ATLANTIS PRESS
*Advances in Computer Science Research, volume 63*
International Conference on Artificial Intelligence and Engineering Applications (AIEA 2016)

# Optimization of Real-Valued Self Set in Immunity-based WSN Intrusion Detection

Weipeng Guo [a], Yonghong Chen [b], Tian Wang [c] and Hui Tian [d]
College of Computer Science & Technology, Huaqiao University, xiamen 361021, China
[a]fyman_gwp@163.com, [b]djandcyhx@163.com, [c]wangtian@hqu.edu.cn, [d]cshtian@126.com

**Abstract.** Real-valued negative selection algorithm (RNSA) has been a main algorithm of immunity-based intrusion detection in wireless sensor networks (WSNs). However, the real-valued initial self-set which is used to train detectors has some defects: boundary invasion and overlapping among the self-samples. Detectors trained by the initial self-set may have the problem of boundary invasion, which will resulted false detection, and due to the redundancy of the self-set, the generation efficiency is low. Therefore, the self-set need to be optimized before training stage. In this paper, we proposed a new improved variable threshold self-set optimization algorithm to optimize the self-set before training the detectors based on a new concept of sample's affinity density, which is used to measure the distribution density of the sample. The experiments based on the Iris data set and wireless sensor networks intrusion detection were used to test the effectiveness of the algorithm. The results show that the optimization of self-set can solve the problem of boundary invasion, improve the detector training efficiency, and reduce the false alarm rate of the abnormal detection system.

**Keywords:** RNSA; Self-set optimization; Immunity; WSN; IDS.

## 1. Introduction

Wireless sensor networks(WSNs),which is a special kind of wireless networks, has been widely used in environmental monitoring, military, health control and so on [1]. However, it has been vulnerable to various attacks since the limitation of energy, computation ability and so on. Therefore, it should pay more attention to the security problem of WSNs.

As an efficient technique of security, intrusion detection system has been become popular way to solve the security problem of WSNs. However, as the WSN is limited by the limited resource, it is necessary to design an intrusion detection system in WSN with low computational complexity, low energy consumption and effective detection performance [2].

Recently, researchers have shown great interest in artificial immunity system based intrusion detection in wireless sensor network, for its features such as self-organization, adaption and fault tolerance are similar to the wireless sensor network desired characteristics [3].

Negative Selection Algorithm (NSA), which is one of the most well-known artificial immune algorithm. It has been widely used in intrusion detection [4]. Since many applications are much natural to be described in real-valued space, Real-Valued Negative Selection Algorithm (RNSA) has been proposed to solve the problem, in which the self-sample or detector is represented by an n-dimensional point and a radius 'r'. Many researches [5, 6] have been focus on the optimization of RNSA. Recently, researches are mainly focus on the optimization of the distribution and reduce the number of detectors [7]. However, lacking the research of the self-set. Nevertheless, the detectors are trained by the self-set, so that the self-set play great effect on the health of the detector. Therefore, the health of self-set is very important to be paid more attention.

It should be mention that there exist various problems of the self-set in real-valued space. In traditional RNSA, when training the detectors, the algorithm setting the radius of the self-sample with the same value without concerning the distribution density of the self-sample set. And it has resulted the problem of mis-classification of the nonlinear sample, which has decreased the detection accuracy. Beside, to represent the self-region more comprehensive, there always need a great number of self-samples, that they are always overlap with each other and resulted redundancy [8]. All these redundant sample overwhelming force of the detector training.

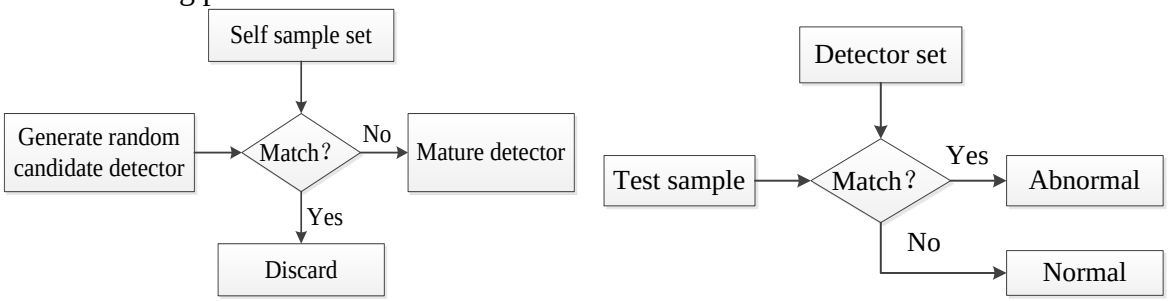
To solve the problems above, this paper proposed a novel method to optimize the initial self-set before training detector. The proposed algorithm firstly evaluating the self-sample's affinity density. And then adjusting the self-radius of the self-sample base on the affinity density. After the second step, discarding the self-samples that are covered by others. Various experiment from Iris data set to wireless sensor network results show that the algorithm proposed can improve the efficiency of the detector and improve the detection performance.

The remainder of our paper is organized as follows: section 2 discussed the preliminaries of the NSA with applications in anomaly detection in WSNs. Section 3 analysis the problem self-sample set. Section 4 details the improved NSA. And in section 5 demonstrates the computer simulations result. In section 6, some conclusions and remarks are given.

## 2. NSA based Intrusion Detection in WSN

### 2.1 Negative Selection Algorithm.

Based on the process of T-cells maturation in thymus in immune system, Forrest et al [4] develop the negative selection algorithm. And summarized the algorithm into two phase: detector training phase and detecting phase.




(a) Detector training phase (b) detecting phase

Fig. 1 The process of Negative Selection Algorithm

As shown in Fig. 1(a), during the detector training phase, the algorithm generate the initial detectors randomly and trained through negative selection. Only detectors that do not match anything self-become mature and be saved. After the detector training phase, detectors must be ready and the system turns into detection phase. As shown in Fig.1 (b), in this phase, the system recognized the abnormal through the detectors. Those samples match the detector is defined as abnormal.

### 2.2 NSA for WSN Intrusion Detection.

As a classical AIS algorithm, NSA has been widely used in wireless sensor networks intrusion detection. In [9], AIS was used in WSN for detecting anomalies. It is a direct one-to-one mapping between a thymus and a sensor node, in which the node is responsible for training the detector set and detecting the intrusion locally. Liu et al. [10] inspired by the immune system, applied the techniques of negative selection algorithm and clonal selection algorithm for the intrusion detection in wireless sensor network. To detect the anomaly, all node in the network are equipped with the detection module. The detection module is divided into four phase: self-acquisition, detector generation, detection phase and clonal selection. The node monitors the behavior of neighboring nodes to train the detectors and utilize the clonal selection technique to store the effective detector in memory to improve the detection performance. Fu et al. [11] proposed a hierarchical anomaly detection framework based on the immune danger theory and negative selection algorithm. The framework consist of three layers: local danger sensor, global detection and detection controller, which is adaptable and flexible in abnormal detection.

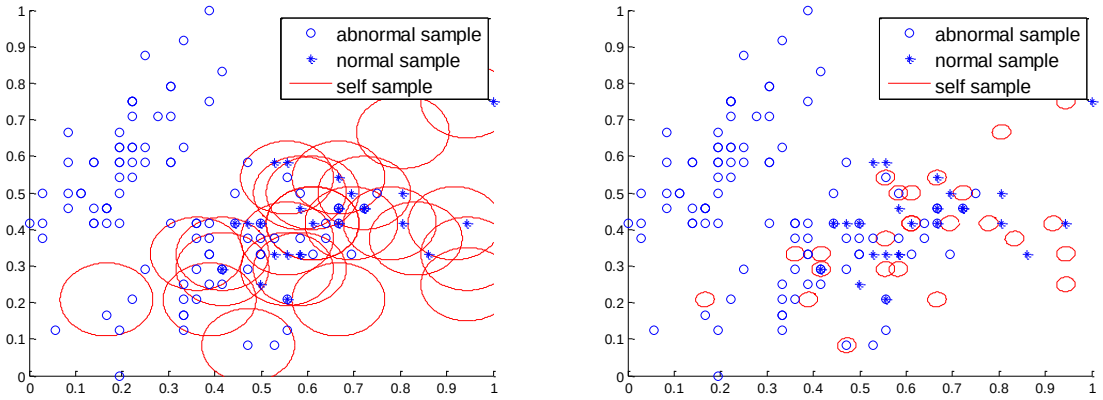
Considering the limited resource of WSN, in this paper, a hierarchical intrusion detection system for WSN that based on the LEACH protocol was proposed. In this framework, the intrusion detection system is a hierarchical-cooperative system with intrusion detection services distributed on the whole network. In the Base Station (BS), it training the detector set according to the normal sample and

distribute the detector to the cluster head, which act as a detecting node. And common node collecting the information that used for anomaly detection.

## 3. Problem of Real-valued Self set

In real-valued based NSA, the real-valued shape-space has been divided into self and non-self-space, and were normalized into a $[0, 1]^n$ super rectangle space or a hyper sphere with diameter of 1. Detectors belong to the non-self-space were trained from the self-set through negative selection. Therefore, it is important to note that the quality of self-set has great effect on the detector set. However, there lacking the research of self-set optimization.

In real-valued based negative selection algorithm, one of the main control parameter is the self-radius "r". The selection of parameter "r" has great effect on the balance of detection rate and false alarm rate. As it shown in fig.2, it is difficult to set the self-radius in an exactly value. For it will result the problem of boundary invasion, that result in high false alarm rate when the value of "r" is too small. And in contrast, low detection rate when the value of "r" is too big.




(a) The result of bigger self-sample (b) The result of smaller self-sample

Fig. 2 The effect on the choose of self-sample radius

In fact, as for multiple dimensional shape-space, the possibility of the self-sample belonging to the high density area is higher than the low density area. Therefore, an ideal method for setting the radius should set the bigger value for self-sample which is in the space of higher density to cover more normal area and in lower density space setting the smaller to avoid the possibility of cover the non-self-space. However, most research set the self-radius without considering the distribution density of the self-set. They set all the self-sample in a same radius, which is contradict with the variable self-radius idea.

Besides, there exist the problem of high overlapping and redundancy of the self-set. In a work situation, the network traffic may be in a concentrated scope that in each self-area, the samples distribute intensively. As shown in fig.3, a large amount of redundant self-samples distribute in the same areas overlap with each other's.

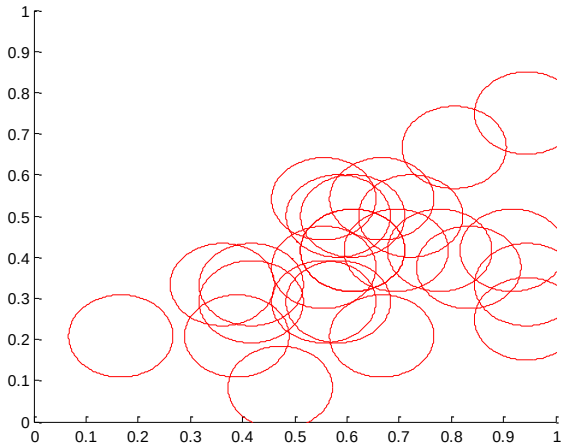



Fig. 3 The redundancy of self-samples

According to the principle of NSA, each initial detector must go through the procedure of negative selection (tolerance). However, great amount of redundant self-sample has resulted great pressure of this progress. Therefore, the overlapping of self-set should be solved before training detector.

## 4. Proposed Method

In view of the analysis above, it is significant to optimize the self-set before training the detectors. The aim of the optimization is to cover the self-space with less number but effectiveness self-sample. In this paper, a new improved method is proposed to optimize the self-set before detector generation.

To optimize the self-set, the algorithm firstly describing the affinity between the self-sample bases on the distance, and proposed a new concept to describe the affinity density of self-sample. Then the algorithm adjusting the self-radius according to the density to solve the problem of single self-radius. Finally, after adjusting the radius, discarding the unnecessary self-sample that are covered by the other samples. Details are as follow:

**Step 1: Describing the affinity density of sample.**

The affinity between two nodes(*X and Y*) always measure by the distance between the nodes. In this paper, we measure the distance through Euclidean distance [5] as follow:

$$d(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{1}$$

Before describing the affinity density of sample, the definition are given as follow: Assume there are number of N self-sample in the condition space, then the affinity density of the sample A in the space is:

$$\rho(A)=\frac{1}{N-1}\sum_{i=1}^{N}\frac{1}{d(A,X_i)} \qquad (A\neq X_i) \tag{2}$$

Affinity density is a measurement that used to describe the dispersion degree of distribution of the training sample with other sample nearby. The bigger density of sample A, means the more similar between sample A with other samples nearby, which means that the sample distribute around the sample A is dense. In contrast, if the density is small means that the sample distribute around is sparse.

**Step 2: Adjusting the self-sample's radius.**

It is clear that the samples in the dense region should have the bigger self-radius to represent more self-area. And the samples in the sparse region should set the smaller self-radius to avoid to cover the non-self-area.

To quantitative this concept, we propose a linear monotonically increasing function with the parameter of self-sample's affinity density to adjusting the self-sample's radius adaptively:

$$r(A) = R \times \frac{\rho(A) - \rho_{\min}}{\rho_{\max} - \rho_{\min}} \tag{3}$$

Where $\rho_{min}$ and $\rho_{max}$ is the min and max value of self-sample's affinity density, and R is the radius value for the max affinity density sample.

**Step 3: Discarding the redundant self-sample.**

After the adjusting of self-radius, samples in the dense area has the bigger radius and cover more self-region. Some of the redundant sample will be covered by others. For those samples that covered by others should be discarded, that is:

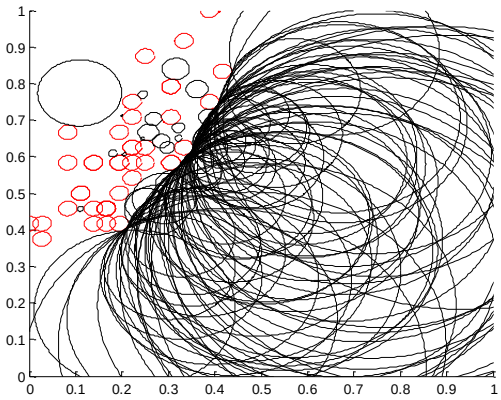
$$\{\forall s_i, s_j \in S, \left\| s_i - s_j \right\| \le r_{s_i} \ or \ \left\| s_i - s_j \right\| \le r_{s_j} \} = \phi \tag{4}$$

## 5. Experiments and Results

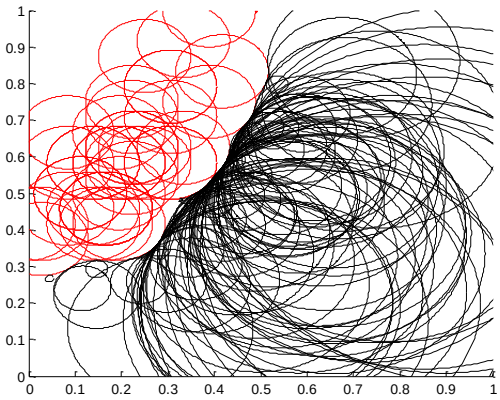
In order to show the effectiveness of the improved algorithm described in the previous section, experiments were carried out using Iris data set and wireless sensor network intrusion. The computational methods of detection rate and false alarm rate in experiments are defined in Ref [12].

### 5.1 Simulation Experiments with Iris Data-set

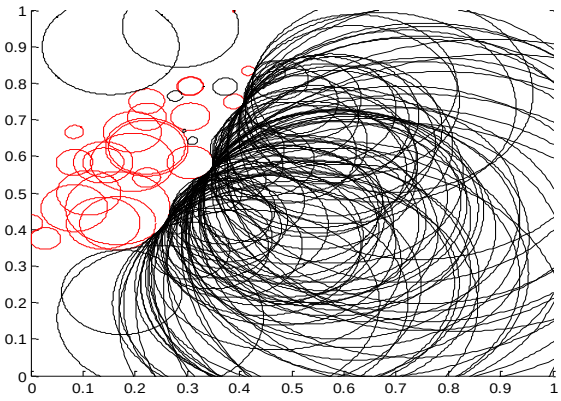
The Iris data set [13] contains 150 of instances, and they are divided into three type of Iris plant: Setosa, Versicolour and Virginica. In each instance, there exist four features: sepal length, sepal width, and petal length and petal width. To show the feasibility of optimization algorithm more obvious, we choose the features of sepal length and width to experiment. As shown in figure 4, we choose the class of setosa as the self and the other classes were treat as the non-self. Before optimization, the data set should be normalized. The comparison results are shown by figure 4.



(a) Detector trained by the initial self-set (with r=0.02)



(b) Detector trained by the initial self-set (with r=0.1)



(c) Detector trained by the optimized self-set

Fig. 4 the comparison result between initial and optimized self-set

In the experiments, 40 setosa sample were used as self to train the detector, and the other 110 sample were used to test the detection performance. In each experiment, 100 number of detector were generated. The figure. 4(a) shows the detector training result of the initial self-set (with smaller self-sample radius of 0.02), and we can see that the boundary invasion and nonuniform distribution problem are extremely severe. And it is clearly that the boundary of the self-region is covered by the detector which is the consequence of the non-self-boundary invasion. On the other hand, the figure. 4(b) shows the result of the initial self-set (with the bigger self-sample radius r=0.1),and the detector set trained by this initial self-set show that the boundary of non-self is not covered completely, and some non-self-sample are not covered by the detectors.

Likewise, the optimized self-set and its detector set are shown by the figure. 4(c). After the step of optimize the self-set, the number of the optimized self-sample is dropped to 26. And we can see that the samples in the dense area has the bigger radius and the samples in the sparse area is the smaller radius so that there are less boundary invasion. As the figure. 4(c) show that after adjusting the sample's radius individually, the boundary invasion can be avoid and the non-self-space's boundary can be covered well.

After the detector training stage, the remaining samples were used to examine the detectors' detection rate. Table 1 show the detection results above.

Table 1 Iris Data Set Detection Comparison

| Self-Radius | R=0.02 | R=0.1 | Optimized method |
|---|---|---|---|
| DR | 93.58% | 88.6% | 95.8% |
| FR | 23.6% | 21.75% | 9.67% |

From the table we can see that at the detection stage, the detection of detectors using optimized self-set is much higher than the others and the false alarm rate is lower than the others because of the solving of boundary invasion.

**5.2 Wireless Sensor Network Intrusion Detection**

In WSN experiment, the NS-2 is used in our simulation to evaluate the performance. It provides an excellent environment to simulate wireless sensor network. Fig.5 show the scenario of simulated wireless sensor network.

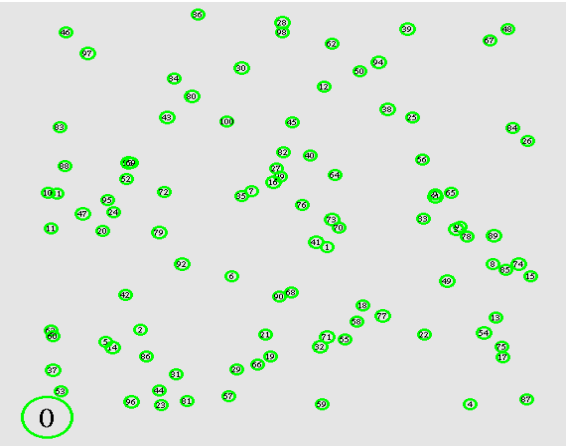



Fig. 5 simulation scenario of WSN

As shown in fig 5. , the WSN consists of 100 sensors, which has a transmitting range of 50m. They randomly distributed in a field of 100m*100m. At the beginning, there has an initial period of 1800sec that without attack for network to learn and train the detector set. After this period, the attackers start their attacks. The intrusion is achieved by a Jamming attack, which is a DoS attack of the wireless sensor network. Every simulation runs 1000 attacks cycles. In each attack cycle, only the first 20% cycle is used as the attack interval. All the results shown in this paper are the average of 10 repeated experiments.

Table 2 WSN Intrusion Detection Comparison

| Self-Radius | R=0.01 | R=0.1 | Optimized method |
|---|---|---|---|
| DR | 86.57% | 82.5% | 91.2% |
| FP | 18.63% | 12.32% | 8.44% |
| FN | 13.43% | 17.5% | 8.8% |

To perform the comparable detection performance of the detectors trained by the initial self-set (with self-radius r=0.1and r=0.01) and the optimized self-set, the detection rate, false positive rate and false negative rate are used to evaluate the detection performance. The detection result is shown in table 2. We can clearly see that the detection rate of the detector set which is trained from the optimized self-set is much higher than the other. And come to the false positive and false negative rate, as the boundary invasion problem has been solved, we can see that the optimized self-set based detector set is remarkable lower than the others.

## 6. Summary

In this paper, we had analyzed the problems of the self-set in real-valued negative selection algorithm. To solve the problems, we had proposed a new self-set optimization algorithm before training the detector set in anomaly detection and wireless sensor network intrusion detection. Experimental results showed that the optimization of self-set is necessary and effective. It improves the effectiveness of generating detectors. What's more, due to the adjusting of the radius of the border self-sample, the problem of boundary invasion is solved and the detection rate is improved obviously with the decrease of false alarm rate.

The optimization of real-valued self-set needs further study, including more strict analysis. The implication of self-radius, or how to interpret each self-sample, is also an important topic to be explored.

## References

[1] C. Y. Chong, S. P. Kumar. Sensor networks: evolution, opportunities, and challenges. Proceedings of the IEEE, Vol. 91(2003) No.8, pp. 1247-1256.

[2] O. Can, S. O. Koray. A survey of intrusion detection systems in wireless sensor networks: Modeling, Simulation, and Applied Optimization (ICMSAO). 2015 6th International Conference on. IEEE, 2015, May 27-29.

[3] N. A. Alrajeh, S. Khan, B. Shams. Intrusion detection systems in wireless sensor networks: a review. International Journal of Distributed Sensor Networks, Vol. 9 (2013) No.5.

[4] S. Forrest, A. S. Perelson, L. Allen, et al. Self-nonself discrimination in a computer. 2012 IEEE Symposium on Security and Privacy (1994), Oakland, CA.1994, May 16-18.

[5] J. Zhou, and D. Dasgupta. Estimating the detector coverage in a negative selection algorithm. Proceedings of the 7th annual conference on Genetic and evolutionary computation. ACM, 2005.

[6] González, Fabio, D. Dasgupta. A study of artificial immune systems applied to anomaly detection. The University of Memphis. 2003.

[7] X. Z. Gao, S. J. Ovaska, X. Wang. Genetic algorithms-based detector generation in negative selection algorithm. Adaptive and Learning Systems, 2006 IEEE Mountain Workshop on. IEEE, 2006.

[8] L. Xi, F. B. Zhang, S. W. Wang, et al. Real-valued self-set optimization algorithm in immunity-based anomaly detection. Application Research of Computers. Vol.28 (2011) No.4, pp.1434-1436.

[9] M. Drozda, S. Schaust, H. Szczerbicka. AIS for misbehavior detection in wireless sensor networks: Performance and design principles. Evolutionary Computation. 2007. IEEE Congress on. IEEE, 2007, Sept 25-28.

[10] Y. Liu, F. Yu. Immunity-based intrusion detection for wireless sensor networks. IEEE World Congress on Computational Intelligence, 2008, June 1-8.

[11] RR. Fu, KF. Zheng, FC. You, et al. Anomaly detection algorithm based on fuzzy and immune theory in wireless sensor networks. Journal of Nanjing University of Science and Technology. Vol.1 (2012) No. 36, pp.137-142.

[12] D. Stopel, R. Moskovitch, Z. Boger, et al. using artificial neural networks to detect unknown computer. Neural Computing & Applications. Vol.7 (2009) No. 18, pp.663-674.

[13] Fisher's iris data is available at ftp://ftp.ics.uci.edu/pub/machinelearning-databases/iris/.