Advanced Science and Technology Letters
Vol.82 (SoftTech 2015), pp.16-20
http://dx.doi.org/10.14257/astl.2015.82.04

A technique for intrusion detection based on real-valued dual negative selection

NIU Ling

Zhou Kou Normal University, Zhoukou 466001, China
Niuling@zknu.edu.cn

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

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].

Moreover, most traditional NSAs often generate candidate detectors randomly to match the whole training sets without considering their overlapping with the current

ISSN: 2287-1233 ASTL

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].

2 Proposed technique

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.

2.1. 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.

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

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

2.2. 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. 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. As a result, it is necessary to optimize the existing set of mature detectors to enhance the detecting efficiency.

Suppose the self-set S= $\{S_1, S_2,\dots, S_i,\dots, S_j, R_{Si}\}$, the detector set D=$\{D_1, D_2,\dots, D_i,\dots, D_j, R_{Di}\}$. 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.

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

$$dis(S_i, D_i) > \alpha \tag{2}$$

Equation (2) 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.

$$dis(D_i, D_j) > \max(R_{D_i}, R_{D_j}) \tag{3}$$

Equation (3) ensures that two random detectors should not coincide with each other.

The concrete optimization scheme is as follows.

Inputs: self-region (S), detector set $D= \{D_1, D_2,\ldots, D_i,\ldots, D_j, R_{Di}\}$, 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 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 (2). If not, the original detectors D_i and D_j will still restore.

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 $[\lambda, R_{Di}+R_{Dj}]$,

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.

$$\beta = \arccos \frac{R_{D_i}{}^2 + dis(D_i, D_j)^2 - R_{D_j}{}^2}{2 \times R_{D_i} \times dis(D_i, D_j)} \tag{4}$$

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

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 1 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.

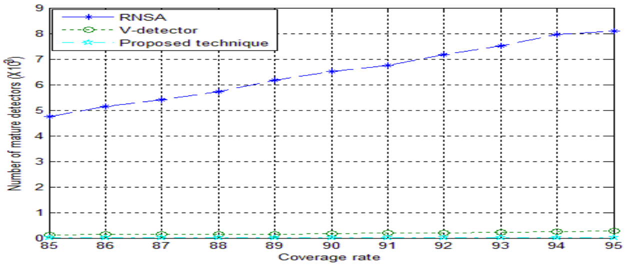

Fig. 1. Comparison of the number of mature detectors of three different NSAs (The size of self-region=0. 05, the constant size of RNSA=0. 10)

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)

	87%	88%	89%	90%	91%	92%	93%	94%	95%
RNSA	3.744	3.852	3.889	3.993	4.056	4.222	4.394	4.715	5.122

V-detector	0.032	0.033	0.034	0.034	0.035	0.037	0.042	0.047	0.057
Proposed technique	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.

4 Conclusion

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.

References

1. Hofmeyr, S., and Forrest, S.: Architecture for an artificial immune system, Evolutional Computation Journal, 4, 8 (2000)
2. de Castro, L. N., and Timmis, J. I.: Artificial immune systems as a novel soft computing paradigm, Soft Computing, 8, 7 (2003)
3. Forrest, A. S. Perelson, L. Allen, et al. Self-nonself discrimination in a computer. Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, (1994); Los Alamitos, USA
4. Dasgupta, D., and Forrest, S.: Novelty detection in time series data using ideas from immunology. Proceedings of the 5th International Conference on Intelligent Systems, (1996); Cancun, Mexico.
5. Dasgupta, D., and Forrest, S.: An anomaly detection algorithm inspired by the immune system, Artificial Immune System and Their Applications, 1, 1 (1999)
6. Balthrop, J., Esponda, F., Forrest, S., et al. Coverage and generalization in an artificial immune system. Proceedings of the Genetic and Evolutionary Computation Conference, (2002); New York, USA