The Method of Anomaly Diagnosis in Image Server Based on the Dendritic Cell Algorithm

Liu Sai, Ke Jie
Computer School of Wuhan University
China
liussaaii@163.com, madfrog.jk@gmail.com



ABSTRACT: *At present WEB application server especially image server lacks of effective methods of anomaly diagnosis. Therefore, this paper presents the method of anomaly diagnosis based on the dendritic cells algorithm of Danger Theory. The method can achieve the purpose of monitoring system status effectively by building a multi-layer indicator system. Experiments prove that the application of the dendritic cells algorithm to anomaly diagnosis in image server is effective.*

Keywords: Artificial Immune System, Image Server, Anomaly Diagnosis, Dendritic Cell Algorithm

Received: 17 May 2015, Revised 20 June 2015, Accepted 25 June 2015

© 2015 DLINE. All Rights Reserved

1. Introduction

Images are widely used in social networking sites, e-commerce sites, which are commonly involved in the high concurrent access and process of the massive images. Image server is the application server used to respond to image access requests of users as well as image process. In this context, it requires image server being able to respond to user requests quickly and process image efficiently, detect anomaly timely and predict when anomaly may occur and repair it. The judgment of these anomaly includes whether the system's response to user access request is delayed, whether IP called remotely in image process is legitimate, whether the image is accessed in the cache, whether the treatment of the thumbnail is successful, whether the throughput of image process per unit time is reasonable, etc. It is important to find in time anomaly that may occur in the system, locate anomaly, and effectively assess the health status of the system.

This paper discusses the anomaly diagnosis method of the software system in image server. Since the software system is a complex system [1], internal software anomaly or external intrusion in the system may cause the system out of balance, reliability and availability. Image server is characterized by high concurrent processing massive image, which means timely and effective detection system anomaly is necessary. Traditional anomaly detection [2] is the method based on feature selection, Bayesian inference, model predictions, machine learning and data mining, etc, They construct a collection of anomaly behaviors, and then find anomaly according to this collection. Traditional methods usually require large calculation, leading to inefficiency, untimely response and sometimes false positives.

Computer immunologists have applied the artificial immune in anomaly detection system [3], for example, port scanning, robot learning and wireless sensor networks. However, the traditional model of artificial immune is self –non self, which causes a high false positive rate in the results of anomaly detection. Bio-immunologist Matzinger proposed the Danger Theory [4], holding that immune recognition is not governed by a self/nonself discrimination but response to the danger signals in the body tissue damage. Greensmith, an immunologist of the Uweteam constructed artificial dendritic cell combined Danger Theory and proposed DCA (Dendritic Cell Algorithm)[5]. He used DCA in antigen-presenting and adopted the result to T cell to immune response. The team took UCI Wisconsin Breast Cancer dataset [6] as the object of study. Green smith applied dendritic cell algorithm to SYN packet detection [7], robot classification [8], and zombie process detection [9]. FengGu, Greensmith and other members of the team analysed and improved dendritic cell algorithm [10] [11].

Due to the indicator system involved in anomaly diagnosis of image server is more complicated, it is necessary to take the indicator of the operating system layer as well as the indicator of the application service layer and network layer into account. It can achieve better detection results in this context using dendritic cell algorithm based on innate immune, using diversity and adaptability of the biological immune.

Currently, there is no uniform definition to health diagnosis of complex software systems. Health diagnosis used for machine diagnostics [12], highway tunnel detection [13] and bridge structure detection [14]. The main work of this paper is anomaly diagnosis for software system of image server (the system in response to user image access and process requests), which is an important part of health diagnosis in image server.

2. Dendritic Cell Algorithm

In immunity, dendritic cells migrate from the lymphatic system to the body tissue, collecting antigen signal and returning to the lymphatic system, to drive T cell immune response. Dendritic cells have three states[5], are iDC (immature DC), mDC (mature DC), smDC (semi-mature DC). When the iDC received signal PAMP (Pathogen-Associated Molecular Pattern), it indicates the signal received at this time is exogenous (indicating external intrusion), or iDC received signal DS (Danger Signal), it indicates the signal received at this time is endogenous (indicating internal lesions), or both received, then iDC will turned into mDC, it is mature dendritic cells, this moment it secrete IL-12 cytokine [6]. When iDC received signal SS (Safe Signal), iDC will turned into smDC, it is semi-mature dendritic cells, this moment it secrete IL-10 cytokine. When the concentration of IL-12 is greater than concentration of IL-10, it indicates that the body tissue is in an abnormal environment, when the concentration of IL-12 is less than the concentration of IL-10, it indicates that the body tissue is in a normal environment. Meanwhile, mature dendritic cells and semi-mature dendritic cells would secrete CD80/86, when the concentration exceeds a threshold value, DC in the organization will be moved back to the lymphatic system, antigen - presenting to T cells cause immune response [6].

Dendritic cell state transition diagram is shown below.

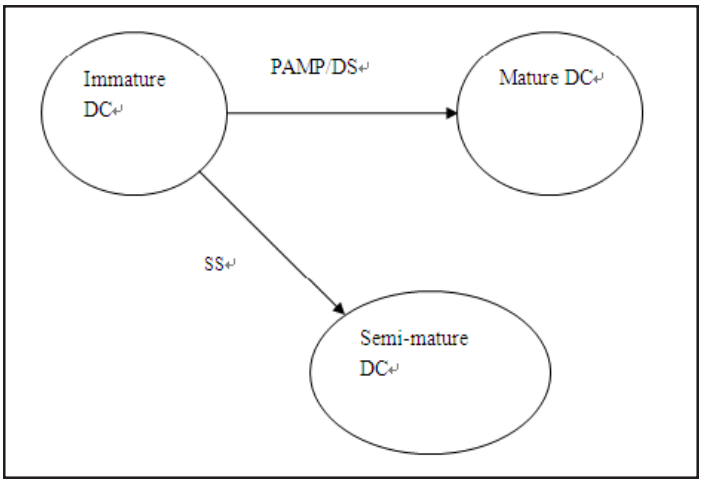

Figure 1. Dendritic Cell State Transition Diagram

Greensmith abstracted DC for antigen-presenting in artificial immune and proposed DCA (Dendritic Cell Algorithm) [5] . DC recorded CSM (Concentration of Co stimulatory Molecules), concentration of Mature cytokine, concentration of Semi-Mature cytokine. CSM threshold value determined the time of state migration and the number of antigen collected while concentration of Mature cytokine and Semi-Mature cytokine determined whether the environment DC in was safe.

Signals DC collected were PAMP, DS,SS and IC (Inflammatory Cytokines).Output signals were CSM, Mature and Semi-Mature. In which, PAMP signals influenced CSM and Mature signals, DS signals influenced CSM and Mature signals. SS signals influenced CSM, Mature and Semi-Mature signals. Inflammatory Cytokines played a role in amplifying the signals[6].

The conversion formula of output signal (1) is as follows [9]. The output signal is presented in Equation 1.

$$O_{ip} = \sum_{j=0}^{2} (S_{ij} * W_{jp}) \quad (1)$$

i indicated the number of collected antigen;

j = 0, 1, 2 indicated separately three input signals PAMP, DS, SS;

p = 0, 1, 2 indicated separately three output signals CSM, Mature, Semi mature;

S_{ij} indicated the concentration of j-th input signal of the collected i-th antigen;

O_{ip} indicated the concentration of p-th output signalof i-th antigen;

In this paper, DCA of computer immunity was used to anomaly detection of image server. It gives the indicator system in health diagnosis environment of image server first, and then designs the algorithm. Finally, the systematic experiment was made and a conclusion was drawn.

3. Image server Anomaly Diagnosis Method

3.1 Image Server Anomaly Diagnosis Indicator System

The software system of image server is a complex system. It needs to classify the collected indicators in carrying out anomaly diagnosis. The indicators can be divided into three layers: operating system layer, application service layer and network layer. The indicators related to system resources acquisition can be classified as the operating system layer, such as CPU usage, memory usage, image cache space situation, etc. The indicators related to application services can be classified as application service layer, such as image processing time, image processing realization, and times of image processing per unit time. The indicators related to network layer can be classified as network layer, involving whether image processing remote calls is in IP white list, whether image request produces 404 error etc.The establishment of the above system can better monitor the status of the image server.

On the basis of the above system, this paper defines PAMP, DS and SS as follows.
Since PAMP indicates abnormal environment, with contrast of image server environment, the following indicators are selected to define PAMP.

PAMP = (I , T , F , δ) in which, the elements are defined as follows:

I: Indicates whether image processing remote calls is in IP white list;

T: Indicates whether image processing is failed;

F: Indicates whether image request produces 404 error;

δ : Is the mapping function meaning the mapping values based on the three indicators above, taken as PAMP.

I records whether image processing remote call is legitimate call. If the call *IP* is not in the white list, indicating that it is illegal call and an anomaly. Trecords whether image processing is failed. If image processing (such as thumbnails processing) is failed, it is an anomaly. F records

whether user image access request produces 404 error. If it produces 404 error, it indicates image links is abnormal.

Since DS indicates the environment it is in has higher likelihood of anomaly, a contrast of image server environments is made to select with the following indicators to define DS.

DS = (P , C , M , R , E , Φ) in which, the elements are defined as follows:

P: Indicates image processing time;

C: Indicates CPU usage;

M: Indicates the memory usage;

R: Indicates the remaining image cache situation;

E: Indicates throughput of image processing in unit time;

Φ: Is mapping function, which indicates the mapping value drawn from the front five indicators, the value is as DS signal.

P records thumbnail processing time of a single image. When processing time is long, the likelihood of anomaly occurrence is high. *C* records *CPU* usage of server, in the case of high CPU usage, image processing speed may be slowed down. *M* records the usage of server memory pages. When available page is inadequate, the frequent server paging would lead to slow image processing. *R* records remaining image cache of server. In lack of remaining cache, image processing may be abnormal. *E* records the throughput of server image processing in unit time. When the throughput is close to the threshold, image processing may be abnormal. We use D_j to representative antigen signal. The *DS* signal is presented in Equation 2.

$$\Phi_j = (\sum_{j=1}^{m} (| D_{ij} - (\sum_{i=1}^{n} D_{ij}) / n |)) / m \qquad (2)$$

In Equation 2, *m* represented the number of elements defined in *DS*. And *n* represented the number of antigens.

Because *SS* signal indicate *sa* high likelihood of normal, a contrast of image server environment is made to select the following indicators to define *SS*.

SS = (I , T , F , P , E , $\mathcal{G}$),in which, the elements are defined as follows:

I: indicates whether the image processing remote calls is in IP white list;

T: indicates whether the image processing is failed;

F: indicates whether the image request produce 404 error;

P: indicates image processing time;

E: indicates throughput of image processing in unit time;

$\mathcal{G}$: is mapping function, which indicates the mapping value drawn from the above five indicators, taken as SS.

For *I*, if image processing remote calls is in IP white list, it is a normal processing request. For *T*, if image processing is successful, then it indicates the current server environment is normal. For *F*, if 404 error has not occurred, it indicates a high likelihood of normal. For P and *E*, if the image processing time and throughput are within a reasonable threshold, it indicates current environment of system has a high likelihood of normal.

	PAMP	DS	SS
Application Layer	T	P , E	T , P , E
Network Layer	F		F
System Layer	I	C , M , R	I

Table 1. Indicator System Table of Image Server Anomaly Diagnosis

Table 1is the three layer indicator system of image server anomaly diagnosis.

3.2 Weights Matrix

This paper refers to weights matrix[6] value of traditional DCA proposed by Greensmith and defines weights matrix of input signals PAMP, DS, and SS with regard to output signals CSM, SemiMature, Mature (as shown in Table 2). In Table 2, PAMP has a positive impact on CSM and Mature. Influence weights are 2. DS has a positive impact on CSM and Mature. Influence weights are 1.SS has a positive impact on CSM and Semi mature, influence weights are2 and 3.SS has a negative impact on Mature. An influence weight is- 3

	CSM	SemiMature	Mature
PAMP	2	0	2
DS	1	0	1
SS	2	3	-3

Table 2. WeightsMatrix Table

3.3 Algorithm

```
Input: antigen signal eigenvectors
Output: security status of environment where antigen is in
1 Initialize antigen pool;
2 isNeedInitDC = true
3 while isNeedInitDC do
4     {
5           Initialize DC cells, set CSM migration threshold;
6           while CSM of the DC does not reach the migration threshold do
7           {
8           Select unlabeled antigen randomly and calculate input signal PAMP,DS,SS;
9           Calculate output signals CSM, SemiMature, Mature and accumulate;
10          If CSM reach migration threshold then set marks;
11          }
12 Calculate and draw DC status;
13 Calculate the determine number(total number, abnormal number) of the antigen
collected              by DC and accumulate the numbers;
14 If there is antigen that has reached the threshold of determine number in the
unlabeled antigen then
15   {
16          Calculate MCAV of the antigen;
17          Compare it with abnormal threshold and mark whether the antigen is abnormal;
18          if there is unlabeled antigen then set is Need InitDC = false;
19    }
20  }
```

Line 1 initialize antigen pool, import the indicator values collected.

Line 5 initialize DC cells, including the life cycle, set random CSM migration threshold (CSM migration threshold has upper and lower limits) of the DC.

Line 6 to 11 show that DC continues antigen acquisition until CSM of it reaches migration threshold.

Line 8 obtains value of the three signals according toPAMP, DS, and SS calculation defined in 2.1.

Line 9 calculates CSM, Semi mature, Mature according to the weights matrix in Table.2, and adds to the cell structure of DC.

Line 12 indicates a state transition when CSM of DC reaches migration threshold. When Semi mature is greater than Mature, DC migrates to the semi-mature status, which indicates that the environment DC in is safe. Meanwhile, mark the antigen collected by DC as normal. When Semi mature is less than mature, DC migrates to a mature state, which indicates that the environment DC in is insecure. Then, mark the antigen collected by DC as abnormal;

Line 13 indicates the total number of each antigen of DC. Update the number of times each antigen is determined to abnormal (if the state of DC is Mature).

Line 14 to 19 shows there is antigen that has reached a determination threshold in antigen pool. Mark the antigen.

Line 16 calculates the ratio of the number of times the antigen is determined abnormal and the total number of times determined;

Line 17 determines whether the antigen is abnormal (comparing voting results of DC with anomaly threshold);

In order to simplify the problem, this paper does not consider the impact of proinflammatory cytokines.

4. System Experiment and Results Analysis

This paper sampled 150 antigen records based on time series from the image server randomly. The number of normal antigen is 105 and the number of abnormal antigen is 45. Each antigen record contains antigen ID, whether image processing remote calls is in IP white list, whether image processing is failed, whether the image access request is 404 error, image processing time, CPU usage, memory usage, server local image cache, image processing throughput in the unit time, and antigenic properties (including whether it is abnormal).

Experiment 1, a single detection for antigen data is set, separate detection three times.

Figure 2 is a schematic diagram of a single detection results. Figure 3 is a schematic diagram of secondary detection results. Figure 4 is a schematic diagram of third detection results. Due to the randomness of algorithm, the detection results are different, but it does not affect the accuracy of detection results.

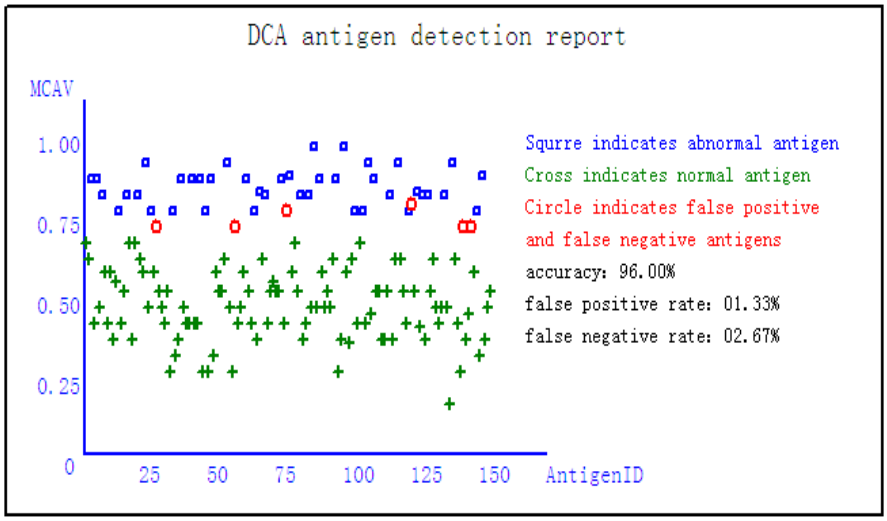

Figure 2. Reporting Figure of Single Detection Results

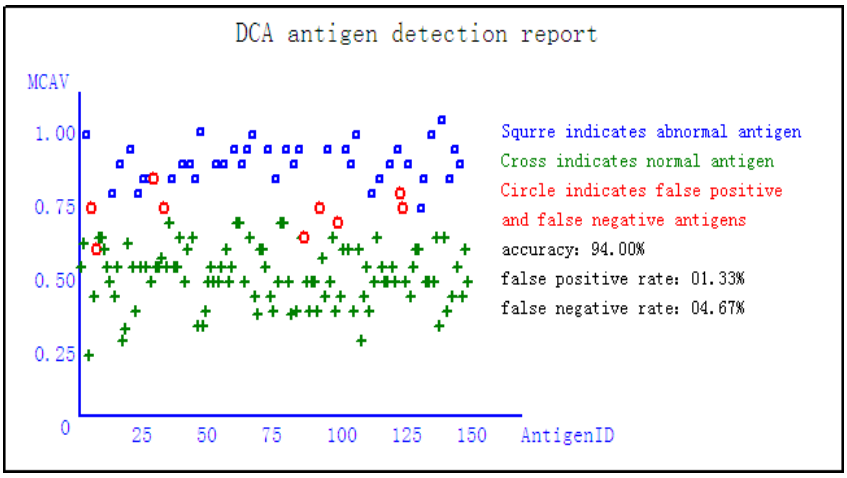

Figure 3. Reporting Figure of Secondary Detection Results

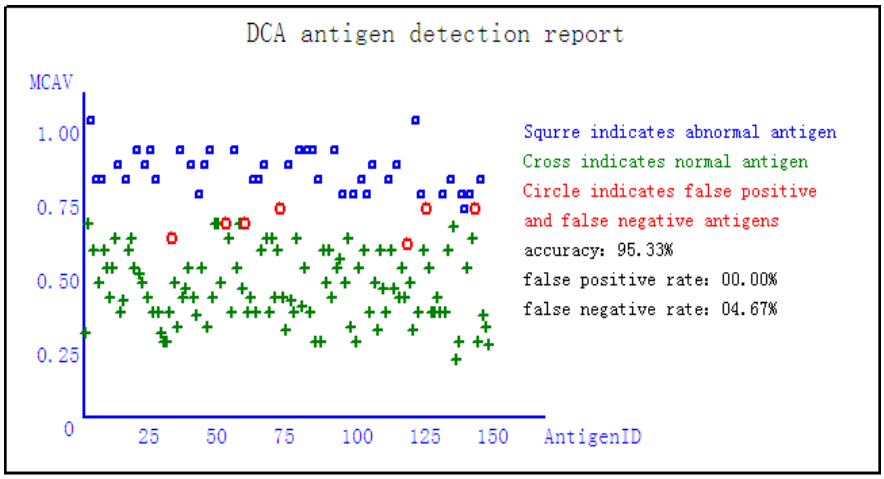

Figure 4. Reporting Figure of Third Detection Results

In the figure above, abscissa indicates antigen ID, ordinate indicates anomaly determination value of antigen. When MCAV value of the detected antigen is greater than or equal to the anomaly threshold, the antigen is labeled as abnormal antigen in blue. When MCAVvalue of detected antigen is less than the anomaly threshold, the antigen is labeled as normal antigenin green. Red indicates false positive(normal antigen is determined to be abnormal) or false negatives (the abnormal antigen is determined to be normal) antigens. The figure also gives statistical value of each detection, including accuracy, false positive rate, and false negative rate, etc.

Due to the randomness of detection algorithm, migration threshold of CSM each DC generates randomly within a range. Each antigen sampled by DC is selected randomly from the unlabeled antigen in the antigen pool. The number of each antigen sampled by DC is not fixed. Each antigen is sampled randomly by several DC, therefore there is some difference in each detection results above. To obtain more accurate results, for the same antigen data set, repeated detections are done 10 times, 50 times, and 100 times detection. Then statistical analysis of the experimental results is made and more accurate detection results can be obtained. In experiment 2, duplicate detection for the antigen data is set. Ten times duplicate detection, 50 times duplicate detection, 100 times duplicate detection is made.

Table 3 shows the statistical value of the experimental results.

	Total antigen	Number of Accurate Determination	Number of False Positive	Number of False Negative	Accuracy	False Positive Rate	False Negative Rate
10 times Detections	150	1430	11	59	95.34%	0.73%	3.93%
50 times Detections	150	7141	51	308	95.21%	0.68%	4.11%
100 times Detections	150	14232	103	665	94.88%	0.69%	4.43%

Table 3. Statistical Analysis Table of Detection Results

In Table 3,the first line was the statistical results of 10 times duplicate detections. The total number of antigens detected was 150.The number of times the antigens determined accurately was 1430. The number of false positives was 11 and the number of false negatives was 59. The accuracy of detection was 95.34%. The false positives rate was 0.73%. The false negatives rate was 3.93%. In Table 4-1,the second line was the statistical results of 50 times duplicate detections. The number of times the antigens determined accurately was 7141 and the number of false positives was 51. The number of false negatives was 308. The accuracy of detection was 95.21% and the false positives rate was 0.68%.The false negatives rate was 4.11%. In Table 4-1,the third line was the statistical results of 100 times duplicate detections. The number of times the antigens determined accurately was 14232.The number of false positives was 103 and the number of false negatives was 665. The accuracy of detection was 94.88%.The false positives rate was 0.69% and the false negatives rate was 4.43%. The detection results above were obtained in the same test. Analysis of the results of experiment 2 showed that the accuracy of image server health diagnosis was about 95% by using dendritic cells algorithm. False positive rate was low, at only about 0.7%, However, the false negative rate was high, at about 4%. Experiments showed that the application of dendritic cells algorithm to image server health diagnosis was valid and accurate.

5. Conclusions and Future Work

Image server system is a complex system. At present, effective application of anomaly diagnosis to the software system is in lack. These methods are dependent on monitoring and management of operation and maintenance personnel or conventional anomaly detection methods.

These methods cannot be adapted to the complex environment of image server. They need large computation but efficiency and accuracy is relatively low. In this paper, dendritic cells algorithm (for antigen-presenting and antigen recognition)of computer immune Danger Theory was introduced into anomaly diagnosis of images server. It defined the three-layer indicators system of image server anomaly diagnosis, and gave related definitions of PAMP, DS, SS under such environment. Finally, accurate and efficient detection results were obtained through the experiments, which indicated that Danger Theory applied to the anomaly diagnosis of information system was feasible and effective. The next job was to analyze the problem of high false negative rate the method caused and make improvements. Meanwhile, it could further establish an efficient evaluation system made quick and proactive defense and to seek ways to help the system return to normal. In the future, the impact of artificial setting parameters and thresholds needs to be reduced but the adaptability of the system needs to be enhanced.

Moreover, this paper has used dendritic cells for antigen-presenting while ignored the role of *T* cells. In biological immune, T cell immune response is done after antigen-presenting. The influence of T cells after the feedback can be considered as follow-up to further improve the algorithm. In this paper, the weights matrix used in calculating CSM, Semimature, Mature according to PAMP, DS, SS is constructed based on experience value. The dynamic automatic optimization weights matrix can be considered according to the feedback of algorithm output in subsequent work.

In addition, further researches into health diagnosis problems of image application server can be done, including the anomaly

diagnosis of systemas well as the coping strategies after anomaly is detected and how to get the system back to health status. Here, when the indicator system is normal and the system status is health; it is available and reliable and can response to user access requests and image processing requests. When part of the indicator system is abnormal, the system status is sub-health. Althoughthe system can be used temporarily, it is more likely to be unavailable. WhenCPU usage is too high, the speed of system response to image processing becomes slow, and response timeout can lead to image processing failure. When the system cannot response to user requests normally, it is no longer available and remains in a non-healthystate, causing frequent failure of image processing, illegal remote using of image processing and so on.

References

[1] Yarushkina, N. (2001). Soft computing and complex system analysis, *International Journal of General Systems*, 30 (1) 71-88.

[2] Goel Radhika., Sardana Anjali., Joshi Ramesh, C. (2012). Parallel misuse and anomaly detection model, *International Journal of Network Security*, 14 (4), 211-222.

[3] Khalkhali Iman, Azmi Reza, Azimpour-Kivi Mozhgan, Khansari Mohammad. (2011). Host-based web anomaly intrusion detection system, an artificial immune system approach, *International Journal of Computer Science*, 8 (55), 14-24.

[4] Matzinger Polly. (2010). The danger model: A renewed sense of self, *Science*, 296 (5566), 301-305.

[5] Greensmith., Julie., Aickelin, Uwe.,Tedesco, Gianni. (2010). Information fusion for anomaly detection with the dendritic cell algorithm, *Information Fusion* 11 (1), 21-34.

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

[7] Greensmith, J., Aickelin , U. (2007). Dendritic cells for SYN scan detection, *In* : Proceedings of the 9th Annual conference on Genetic and evolutionary computation. London, England; ACM. 49-56.

[8] OATES, R., Greensmith, J., Aickelin, U. (2007). The Application of a Dendritic Cell Algorithm to a Robotic Classifier . 204-215.

[9] Al-Hammadi , Y., Aickelin, U., Greensmith, J. (2008). DCA for bot detection. *In* : Proceedings of IEEE Congress on Evolutionary Computation, CEC .

[10] GuFeng, Greensmith, J., Aickelin, U. (2009). Integrating real-time analysis with the dendritic cell algorithm through segmentation [M], *In* : Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation. Montreal, Canada; ACM. 1203-1210.

[11] Greensmith, J., Aickelin, U. (2008). The Deterministic Dendritic Cell Algorithm, 291-302.

[12] Wang Jinjiang., Gao Robert, X., Yan Ruqiang. (2013). Multi-scale enveloping order spectrogram for rotating machine health diagnosis, Mechanical Systems and Signal *In* : Processing.

[13] Xiekai-zhong, Song bing-hui, Wang hong-wei, Lv lin-hai,(2013). The neural network method of index system and weights for highway tunnel health diagnosis, Applied Mechanics and Materials, 303-306, 794-798.

[14] Yuan Shen-Fang, Liang Dong, Gao Ning, (2013). The bridge data diagnosis research based on structural health monitoring system, *Journal of the University of Electronic Science and Technology of China*, 42 (1), 69-74.