant, so that the performance of RNSA is not very good. Based on the above, Gonzalez presented an improved version of RNSA called randomized real-value negative selection algorithm (RRNSA) [8], which can adaptively determine the number of detectors. Reference [9] proposed a novel technique for intrusion detection in which the size of the detectors is all variable, so the number of detectors declines a lot. However, the radius size is offered as the only standard to evaluate the performance of detectors in [9]. Zhou developed the V-detector algorithm [10, 11]. Hyper-sphere shaped detectors with different sizes are generated via iteration, the larger ones of which are responsible for covering the non-self-region, while the others are used to capture the cracks located at the borderline between self-region and non-self-region. Furthermore, the scales of detectors and black holes have been both properly controlled. Nevertheless, in V-detector algorithms, the detecting radius is determined by computing the distance between the core and the self-border, which shares the largest affinity level with the detector for the sake of covering the area between self-region and non-self-region as much as possible, so that there exists an overlapping phenomena among different detectors. Although the improved V-detector algorithm [12] greatly eased the overlapping issues, it also increased the false-alarm rate to some extent.

Moreover, most traditional NSAs often generate candidate detectors randomly to match the whole training sets without considering their overlapping with the current existing detector sets. It has directly resulted in the unnecessary self-tolerance of candidate detectors, an excessive count of detectors and much lower efficiency of detector generation [13].

In this paper, a novel technique for intrusion detection based on real-valued dual negative selection scheme is proposed. The core framework of the proposed technique consists of three parts. First, the candidate detectors generated randomly are used to match the existing mature ones. If the match process does not success, the candidate detector enters the next round. Second, the training self-region is chosen to match the candidate detector which is not covered by the existing mature detectors. Similarly, if the match process fails, the candidate detector is added into the set of mature detectors. The reason for the behavior mentioned above is to enhance the efficiency of the successful candidate detectors. Thirdly, it is necessary to optimize the set of mature detectors. The number or size of the detectors may be adjusted according to the actual conditions. Experimental results demonstrate that the proposed technique has much less black holes, fewer detectors and higher detecting rates.

2. Traditional Real-valued Negative Selection Algorithm

In traditional RNSAs, whether the candidate detector can recognize the self-region or not relies on the affinity extent, which is determined by computing the Minkowski distance between the detector core and the self-core. If the distance is less than the radius of the self-region, the candidate detector is considered to be recognizing the self-region. The main categories of RNSAs include constant-sized radius RNSAs and variable-sized radius RNSAs, namely V-detector algorithms.

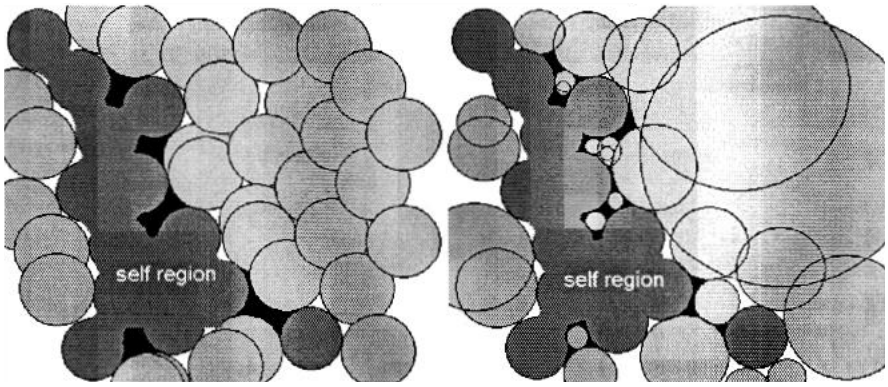
In constant-sized radius RNSAs, the radius r_d of the detectors is a constant, and the end condition of the algorithm is the number of detectors. The concrete steps are as follows.

(a) A candidate detecting center point denoted by X (x_1, x_2,…, x_n) is generated randomly;
(b) Computing the shortest distance dis_{min} between X and the self in the training set;
(c) If $dis_{min}>r_s+r_d$, the candidate detector with the center X and the radius r_d is regarded as a member of mature detectors. Where r_s is the radius of self-region.

In V-detector algorithms, the radius r_d of the detectors is a variable, and the end condition of the algorithm is the expectation coverage. The concrete steps are as follows.

(a) A candidate detecting center point denoted by X (x_1, x_2,…, x_n) is generated randomly;
(b) Computing the shortest distance dis_{min} between X and the self in the training set;
(c) If $dis_{min}>r_s$, the candidate detector with the center X and the radius $r_d=dis_{min}-r_s$ is regarded as a member of mature detectors.

Figure 1 shows the core idea of the above two categories of RNSAs.



(a) Constant-sized detectors (b) Variable-sized detectors

Figure 1. Core Idea of RNSAs and V-detector Algorithms

As shown in Figure 1, the dark gray area denotes the self-region, which is usually given as the training data. The light gray circles represent the detectors covering the non-self-region. The black holes are marked in black. If we wish to cover the non-self-regions with the same size, much more detectors are required in Figure 1 (a) than those in Figure 1 (b). However, the rising number of detectors doesn't necessarily imply the high ability of detecting. Since the radius in RNSAs is a constant, the detectors cannot adjust the size of radius to cover the black holes as much as possible. On the contrary, the superiorities of the V-detector algorithms are very clear. Firstly, the total number of detectors declines a lot at the same time the non-self-region coverage rate rises. Secondly, due to the radius size variability, the detectors can change their own sizes to even cover smaller black holes.

Unfortunately, only one negative selection process is conducted to judge the candidate detectors regardless of RNSAs or V-detector algorithms. Although the above process could guarantee that the newly created detectors are outside of the self-region and its coverage scope is confined to the non-self-region, the possibility of the coincidence between the newly created detectors and existing mature ones has never been considered. No wonder the rise of the repetition rate will produce too many mature detectors, which directly result in the high computational costs.

3. Proposed Technique

In order to increase the detecting performance, we present a novel technique for intrusion detection based on real-valued dual negative selection scheme. The proposed

one is composed of three phases, including two negative selection processes and the optimization of the mature detector set.

In order to deal with the drawbacks of RNSAs, the variable-sized radius mechanism has been utilized, and the expectation coverage rate is seen as the end condition in the process of detectors generating.

3.1. The First Process of Negative Selection

The purpose of the first negative selection process is to guarantee that the possible mature detector newly generated is necessary. In other words, the effect of the new detector should not be neglected or replaced by any existing one. The concrete steps are as follows.

(a) A candidate detecting center point denoted by X (x_1, x_2,…, x_n) is generated randomly;

(b) Computing the Euclidean distance *dis* between X and each member in the mature detector set;

(c) If the following expression is satisfied, the candidate detector will become the one-level detector, or the candidate one is out. Go back to (a).

$$dis(d_{new}, d_i) > r_{d_i} \quad i = 1, 2, ..., N_d \tag{1}$$

3.2. The Second Process of Negative Selection

The objective of the second negative selection process is to ensure the fundamental function of the newly generated detector is effective, namely it must cover the non-self-region rather than the self-region. As a result, the concrete steps of the second negative selection process are similar to the one of traditional RNSAs.

(a) Computing the Euclidean distance dis_{min} between the one-level detector and each self-body in the training set;

(b) If the following equation is satisfied, the one-level detector will be added to the set of mature detectors, or the one is out. Go back to step (a) in Section 3.1.

$$dis_{\min}(d_{one-rank}, s_j) > r_s \quad s_j \in Self, j = 1, 2, ..., N_s \tag{2}$$

The radius of the newly mature detector is dis_{min}-r_s.

3.3. Optimization of the Mature Detector Set

Figure 2 shows a distribution example of the mature detectors. The colorful circles represent the self-region, while hollow circles denote the mature detectors covering the non-self-region. Although the six detectors all abide by the fundamental principle of the NSA namely the detector can only cover the non-self-region rather than the self-region. The excessive overlapping phenomena are very serious as shown in Figure 2. As clearly shown, detector 3 is contained by detector 2, in other words, the effects of detector 3 can be entirely substituted by those of detector 2. In addition, there are too many overlapping layers of the three detectors in the lower left corner of Figure 2. As a result, it is necessary to optimize the existing set of mature detectors to enhance the detecting efficiency.

Suppose the self-set S= {S_1, S_2,…, S_i,…, S_j, R_{Si}}, the detector set D={D_1, D_2,…, D_i,…, D_j, R_{Di}}. Where R_{Si} and R_{Di} represent the radius of S_i and D_i respectively. The Euclidean distance is chosen as the computing index. α is the threshold relevant to the self-region. The restrictive conditions are given as follows.

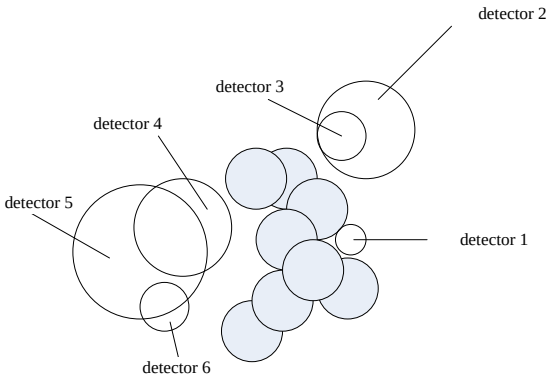


Figure 2. Excessive Overlapping Example of Detectors

(a) Towards each pair of elements in S_i and D_i, the following equality should be satisfied.

$$dis(S_i, D_i) > \alpha \tag{3}$$

Equation (3) guarantees that each mature detector D_i in set D indeed covers the non-self-region.

(b) Towards two random elements named D_i and D_j in set D, the following equality should be satisfied.

$$dis(D_i, D_j) > \max(R_{D_i}, R_{D_j}) \tag{4}$$

Equation (4) ensures that two random detectors should not coincide with each other.

The concrete optimization scheme is as follows.

Inputs: self-region (S), detector set D= {D_1, D_2,…, D_i,…, D_j, R_{Di}}, a variable λ;

Outputs: OD (the optimizing detector set);

Steps:

(a) Choosing D_i as the benchmark, and find another detector set D_j with the largest affinity value $dis(D_i, D_j)$ between D_i and D_j;

(b) If $dis(D_i, D_j) <= |R_{Di}-R_{Dj}|$, then certain detector set is entirely contained by the other one. The diagram is shown in Figure 3.

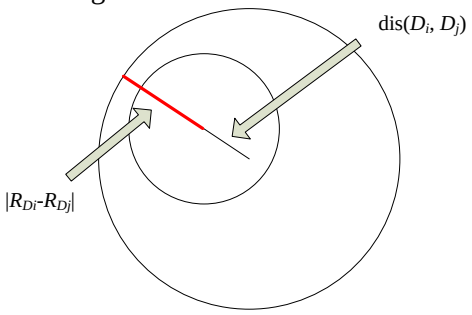


Figure 3. Diagram of Step (b)

The red line and the black line denote the length of $|R_{Di}-R_{Dj}|$ and $dis(D_i, D_j)$ respectively. In this case, the detector with the small size can be eliminated from the set D.

(c) If $dis(D_i, D_j) > R_{Di}+R_{Dj}$, then another new detector named $D(\text{new})_k$ will exist with the center Mean and the radius size of $dis(D_i, D_j)-(R_{Di}+R_{Dj})$. The detectors D_i and D_j will be eliminated from the set D. Mean denotes the midpoint of the two detectors D_i and D_j. Of course, $D(\text{new})_k$ must fulfill Equation (3). If not, the original detectors D_i and D_j will still restore. The diagram is shown in Figure 4.

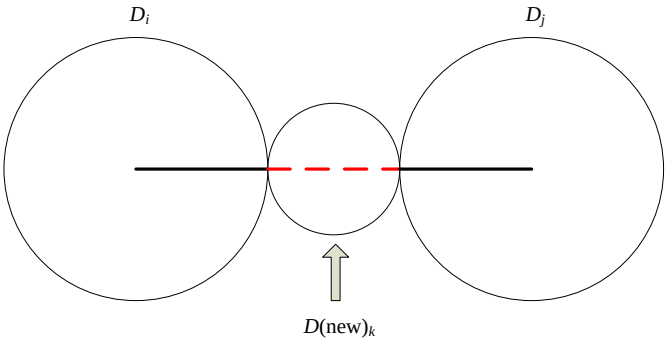


Figure 4. Diagram of Step (c)

The circle in the middle of the figure denotes the newly generated one. The purpose of step (c) is to generate a new detector with a smaller size to cover the black holes with small scale as well as possible, which is also a remarked superiority compared with the constant-sized NSAs.

(d) When $dis(D_i, D_j)$ locates the interval $[\lambda, R_{Di}+R_{Dj}]$, as shown in Figure 5, what should we do?

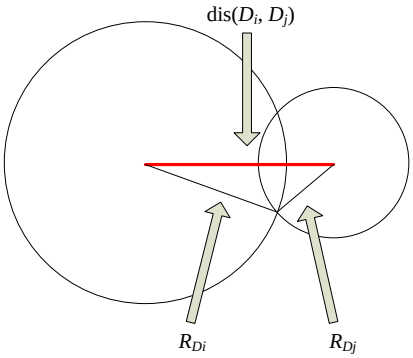


Figure 5. Diagram of Step (d)

Obviously, the value of λ directly decides the coincidence degree between two different mature detectors. More original detectors are preserved with the value of λ increasing, but the probability of the occurrence of newly detector will drop accordingly so that the black holes with small size may be not covered. In this paper, $\lambda = (R_{Di}+R_{Dj})/2$.

Moreover, the cosine law is used to evaluate the coverage extent between different detectors. As shown in Figure 5, the angle between R_{Di} and $dis(D_i, D_j)$ is denoted as β. Similarly, the angle between R_{Dj} and $dis(D_i, D_j)$ is denoted as γ. According to the cosine law, the following expressions can be obtained.

$$\beta = \arccos \frac{R_{D_i}{}^2 + dis(D_i, D_j)^2 - R_{D_j}{}^2}{2 \times R_{D_i} \times dis(D_i, D_j)} \tag{5}$$

$$\gamma = \arccos \frac{R_{D_j}{}^2 + dis(D_i, D_j)^2 - R_{D_i}{}^2}{2 \times R_{D_j} \times dis(D_i, D_j)} \tag{6}$$

If $\max(\beta, \gamma) \geq \pi/3$, we think the coverage rate is considerably high. As a result, If $dis(D_i, D_j)$ locates the interval $[\lambda, R_{Di}+R_{Dj}]$ or $\max(\beta, \gamma) < \pi/3$, the two detectors D_i and D_j should be preserved.

(e) If $dis(D_i, D_j) < \lambda$, step (c) is conducted.

(f) If all detectors in set D have been discussed, the algorithm ends, or goes back to step (a).

4. Experimental Results and Analysis

Simulation experiments are conducted to verify the effectiveness of the proposed technique in this section. Iris dataset is used to do the performance evaluation and efficiency analysis. The properties of all dimensions have been normalized into the interval $[0, 1]^n$, and the self-radius is set as 0.05. Two traditional NSAs including RNSA and V-detector has been used to compare with the proposed one.

Figure 6 shows the comparison of the number of mature detectors of three algorithms. With the coverage rate increasing, the number of mature detectors required of the three algorithms rises accordingly. However, it is not difficult to see that the two traditional algorithms are much more sensitive to the coverage rate than the proposed one. For example, when the coverage rate equals to 95%, the numbers of mature detectors of RNSA, V-detector and the proposed technique are 8105, 265.20 and 10.35, respectively. Consequently, the efficiency of the detectors in the proposed technique is distinctly higher than those of other two ones.

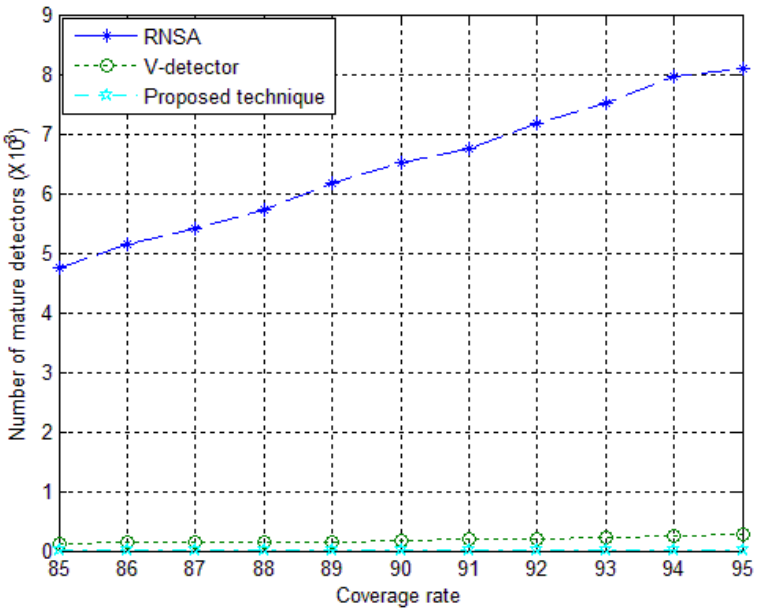


Figure 6 Comparison of the Number of Mature Detectors of Three Different NSAs (The size of self-region=0. 05, the constant size of RNSA=0. 10)

The detecting performance has been researched in this section. Figure 7 depicts the comparison of the detection rate of three different algorithms.

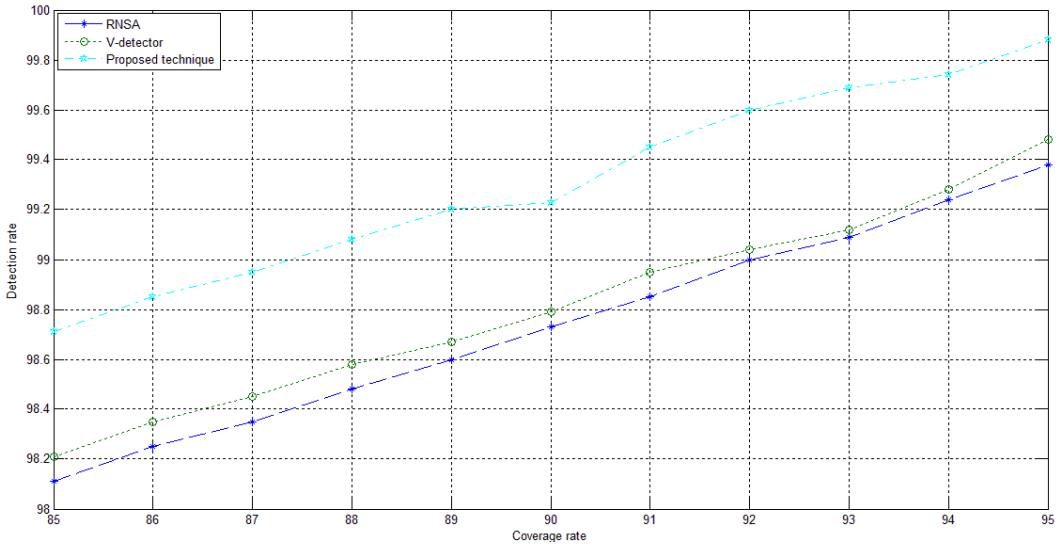


Figure 7. Comparison of the Detection Rate of Three Different NSAs (The Size of Self-region=0. 05, the Constant Size of RNSA=0. 10)

As shown in Figure 7, the detection rates of three algorithms all beyond 98%. By careful comparison, we can find that the detection rate of the proposed technique is sharply larger than those of other two ones.

In addition to the performance of the number of mature detectors and the detection rate, the index of the average computational costs is also a facet need to be noticed in practical application process. Therefore, the performance of the three techniques is described in Table 1.

Table 1 Comparison of Average Computational Costs of Three Different NSAs (The Size of Self-region=0.05, the Constant Size of RNSA=0.10)

	85%	86%	87%	88%	89%	90%	91%	92%	93%	94%	95%
RNSA	3.552	3.683	3.744	3.852	3.889	3.993	4.056	4.222	4.394	4.715	5.122
V-detector	0.033	0.034	0.032	0.033	0.034	0.034	0.035	0.037	0.042	0.047	0.057
Proposed technique	0.007	0.008	0.009	0.010	0.012	0.015	0.017	0.018	0.019	0.019	0.020

From Table 1, we can find that RNSA is the most time-consuming, while V-detector and the proposed technique are comparatively time-saving. For example, when the coverage rate rises to 95%, the computational cost of RNSA is 5.122, which is greatly more than the other two algorithms. Furthermore, the proposed technique still has much more efficiency than V-detector.

5. Conclusion

This paper proposes a novel technique for intrusion detection based on real-valued dual negative selection scheme. Experimental results and analysis indicate that the proposed technique owns remarked superiorities over current existing algorithms as follows.

(a) Dual negative selection scheme is utilized in the novel model. These two processes guarantee that the newly mature detector is necessary and effective.

(b) Detecting efficiency is enhanced a lot by optimizing the existing set of mature detectors.

(c) The number of mature detectors of the proposed technique is greatly less than RNSA and V-detector.

(d) The detection rate of the proposed technique is obviously higher than those of RNSA and V-detector.

(e) The computational cost of the proposed technique is remarked lower than those of RNSA and V-detector.

Accordingly, we expect that the proposed technique will have extensive application prospects in future. Of course, the further optimization of the proposed technique is still our focus on work.

References

[1] S. Hofmeyr and S. Forrest, "Architecture for an artificial immune system", Evolutional Computation Journal, vol. 4, no. 8, **(2000)**.
[2] L. N. de Castro and J. I. Timmis, "Artificial immune systems as a novel soft computing paradigm", Soft Computing, vol. 8, no. 7, **(2003)**.
[3] S. Forrest, A. S. Perelson and L. Allen, "Self-nonself discrimination in a computer", Proceedings of the IEEE Symposium on Research in Security and Privacy, **(1994)**; Los Alamitos, USA.
[4] D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology", Proceedings of the 5th International Conference on Intelligent Systems, **(1996)**; Cancun, Mexico.
[5] D. Dasgupta and S. Forrest, "An anomaly detection algorithm inspired by the immune system", Artificial Immune System and Their Applications, vol. 1, no. 1, **(1999)**.
[6] J. Balthrop, F. Esponda and S. Forrest, , "Coverage and generalization in an aritificial immune system", Proceedings of the Genetic and Evolutionary Computation Conference, **(2002)**; New York, USA.
[7] F. Gonzalez and D. Dasgupta, "Anomaly detection using real-valued negative selection", Genetic Programming and Evolvable Machines, vol. 4, no. 4, **(2003)**.
[8] F. Gonzalez, D. Dasgupta and L. F. Nino, "A randomized real-value negative selection algorithm", Proceedings of Second International Conference on Artificial Immune System, **(2003)**; Edinburgh, UK.
[9] D. Dasgupta and K. Krishna, "Negative selection algorithm for aircraft fault detection", Proceedings of Third International Conference on Artificial Immune Systems, **(2004)**; Catania, Italy.
[10] J. Zhou and D. Dasgupta, "Real-valued negative selection algorithm with variable-sized detectors", Proceedings of GECCO, **(2004)**; Berlin, Germany.
[11] J. Zhou and D. Dasgupta, "Augmented negative selection algorithm with variable-size detectors", IEEE Congress of Evolutionary Computation, **(2004)**; Washington, USA.
[12] Z. Hong, L. F. Wu and Y. Y. Wang, "Worm detection with improved V-detector algorithm", Journal of Beijing University of Posts and Telecommunications, vol. 2, no. 30, **(2007)**.
[13] X. F. Zheng, Y. H. Fang and T. Li, "Dual negative selection algorithm", Science China F-Information Sciences, vol. 4, no. 43, **(2013)**.

Author

Niu Ling, she received the B.Eng degree in Computer science from Henan normal university and M.Eng degree in Computer science from Chengdu University of Technology. She is currently researching on computer application technology.

Feng Gaofeng, he received the computer science degree from Henan Normal University, China, and the master degree in computer science from Beijing University of Posts and Telecommunications, He is a member of China Computer Federation and Association of Fundamental Computing Education in Chinese Universities in Beijing, China.