International Journal of Security and Its Applications
Vol.9, No.9 (2015), pp.95-104
http://dx.doi.org/10.14257/ijsia.2015.9.9.09

Technique for Intrusion Detection based on Cutting-based Real-valued Negative Selection

[1]Niu Ling, [2]Feng Gao Feng and [3]Ma Jing

[1]Zhou Kou Normal University, Zhoukou 466001, China;
[2]JiYuan Vocational And Technical College, JiYuan Henan 454650, China
[3]Key Laboratory of Information Assurance Technology, Beijing, 100072, China
Niuling@zknu.edu.cn[1], fengjyjava@126.com[2], mrsma919@163.com[3]

Abstract

A novel technique for intrusion detection based on cutting-based real-valued negative selection scheme is proposed in this paper. Different from the current typical techniques, the proposed one sets a much more strict and reasonable mechanism to generate and optimize the set of mutual detector. Concretely, firstly, the new generating detector must be necessary and it should not be detected by the current existing mutual ones. Besides, those detectors coinciding with the self-set will be cut and optimized into several qualified ones which have better detecting abilities. Experimental results demonstrate that the proposed technique has much higher detecting rates.

Keywords: *intrusion detection, cutting-based, negative selection, detector*

1. Introduction

Recently, computer viruses have been causing increasing headaches for all of us, and more and more commercial secrets and individual privacy is currently tackling growing threats. Against this background, the security of computer information has become the research hotspot both at home and abroad, and a variety of technical models and techniques [1-3] have been proposed. Inspired by the biological immune system, the theory of artificial immune system is established and used to deal with the issue of intrusion detection.

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. T cells in the immune system are responsible for 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, T cells should own the following properties.

(a) Mutual T cells have the ability to discriminate self from non-self.
(b) Mutual T cells should recognize the non-self ones as much as possible.
(c) The list of non-self ones mutual T cells have should not include any self ones, otherwise the immune system will provoke harmful responses to the body itself.

Similarly, the main task of the computer security system is also to distinguish between the normal behaviors and abnormal ones. In other words, a qualified intrusion detection system should recognize all non-self behaviors, namely the "negative" mechanism. Based on the above, Forrest et al. proposed negative selection algorithms (NSA) [4, 5] in 1994. NSA only requiring normal data to train [6] simulates the training course of immune cells, and it has been widely and successfully used in a great many of areas such as the network security due to its inherent special characteristics.

String and binary presentation [7] were used to encode the samples and detectors which are very convenient for computer processing, but their intrinsic drawbacks are still obvious.

ISSN: 1738-9976 IJSIA

(a) The detector efficiency is low.
(b) The false-alarm rate is high.
(c) It is more reasonable to use real values rather than strings or binary presentations in some situations.
(d) Constant-sized detector cannot adaptively adjust the search radius so that a number of detectors are needed to cover the black holes.
(e) The computational costs are very high.
(f) 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) [8]. 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) [9], which can adaptively determine the number of detectors. Reference [10] proposed a novel technique for intrusion detection in which the size of detectors are 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 [10]. Zhou developed the V-detector algorithm [11-12]. 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 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 [13] greatly eased the overlapping issue, 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 directly results in the unnecessary self-tolerance of candidate detectors, an excessive count of detectors and much lower efficiency of detector generation [14]. In order to enhance the detecting performance of detectors, Zheng *et al.* presented a novel NSA called dual NSA (DNSA) [14]. DNSA consisted with two phases. 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. The DNSA avoids the time-consuming self-tolerance process of the candidate detector within the coverage of existing mature detectors, and thus greatly reduces the size of the detector set and improves detector generation efficiently. However, DNSA performs poorly on balancing the scale of detectors and detecting efficiency. On the one hand, excessively tough rules lead to the sharp decline of the detectors. On the other hand, the remaining few detectors cannot detect the non-self-region efficiently.

In order to overcome the drawbacks of the above techniques, a novel technique for intrusion detection based on cutting-based real-valued negative selection scheme is proposed in this paper, whose core framework is composed of two phases. Firstly, the new generating detector must be necessary and it should not be detected by the current existing mutual ones. Secondly, those detectors coinciding with the self-set will be cut and optimized into several qualified ones which have better detecting abilities.

Experimental results demonstrate that the proposed technique has much 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.

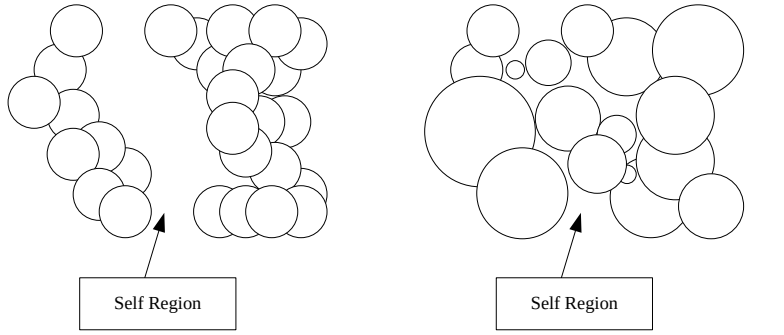
In constant-sized radius RNSAs, the radius r_d of the detectors is a constant, and the end condition of the algorithm is the number of detectors. The concrete steps are as follows.

(a) A candidate detecting center point denoted by $X(x_1, x_2,\ldots, x_n)$ 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 $X(x_1, x_2,\ldots, x_n)$ 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 middle area denotes the self-region, which is usually given as the training data. The circles represent the detectors covering the non-self-region. It is noteworthy that the non-self-region which is not covered by detectors is known as the black hole. 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 rise 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 the small 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 been never 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 cutting-based real-valued negative selection scheme. The proposed one is composed of two phases. Firstly, the new generating detector should not be detected by the current existing mutual ones. Secondly, those detectors coinciding with the self-set will be cut and optimized into several qualified ones which have better detecting abilities.

In order to deal with the drawbacks of current typical 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 Phase of Proposed Technique

The purpose of the first phase 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 detector X 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).

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

Where d_{new} and d_i denote the newly generated detector and the i^{th} one in the mature detector set MDS, respectively. r_{di} is the radius of the detector d_i. N_d is the number of the elements in MDS.Figure 2 shows the core idea of the first phase.

Obviously, if Eq. (1) is satisfied, there is no intersection of d_{new} and d_i, which means that the newly generated detector cannot be replaced by any existing one and is necessary. Conversely, as shown in Figure 2, if Eq. (1) cannot be satisfied, it must lead to two position relations namely ‘inclusion relation’ and ‘intersection relation’. Under the above two cases, the effect of the newly generated detector would undermine even be covered completely. As a result, any newly generated detector should meet the requirement in Eq. (1).

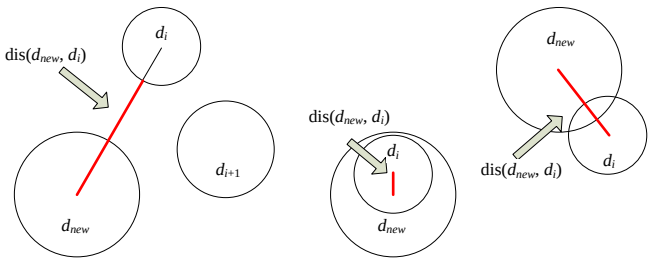


Figure 2. Core Idea of First Phase

3.2. The Second Phase of Proposed Technique

In this paper, we suppose that the shape of the detector is hyper-cube. Although the fundamental principle in NSA is that the detector must cover the non-self-region rather than the self-region, it doesn’t mean that those detectors covering self-region must be eliminated thoroughly. Moreover, these flawed detectors can provide us with much

location information of the self-region and non-self-region especially the boundary of them. Therefore, we could cut the flawed detectors and adjust their boundary to guarantee that there is no intersection between the new ones and the self-regions. Furthermore, for sake of increasing the coverage rate, the reasonable location relation is that the new detector should lay tangent to the self-region after cutting. In order to achieve the above state, the cutting-based mechanism is proposed in this paper whose concrete procedures are shown in Figure 3.

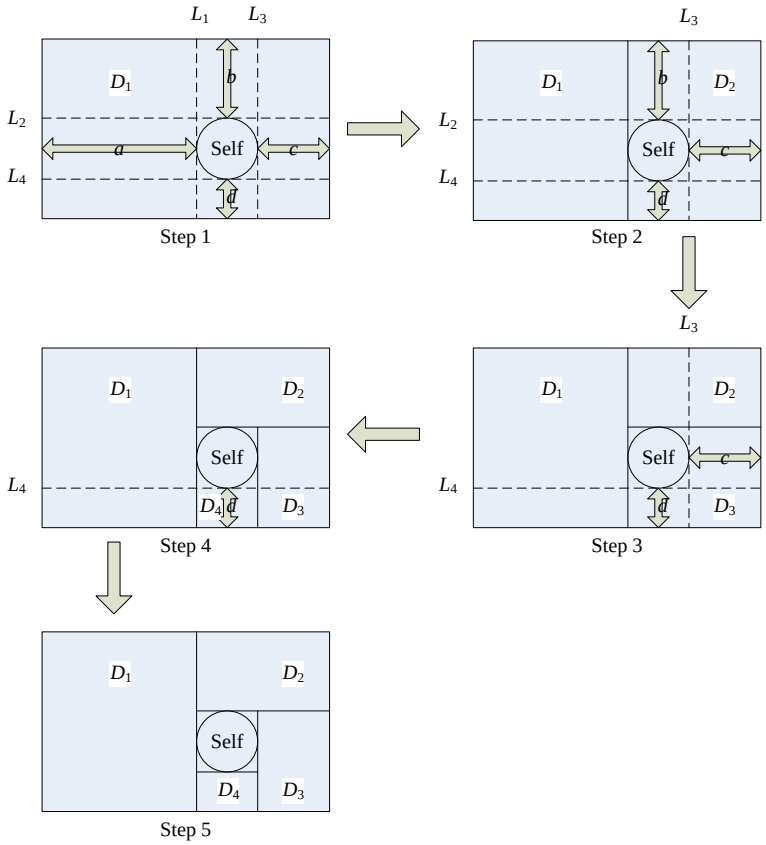


Figure 3. Cutting-based Mechanism

In Figure 3, we take the two-dimensional space for example.

In Step 1, the original detector D_1 covers the self-region called 'Self'. Thus, D_1 is a flawed detector and should be cut accordingly. By measuring the distances of the four directions between D_1 and 'Self', it can be concluded that $a>b>c>d$. In order to facilitate conducting the following cutting, four guides named L_1, L_2, L_3 and L_4 are also given, which correspond to the distances of the four directions respectively.

In Step 2, the longest distance a is considered. The original detector D_1 is cut into two parts including D_1 and D_2 along guide L_1. Obviously, the newly generated detector D_1 lays tangent to 'Self', but D_2 is a flawed one due to the location relation between it and 'Self'. Therefore, 'cutting' would still be needed next.

In Step 3, the longest distance among b, c and d is considered. The original detector D_2 is cut into two parts including D_2 and D_3 along guide L_2. Similar to the detector D_1, the newly generated detector D_2 is also a perfect one because it lays tangent to 'Self', but D_3 covers 'Self' entirely.

In Step 4, with the help of guide L_3, the original detector D_3 is cut into a perfect one D_3 and a flawed one D_4. Clearly, 'cutting' is still required.

In Step 5, the last defective detector D_4 is cut into a reasonable one.

After four such cuttings, the original flawed detector D_1 has been transformed into four smaller qualified detectors. Moreover, there are only four small black holes between the detector and 'Self'.

Several points should be noted in this section as follows.

(a) The cutting should be conducted according to the distance from the max to the min. The reason is that it can generate the detector with as large space volumes as possible, which is helpful for the detector to increase the coverage performance on the non-self-region. Besides, owing to the inclusion of some efficient detectors, it accordingly leads to the fast optimization and convergence of the mutual detector set.

(b) The space relation between the guide and the relative distance should be vertical. For example, L_1 is perpendicular to a.

(c) Figure 3 only describes an ordinary and simple example in the proposed technique. After cutting the original flawed detector, the four newly generated detectors are still needed to be detected whether there are intersections between them with other self-regions. If the coverage area exists, we still have to cut certain defective detector into four smaller ones until they meet the fundamental principle of the detector.

3.3. The Overall Algorithm of the Proposed Technique

In this section, the overall algorithm of the proposed cutting-based real-valued negative selection scheme is proposed. The concrete steps are as follows.

Inputs: self-region set S=$\{S_1, S_2,\ldots, S_N\}$, the existing mature detector set MDS=$\{D_1, D_2,\ldots, D_{Nd}\}$, i=1, Coverage rate λ.

Outputs: FMDS (the final mature detector set).

Steps:

(a) A candidate detector X is generated randomly.

(b) If Eq. (1) is not satisfied, delete X and go back to Step (a), else go to Step (c).

(c) If there is no intersection between X and S_i, go to Step (d), else go to Step (e).

(d) If i=N, put X into MDS, else i=i+1, and go back to Step (c).

(e) Cutting X into four new detectors named X_1, X_2, X_3 and X_4. i=1. Go back to Step (c) and replace X with X_1, X_2, X_3 and X_4 respectively.

(f) If the coverage rate is greater than or equal to λ, the algorithm stops, else go back to Step (a).

4. Experimental Results and Analysis

In this section, several simulation experiments are conducted to verify the effectiveness of the proposed technique. The running platform is a person computer with Linux OS, 2.3GHz CPU, 4G memory and VC++ programming language. For simplicity, the ring shaped dataset is used for the self-set to be trained. The properties of all dimensions have been normalized into the interval $[0, 1]^n$, and the self-radius is set as 0.05. The mature detectors are utilized to detector the non-self-region.

Two typical algorithms including RNSA and V-detector have been also used to be compared with the proposed technique. As known to us, the radius of the detector in RNSA is a constant, which is set as 0.10 here. Conversely, the radius in V-detector is a variable, which can be adaptively adjusted according to the distance between the core and the self-border.

Suppose the number of self-set ranges from 10 to 100. The first experiment is to analyze the number of mature detectors of three involved techniques in the case of the 99% detecting rate. The simulation result is shown in Figure 4.

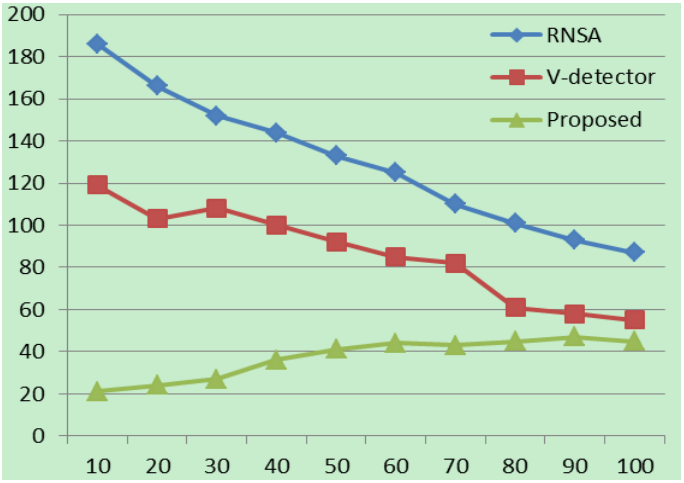
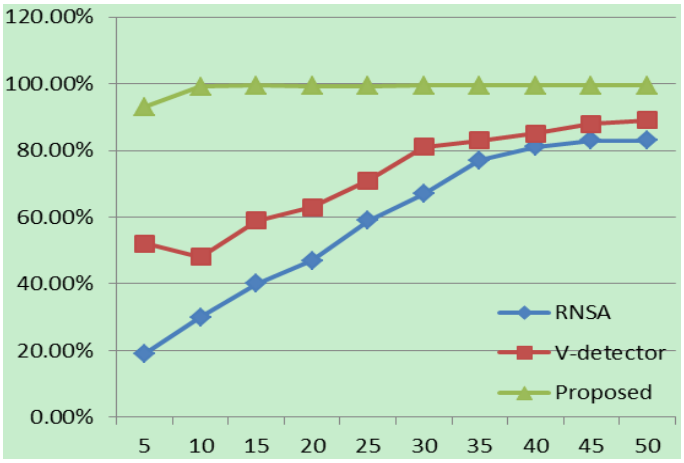


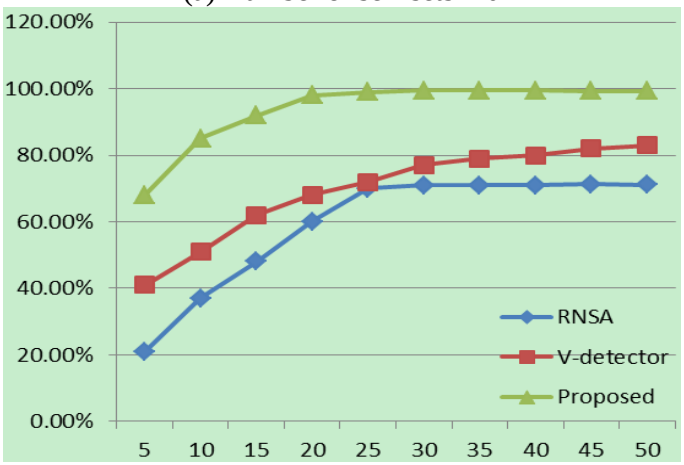
Figure 4. Results of the First Experiment

As shown in Figure 4, it is understandable that the numbers of mature detectors of three techniques vary according to those of self-sets, since the generation of mature detectors mainly rely on the distribution of self-sets. Obviously, with regard to the proposed technique, the increasing number of self-sets leads to the more number of matching the algorithm requires. Correspondingly, the splitting frequency and the number of detectors also rise. It is noteworthy that the number of detectors is not directly proportional to that of self-sets, because when the number of self-sets exceeds certain threshold, some newly generated detectors cannot be transformed into the final mature ones, which results in mature detectors reduction. The above point has been well vindicated in other two techniques. Firstly, with the growing number of self-sets, the constant-sized detector in RNSA directly results in the decline of the mature detectors' number. Secondly, although the variable-sized mechanism in V-detector enhances the detecting ability compared with NSA, poor performance on the detector optimization makes that the number of qualified detectors becomes fewer and fewer. As a result, the curves indicating the number of mature detectors mainly present a downward trend.

In traditional techniques, the detector will be deleted immediately if there are intersections between it and certain self-set. It is an easy and simple scheme, but it is not a good one, because too few detectors cannot guarantee a high detecting rate and more black holes which cannot be covered effectively will appear. Different from current typical techniques, the core idea of the proposed technique is to split the detector according to the spatial distribution of different self-sets, which is helpful to reduce the number of black holes with fewer and more efficient mature detectors. The second experiment is as follows. First, several mature detectors with three techniques are generated. Second, these detectors are chosen to conduct the detecting. The corresponding simulation result is shown in Figure 5.



(a) Number of self-sets=10



(b) Number of self-sets=40

Figure 5. Results of the Second Experiment

In Figure 5, the X-axis indicates the number of mature detectors, while the Y-axis demonstrates the detecting rate. Due to the good performance of the proposed technique, relatively high detecting rate can be achieved even with few detectors. For example, when the number of mature detectors equals to 5, the detecting rate of proposed technique is much larger than that of other two ones. Of course, with the growing number of mature detectors, the curves based on three techniques all present upward trends. However, due to the constant-sized mechanism, the largest detecting rate based on RNSA is clearly less than that of other two techniques. V-detector could vary the radius of detectors adaptively, so it achieves larger detecting rates than RNSA. Compared with other two, the proposed technique sharply enhances the detecting ability, and thus the relatively high detecting rate can be achieved even with several few detectors. For example, the curve based on proposed technique trends to be stable only with ten mature detectors in Figure 5(a). In Figure 5(b), the figure is around 20.

5. Conclusion

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

(a) Two phases are utilized which effectively guarantee that the newly mature detector is necessary and efficient.

(b) Detecting efficiency is enhanced a lot by introducing the splitting mechanism towards the newly generated 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.

Accordingly, we expect that the proposed technique will have extensive application prospects in future. Of course, how to further optimize the proposed technique is our main research work next.

Acknowledgements

The authors thank the anonymous reviewers and editors for their invaluable suggestions. This work was supported by the Foundation of Science and Technology on Information Assurance Laboratory (Grant Number: KJ-13-108).

References

[1] J. H. Yang and D. Woolbright, "Correlating TCP/IP Packet contexts to detect stepping-stone intrusion", Computers & Security, vol. 30, **(2011),** pp. 6-7.
[2] J. H. Yang and S. H. S. Huang, "Mining TCP/IP packets to detect stepping-stone intrusion", Computers & Security, vol. 26, **(2007),** pp. 7-8.
[3] Y. Wang, I. Kim, G. Mbateng and S. Y. Ho, "A latent class modeling approach to detect network intrusion", Computer Communications, vol. 30, **(2006),** p. 1.
[4] S. Forrest, A. S. Perelson and L. Allen, "Self-nonself discrimination in a computer", Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, **(1994)**; Los Alamitos, USA.
[5] 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.
[6] 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).**
[7] 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.
[8] F. Gonzalez and D. Dasgupta, "Anomaly detection using real-valued negative selection", Genetic Programming and Evolvable Machines, vol. 4, no. 4, **(2003).**
[9] 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.
[10] D. Dasgupta and K. Krishna, "Negative selection algorithm for aircraft fault detection", Proceedings of Third International Conference on Artificial Immune Systems, **(2004)** September 13-16; Catania, Italy
[11] J. Zhou and D. Dasgupta, "Real-valued negative selection algorithm with variable-sized detectors", Proceedings of GECCO, **(2004)**; Berlin, Germany.
[12] J. Zhou and D. Dasgupta, "Augmented negative selection algorithm with variable-size detectors", IEEE Congress of Evolutionary Computation, **(2004)**; Washington, USA.
[13] 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, **(2007),** p. 30.
[14] X. F. Zheng, Y. H. Fang and T. Li, "Dual negative selection algorithm", Science China F-Information Sciences, vol. 4, no. 43, **(2013).**

Authors

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.