doi:10.3772/j.issn.1006-6748.2016.04.005

Intrusion detection based on rough set and artificial immune①

Zhang Ling(张 玲)②, Sun Haiyan, Cui Jiantao, Yang Hua, Huang Yan
(Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450001, P. R. China)

Abstract

In order to increase intrusion detection rate and decrease false positive detection rate, a novel intrusion detection algorithm based on rough set and artificial immune (RSAI-IDA) is proposed. Using artificial immune in intrusion detection, anomaly actions are detected adaptively, and with rough set, effective antibodies can be obtained. A scheme, in which antibodies are partly generated randomly and others are from the artificial immune algorithm, is applied to ensure the antibodies diversity. Finally, simulations of RSAI-IDA and comparisons with other algorithms are given. The experimental results illustrate that the novel algorithm achieves more effective performances on anomaly intrusion detection, where the algorithm's time complexity decreases, the true positive detection rate increases, and the false positive detection rate is decreased.

Key words: rough set, artificial immune, anomaly intrusion detection, rough set and artificial immune (RSAI-IDA)

0 Introduction

Intrusion detection system (IDS) is to detect unwanted attempts at accessing, manipulating and disabling computer system by either authorized users or external perpetrators, mainly through the network, such as the Internet. The true positive detection rate and false positive detection rate are two important evaluation criteria for judging the performances of intrusion detection system[1].

Recently, many techniques have been proposed on IDS to reduce false positive detection rate and increase true positive detection rate, such as neural network, fuzzy theory, genetic algorithm, rough set algorithm (RSA), artificial immune theory, etc[2,3].

While most solutions for network security are static, which collect, analyze and extract evidences after attacks[4,5]. Therefore, they lack the abilities of self-learning and self-adapting. However artificial immune system (AIS) is evolved from a biological immune system, which has some particular features: distribution, adaptability and real-time capability.

In 1974, American biologist, therapist and immunologist, Jerne (Jerne N K) put forward an immune network theory, and won the Nobel Prize in medicine in 1986. Matzinger firstly put forward a danger theory, who proposed that the immune response in a human body was not caused by nonself antigens, but by danger signals when the cells were dead unnaturally in body. In AIS, dendritic cells (DC) can anomaly detect the antigens in real-time. DCs are sensitive to both signal types and have the ability to stimulate or suppress the adaptive immune system. DCs are the intrusion detection agents of the artificial immune system, policing the tissue for potential sources of damage in the form of signals and for potential culprits responsible for the damage in the form of 'antigen'. Antigens are proteins which can be 'presented' to adaptive immune system by DCs, and belong to pathogens or the host. Many scholars have carried out many studies on AIS in actual engineering. Thus, AIS is applied in anomaly detection.

Aickelin of University of Nottingham introduced the danger theory into AIS for the first time. On the 4th International Conference of Artificial Immunology in 2005, Greensmith et al put forward a new algorithm based on the "danger theory"——dendritic cell algorithm (DCA). DCA was applied in the intrusion detection system[6].

Wang designed a new multi-agent immune detection model using vaccine based on the danger theory and applied the model to deal with problems on cooperative air-defense of surface warship formation[7].

① Supported by the National Natural Science Foundation of China (No. 61502436), the Science and Technology Project of Henan Province (No. 152102210146) and the Doctoral Fund for the Central Universities (No. 2014BSJJ084).

② To whom correspondence should be addressed. E-mail: ll790217@163.com

Received on May 18, 2015

The main contribution of Ref. [7] is to use the danger theory to enhance system adaptability[7].

In Ref. [8], Suha applied a genetic algorithm and a dynamic clonal selection algorithm to detect virus in computer system and achieved good result.

In Ref. [9], enlightened by the similarity between the biological immune and antibody concentration when invasions happened in the network flow model, Zeng put forward a new intrusion detection method to reduce the rate of false positives while not affecting the detection rate.

In Ref. [10], agent-based artificial immune system (ABAIS) was used in the intrusion detection system. The agent-based IDS (ABIDS) was inspired by the danger theory of human immune system. In ABIDS, multiple agents were proposed, meanwhile, agents coordinated each other to calculate mature context antigen value (MCAV). The experiments proved that malicious attacks were detected by ABIDS from system calls directly by the agent-based dendritic cell algorithm.

Ref. [11] proposed a self set optimization algorithm using the modified clustering algorithm and Gaussian distribution theory. The results showed that the optimized method could solve the problem of boundary holes, increase the efficiency of detector generation effectively, and improve the system's detection rate. In Ref. [12], an integrated artificial immune system (IAIS) was proposed. The DCA and negative selection algorithm (NSA) were embedded in IAIS. DCA was used to detect behavioral features. NSA was used to detect structural features.

The problems of traditional detection algorithms are as follows: the computational complexity is high for the anomaly detection, so it is necessary to reduce the unimportant attributes; the method by which detectors are randomly generated may lead to slow function convergence. To solve these problems, some improvements will be done with the traditional algorithms. In this paper, first, the decision information table based on the immune theory will be redefined; second, the method of discernibility matrix is adopted to obtain effective antibody; third, in order to ensure the diversity of the antibody, part of the antibody are generated randomly.

The arrangement of this paper is as follows. In Section 1, the basic definitions such as antigen, antibody, affinity of antigen and antibody and the redefinition of antibody information decision table are introduced. In Section 2, an improved immune intrusion detection algorithm is given, the algorithmic complexity is shown and the comparison with other algorithms is discussed. In Section 3, some parameters are given by the experiment. Finally, the true detection rate and false positive detection rate are compared with other algorithms. In Section 4, conclusions are given.

1 The basic definitions

RSAI-IDA designed in this paper is inspired by AIS and RSA. In this section, the related definitions are given in Subsection 1.1, and the antibody information decision table and antibody information discernibility matrix are defined based on Rough Set theory in Subsection 1.2.

1.1 Preliminary knowledge

Definition 1 Antigen $a_g \in A_g$, $A_g \subset D$, $D = \{0, 1\}^l (l \in N, l > 0)$, where Ag is the set of antigens. D is the binary character string of length l, the value of antigen (A_g) denotes the characteristics of behavior.

Definition 2 The antibody set is defined as: $A_b\{< d, s, g, c >\}$, $d \in D$, $a_g \in N$, $a_b \in A_b$, $s \in \{00, 01, 10\}$, where Ab is the set of antibodies, $a_b.s$ is state of antibody, whose values are 00, 01 or 10, $a_b.g$ is the age of antibody, $a_b.c$ is the matching numbers of antibody and antigen, N is positive integer. $A_b = A_{bI} \cup A_{bT} \cup A_{bM}$. A_{bI} is the immature antibodies set , A_{bT} is the set of mature antibodies, A_{bM} is the set of memory antibodies.

The set of S_f combines all data of normal behavior characteristics, N_f set combines all the data of abnormal behaviors' characteristics, $S_f \cap N_f = \varnothing$.

The affinity represents the bonding strength between antigen and antibody, the function of affinity is shown as Eq. (1). The methods of calculating affinity between antibodies and antigens include: hamming distance, Euclidean distance, Manhattan distance, r-contiguous bites matching, etc.

The affinity between antigen and antibody is calculated with Euclidean distance, shown as

$$D = \sqrt{\sum_{i=1}^{L} (x_i - y_i)^2} \tag{1}$$

In Eq. (1), x_i is the ith feature of antigen x, y_i is the ith feature of antibody.

DCs receive four types of antigen signals: pathogen associated molecular pattern (PAMP), which is the abnormal behavior; safe signal which is representative of normal mode, the enhancement of this signal represents that it is more possible that the behavior is normal; danger signal, which indicates that the behavior may be abnormal, if the signal increases, the more possibility of abnormal behavior may occur; inflammation signal, the signal is used for signal amplification.

The preprocessing of the input signals is referred to Ref. [12].

In organization, DCs collect antigen signals (PAMP, safe signals, danger signals), and work out the three output signals: coordinating stimulus signals (CSM), semi mature signals (SEMI) and mature signals (MAT). The formula is shown in Eq. (2), by which the input signals are converted to the output signals. The recommended weight values of formula are given in Section 2.

$$C_{[CSM,SEMI,MAT]} = \frac{(W_P \times C_P) + (W_S \times C_S) + (W_D \times C_D)}{|W_P| + |W_S| + |W_D|} \times \frac{1+I}{2} \tag{2}$$

If the cumulative value of CSM is greater than the DC's migration threshold which is defined beforehand, DCs migrate to the lymph nodes, and mature dendritic cells are capable of presenting antigen to T cells and activated T cells leading to release of cytokines. The antigens and DCs' states are recorded, and the new DC is added into organization. Each antigen's mature context antigen value (MCAV) is calculated by Eq. (4). The abnormal antigens whose matching radius excesses the threshold will be dealt with by the security operation center (SOC).

Definition 3. Dynamic anomaly index is calculated with

$$C_a = \frac{m_a}{\sum_{i=1}^{A} A_i} \tag{3}$$

In Eq. (3), α is the antigens dataset in which all the decision attribute values are same, m_a is the total number of mature antigens whose type is α and presented to the lymph nodes. A_i is the sum of antigens which is the ith type and presented to mature cells, A is the total sum of antigens' types.

Definition 4 Antigen's matching threshold is defined in Eq. (4)[12].

$$Y_a = (L - \varepsilon) \times e^{-\partial(C_a - \delta)} \tag{4}$$

In Eq. (4), L is the length of detector, the type of ∂ is constant, δ is the anomaly threshold in anomaly detection, ε is self's radius, C_a is the dynamic anomaly indicators, for memory detector, the value of ∂ is set at a greater value.

1.2 Antibody information decision table

In 1982, rough set theory was put forward by Pawlak, who was a scientist of Poland Warsaw Polytechnic University. Attribution reduction is one of the most important contents[13]. After many years' development, rough set and other disciplines are combined in the areas of research such as knowledge discovery[14,15].

In rough set theory, the information system is also called a decision table. Referring to Definition 1 and Definition 2, antibody information decision table, the equivalence relation, approximation, antibodies attribute dependency and attribute importance are defined as follows:

Definition 5 An antibody information system is defined as $D_T = < U, C \cup D, V, f >$, where:

$U = \{a_{b_1}, a_{b_2}, \cdots, a_{b_n}\}$, $U \neq \varphi$, $a_{b_1}, a_{b_2}, \cdots, a_{b_n}$ are antibodies, U is finite set.

$C = \{a_1, a_2, \cdots a_m\}$, C is conditional attributes subset.

D is decision attributes subset, and $A = C \cup D$; $C \cap D = \varnothing$.

$V = \bigcup_{a \in C \cup D} V_a$, V_a is called the domain of a;

$f: U \times A \rightarrow V$, f is an information function, for any $\forall a \in C \cup D$; $x \in U$ and $f(x, a) \in V_a$.

Definition 6 An antibody information decision table $DT = < U, C \cup D, V, f >$, $A = C \cup D$, $\forall B \subseteq A$, $A_b \subseteq U$, $B*(A_b)$ is the lower approximation of the antibody set A_b, the positive domain of decision attribute D is defined as

$$P_B(D) = \bigcup_{A_b \in U/D} B*(A_b) \tag{5}$$

Definition 7 An antibody information decision table is defined as $D_T = < U, C \cup D, V, f >$, whose discernibility matrix is $\boldsymbol{M}_D(B)$.

$$\boldsymbol{M}_D(B) = \{\boldsymbol{M}_D(i, j)\} n \times n.$$
$$1 \leqslant i, j \leqslant n = |U/IND(B)| \tag{6}$$

$M_D(i, j)$ is the ith column and the jth row element, $IND(B)$ is referred to Ref. [14].

$$MD(i, j) = \begin{cases} \{a_k \mid a_k \in B \wedge a_k(a_{b_i}) \neq a_k(a_{b_j})\}, \\ \quad d(a_{b_i}) \neq d(a_{b_j}) \wedge a_k(a_{b_i}) \neq a_k(a_{b_j}) \\ 0, \quad d(a_{b_i}) = d(a_{b_j}) \\ \varphi, \quad a_k(a_{b_i}) = a_k(a_{b_j}) \wedge d(a_{b_i}) \neq d(a_{b_j}) \end{cases} \tag{7}$$

Symbol $a_k(a_{b_j})$ is the value of sample ab_j about attribute a_k, $d(a_{b_j})$ is the value of sample a_{b_j} about decision attribute a_k, $i, j = 1, 2 \cdots, n$, $\boldsymbol{M}_D(i, j)$ is the attributes set, which shows that two values on the classes are different to the attributes set. Any discernibility matrix $\boldsymbol{M}_D(B)$ can uniquely identify a discernibility function by the element $\boldsymbol{M}_D(i, j)$.

$$F_{M_D}(B) = \bigwedge_{M_D(i,j) \neq 0,\, M_D(i,j) \neq \varphi a_i \in M_D(i,j)} \bigvee a_i, \quad i, j = 1, 2, \cdots, n. \tag{8}$$

2 RSAI-IDA

Synthesizing the dendritic cell algorithm and nega-

tive selection algorithm, two measures are applied to improve intrusion detection's performances. Rough set theory is introduced to train data and get decision rules. In order to improve the diversity of antibodies, the improved algorithm includes two methods: getting the detectors partly by training with random function.

RSAI-IDA.

Input: connection instances: $I = \{i_1, i_2, \cdots, i_n\}$

Output: clusters: $C = \{c_1, c_2, \cdots, c_k\}$

1. Preprocess log information to get decision rules table $U = \{a_{b_1}, a_{b_2}, \cdots, a_{b_n}\}$
2. Reduce redundant attributes
3. Generate intrusion detection rules with RSA
4. Rules are divided into self-rules and non-self-rules according to decision attribute
5. Misuse detection
6. Generate detectors with NSA
7. DCs get antigen signals
8. Process antigen and signals with DCA
9. Analyse signals in real-time to gain abnormal index
10. According to the dynamic anomaly index calculate matching threshold value
11. Perform anomaly detection with NSA
12. Produce the alarm signals

2.1 Key sections

There are four key sections in RSAI-IDA, which are described as below:

(1) First, according to Definition 1, Definition 2 and Definition 5, preprocess the training data to obtain the antibody information decision table, and judge whether the decision table information is complete. If there are some elements which are absent, then fill the absence to delete duplicates. Second, generate equivalence classes using Definition 6 and Definition 7, and calculate with identification function to get the discernibility matrix. Third, judge whether the attributes in the decision table need reduction according to attributes class projection. If the attribute is not important, then reduce the unimportant attributes. Fourth, get the simplest antibody decision table, and generate decision rules, finally, select the decision rules whose decision attribute value is '0', and sort the rules by numbers of rule matching, and store the autologous rules. The flow chart of generating rules is shown in Fig. 1.

(2) Generate candidate detectors randomly, calculate the affinities between each detector and all mature detectors. If the affinity is smaller than the self-radius, then remove the candidate detector, otherwise the candidate detector is added to the mature detectors set. If the total number of detectors is equal to the self given threshold value, then obtain a set of mature detectors. The flow chart of generating detectors is shown in Fig. 2.

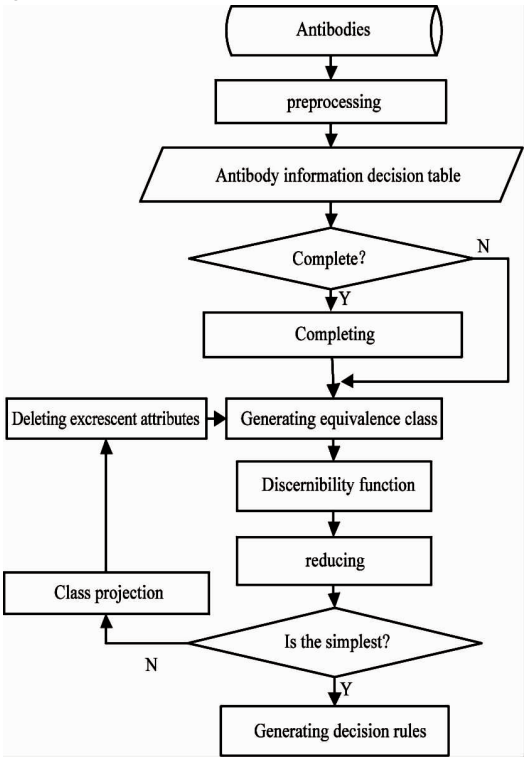

Fig. 1 The flow chart of generating rules

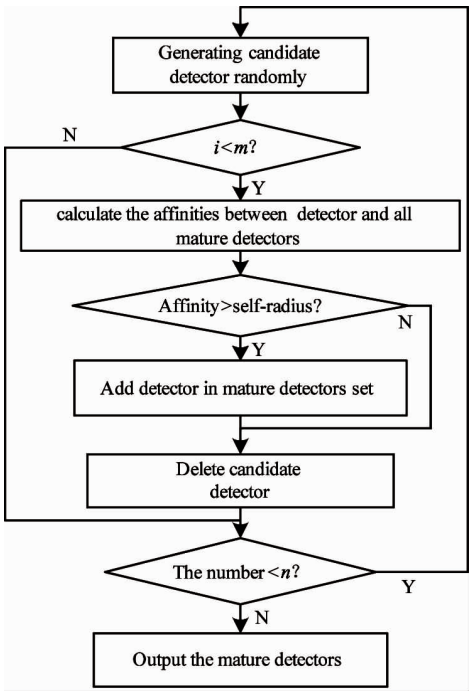

Fig. 2 The flow chart of generating detectors

(3) At First, obtain and normalize signals of antigens, store in the antigen input signal matrix. And

then, calculate the input signals and get the output signals with Eq. (2), if the cumulative value of CSM is greater than migration threshold, and the cumulative mature value is greater than the semi-mature value, DC becomes mature, and migrates to the lymph nodes, at the same time, a new dendritic cell is added into the organization, else it continues to collect new signals. At last, statistics is done to mature DCs and calculation is given to the anomaly index with Eq. (3). By Eq. (4), work out matching threshold value. The flow chart of calculating anomaly index is shown in Fig. 3.

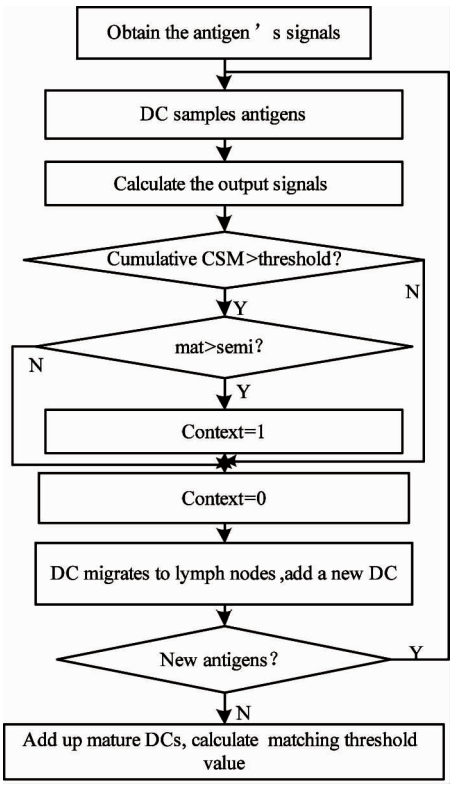

Fig. 3 The flow chart of calculating anomaly index

(4) Collect antigens and calculate the affinity of the antigens and detectors with Eq. (1). If the affinity is greater than the threshold value which is given in section 3 , then the antigen is judged as nonself, and generate an alarm in Security Operation Center. If the affinity is less than the threshold value, then the antigen is self-one, and replace the mature detector using the principle of least recently used.

2.2 Algorithm's complexity

According to the improved algorithm in this Subsection 2.1, the algorithm's time complexity and space complexity are given in this subsection.

The parameters are given as: let the number of condition attributes of decision table be M, the number of decision attribute value be 2, the number of training samples be N_1, total number of antigens which are to be detected be N, the length of detector be L, and L_1 is the length of input antigen signal.

Table 1 The complexity of algorithms

Algorithm	Time complexity	Space complexity
DCA	$O(35N_2^2)$	$O(500(L+1))$
NSA	$O(LNK)$	$O((L+1)K)$
IAIS	$O(N(L_1N^2+LNK))$	$O(MNK_1L)$
RSAI-IDA	$O(N(L_1N^2+LNK))$	$O(MN_1K_1)$

From Table 1, in comparison with DCA, NSA and IAIS, the time complexity of the algorithm is lower than DCA, NSA and IAIS, but space complexity is a little higher than other algorithms. In the training stage, space complexity increases, but the detection speed is a higher. So, it is worth consuming the space. While in the process of detecting the antigens, the space complexity is the same.

3 Simulations and experimental results

The experimental results are evaluated by the true positive detection rate (TP) and the false positive detection rate (FP). They are two important evaluation criterions for judging the performances of intrusion detection system.

In this section, the processing of experiment data sets and some simple parameters are given in Subsection 3.1. The two important parameters are shown in Subsection 3.2 and Subsection 3.3. With RSA, the length of detectors is obtained in Subsection 3.2. When the values of detector's radius are between 0 and 16, TP and FP are shown in Subsection 3.3, in which, the detector's radius is determined. In subsection 3.4, the anomaly detection results are given using the parameters above. In Subsection 3.5, comparisons of performances of different algorithms are shown.

The algorithm's code is written with C language, and all the experiments are simulated in the Linux system (Intel Pentium Dual CPU E2180, 2G RAM).

3.1 Experimental datasets

In this study, the experiential data are from "10% KDD 99" dataset. 300 million data are adopted in the simulations, the definition of data fields are referred to Ref. [15]. About 10% of the training data are denoted as S. Dataset S is divided into two parts, 20% of the data are combined in set S_1, which are used for training to get decision rules, others (80% of

the data set) are included in set S_2, which are used as testing data divided into 5 groups, as T_1, T_2, T_3, T_4 and T_5. Each set has 4800 data. The timestamp will be added to each record in order to simulate the detection that one record is sampled per second.

1) Antigens

The data domains are converted to binary strings which are used to represent antigen, and the methods of conversation are shown in Table 2.

Table 2 The table of antigens' construction

Item	Conversion method	Bits
2	TCP, UDP and ICMP are replaced by 01,11,11	2
3	According to the initials, the strings are converted into 1 to 66, and converted into binary string	7
4	According to the initials, the strings are converted into 1 to 11, and transformed to binary string	4
7,12,4, 15,21,22	0 and 1	1
8	Converted into binary string	2
9	Converted into binary string	2
11	Represented as 000, 001, 010, 011	3
17	Converted into binary string	4
18	Converted into binary string	2
19	Converted into binary string	3
31	Multiple 100, and converted into binary string	7
Others	Lower is 00, middle is 01, high is 10 and higher is 11	2

2) Input antigen signals

The method which is used to obtain the input antigen signals is referred in Refs[12,14]. The 10 types of data domains are divided into three types as follows:

PAMP: serror rate, srv serror rate, same srv rate, dst host same src port rate and dst host serror rate.

Danger signals: count and srv count.

Safe signals: logged in, srv diff host and dst host count.

Let x be the value of data domain. If the system is abnormal and the value of x is in $[m, n]$, then the signal is safe. The method of processing the data is shown as Eq. (9), which is adopted to normalize the antigens signals, which are converted into $[0,100]$.

$$f(x) = \begin{cases} 0, & x \in [0,m) \\ 100\dfrac{x}{n-m}, & x \in [m,n] \\ 100, & x \in (0,+\infty) \end{cases} \tag{9}$$

In Eq. (9), m is the smallest one of the data domains, and n is the biggest one.

3) The parameters

In experiments, the number of dendritic cells is set as 10; each dendritic cell collect 500 antigens; migration threshold is 300; the anomaly threshold is 0.35; the value of ∂ is constant, when the detector is memory one, ∂ is 1, ordinary detector has a value of 0.5; the length of detector L is given in subsection 3.2; the recommended weights of Eq. (2) are shown in Table 3.

Table 3 Recommend weights

W(weigh)	CSM	SEMI	MAT
PAMP(P)	2	0	2
Safe signal(S)	1	0	1
Danger signal(D)	2	3	3

3.2 The length of detectors

There are 41 conditional attributes in the decision table, while some attributes are unimportant, if all the attributes are considered in detection, the time complexity and space complexity will increase. So it is necessary to delete the unimportant attributes.

The reduction algorithm based on rough set theory is used to deal with the antibody information decision table S_1, and to get the simplest decision table.

After reduction, there are 11 attributes, which are: 2,3,4, 25,26,28,29,30,31,35,36. In Table 4, the decision rules before and after reduction are given, after reduction, the length of antibodies L is 30.

Table 4 The number of decision rules

Type	Before reduction	After reduction
Self	1258	315
Nonself	4742	896

In Table 4, before reduction, there are 1258 self-antigens and 4742 nonself-antigen. With RSA, the number of self-antigens is 315, the number of nonself-antigens is 896.

After reduction, the number of antibody decision rules obviously decreases. In accordance with the important degree of rules, all the decision rules whose decision attribute value is 0, and all the rules are sorted by matching number from high to low.

3.3 The detectors' radius

Receive operation curve (ROC) is applied to describe the intrusion detection performances and the relationship of the true positive detection rate and the

false positive detection rate. Each point represents a value of detector's radius length, plotted by the false positive detection rate on the horizontal axis and the true positive detection rate on the vertical. The datasets S_2 of KDD99 are employed in this experiment for testing, and each test is repeated for 10 times, all the results are averaged, which are shown by ROC curves.

The value of self's radius ε determines the coverage of each detector. When ε is 0, the antigen and detector are matching exactly. When the affinity between antigen and the detector is greater than "$L-\varepsilon$", the detector and antigen are considered to be matching.

In order to study how the artificial immune performances varies when the self's radius have different values, the self's radius are set to different values separately, and calculate the corresponding TP and FT. The results are shown in Fig. 4.

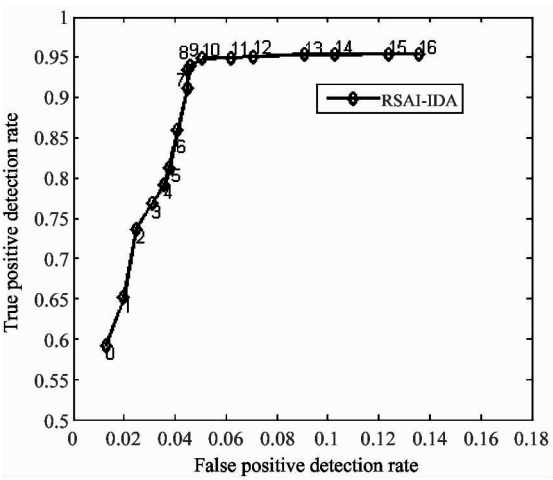

Fig. 4 The receive operation curve

The results are shown in Fig. 4, when ε is 0, TP is about 0.592, and FP is 0.013; when ε is 10, TP is about 0.9483, and FP is 0.051; when ε is 11, TP is about 0.9491, and FP is 0.062. With increasing of value ε, FP increases too. When ε is 10, the area of the curve is the greatest, but when the value of ε is higher, TP is nearly unchanged, while FT grows. According to ROC curve in Fig. 4, the value of ε is set as 10, namely self's radius is 10.5.

3.4 True positive detection rate and false positive detection rate

In Subsection 3.3, the testing dataset is S_2, which is from 80% KDD99 datasets. In this section, the algorithm is to be emulated which has been proposed in Section 2. In the experiment, the value of L is 30, which is obtained in Subsection 3.2, ε is 10.5, which is shown in Subsection 3.3. Datasets S_2 are divided into five groups, and each group is applied for testing 10 times. And the values of TP and FP are calculated to get the average, which are used to evaluate the performances of RSAI-IDA. The simulation results are shown in Table 5.

Table 5 The TP and FP of detection

Type	TP		FP	
	mean	deviation	mean	deviation
T_1	0.9521	0.0035	0.0349	0.0028
T_2	0.9783	0.0027	0.0483	0.0020
T_3	0.9735	0.0023	0.0458	0.0017
T_4	0.9686	0.0025	0.0321	0.0023
T_5	0.9726	0.0023	0.0345	0.0032
mean	0.96902	0.00266	0.03912	0.0024

As can be seen from Table 5, the average value of TP is above 0.96902, and the mean variance of TP is 0.00266; the average value of FP is above 0.03912, and the mean variance of FP is 0.0024. The five simulation results are similar, so it is concluded that the algorithm is stable, to detect the abnormal behaviors in real time.

3.5 Comparison of performances

In Section 2, the various detection algorithms' time complexity and space complexity are compared. This subsection, for dataset S_2, the average of TP and FP is taken and the two are compared with the negative selection algorithm, the dendritic cell algorithm and integrated immune algorithm. Though the methods of each experiment's settings are not identical, the true positive detection rate and false positive detection rate are shown in Table 6.

Table 6 Comparisons of different algorithms

No	Method	References	TP	FP
1	DCA	Ref. [7]	≈0.75	≈0
2	Optimization algorithm	Ref. [11]	98.1	0.036
3	IAIS	Ref. [12]	0.9691	0.0321
4	RSAI-IDA		0.9768	0.0317

From Table 6, the improved algorithm is compared with the optimization algorithm with Gaussian distribution optimization algorithm, the TP decrease by 0.42%, false alarm rate increase by about 0.43%, but the detection data in Ref. [11] is much less. Compared with DCA, TP is increased by more than 22%, and FP is increased by 2.68%. Through comparison with the above two algorithms, both of the true positive

detection rate and the false positive rate are increased. Compared with IAIS, the true positive detection rate reduces and the false positive detection rates are equal.

4 Conclusions

The artificial immune algorithms (NSA, DCA, IAIS) are proved to be effective methods which are used in real-time intrusion detection. While there are two problems of the traditional detection algorithms: the detection rate of DCA is low; the other two algorithms' time complexities are high. In this paper, after analyzing DCA and IAIS algorithm, an improved algorithm (RSAI-IDA) is proposed, and it's time complexity and space complexity are discussed. The time complexity is lower but space complexity is higher. In the training stage, space complexity increases, but the detection speed is higher. So, it is worth sacrificing some memory, and in the process of detection, the space complexity is equal.

The detection performances of RSAI-IDA are validated on KDD 99 dataset. Since there are many parameters of algorithm, only the values of some key parameters are analyzed. The length of the detector and self's radius are discussed with the simulations respectively. Finally, comparisons with the classical algorithms are given. The results show that RSAI-IDA algorithm, integrated the NSA and DCA, is effective for real-time detection. Rough set method is applied to obtain effective antibody, increase antigen detection rate, and the method of the random function to obtain parts of antibodies is to ensure the diversity of antibodies and decrease FP. All in conclusion, RSAI-IDA can be effectively used for anomaly detection, which can improve the speed of the intrusion detection and reduce the rate of false positives.

References

[1] Denning, Dorothy E. An intrusion detection model. *IEEE Transaction on Software Engineer*, 1987, 13: 222-232

[2] Sperotto A, Sadre R. Autonomic parameter tuning of anomaly-based IDSs: an SSH case study. *IEEE Transactionson Network and Service Management*, 2012, 9(2): 128-141

[3] Jiang F, Sui Y F, Cao C G. An incremental decision tree algorithm based on rough sets and its application in intrusion detection. *Artificial Intelligence Review*, 2013, 40(4): 517-530

[4] Kabir M M, Shahjahan M, Murase K. A new hybrid ant colony optimization algorithm for feature selection. *Elsevier-Expert Systems*, 2012, 39(3): 3747-3763

[5] Li S C, Yun X C, Zhang Y Z. Anomaly-based model for detecting HTTP-tunnel traffic using network behavior analysis. *High Technology Letters*, 2014, 20(1): 63-69

[6] Greensmith J, Aickelin U, Cayzer S. Introducing dendritic cells as a novel immune-inspired algorithm for anomaly detection. In: Proceedings of the 4th International Conference on Artificial Immune Systems (ICARIS), Springer, LNCS 3627, 2005. 153-167

[7] Wang J, Zhao X Z, Xu B, et al. Immune multi-agent model using vaccine for cooperative air-defense system of systems for surface warship formation based on danger theory. *Journal of Systems Engineering and Electronics*, 2013, 24(6): 946-953

[8] Suha A, Raed A, Alaa A. Virus detection using clonal selection algorithm with genetic algorithm. *Applied Soft Computing*, 2013, 13(1): 239-246

[9] Zeng J, Liu X J, Li T, et al. A novel intrusion detection approach learned from the change of antibody concentration in biological immune response. *Springer Applied Intelligence*, 2011, 35(1): 41-62

[10] Ou C M. Host-based intrusion detection systems adapted from agent-based artificial immune systems. *Neuro computing*, 2012, 88(1): 78-86

[11] Zhang F B, Xi L, wang S W. Real-valued multi--area self set optimization in immunity-based network intrusion detection system. *High Technology Letters*, 2014, 18(1): 1-6

[12] Chen Y B, Feng C, Zhang Q. Integrated artificial immune system for intrusion detection. *Journal on Communications*, 2012, 33: 125-131 (In Chinese)

[13] Pawlak Z. Rough sets. *International Journal of Computer and Information Science*, 1982, 11: 341-356

[14] Zhang L, Bai Z, Lu Y, et al. Integrated intrusion detection model based on artificial immune. *Journal of China Universities of Posts and Telecommunications*, 2014, 21(2): 83-90

[15] Qian J, Miao D Q, Zhang Z H. Knowledge reduction algorithms in cloud computing. *Chinese Journal Computers*, 2011, 34: 2332-2343 (In Chinese)

Zhang Ling, born in 1979. She received the Ph. D degree from Beijing University of Posts and Telecommunications in 2014. Her main researches are network security, artificial intelligence and cloud computing.