Acta Technica 62 No. 1B/2017, 509–520

Computer intrusion factor decomposition detection based on biotechnology

Tang Guoqiang[1]

Abstract. Computer monitoring system has been widely used in hydropower plants, which has provided a guarantee for the safe and economical operation of hydropower plants. In order to study the computer intrusion factor decomposition of biotechnology, the biological immune mechanism was used to establish an intrusion detection system based on artificial immune as a research hotspot. The intrusion detection system, biological immune system, artificial immune system, risk theory and other related research were adopted. The characteristics and limitations of the cellular algorithm were studied. The dendritic cell algorithm was improved and the dendritic cell algorithm was applied to the intrusion detection system. The final experimental results show that this method can better detect the computer intrusion factor and obtain less error rate, and the dendritic cell algorithm is suitable for the field of vibration anomaly detection.

Key words. Dendritic cell algorithm, intrusion detection, anomaly detection.

1. Introduction

As a proactive security technology, intrusion detection is considered to be the second safety valve after the security measures such as firewalls, it can monitor the network system in real time without affecting the network, thus providing real-time defense against external attacks, internal attacks and misoperation. Traditional intrusion detection techniques basically start with defining intrusion patterns or normal behavior patterns, and then compare detected data with these known patterns to detect intrusions. There is a large number of missing and false positives in the result, which lacks diversity, real-time and scalability. The biological immune system is naturally endowed with magical powers of self-preservation, which is not only able to resist a variety of foreign known or unknown pathogens threat, but also to repair the body's own damage and maintain the stability and balance of the body. This is what the current intrusion detection research institute expects to achieve, which is

[1] Department of Information Engineering, Guangzhou Institute of Technology & Engineering, 510925, China

conducive to overcome the shortcomings of the traditional intrusion detection system. There are many similarities between computer network security defense system and biological immune system in system function and architecture, which protect the system's security and maintain a stable balance of the system in complex environments. The specific immune defense, immune surveillance, immune recognition, immune tolerance, immune memory, immune regulation and immune homeostasis of the biological immune system, and stronger adaptability, diversity, distribution, dynamic and robust characteristics are the dream of the current intrusion detection system.

2. State of the art

The development of modern science cannot be separated from each other. In the cross field of information science and life science, artificial immune system developed by mimicking the immune system of the biological system has gradually become the focus of research, and has also provided new ideas and methods for intrusion detection [1]. As the advantage and weakness of the classifier, the dendritic cell algorithm indicated the future development direction of the dendritic cell algorithm, and discussed the applicability and limitations of the network intrusion detection system based on the dendritic cell algorithm. In view of the danger theory, some suggestions were given for the need for modification, and an improved intrusion detection method was proposed to apply the fuzzy set theory to the dendritic cell algorithm [2]. The blunt boundaries between mature and semi-mature dendritic cells in the original dendritic cell algorithm were changed by fuzzy decision. The experimental results show that the accuracy of the improved decision has been improved [3]. More and more researchers draw on and simulate the information processing ability of the immune system to design intelligent algorithms, establish artificial immune model, and apply to solve engineering and scientific problems, so as to make a lot of progress. But relative to the development of artificial neural networks, fuzzy systems and evolutionary computation, this is only the beginning [4]. Dendritic cell algorithm has been taken as the latest research in the theory of risk in artificial immunology, but the combination of many immune mechanisms and application problems of dendritic cell algorithms has not yet been fully understood. Whether it is a dendritic cell algorithm itself or its popularization, there are still many areas worth exploring and improving [5].

3. Methodology

As a full-time antigen-presenting cell in the innate immune system, dendritic cells (DC) can fuse a variety of environmental signals and correlate the signals with antigens to analyze the abnormalities of the antigen [6]. Inspired by the DC function, the dendritic cell algorithm (DCA) is designed by abstract modeling the antigen presentation behavior, which creates a new immune algorithm.

DC is currently known as the strongest antigen-presenting cells, as shown in

Fig. 1, when it matures, the cell membrane sticks out many long dendritic processes, so it gets its name. It belongs to an immune cell in the innate immune system, which plays a major role in perceiving the risk of the organism and controlling the immune response. DC collects antigen by endocytosis, or uses its dendrites to capture or retain antigens to process the collected antigens, presents antigens for cell recognition, and stimulates or inhibits cellular immune responses to antigens according to environmental signals [6].

Fig. 1. Dendritic cell

DCA is the latest research result of dangerous theory in artificial immunology. Its basic principle is to simulate the state transition process of DC in biological immune system, abstract it into a data structure similar to the signal processor, and calculate the input signal and obtain the state of the output signal through the signal processing process simulated by the linear signal processing model [7].

As shown in Table 1, the names and functions of the biological signals and the corresponding abstract signals and meanings are summarized.

When the immature DC processes a set of input signals, it sums the three output signals, respectively. When $\sum$csm in DC reaches the migration threshold (MT), it compares $\sum$semi and $\sum$mat: if $\sum$semi $>$ $\sum$mat, DC is converted to semi-mature state, the antigenic environment value $= 0$, which indicates that the cell environment is safe, and which means that the antigen is collected in the normal state; otherwise the DC is transformed into a mature state, and the presenting antigenic environment value $= 1$, the cellular environment is dangerous, which means that the antigen is collected in an abnormal state [8].

Table 1. Function of biological signals and their abstract signals and meanings

Biological signal	Features	Abstract signal	Meaning
PAMP	Indicating the presence of pathogens	PAMP	Indicating abnormal features
Necrosis signal	Indicating tissue damage	DS	Indicating high likelihood of abnormality
Apoptotic signal	Indicating tissue health	SS	Indicating a high likelihood of normal
Proinflammatory cytokine	Indicating that the tissue is damaged in general	IC	Zoom in other input signals
Costimulatory signal	Co stimulatory molecule	csm	Determines whether the iDC is converted
Mature output signal	Cytokines secreted by mDC	mat	Abnormal signal
Semi mature output signal	Cytokines secreted by smDC	semi	Normal signal

The basic principle map of DCA is shown in Fig. 2.

DCA can combine the treatment of a variety of environmental signals associated with the antigen, and analyze the anomalous index of antigen, which has the advantages of small scale calculation, fast response speed and strong recognition ability. In particular, there is no difference between known and unknown data for DCA, it recognizes invasion as long as it is aware of danger, even if it is the first encounter, there are no extensive training and centralized control [9]. Therefore, DCA is suitable for the distributed real-time intrusion detection between the internal network and the external network, the subnets and the nodes in the computer monitoring system of the hydropower plant. Of course, DCA will also produce false positives and omission. In order to meet the stringent requirements of network security for computer monitoring systems in hydropower plants, an intrusion detection system model combining with innate immunity and adaptive immunity is designed and the intrusion detection system which is connected to DCA is tested through the KDD Cup99 data set [10].

The system mainly includes antigen and signal acquisition module, DCA detection module, detector module, intrusion comprehensive evaluation module, administrator confirmation module [11]. Since this paper focuses on DCA-based intrusion detection, the adaptive immune part is simplified.

Antigen and signal acquisition module is mainly to simulate the collection of antigen and various signals in the system. On the one hand, it provides DCA with antigen flow and signal flow; on the other hand, it provides the antigenic flow directly to the detector module. Because intrusion often leads to host and network anomalies,

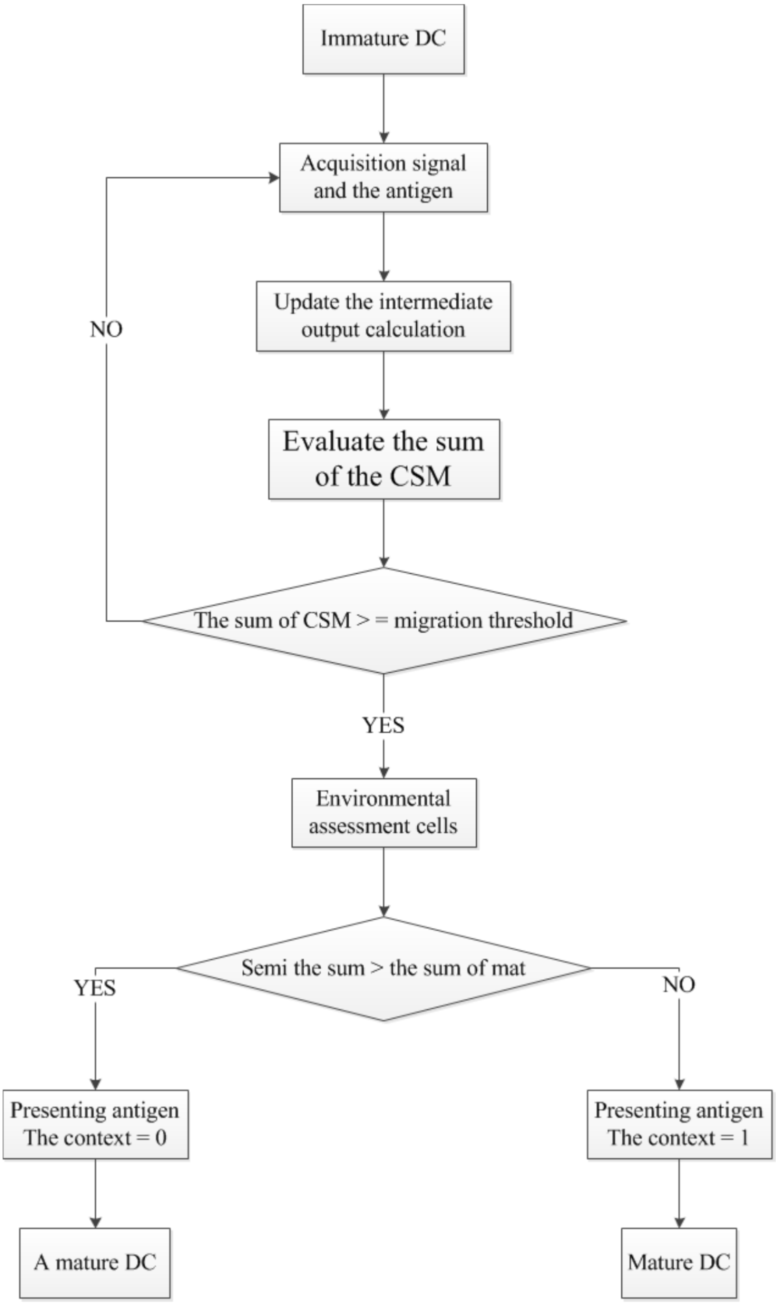
Immature DC
Acquisition signal and the antigen
Update the intermediate output calculation
Evaluate the sum of the CSM
The sum of CSM > = migration threshold
NO
YES
Environmental assessment cells
Semi the sum > the sum of mat
YES
NO
Presenting antigen The context = 0
Presenting antigen The context = 1
A mature DC
Mature DC

Fig. 2. DCA basic schematic diagram

this can be used as a basis for DCA signal extraction. PAMP indicates that a clear exception can be described by repeated login failures, connection errors, and so on in the intrusion detection. Danger signal DS indicates that the possibility of anomaly is very large. In the intrusion detection, it can be expressed by the high utilization ratio of CPU and memory, the rapid change of network traffic, the exhaustion of bandwidth, the abnormal data receiving and sending and the high request of host connection and so on [12]. The security signal SS indicates that the organizational environment is in a normal state and can be represented by stable network traffic, low memory usage in intrusion detection.

Step 1: An antigen cell with the length of m is initialized, and the first m entries of the dataset are put into the pool.

Step 2: The DC is initialized, and the random migration threshold and life span in a certain range are set.

Step 3: DC randomly takes samples from the antigen signal pool within the lifetime, accumulates three output values: csm, semi, mat, DC which is over lifetime is re-initialized and transferred to Step2.

Step 4: Whether the cumulative csm reaches the migration threshold is determined, if it is not, which needs to be transferred to Step3 to continue sampling.

Step 5: The antigens used in DC are counted, the number of times and the number of exceptions for each antigen are recorded.

Step 6: Whether there is an antigen that has reached the number of times of determination N is determined, if so, the antigen of the MCAV is calculated, the abnormality of the antigen and output it in real time is evaluated, and then the evaluated antigen is removed from the antigen cell, otherwise, it is needed to turn Step2.

Step 7: Whether there is a new antigen is determined, if so, the new antigen will be added to the sampling pool to fill the vacancies removed, turn Step2.

Step 8: Whether any antigens in the pool have not been evaluated is determined, if so, then it is needed to turn Step2, otherwise the algorithm terminates.

The whole system model is based on the combination of congenital and adaptive immunity, and congenital immunity and adaptive immunity are not isolated, but are interrelated. Innate immunity is instructive for more accurate body non-auto recognition [13]. In this module, in addition to the realization of DCA detection function, the detector set is dynamically updated according to the detection results.

The detector module consists of two sets of detectors; one is the general set of detectors, which contain a large number of detectors, and can cover as large a detection space as possible. The other is a collection of memory detectors that contain relatively few detectors, but are aimed at typical invasion features. The number of memory detectors is relatively small, which is designed to reduce computational overhead and expect fast response. For typical intrusion features, it fully reflects its specificity and strengthens the detection accuracy. The antigenic flow provided by the antigen collection module first passes through the memory detector module and the antigen detected by the memory detector has a high confidence in the results [14]. The detected antigen detected by DCA will pass through the general detector module, and the results of the test and the results of the DCA test are sent to the comprehensive evaluation module for intrusion. The detector set in the module is dynamically updated. On the one hand, it comes from the assistance of DCA test results; on the other hand, it is adaptive development through its own cloning, mutation, life cycle, elimination mechanism.

The general responsibility of the integrated evaluation module is to determine the final test results, the general principle is: firstly, the detection results of memory detectors are used as standard, unless the detector set is changed by the results of the test and the confirmation of the administrator, therefore, the detection results of memory detectors are generally not processed and can be output directly; secondly, the determination detection of DCA is taken as the primary. DCA has the strong real-time and high detection accuracy. The determination of the DCA test results provides assistance for the dynamic update of the detector with the participation of the administrator; thirdly, the antigen in the interval of DCA is detected by the general detector and then evaluated with the results of DCA test.

The administrator confirmation module is set up for the decision of the special case, for example, the detection result updates the control of the detector as well as the final verdict of the intrusion synthesis evaluation.

KDD99 data set is a benchmark database used in intrusion detection field. So far, it is the most used and most authoritative data set for intrusion detection researchers. The experiment used a subset of the KDD99 data set of 10 % (kddcup.data_10_percent.gz), including a total of 494021 data, of which 97278 were normally connected and 396743 attacks. The 10 % subset had similar statistical properties to the complete data set, maintained a similar normal connection and attack ratio. Each connection in the KDD99 data set had 42 data items, and the last item was a sign to mark the journey. Anomaly detection based on DCA was a two classification algorithm, which only detected whether it was intrusion or not, and could not tell which was the invasion. Therefore, all normal connected tags were set at 1 (Class1), which indicated normal; and all attack markers were set at 2 (Class2), which indicated an exception. Then, the remaining 41 items were examined, and some of the data items had little effect on the detection and could be removed to reduce the dimension of the data set.

The information gain of the property shows the statistical correlation of the attribute to the classification result. In the information gain, if the feature can bring more information for the classification system, then it is more important. The information gain of each attribute was calculated and analyzed, and the attribute of the low information gain was deleted from the data set. Eventually, 10 data items were selected to extract the three signals required for DCA:

PAMP:serror_rate?srv_serror_rate;same_srv_rate;dst_host_serror_rate?dst_srerror_rate?

Dangerous signal: count;srv_count?

Safety signal: logget_in;srv_diff_host_rate;dst_host_count?

The specific information of these 10 data items is shown in Table 2. The value of each attribute was normalized to the [0,100] interval according to the linear function conversion method. The average value of each type of signal was the value of such signal.

Table 2. Data item information for KDD99 data

Features	Description	Types	Ranges
serror_rate	Over the past 2 seconds, the percentage of connections with "SYN" errors occurred in the connection with the same target host as the current link.	continuous	[0.00,1.00]
srv_serror_rate	Over the past 2 seconds, the percentage of connections with "SYN" errors in the connection with the same service as the current link	continuous	[0.00,1.00]
same_srv_rate	In the last 2 seconds, in the connection with the same target host as the current link, the percentage of connections with the same service as the current connection	continuous	[0.00,1.00]
dst_host_serror_rate	In the first 100 connections, the percentage of connections with "SYN" errors in the connection with the same destination host in the current connection	continuous	[0.00,1.00]
dst_srerror_rate	In the first 100 connections, the percentage of connections with "REJ" errors in the connection with the same destination host in the current connection	continuous	[0.00,1.00]
count	The number of connections for the target host that has the same current connection as the current connection over the last 2 seconds	continuous	[0,511]
srv_count	The number of connections that have the same service as the current connection over the last 2 seconds	continuous	[0,511]
logget_in	Successful landing is 1, otherwise 0	dispersed	0 or 1
srv_diff_host_rate	In the last 2 seconds, the percentage of connections with different target hosts that connect to the current connection in a connection that has the same service as the current connection	continuous	[0.00,1.00]
dst_host_count	The number of connections in the first 100 connections that have the same target host as the current connection	continuous	[0,255]

4. Result analysis and discussion

The experimental data was randomly selected from the entire data set according to the attack type.

Experiment 1: first of all, DCA experienced a wave attack after a stable normal environment. 100 normal connections +100 smurf attacks were taken as the experimental data. The experimental results are shown in Table 3.

Table 3. Statistics of results of DCA running 10 times

Numbering	Number of false positives	The number of missed	False alarm rate (%)	False negative rate (%)	Accuracy (%)
1	2	0	1.00	0.00	99.00
2	1	0	0.50	0.00	99.50
3	4	0	2.00	0.00	98.00
4	0	0	0.00	0.00	100.00
5	4	0	2.00	0.00	98.00
6	3	0	1.50	0.00	98.50
7	2	0	1.00	0.00	99.00
8	5	0	2.50	0.00	97.50
9	2	0	1.00	0.00	99.00
10	1	0	0.50	0.00	99.50
average	2.4	0	1.20	0.00	98.80

Table 3 shows the results of the experiment run statistics. It can be seen that the average detection accuracy was 98.80 %, missing was 0, false positives were 0–5, and all of them occurred in the transition period of the two types of environmental change, which conformed to the characteristics of DCA. The experimental error threshold was set to 0.6, if it was set to 0.8 or 0.9, the experimental results were better. The detection accuracy was more than 99.50 %, which showed that DCA had good classification effect on normal connections and smurf attacks. If the abnormal threshold was set to 0.5, the detection accuracy would be reduced to 97.5 %.

It is necessary to note that since the exception threshold is set to 0.8 or 0.9 to achieve the best detection results, why the exception threshold is set to 0.5, the reason is that in addition to detecting smurf attacks, it is necessary to face other more attack types, Some of the attack threshold is low, for example, it can be detected and only it is set to 0.3, otherwise the false negative rate will be high. Experiments with multiple abnormal thresholds are prepared for the final DCA's comprehensive experiment against multiple attacks.

Experiment 2: the order of the experimental data used the normal connection data + attack data alternately placed form: 100 normal connection +100 smurf attack +100 normal connection +100 neptune attack + ... The test results are shown in Table 4.

Table 4 shows the comprehensive detection data of the different anomaly threshold. As can be seen from the experimental results, when the abnormal threshold was set to 0.5, the detection accuracy was the highest, which reached 88.8 %. It should be noted that the multiplication mechanism for the disordered data set is not used in the experiment because the data stream with a length of 100 can form a stable environment compared to an antigen signal pool with a length of 20. But in the actual environment, it will appear intermittent and brief attack. In order to detect such attacks, MMDCA with multipath merge mechanism can be used.

Table 4. Comprehensive detection data for different abnormal thresholds

Numbering	Abnormal threshold	Average false positive rate (%)	Average false negative rate (%)	Average inspection accuracy (%)
1	0.8	0.3	15.2	84.5
2	0.7	0.6	13.0	86.4
3	0.6	1.2	10.9	87.9
4	0.5	1.9	9.3	88.8
5	0.4	3.3	8.1	88.6
6	0.3	4.8	6.7	88.5

5. Conclusion

The computer monitoring system of hydropower plant is responsible for the supervision and control of the main auxiliary equipment of hydropower plants. In order to ensure its safety, the traditional DCA off-line analysis process was improved, and an online analysis mechanism parallel to the detection process was designed, an antigen that was evaluated for a sufficient number of times was exported, thus achieving the goal of real-time or near real-time analysis. The conclusions were drawn as follows: the distributed real-time self-protection mechanism of biological immune system provides a new idea for the study of intrusion detection. DCA is a congenital immune algorithm based on the dangerous theory, and it is not based on pattern matching of antigenic features, but rather associates antigen with signals and assesses the degree of abnormalities of the antigen according to the degree of environmental risk, which has the characteristics of small operation scale, fast response speed and strong recognition ability. For DCA, there is no difference between known and unknown data, so it does not need a lot of training and centralized control, which is suitable for the distributed real-time intrusion detection in the computer monitoring system of hydropower plants. However, the study still has some limitations, for example, intrusion detection is only the first step in intrusion prevention, which should combine intrusion detection with firewall, vulnerability scanning, antivirus software and other security products to establish a more complete intrusion prevention system.

References

[1] J. C. Ni, Z. S. Li, J. R. Sun, L. P. Zhou: *Research on differentiation model and application of dendritic cells in artificial immune system.* Acta Electronica Sinica *36* (2008), No. 11, 2210–2215.

[2] Z. Chelly, Z. Elouedi: *Hybridization schemes of the fuzzy dendritic cell immune binary classifier based on different fuzzy clustering techniques.* New Generation Computing *33* (2015), No. 1, 1–31.

[3] T. Schulze, S. Golfier, C. Tabeling, K. Räbel, M. H. Gräler, M. Witzenrath, M. Lipp: *Sphingosine-1-phospate receptor 4 (S1P*$_4$. FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology *25* (2011), No. 11, 4024–4036.

[4] I. Hansenne, C. Renard-Charlet, R. Greimers, V. Geenen: *Dendritic cell differentiation and immune tolerance to insulin-related peptides in Igf2-deficient mice.* Journal of Immunology *176* (2006), No. 8, 4651–4657.

[5] S. Liu, J. Ke: *Method of locating anomaly source in software system based on dendritic cell algorithm.* Applied Mechanics and Materials *556–562* (2014), 6255–6258.

[6] A. A. Al-Hasan, E. S. M. El-Alfy: *Flexural vibrations of non-homogeneous elliptic plates.* Procedia Computer Science *52* (2015) 244–251.

[7] J. Y. Zhao, H. Guo, L. F. Wang: *Chinese spam filtering algorithms based on case base reasoning.* Microelectronics & Computer *26* (2009), No. 12, 64–66.

[8] Z. H. Ouyang, X. Y. Zhang, S. Tu, J. J. Gu: *Research on problem of application of "Cloud security" in computer anti-virus.* Computer and Modernization (2010), No. 12, 72–75.

[9] S. K. Choi, J. H. Song, S. H. Kim, Y. D. Lee: *Characteristics of the load of small hard body used for impact resistance test of the lightweight wall.* Journal of the Korea Institute of Building Construction *14* (2014), No. 4, 350–358.

[10] D. Guo, X. L. Hong, X. Hu, X. Wu: *A bit-parallel algorithm for sequential pattern matching with wildcards.* Cybernetics and Systems *42*, (2011), No. 6, 382–401.

[11] M. Asaka, T. Onabuta, T. Inoue, S. Okazawa, S. Goto: *A new intrusion detection method based on discriminant analysis.* IEICE Transactions on Information and Systems *E84-D* (2001), No. 5, 570–577.

[12] E. F. Fernández, F. Almonacid, N. Sarmah, P.Rodrigo, T. K. Mallick, P.Pérez-Higueras: *A model based on artificial neuronal network for the prediction of the maximum power of a low concentration photovoltaic module for building integration.* Solar Energy *100* (2014), 148–158.

[13] A. Sano: *Generating novel memories by integration of chaotic neural network modules.* Artificial Life and Robotics *4* (2000), No. 1, 42–45.

[14] J. Su, P. L. Qiao, Y. H. Liu: *Distributed intrusion detection method based on immune evolution computing.* IJ Solids and Structures *36* (2010), No. 6, 163–165.

Received July 12, 2017